UNCLASSIFIED

AD NUMBER

AD450443

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational use; Dec 1964. Other requests shall be referred to Army Research Office, Durham NC.

AUTHORITY

CFSTI per USARO ltr, 21 Sep 1965.

THIS PAGE IS UNCLASSIFIED

AD- 4 5 0 4 4 3

ARO-D Report 64-2

PROCEEDINGS OF THE NINTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by
U.S. Army Mathematics Steering Committee
on Behalf of
[illegible] RESEARCH AND DEVELOPMENT

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 64-2
December 1964

PROCEEDINGS OF THE NINTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by the Army Mathematics Steering Committee

conducted at

Directorate of Research and Development
U. S. Army Missile Command
Redstone Arsenal, Alabama
23-25 October 1963

U. S Army Research Office-Durham
Box CM; Duke Station
Durham, North Carolina

TABLE OF CONTENTS

TABLE OF CONTENTS (contd) | Page

* This paper was presented at the conference. It does not appear in these Proceedings.

TABLE OF CONTENTS (contd) Page

---

*This paper was presented at the conference. It does not appear in these Proceedings.

TABLE OF CONTENTS (contd)

---

* This paper was presented at the conference. It does not appear in these Proceedings.

**The author was unable to present his paper at the conference. It does not appear in this technical manual.

TABLE OF CONTENTS (contd) Page

*This paper was presented at the conference. It does not appear in these Proceedings.

FOREWORD

Professor S. S. Wilks outlined the form of the first conference on the Design of Experiments in Army Research, Development and Testing. Basically, this meeting was composed of three parts; an invited speaker phase in which experts were invited to discuss recent developments and applications in the field of design of experiments; the second phase consisted of technical papers presented with Army personnel discussing recent successes in handling their problems, and finally the third phase consisted of so-called clinical sessions. The speakers for these sessions were also Army personnel, but now they were presenting unsolved problems. Nationally known scientists served as panelists. These individuals, as well as members of the audience, discussed the unsolved design problems and suggested methods of attack on them. Professor Wilks agreed to serve as chairman of the first conference. It was such a success that the Army has continued the meetings on a yearly basis. Dr. Wilks served as their chairman until his recent untimely death. The Army Mathematics Steering Committee, sponsors of the design symposia, fully realize the debt they owe to Professor Samuel S. Wilks of Princeton University and have asked that the next conference, the tenth in this series, be dedicated to him.

At the Ninth Conference on the Design of Experiments, Drs. David Duncan, Churchill Eisenhart, H. O. Hartley, Solomon Kullback and Frank Proschan delivered invited addresses. Estimation of missile trajectories, precision and accuracy, non-linear estimation, communication theory, and monotone hazard rates in systems reliability were, respectively, the topics treated by these specialists. Dr. Craig Crenshaw was invited by the host to be the After Dinner Speaker. Boyd Harshbarger served as the Chairman of the Panel Discussion on what types of Statisticians are needed in Research and Development Laboratories. He asked the following persons to serve on this panel and discuss various aspects of the topic: E. L. Cox, Churchill Eisenhart, Donald Gardner, Frank E. Grubbs, John L. McDaniel, Paul I. Rider, and William Wolman. In addition to these phases of the program, six papers were given in the Clinical Sessions and twenty-six papers in the Technical Sessions.

In order to contribute to a wide dissemination of knowledge and use of modern statistical principles in the design of experiments, the AMSC is making these proceedings available to all interested Army personnel. The present volumn contains twenty-nine of the papers which were presented at this meeting.

Two hundred and twenty-nine registrants and participants from over seventy different organizations attended the Ninth Conference. Speakers and panelists came from the Atlantic Missile Range; Atomic Energy Commission; Biometrical Services, Agricultural Research Services, Plant Industry Station; Boeing Scientific Research Laboratories; Brigham Young University; C-E-I-R, Inc.; George Washington University; Goddard Space Flight Center; Johns Hopkins University; NASA Marshall Space Flight Center (George C. Marshall Space Flight Center); National Bureau of Standards; National Institutes of Health; North Carolina State University of the University of North Carolina at Raleigh; Oakridge National Laboratories; Pan American World Airways; Patrick Air Force Base; Pratt and Whitney Aircraft; Princeton; Research Analysis Corporation; Research Triangle Institute; Sandia Corporation; Texas A-M University; University of Wisconsin; War Office, United Kingdom; Wright Patterson Air Force Base; and ten Army facilities.

The members of the Army Mathematics Steering Committee take this opportunity to express their thanks to the many speakers and other research workers who participated in the conference; to Brigadier General John Zierdt for having the U. S. Army Missile Command serve as host and for making available the facilities under his command at Redstone Arsenal for the meeting; and to W. H. Ewart who served as Chairman on Local Arrangements. Those in attendance are indeed indebted to him for the excellent handling of the many details needed to make such a meeting profitable and enjoyable. Mr. Ewart was ably assisted in these tasks by members of his committee: E. L. Bombara, Eleanor Colbert, Henry A. Dihm, Don Fulton, Siegfried H. Lehnigk, and Clyde R. Ward.

A complete history of the U. S. Army Missile Command or details of the work in progress at Redstone Arsenal will not be given here, but the following information about the host of the Ninth Conference on the Design of Experiments should prove very interesting to those in attendance. These remarks and the following picture were taken from the pamphlet entitled U. S. Army Missiles and Rockets.

"Redstone Arsenal, Ala., is the home of U. S. Army missiles and rockets where a dedicated military-civilian team of highly trained specialists see to it that American soldiers have missile and rocket weapons ready for use, wherever and whenever they may be needed in the defense of freedom.

The U. S. Army Missile Command, which is a major commodity command of the U. S. Army Materiel Command, directs world wide Army missile activities from its headquarters here. Also located at Redstone Arsenal is the Army Ordnance Guided Missile School where troops of the U. S. Army, the Army's sister services, and allied countries, learn to be missilemen. The Nation's top priority program of perfecting an effective defense against attack by intercontinental ballistic missiles is directed from Redstone Arsenal by the Nike Zeus Project Office. In logistical support of all these organizations is the Army Missile Support Command, which is a major element of the Army Missile Command.

U. S. Army missiles are the proud products of the people who stand behind them -- more than 12,000 men and women in and out of uniform who are engaged in this vital work at Redstone Arsenal. They are responsible for managing, developing and supporting the Free World's most versatile missile systems. The American soldier depends on the long experience and know-how of these scientists, technicians, engineers, secretaries and shopmen to bring a missile system successfully to life. They work in partnership with defense contractors across the nation to form a true Army-Industry team.

It takes money to provide up-to-date weapons for your modern Army. One of the major responsibilities of the Army missile team is to make sure that the taxpayer gets a fair return for the portion of his tax dollar allotted Army missiles. This means a billion dollar a year investment which is split among universities, private research institutions and American industry and which involves more than 40 prime contractors, 300 first tier subcontractors and more than 5,400 subcontractors in almost every state in the union.

Much of the work at Redstone is done behind closed doors under tight security wraps. In the Missile Command's eight major research and development laboratories, tomorrow's missiles are today's work. The Army calls it "in-house" research. It ranges across the entire spectrum of missile technology including such activities as experimentation with new high energy rocket fuels, investigation of new means of steering missiles and advanced work in metals, chemicals, and the physical sciences. The tools of the trade include massive captive test stands where missiles are held firmly to earth as their rocket motors roar, lightning fast computers analyzing miles of test data in their electronic brains, and ever present slide rule in the hands of engineers. The single word facilities is used to describe land, buildings, improvements and specialized equipment. The Army investment in "facilities" at Redstone is approximately $300 million dollars.

Missile hardware is the end product of our resources of people, know-how, money and facilities. These resources combine to form the Army team at Redstone, which turns out superior missiles and their supporting equipment that do a variety of jobs for the soldier.

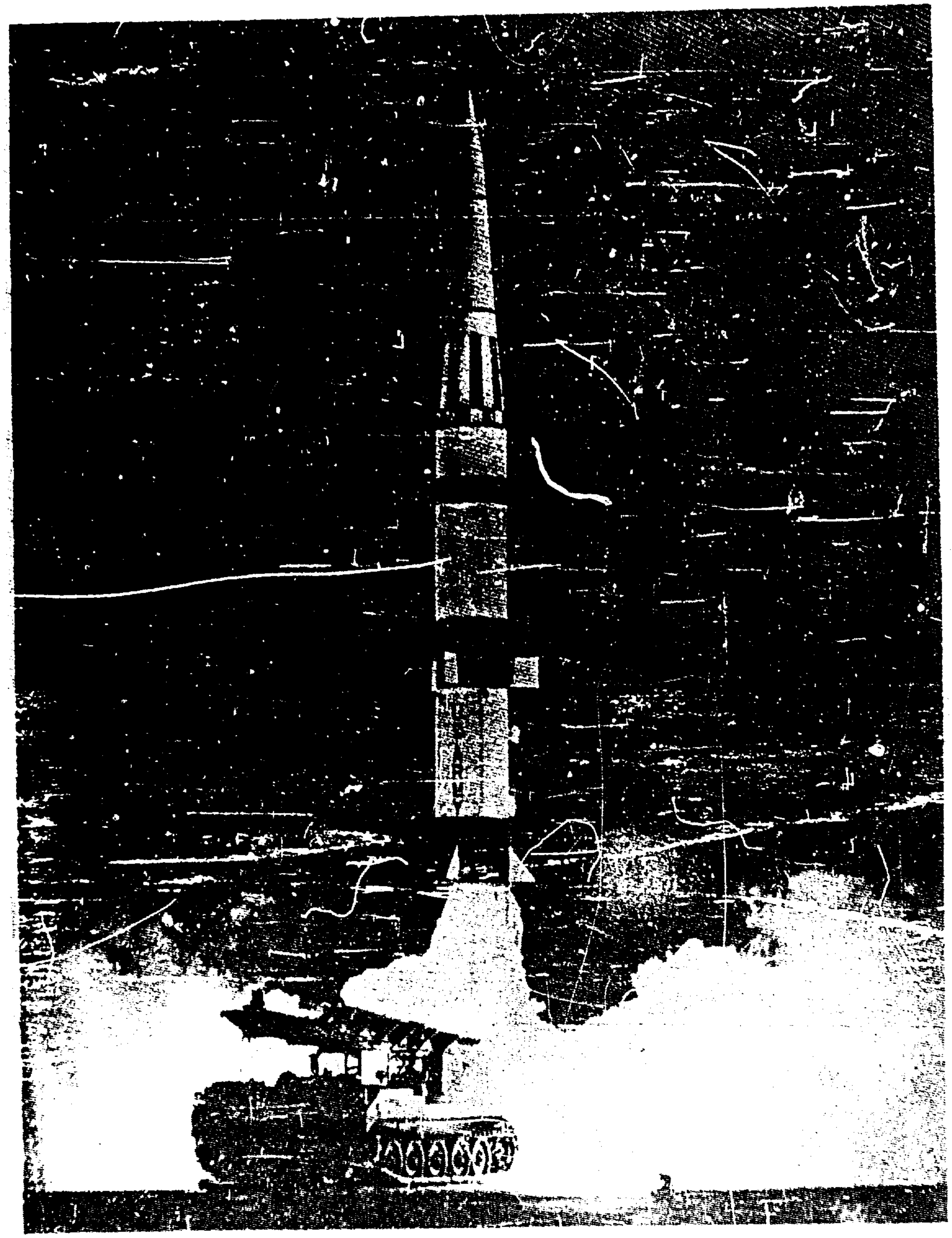
This missile hardware covers the entire spectrum of the Soldier's needs from a shoulder fired, two-pound, tank killer to the two-stage Pershing ballistic missile with a range of several hundred miles. It includes missiles like Sergeant, a quick reacting, solid fuel nuclear punch for the field Army, and the Nike Hercules and Hawk, which guard Army units overseas and American cities at home from aircraft attack.

Their shapes may differ but their pedigree is the same."

V

ARMY'S MISSILE FAMILY

"THE ARMY'S MISSILE FAMILY -- Here stands the end products of the Army's research, development, and testing in missile weapons. The larger missiles, left to right, are: Nike Hercules, Sergeant, Nike Zeus, Pershing, and Nike Ajax. In the background is Lacrosse. In the foreground is Hawk. Soldiers in the foreground hold, left to right, the M-72 Rocket Grenade, the Redeye, and the ENTAC."



PERSHING

NINTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
ARMY RESEARCH, DEVELOPMENT AND TESTING
23-25 October 1963

U. S. Army Missile Command
Redstone Arsenal, Alabama

<u>Wednesday, 23 October 1963</u>

0815 REGISTRATION — Bldg. 7120, Lobby of Rocket Auditorium

0900 CALLING OF CONFERENCE TO ORDER — Rocket Auditorium
Colonel Daniel F. Shepherd
Director, Research and Development
U. S. Army Missile Command

WELCOME
Brigadier General John G. Zierdt
Commanding General
U. S. Army Missile Command

MOVIE
"This is Redstone"

ANNOUNCEMENTS
Wade H. Ewart, Chairman
Local Arrangements Committee

0930 GENERAL SESSION I
Chairman: Colonel Nils M. Bengston
Commanding Officer, Army Research
Office, Durham, N. C.

"COMMUNICATION THEORY"
Solomon Kullback, Dept. of Statistics,
The George Washington University

(1040-1115 COFFEE BREAK - Lobby of Rocket Auditorium)

"THE CONCEPT OF MONOTONE HAZARD RATE
IN SYSTEMS RELIABILITY"
Frank Proschan, Mathematics
Research Laboratory, Boeing
Scientific Research Labaoratories

x

1230 LUNCH — Bldg. 7101, Cafeteria

1330 TECHNICAL SESSION I — Bldg. 7101, Control Room

Chairman: Kenneth H. Abbott, Army Materials Research Agency, Watertown Arsenal, Watertown, Mass.

"UNBIASED ESTIMATES OF RELIABILITY WHEN TESTING AT ONE EXTREME STRESS LEVEL"
A. Bulfinch, Quality Assurance Directorate, Picatinny Arsenal

"THE COLLECTION, PROCESSING, AND STORAGE OF DATA ON SERIALIZED NUCLEAR WEAPONS ITEMS"
Manfred W. Krimmer Statistics and Analysis Section, Reliability Branch, QAD, Picatinny Arsenal

1330 TECHNICAL SESSION II — Bldg. 7101, Conference Room

Chairman: Siegfried H. Lehnigk, Physical Sciences Laboratory Army Missile Command

"PROBABILITY THAT STRESS IS LESS THAN STRENGTH AT PRESCRIBED CONFIDENCE LEVELS, FOR NORMALLY DISTRIBUTED DATA"
E. L. Bombara, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

"STATISTICAL TESTING TECHNIQUES USED IN THE DEVELOPMENT OF THE RL10 ROCKET ENGINE"
Harold J. Tiedemann, Pratt & Whitney Aircraft, representing NASA Marshall Space Flight Center

"COMPUTER SIMULATION STUDY OF BRUCETON AND PROBIT METHODS OF SENSITIVITY TESTING"
J. B. Gayle, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

1330 CLINICAL SESSION A Rocket Auditorium

Chairman: Bertram W. Haines,
Biometrics Division, Army Biological
Warfare Laboratories, Fort Detrick,
Frederick, Md.

Panelists:

Dr. Walter D. Foster, Biomathematics Division, Army Biological Warfare Laboratories

Professor H. O. Hartley, Institute of Statistics, Agricultural and Mechanical College of Texas

Professor H. L. Lucas, Jr., Institute of Statistics, North Carolina State College

Dr. Clifford J. Maloney, Division of Biologics Standards, National Institute of Health

"THE IMPACT OF ADMINSTRATIVE LIMITATIONS ON THE DESIGN OF EXPERIMENTS"
Gerhard J. Isaac, Army Medical Research and Nutrition Laboratory, Denver, Colorado

"USE OF THE MITCHERLICH EQUATION IN THE THREE-PHASE EXPERIMENTAL DESIGN"
Ardie Lubin, Dept. of Clinical and Social Psychology, Neuropsychiatry Div., Walter Reed Army Institute of Research, Walter Reed Army Medical Center

(1450-1530 COFFEE BREAK - Lobby, Rocket Auditorium and Cafeteria, Bldg. 7101)

1530 TECHNICAL SESSION III Bldg. 7101 Control Room

Chairman: Alan S. Galbraith,
Army Research Office, Durham, N. C.

"VERIFICATION OF PRODUCT ACCEPTANCE INSPECTION BY VARIABLES"
Henry Ellner, Directorate for Quality Assurance, Edgewood Arsenal, Md.

"VERIFICATION OF PRODUCT ACCEPTANCE INSPECTION BY ATTRIBUTES"
Joseph Mandelson, Directorate for Quality Assurance, Edgewood Arsenal, Md.

1530 TECHICAL SESSION IV — Bldg. 7101 Conference Room

Chairman: Frank L. Carr, Army Materials Research Agency, Watertown Arsenal, Watertown, Mass.

"ADDITIONAL ANALYSIS OF MISSILE TRAJECTORY MEASURING SYSTEMS"
Oliver Lee Kingsley and Bernie R. Froe, Range Instrumentation Systems Office, White Sands Missile Range, New Mexico

"ASPECTS TO CONTROL LIQUID PROPELLANT SLOSHING BASED UPON EXISTING THEORY"
Werner R. Eulitz, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

"BASIC CONSIDERATIONS FOR THE PRELIMINARY DESIGN OF A SHOCK TUBE FOR THE EXPERIMENTAL INVESTIGATION OF THE ACTION OF A NUCLEAR WAVE UPON A MISSILE"
Dietrich E. Gudzent, Structures and Mechanics Laboratory, Army Missile Command

1530 CLINICAL SESSION B — Rocket Auditorium

Chairman: Henry A. Dihm, Advanced Systems Laboratory, Army Missile Command

Panelists:
- Professor David B. Duncan, Johns Hopkins Univ. and Atlantic Missile Range
- Professor Solomon Kullback, Dept. of Statistics, The George Washington Univ.
- Dr. Frank Proschan, Boeing Scientific Research Laboratories
- Professor G. S. Watson, Dept. of Statistics, Johns Hopkins University

"RELIABILITY CONCEPTS FOR MISSILE BATTERIES"
Nicholas Wilburn, Army Electronics Research and Development Laboratory, Fort Monmouth, N. J.

"MONTE CARLO APPLICATION FOR DEVELOPING A DESIGN RELIABILITY GOAL COMPATIBLE WITH SMALL SAMPLE REQUIREMENTS"
Roy Heathcock and Dale L. Burrows, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

Thursday, 24 October 1963

0830 TECHNICAL SESSION V — Bldg. 7101, Confer. Rm.

Chairman: Lt. Colonel Stefano Vivona, Division of Communicable Diseases, Walter Reed Army Institute of Research

"EXACT MULTIPLE CONTINGENCY TABLE ANALYSIS"
Dorothy Berg, Morley Leyton, and Clifford Maloney, Division of Biologics Standards, National Institute of Health

"MICROSPECTROSCOPY OF TISSUES"
George I. Lavin, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.

(0950-1000 SHORT BREAK)

"RELATIONSHIP OF AEROSOL STABILITY TO VIRULENCE IN MICRO-ORGANISMS"
Howard O. Nielson, Dugway Field Office, C-E-I-R, INC., and Brigham Young Univ., rep. Dugway Proving Ground

0830 TECHNICAL SESSION VI — Rocket Auditorium

Chairman: Badrig Kurkjian, Harry Diamond Laboratories

"DESIGN FOR THE SEQUENTIAL APPLICATION OF FACTORS"
Sidney Addelman, Research Triangle Inst., rep. Army Research Office, Durham

"2^p FACTORIAL EXPERIMENTS WITH THE FACTORS APPLIED SEQUENTIALLY"
R. R. Prairie and W. J. Zimmer, Sandia Corp., rep. Atomic Energy Commission

(0950-1000 SHORT BREAK)

"ESTIMATION OF ERROR-SPECTRA FROM THE CROSS-COVARIANCE FUNCTIONS OF DIFFERENCES"
D. B. Duncan, Johns Hopkins Univ., and W. T. Wells, Pan American World Airways, Guided Missile Range Div., PAFB, Fla.

0830 TECHNICAL SESSION VII — Bldg. 7101, Control Rm.
Security Classification: CONFIDENTIAL

Chairman: John Purtell, Watervliet Arsenal, Watervliet, N. Y.

"A COMPARISON OF DIFFERENT METHODS OF WEAPONS EVALUATION"
Andrew J. Eckles, III, Research Analysis Corp.

"DESIGN OF A DATA GATHERING MODEL FOR EVALUATING SURFACE-TO-AIR MISSILE SYSTEM SUPPORT REQUIREMENTS"
Leon Miller, Research Analysis Corporation, presented by N. Ray Sumner

(0950-1000 SHORT BREAK)

"AN ANALYSIS OF HELICOPTER RECONNAISSANCE TECHNIQUES"
Arthur P. Woods, Research Analysis Corp.

(1040-1120 COFFEE BREAK - Lobby of Rocket Auditorium)

1120 GENERAL SESSION 2 — Rocket Auditorium

Chairman: Dr. Walter D. Foster, Biometrics Div., Army Biological Warfare Labs, Fort Detrick, Md.

"REALISTIC EVALUATION OF THE PRECISION AND ACCURACY OF INSTRUMENT CALIBRATION SYSTEMS"
Churchill Eisenhart, National Bureau of Standards, Washington, D. C.

1230 LUNCH Bldg. 7101, Cafeteria

1330 TECHNICAL SESSION VIII Bldg. 7101, Control Rm.

Security Classification: CONFIDENTIAL

Chairman: Harold Fassberg, Research Analysis Corporation

"STATISTICAL STUDY OF RELIABILITY AND ACCURACY OF SURFACE-TO-AIR MISSILES"
Bruce Sterner, Surveillance Group, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.

1330 TECHNICAL SESSION IX Bldg. 7101, Confer. Rm.

Chairman: Miss Arla Weinert, Research Analysis Corporation

"NOTES ON FLEET HOMOGENEITY AND HETEROGENEITY"
G. E. Cooper, Research Analysis Corporation

1330 TECHNICAL SESSION X Rocket Auditorium

Chairman: E. L. Bombara, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

"STATISTICAL DECISION THEORY"
Lionel Weiss, Army Mathematics Research Center, University of Wisconsin

(1415-1450 COFFEE BREAK - Lobby of Rocket Auditorium)

1450 GENERAL SESSION 3 Rocket Auditorium

PANEL DISCUSSION ON WHAT TYPE OF STATISTICIANS ARE NEEDED IN RESEARCH AND DEVELOPMENT LABORATORIES

Chairman: Professor Boyd Harshbarger, Virginia Polytechnic Institute

Panelists:

Dr. E. L. Cox, Biometrical Services, ARS, Plant Industry Station

Dr. Churchill Eisenhart, National Bureau of Standards

Dr. Frank E. Grubbs, Army Ballistics Research Laboratories, Aberdeen, Md.

Mr. John L. McDaniel, Directorate of R&D, Army Missile Command

Dr. Paul R. Rider, Office of Aero-Space Research, Wright-Patterson AFB

Dr. William Wolman, Goddard Space Flight Center, NASA

Dr. Donald A. Gardiner, Mathematics Div., Oakridge National Laboratories

1900 SOCIAL HOUR AND DINNER — Redstone Arsenal Officers' Open Mess

Evening Session Chairman:
Mr. John L. McDaniel, Technical Director of R&D, Army Missile Command

After Dinner Speaker:
Dr. Craig M. Crenshaw, Chief Scientist, U. S. Army Materiel Command

Friday, 25 October 1963

0830 TECHNICAL SESSION XI — Rocket Auditorium

Chairman: Erwin Biser, System Division, Surveillance Department, Army Electronics Research and Development Laboratory, Fort Monmouth, N. J.

"AN ANALYSIS OF FACTORIAL EXPERIMENTAL DESIGNS"
L. W. Kesting, Propulsion Laboratory, Army Missile Command

"SOME ASPECTS OF ANALYSIS OF PARTIALLY FACTORIAL EXPERIMENTS"
Scott A. Krane, Dugway Field Office, C-E-I-R, INC., representing Dugway Proving Ground

0830 TECHNICAL SESSION XII — Bldg. 7101, Control Room

Chairman: Robert E. Weigle, Watervliet, Arsenal, Watervliet, N. Y.

"RELIABILITY ESTIMATION FOR MULTI-COMPONENT SYSTEMS"
James R. Kniss, Surveillance Group, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.

"STATISTICAL STUDY OF AGING CHARACTERISTICS OF ARTILLERY MISSILES"
Raymond Bell, Surveillance Group, Army Ballistic Research Laboratories, Aberdeen Proving Ground, Md.

0830 CLINICAL SESSION C — Bldg. 7101, Conference Room

The paper by D. H. Chaddock carries a security classification of CONFIDENTIAL; the paper by Lester Katz is not classified.

Chairman: Fred Frishman, U. S. Army Research Office, Washington, D. C.

Panelists:
- Dr. O P. Bruno, Surveillance Group, Army Ballistics Research Laboratories
- Dr. F. E. Grubbs, Army Ballistics Research Laboratories
- Dr. Lionel Weiss, Army Mathematics Research Center, University of Wisconsin
- Professor S. S. Wilks, Princeton University

"THE MEASUREMENT OF THE MORALE AND SUPPRESSIVE EFFECTS OF WEAPONS"
D. H. Chaddock, Esq., C. B. E., Director of Artillery Research and Development, The War Office, United Kingdom (presented by G. F. Komlosy, presently attached to the Research Analysis Corporation)

"THE DESIGN OF AN EXPERIMENTAL PROGRAM TO INVESTIGATE BUCKLING IN THE HOOP DIRECTION OF PRATIALLY FILLED BULKHEADS UNDER LOADING"
Lester Katz, Propulsion and Vehicle Engineering Division, NASA Marshall Space Flight Center

(0950-1020 COFFEE BREAK - Lobby of Rocket Auditorium)

1020 GENERAL SESSION 4 Rocket Auditorium

Chairman: Professor S. S. Wilks, Princeton University

"NONLINEAR ESTIMATION"
H. O. Hartley, Institute of Statistics, Agricultural and Mechanical College of Texas

"ON THE SIMULTANEOUS ESTIMATION OF A MISSILE TRAJECTORY AND THE ERROR VARIANCE COMPONENTS INCLUDING THE ERROR POWER SPECTRA OF SEVERAL TRACKING SYSTEMS"
David B. Duncan, Professor of Statistics and Biostatistics, Johns Hopkins University, and Statistical Consultant, Atlantic Missile Range

1230 LUNCH Bldg. 7101, Cafeteria

1330-1530 TOUR OF ARSENAL FACILITIES

COMMUNICATION THEORY

Solomon Kullback
Department of Statistics
The George Washington University

This presentation will cover one aspect of communication theory and I shall try to relate some of the concepts of interest to the communication engineer with the general subject of this conference. In particular I shall talk about certain facets of information theory, an important field of contemporary probability theory and statistics.

I shall not take the time necessary to engage in a detailed discussion motivating the technical definitions of measures of information and information theory, except to remark that among the various definitions which have been given is one that states "information is a measure of time or cost of a sort, which is of particular use to the engineer in his role of designer of an experiment".

Suppose a system (or information source) has c different possible events or categories or messages $A_1, A_2, \ldots A_c$ with respective probabilities of occurrence $p_1, p_2, \ldots, p_c$, $\sum_{i=1}^{c} p_i = 1$.

$$(1) \qquad \phi : \begin{pmatrix} A_1, & A_2, & \ldots, & A_c \\ p_1, & p_2, & \ldots, & p_c \end{pmatrix}$$

The expression

$$(2) \qquad H(\phi) = -p_1 \log p_1 - p_2 \log p_2 - \ldots - p_c \log p_c$$
$$= E(h)$$

where h is a random variable which takes on the values $\log \frac{1}{p_1}$, $\log \frac{1}{p_2}$, $\ldots$, $\log \frac{1}{p_c}$, is called the entropy of the system ϕ, by analogy with a similar concept and mathematical expression in statistical mechanics. The entropy is interpreted as the mean uncertainty about ϕ prior to an observation, or the mean information about ϕ provided by an observation. The base of the logarithm is quite arbitrary and is just a unit of measurement.

The fact that

$$(3) \quad n = b^x = a^y, \; x = \log_b n, \; y = \log_a n, \; \log_b n = \log_b a \cdot \log_a n$$

permits ready change from one logarithmic base to another. The following Table 1 indicates the unit in which H (ϕ) is measured for the more common bases of logarithms

Table 1

Base	Unit
2	bit
e	nat, nit
10	dit, Hartley

where the unit name Hartley is to honor the communication engineer R. V. L. Hartley who in 1928 introduced a logarithmic measure of information for use in communication engineering.

Let us consider some of the properties of H (ϕ).

(1) The value of H (ϕ) is independent of the numerical value, name, quality category, or other disignation of the events A_i, but depends only on the probabilities of their occurrence.

(2) H (ϕ) is a symmetric function of the probabilities

(3) H (ϕ) ≥ 0, with equality if and only if some $p_i = 1$ and all other $p_j = 0$, $j \neq i$ (we define 0 log 0 = 0).

(4) For two events, that is c = 2, with $p_1 = 1/2 = p_2$, H (ϕ) = log 2 = 1 bit, that is

<u>one bit of information is the capability of resolving the uncertainty in a choice between two equally likely alternatives.</u>

(5) H (ϕ) is a maximum if all the p_i are equal to 1/c in which case H (ϕ) = log c, Hartley's measure.

The maximum value of H (ϕ) corresponds to the resolution of the greatest uncertainty and the zero value of H (ϕ) corresponds to no uncertainty. From this we may infer an important principle. The determination of one of a group of objects by a sequence of selections will be most efficiently accomplished in the sense of minimum mean effort if each selection is made from equally probable groupings, that is, with maximum information, or maximum uncertainty resolved for each selection.

Suppose one of k objects is to be determined by the answers to m questions as to which group contains the object in question. The maximum information is provided by an answer if the objects are grouped into r groups of equal size yielding $\log_2 r$ bits of information per answer. The answers to m independent questions may be considered as the analogue of a communication channel of capacity $C = m \log_2 r$ bits. The maximum uncertainty in a source of k objects is $H = \log_2 k$ bits. If $H \leq C$, a predetermined object can be uniquely identified without error, that is, the uncertainty can be completely resolved. However, if $H > C$, the identification cannot be made without some possibility of error, that is, the uncertainty cannot be completely resolved. For example, suppose there are 27 coins of equal value of which 26 are of the same weight and one, a false coin, is heavier than the others. How many weighings on a two pan balance, without weights, are necessary always to determine the false coin? The uncertainty that must be resolved is $H = \log 27$. There are 3 possible outcomes at a weighing, the left pan is heavier, the right pan is heavier, both pans balance. Thus each weighing provides information of measure at most log 3. To be sure to determine the false coin at least three weighings are therefore necessary since $3 \log 3 = \log 27$. We leave it to you to develop the appropriate sequence of operations. What if there are 26 coins?

Closely related to the preceding ideas is the coding theorem which states that given a source of mean information of H bits per observation it is possible to encode or translate sequences of observations into sequences of two elements only, say the binary digits 0 and 1, such that on the average, a sequence of N original observations (information content NH and information content NH bits. We note this as a law of conservation of information.

Suppose there are k possible observations with probabilities of occurrence $p_1, p_2, \ldots, p_k$, $\sum_{i=1}^{k} p_i = 1$. Each of the k possible observations could be represented by a sequence of r binary digits, where $k \leq 2^r$ or $r \geq \log_2 k$. A sequence of N observations would thus be translated into a sequence of $rN \geq N \log_2 k$ binary digits. The coding theorem says in effect, that by taking advantage of the different probabilities of occurrence of the observations since $-p_1 \log_2 p_1 - p_2 \log_2 p_2 - \cdots - p_k \log_2 p_k = H \leq \log_2 k$, there exists a more economical translation, $NH \leq N \cdot \log_2 k \leq rN$, and that it is the best possible translation. A similar notion underlies the technique of sequential analysis in statistics which achieves a test of a certain strength with a mean number of observations smaller than a fixed sized sample of similar strength.

In particular, suppose there are four possible observations, say A, B, C, D, with respective probabilities of occurrence $p(A) = 1/2$, $p(B) = 1/4$, $p(C) = 1/8$, $p(D) = 1/8$. For this source it is found that

$$
\begin{aligned}
H &= -1/2 \log 1/2 - 1/4 \log 1/4 - 1/8 \log 1/8 - 1/8 \log 1/8 \\
&= 1/2 \log 2 + 1/2 \log 2 + 3/8 \log 2 + 3/8 \log 2 \qquad (4) \\
&= 1.75 \log 2 = 1.75 \text{ bits}
\end{aligned}
$$

A coding that will achieve a mean of 1.75 binary digits per original observation is given by the <u>Fano coding</u>

$$
\begin{aligned}
A &= 0 \\
B &= 1\,0 \\
C &= 1\,1\,0 \qquad (5) \\
D &= 1\,1\,1
\end{aligned}
$$

Note that this coding is derived by a sequence of groupings into two equally probable groups, identifying the elements remaining in each group in each step by the binary digit 0 or 1: first into the grouping (A), (B, C, D); then into the grouping (B), (C, D); and finally into the grouping (C), (D). This Fano coding will uniquely convert a sequence of A, B, C, D's into a sequence of binary digits and back to the A, B, C, D's.

As an example,

(6) 1 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 0 1 1 1 0 1 1 0 . . .

is grouped uniquely according to the given Fano coding into

(7) 1 0/ 1 1 0/ 1 0/ 1 0/ 0/ 1 1 1/ 0/ 1 0/ 0/ 0/ 1 1 1/ 0/ 1 1 0/ . . .

and converted to

(8) B C B B A D A B A A D A C . . .

The mean number of binary digits per original observation is

(9) $1 \times 1/2 + 2 \times 1/4 + 3 \times 1/8 + 3 \times 1/8 = 1.75$

of which

(10) $1 \times 1/2 + 1 \times 1/4 + 1 \times 1/8 = 7/8 = 0.875$

are "0", and 0.875 are "1".

For the binary coding,

(11)
A = 0 0
B = 0 1
C = 1 0
D = 1 1

The converted binary sequence corresponding to N original observations would consist of 2 N binary digits rather than 1.75 N binary digits.

This sort of idea is clearly of importance in considerations of efficiency and economy not only in communications but the storage and rapid scanning and retrieval of large volumes of data.

The facsimile transmission of typewritten material by two signal levels, white or black, may be accomplished by dividing the typewritten material into elements by rows and columns 100 to the inch. A particular small sample of typewritten text yielded the following data on the four possible pairs of consecutive elements:

Table 2

	Frequency	Probability	Binary Coding	Fano Coding
White White	14,381	0.8209	0 0	0
Black Black	1,786	0.1020	1 1	1 0
Black White	709	0.0405	1 0	1 1 0
White Black	642	0.0366	0 1	1 1 1
	17,518	1.0000		

The number of binary digits required to record the sample in binary coding, 2 digits for each pair of elements is 2 x 17518 = 35,036; for Fano coding the number of binary digits required is

(12) $$1 \times 14,381 + 2 \times 1,786 + 3 \times (709 + 642) = 22,006$$

The number of binary digits given by the theoretical H is still smaller, 16,320, obtained from

(13) $$\begin{aligned} NH &= 17518\ (-0.8209 \log_2 0.8209 - 0.1020 \log_2 0.1020 - \\ &\qquad 0.0405 \log_2 0.0405 - 0.0366 \log_2 0.0366) \\ &= 17518\ (0.2335 + 0.3359 + 0.1875 + 0.1747) \\ &= 17518\ (0.9316) = 16,320. \end{aligned}$$

For a grouping of the elements into the sixteen sets of four consecutive elements, the data gave:

Table 3

	Frequency	Probability	Binary Coding	Fano Coding	$-p \log_2 p$
WWWW	6,586	0.752	0 0 0 0	0	0.3092
BBBB	387	.044	1 1 1 1	1 0 0 0	.1983
WWBB	295	.034	0 0 1 1	1 0 0 1	.1659
BWWW	283	.032	1 0 0 0	1 0 1 0	.1589
BBWW	282	.032	1 1 0 0	1 0 1 1	.1589
WWWB	264	.030	0 0 0 1	1 1 0 0	.1518
BBBW	211	.024	1 1 1 0	1 1 0 1	.1291
WBBB	171	.020	0 1 1 1	1 1 1 0 0	.1129
WBBW	105	.012	0 1 1 0	1 1 1 0 1	.0766
WWBW	54	.006	0 0 1 0	1 1 1 1 0 0	.0443
BBWB	34	.004	1 1 0 1	1 1 1 1 0 1	.0319
BWWB	31	.004	1 0 0 1	1 1 1 1 1 0 0	.0319
WBWW	31	.004	0 1 0 0	1 1 1 1 1 0 1	.0319
BWBB	19	.002	1 0 1 1	1 1 1 1 1 1 0	.0179
WBWB	3	.000	0 1 0 1	1 1 1 1 1 1 1 0	.0000
BWBW	3	.000	1 0 1 0	1 1 1 1 1 1 1 1	.0000
	8,759	1.0000			1.6195

The number of binary digits required to record the sample in binary coding, 4 digits for each of the sixteen possible groups of elements is $4 \times 8759 = 35,036$. For Fano coding the number of binary digits required is

$$(14) \qquad 1 \times 6586 + 4 \times (387 + 295 + 283 + 282 + 264 + 211) + 5 \times (171 + 105) + 6 \times (54 + 34) + 7 \times (31 + 31 + 19) + 8 \times (3 + 3) = 15,997$$

The number of digits given by the theoretical H is $8759 \times 1.6195 = 14,185$.

For a grouping of the picture elements into the 256 sets of eight consecutive elements binary coding would still require 35036 digits, but Fano coding requires only 13251 digits, a figure that more nearly approaches the theoretical NH for this situation, 12869 digits. Clearly, as the number of elements grouped increases, the results of the Fano coding approach the theoretical value determined by H. This is true because grouping into equally likely groups can be more nearly accomplished with the smaller probabilities. To achieve the resultant saving in transmission, unfortunately requires extra complexity in the equipment for storage, scanning, and reconversion.

In transmitting sequences of binary digits whether in a communication channel or a computer, errors do occur and it is therefore of interest to be able to detect and correct such errors and most desirably by the equipment itself. The basic principle is to limit the number of possible signals that may be transmitted by the use of particular patterns or various kinds of check or parity digits. The following problem is mathematically equivalent with an error detecting and correcting code. Suppose that no more than one of seven electrical components say a resistor may be defective, say shorted. How can the possible defective one always be detected in a minimum number of tests? Here there are eight possibilities, none defective or $R_1, R_2, \ldots, R_7$ defective. We thus need $\log 8 = 3$ bits of information. The test can be accomplished by three circuits with resistors in series and testing each for operability.

$$
\begin{aligned}
& R_1,\ R_2,\ R_3,\ R_5 \quad : C_1 \\
& R_1,\ R_2,\ R_4,\ R_6 \quad : C_2 \\
& R_1,\ R_3,\ R_4,\ R_7 \quad : C_3
\end{aligned}
\tag{15}
$$

The defective resistor, if any, is deduced in accordance with the following table:

Table 4

Nonoperable	Defective
C_1	R_5
C_2	R_6
C_3	R_7
C_1 and C_2	R_2
C_1 and C_3	R_3
C_2 and C_3	R_4
C_1 and C_2 and C_3	R_1

Of course if all the circuits are operable then there is no defective resistor.

Let us return to the c-category system ϕ. Clearly $H = -\Sigma p_i \log p_i$ is a parameter of this system of probabilities or of the corresponding multinomial distribution were we to take a random sample of n observations from the c-category population. A measure of the mean divergence per observation of the distribution $(p_1, p_2, \ldots, p_c)$ from the uniform distribution $(1/c, 1/c, \ldots, 1/c)$ or the redundancy of the communication engineer in the signals $A_1, \ldots, A_c$ is

$$(16) \quad \log c - H = \log c + \sum_{i=1}^{c} p_i \log p_i = \sum_{i=1}^{c} p_i \log p_i/(1/c),$$

and for n independent observations it is

$$(17) \qquad I_n = n \sum_{i=1}^{c} p_i \log p_i/(1/c).$$

We may verify this as follows. There are c^n possible samples. The entropy for the sample distribution from the population $(p_1, p_2, \ldots, p_c)$ is

$$\begin{aligned}(18) \qquad H_n &= -\Sigma\, p_{i1} \cdots p_{in} \log p_{i1} \cdots p_{in} \\ &= -\Sigma \frac{n!}{n_1! \ldots n_c!} p_1^{n_1} \ldots p_c^{n_c} \log p_1^{n_1} \ldots p_c^{n_c} \\ &= -n\,(p_1 \log p_1 + \ldots + p_c \log p_c) = nH\end{aligned}$$

where in the first summation $(i_1, \ldots, i_n)$ varies over all possible permutations of $(1, 2, \ldots, c)$ taken n at a time and $n_1 + n_2 + \ldots + n_c = n$ with n_i the number of occurunces of category i in a sample. The entropy for the sample distribution for the uniform population is $\log c^n = n \log c$

Given a sample of n observations, with $n_1, n_2, \ldots, n_c$ occurrences respectively in each of the c categories, $n = \sum_{i=1}^{c} n_i$, we may estimate the p's and I_n in (17) above by

$$(19) \qquad \hat{p}_i = n_i/n, \quad \hat{I}_n = n \sum \hat{p}_i \log \hat{p}_i/(1/c).$$

It turns out that $2\hat{I}_n$, using natural logarithms, asymptotically has a chi-square distribution with c-1 degrees of freedom under the null hypothesis of a uniform distribution. If the sample were drawn from the population $(p_1, \ldots, p_c)$ then $2\hat{I}_n$ asymptotically has a noncentral chi-square distribution with c-1 degrees of freedom and noncentrality parameter $2 I_n$.

Let us now consider the case of the system ϕ in which each event corresponds to the joint occurrence of a pair of values say $A_i = (\alpha_i, \beta_j)$ $i = 1, \ldots, r$, $j = 1, \ldots, c$, with the corresponding joint probabilities $p(\alpha_i, \beta_j)$, with

$$(20) \qquad \sum_i \sum_j p(\alpha_i, \beta_j) = 1.$$

There now enter into consideration the marginal probabilities

$$(21) \qquad p(\alpha_i) = \sum_j p(\alpha_i, \beta_j), \quad p(\beta_j) = \sum_i p(\alpha_i, \beta_j)$$

and the conditional probabilities

$$(22) \qquad p(\alpha_i \mid \beta_j) = p(\alpha_i, \beta_j)/p(\beta_j) \;:\; p(\beta_j \mid \alpha_i) = p(\alpha_i, \beta_j)/p(\alpha_i).$$

We define the entropy for each of the sets of joint, marginal and conditional probabilities by

(23)

$$H(\alpha, \beta) = -\sum_i \sum_j p(\alpha_i, \beta_j) \log p(\alpha_i, \beta_j),$$

$$H(\alpha) = -\sum_i p(\alpha_i) \log p(\alpha_i) ; \quad H(\beta) = -\sum_j p(\beta_j) \log p(\beta_j),$$

$$H(\alpha \mid \beta_j) = -\sum_i p(\alpha_i \mid \beta_j) \log p(\alpha_i \mid \beta_j)$$

$$H(\alpha \mid \beta) = \sum_j p(\beta_j) H(\alpha \mid \beta_j)$$

$$H(\beta \mid \alpha_i) = -\sum_j p(\beta_j \mid \alpha_i) \log p(\beta_j \mid \alpha_i)$$

$$H(\beta \mid \alpha) = \sum_i p(\alpha_i) H(\beta \mid \alpha_i)$$

It may be shown that

(24)
$$H(\alpha, \beta) = H(\alpha) + H(\beta \mid \alpha) = H(\beta) + H(\alpha \mid \beta)$$

that is, the mean uncertainty about the pair (α, β) is the mean uncertainty about α plus the mean uncertainty about β given α, with a similar interpretation for the second equation. It may also be shown that

(25)
$$H(\beta \mid \alpha) \leq H(\beta) ; \quad H(\alpha \mid \beta) \leq H(\alpha)$$

with equality if and only if α and β are independent, that is $p(\alpha_i, \beta_j) = p(\alpha_i) p(\beta_j)$. Note that (25) means that the mean uncertainty about β given α is less than the mean uncertainty about β and the two are equal if and only if α is independent of β and contributes no information.

From (24) and (25) we see that

(26)
$$H(\alpha, \beta) \leq H(\alpha) + H(\beta)$$

with equality if and only if α and β are independent.

If α is the input to a communication channel and β is the output, then

(27) $$H(\alpha) - H(\alpha \mid \beta) = I(\alpha, \beta) \geq 0$$

is a measure of the information transmitted via the channel or of the statistical dependence between output and input. It may be seen that

(28) $$\begin{aligned} I(\alpha, \beta) &= H(\alpha) + H(\beta) - H(\alpha, \beta) \\ &= -\sum_i p(\alpha_i) \log p(\alpha_i) - \sum_j p(\beta_j) \log p(\beta_j) \\ &\quad + \sum_i \sum_j p(\alpha_i, \beta_j) \log p(\alpha_i, \beta_j) \\ &= \sum_i \sum_j p(\alpha_i, \beta_j) \log\left(\frac{p(\alpha_i, \beta_j)}{p(\alpha_i)\, p(\beta_j)}\right). \end{aligned}$$

Mathematically, the channel is characterized by the conditional probabilities $p(\beta_j \mid \alpha_i)$. We note that

(29) $$I(\alpha, \beta) = \sum_i p(\alpha_i) \sum_j p(\beta_j \mid \alpha_i) \log \frac{p(\beta_j \mid \alpha_i)}{p(\beta_j)}$$

and the capacity of the channel is defined as

(30) $$C = \sup I(\alpha, \beta)$$

where the sup is over all possible inputs $\{p(\alpha_i)\}$ given $\{p(\beta_j \mid \alpha_i)\}$, and we recall that $p(\beta_j) = \sum_i p(\beta_j \mid \alpha_i)\, p(\alpha_i)$.

The expression for $I(\alpha, \beta)$ as written in (28) is a special case of the more general result

(31) $$I(X; Y) = \int f(x, y) \log \frac{f(x, y)}{g(x)\, h(y)} \, d\lambda(x)\, d\mu(y)$$

which is itself a special case of

(32) $$I(1:2) = \int f_1(\xi) \log \frac{f_1(\xi)}{f_2(\xi)} \, d\lambda(\xi)$$

where $f_1(\xi)$ and $f_2(\xi)$ are the generalized densities corresponding to the distribution of the random variable ξ under the hypotheses H_1 and H_2. In the expression for $I(X, Y)$, H_1 is the hypothesis that the component random variables in $\xi = (X, Y)$ are dependent and H_2 is the hypothesis of independence. We note that ξ may be a random vector of n components or even a stochastic process and the same remark holds for X and Y in (31).

In the remainder of the discussion we limit ourselves to natural logarithms. We mention two interesting particular cases of (31).

(1) if X and Y are normally distributed, then it is found that

$$(33) \qquad I(X, Y) = -1/2 \log (1 - \rho^2)$$

where ρ is the correlation coefficient of X and Y and we note that (33) does not depend on the means or variances.

(2) Let Y be a parameter θ ranging over a space Θ, so that $f(x, \theta)$ is the joint probability density of x and θ, $h(\theta)$ is the prior probability density of θ, $g(x \mid \theta)$ is the conditional probability density of x given θ, and the marginal probability density of x is $g(x) = \int_\Theta g(x \mid \theta)\, h(\theta)\, d\theta$. An experiment is the ordered triple $(X, \Theta, g(x \mid \theta)) = \xi$ and the information provided by the experiment ξ, with prior knowledge $h(\theta)$ is

$$(34) \quad I(\xi) = \iint f(x, \theta) \log \frac{f(x, \theta)}{g(x)\, h(\theta)}\, dx\, d\theta.$$

We shall not consider the many interesting and useful properties of $I(1;2)$ in (32).

We shall now give a simple illustration of the additive analysis of the information measures and its application in tests of hypotheses. It may be easily shown that

$$(35) \qquad I(X, Y, Z) = I(X, (Y, Z)) + I(Y, Z)$$

where

$$(36) \qquad I(X, Y, Z) = \int f(x, y, z) \log \frac{f(x, y, z)}{f(x) f(y) f(z)} \, d\lambda(x) \, d\mu(y) \, d\nu(z)$$

and

$$(37) \qquad I(X, (Y, Z)) = \int f(x, y, z) \log \frac{f(x, y, z)}{f(x) f(y, z)} \, d\lambda(x) \, d\mu(y) \, d\nu(z)$$

that is, the measure of the mutual independence of X, Y, and Z is analysed into a measure of the independence of X and the pair (Y, Z) and the measure of the independence of Y and Z.

For samples of size n from normal populations the estimates of the values in (35) lead to the analysis of information table 5

Table 5

Information Component	D. F.	σ^2
$-(n-1) \log (1 - r^2_{yz})$	1	3/2N
$-(n-1) \log (1 - r^2_{x.yz})$	2	8/2N
$-(n-1) \log \begin{vmatrix} 1 & r_{xy} & r_{yz} \\ r_{yx} & 1 & r_{yz} \\ r_{zx} & r_{zy} & 1 \end{vmatrix}$	3	11/2N

where the degrees of freedom are those of the noncentral chi-square distribution with noncentrality parameter λ^2 under the null hypothesis of independence.

For a three-way r x c x d contingency table the corresponding analysis becomes that in table 6.

Table 6

Information Component	D. F.
$2 \sum_j \sum_k f_{.jk} \log \frac{n f_{.jk}}{f_{.j.} f_{..k}}$	(c-1) (d-1)
$2 \sum_i \sum_j \sum_k f_{ijk} \log \frac{n f_{ijk}}{f_{i..} f_{.jk}}$	(r-1) (cd-1)
$2 \sum_i \sum_j \sum_k f_{ijk} \log \frac{n^2 f_{ijk}}{f_{i..} f_{.j.} f_{..k}}$	rcd-r-c-d+2

where f_{ijk} is the frequency of occurrence in the cell at the i-th row, j-th column, and k-th depth with $f_{.jk} = \sum_i f_{ijk}$, $f_{i..} = \sum_j \sum_k f_{ijk}$, $n = \sum_i \sum_j \sum_k f_{ijk}$ etc., and the degrees of freedom are those of the asymptotic chi-square distribution under the null hypothesis of independence.

We remark that there are many other applications of these last concepts but we limit ourselves to these two because of time. Also that more extensive tables of the noncentral chi-square distribution have recently been computed than those first published by R. A. Fisher. It may be noted that all the expressions in the contingency table analysis may be expanded into sums and differences of terms of the form 2nlog n. Tables of 2nlog n are available for n = 1 (1) 10,000 so that the arithmetic is one of table look-up addition and subtraction.

I hope that this exposition has stimulated some of you to look further into these matters as possibly useful in your areas of application.

THE CONCEPT OF MONOTONE FAILURE RATE IN RELIABILITY THEORY

Frank Proschan
Boeing Scientific Research Laboratories

1. INTRODUCTION. Most analyses of reliability problems assume that the form of the underlying failure distribution(s) is known; the parameter may be assumed either known or unknown. Popular families of failure distributions are the exponential, Weibull, gamma, normal, and lognormal.

The weakness of such analyses is that the conclusions reached may be grossly in error if the assumption as to underlying failure distributions is incorrect. In fact, the error in the original assumption may be greatly compounded in arriving at the final conclusion, especially if the conclusion concerns a tail probability or if the system analysed is complex. A good example of the pitfalls of an erroneous assumption as to the form of the underlying failure distribution is furnished by Zelen and Dannemiller (1961). They investigated the robustness of four representative acceptance sampling procedures derived from the exponential distribution when in fact the failure times followed a Weibull distribution with the same mean life. This was done by constructing the operating characteristic (O. C.) curves for these procedures when the parent distribution of failure times was the Weibull distribution with survival probability $\exp(-t^p/\Theta)$ for values of the shape parameter $p = 3/2$, 2, and 3. To illustrate how thoroughly erroneous a conclusion can be when based on an incorrect assumption as to the form of the underlying failure distribution, we reproduce the O. C. curves for a censored non-replacement plan ($n = 28$, $r = 14$) based on the exponential distribution when the parent distribution is actually Weibull with survival probability $\exp(-t^p/\theta)$. A censored non-replacement plan consists of placing n items on test simultaneously and stopping the test after the r^{th} failure.

Note that while the probability of acceptance assuming an exponential distribution with mean 500 is only .10, the corresponding probability assuming a Weibull distribution with the same mean is essentially 1. Thus if a mean life of 500 hours were unacceptable, "bad" lots would *always* be accepted under the sampling plan if the distribution were Weibull with shape parameter 3, instead of being rejected 90% of the time as they should be if the distribution were exponential with the same mean.

A further embarrassment may face the reliability analyst who attempts to make specific assumptions as to the form of the underlying failure distributions. He just may not have sufficient relevant information, since in general many observations are required to make reasonable inferences about the form of a distribution, as compared to the number required to estimate a parameter, such as the mean. As an example, we summarize the results of an investigation undertaken by Zelen and Dannemiller (1961) to see how well two statistical tests described by Epstein (1960) would distinguish between the exponential and the Weibull distributions. Consider a non-replacement censored situation where the life test starts with n items and ends as soon as the first r failures occur. Assuming failures occur at times $T_1 \leq T_2 \leq \cdots \leq T_r$, the total time on test up to the i^{th} failure is

$$T(T_i) = T_1 + \cdots + T_{i-1} + (n - i + 1)T_i, \quad i = 1, \ldots, r.$$

Then the conditional distribution of $T(T_1), T(T_2), \ldots, T(T_{r-1})$ is uniform over $[0, T(T_r)]$ for fixed $T(T_r)$. Hence

$$Z = \frac{\sum_{i=1}^{r-1} T(T_i) - \frac{r-1}{2} T(T_r)}{T(T_r)\sqrt{(r-1)/12}}$$

is an approximate normal deviate. Epstein's Test 3 calls for rejection at level α if $|Z| > Z_\alpha$ where Z_α is the $100(1 - \alpha/2)$ percentage point of the standard normal distribution.

Epstein's Test 8 uses the Bartlett statistic

$$W = 2r \frac{\ln \frac{T(T_r)}{r} - \frac{1}{r}\left[\ln T(T_1) + \ln\{T(T_2) - T(T_1)\} + \cdots + \ln\{T(T_r) - T(T_{r-1})\}\right]}{1 + \frac{r+1}{6r}}$$

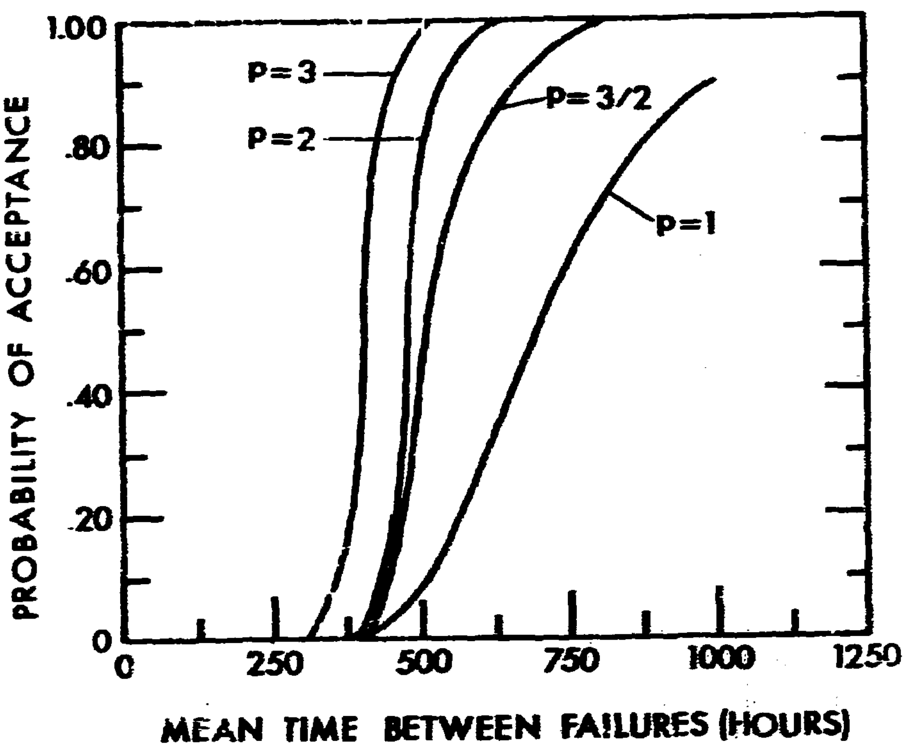


Figure 1.1 Reproduced from "The robustness of life testing procedures derived from the exponential distribution" by M. Zelen and M. C. Dannemiller, Technometrics, Vol. 3, No. 1, February, 1961.

asymptotically distributed as a χ^2 variable with $r - 1$ degrees of freedom. The test procedure is to reject at level α if $W > \chi^2_\alpha(r-1)$.

To determine the proportion of times the tests would distinguish between the exponential and Weibull distributions, Zelen and Dannemiller ran an empirical sampling study on a computer. Samples were drawn from each of the Weibull distributions with shape parameter $p = 3/2$, 2, and 3 and from the exponential distribution, all having mean one. The two tests were applied at level of significance $\alpha = .05$. The results were:

† Proportion of times exponential assumption rejected at .05 level of significance

Test	p	n = 14, r = 14	n = 25, r = 14	n = 25, r = 25
3	1	.057	.050	
	3/2	.340	.253	
	2	.837	.643	
	3	1.000	.983	
8	1	.043*	.063	.045*
	3/2	.063*	.080	.120*
	2	.178*	.137	.328*
	3	.575*	.420	.847*

*Starred values are based on 600 samples or simulations. All other values are based on 300 samples.

† Reproduced from M. Zelen and M. C. Dannemiller, "The robustness of life testing procedures derived from the exponential distribution," *Technometrics*, Vol. 3, No. 1, February 1961.

The conclusion seems clear that for small sample sizes it is difficult to distinguish between the Weibull and the exponential.

What is needed then in reliability are analyses based not on specific families of underlying failure distributions, but rather on broad assumptions concerning failure corresponding to the actual physical situation. One very natural assumption of this type is that corresponding to a failure distribution F with density f and survival probability $\overline{F} = 1-F$, the conditional failure rate

$$r(t) = \frac{f(t)}{\overline{F}(t)} \tag{1.1}$$

defined for $\overline{F}(t) > 0$ is increasing (decreasing) with time t, $0 \leq t < \infty$. Physically, $r(t)dt$ is the conditional probability of failure in $t, t+dt$, given survival until time t. We may verify by differentiation, that under the assumption of increasing (decreasing) failure rate the quantity $\ln \overline{F}(t)$ is concave (convex) on $[0, \infty)$. More generally, whether a density exists or not, we shall say a distribution has an increasing failure rate (IFR) if $\ln \overline{F}(t)$ is concave on $[0, \infty)$, and a decreasing failure rate (DFR) if $\ln \overline{F}(t)$ is convex on $[0, \infty)$.

Examples of IFR and DFR distributions are:

(i) The exponential:

$f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.$
Both IFR and DFR.

(ii) The gamma:

$f(t) = \lambda(\lambda t)^{\alpha-1} \exp(-\lambda t^{\alpha}), \; \alpha > 0, \; t \geq 0.$
IFR for $\alpha \geq 1$, DFR for $\alpha \leq 1$.

(iii) The Weibull:

$f(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^{\alpha}), \; \alpha > 0, \; t \geq 0.$
IFR for $\alpha \geq 1$, DFR for $\alpha \leq 1$.

(iv) The truncated normal:

$$f(t) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} \quad \sigma > 0, \; 0 \leq t < \infty,$$

where a is a normalizing constant.

Physically, IFR might correspond to wearout, so that the older the item gets, the greater its chance of failure. Examples are rubber tires, human beings past some initial period, and many mechanical parts which gradually wear out. Physically, DFR might correspond to work hardening, so that the older the item gets the tougher it gets, and hence the less chance it has of failing. Certain metals act in this fashion. In Proschan (1963) is discussed another commonly occurring process producing DFR distributions.

How far can we get in reliability analyses assuming IFR (DFR)? In the following sections we hope to show that a rather surprising number of useful results and methods follow from this modest and natural assumption.

2. <u>BOUNDS ON QUANTITIES OF INTEREST IN RELIABILITY THEORY ASSUMING IFR (DFR)</u>. A number of interesting and informative bounds can be obtained on quantities arising frequently in reliability problems if we assume IFR (DFR) distributions. In this section we survey some of the more simple and useful bounds. The reader is referred to Barlow and Marshall (1963a and b), Barlow and Proschan (1963), and Barlow and Proschan (1964), for details and further bounds.

Since the exponential distribution with constant failure rate is boundary distribution between IFR and DFR distributions, it provides natural bounds on the survival probability of IFR and DFR distributions.

<u>Theorem 2.1. If</u> F <u>is</u> IFR (DFR) <u>and</u> $F(\xi_p) = p$, i.e., <u>ξ_p is a p^{th} percentile, then</u>

(2.1)
$$\bar{F}(t) \begin{cases} \geq e^{-\alpha t} \\ (\leq) \end{cases} , \quad t \leq \xi_p , \qquad \bar{F}(t) \begin{cases} \leq e^{-\alpha t} \\ (\geq) \end{cases} , \quad t \geq \xi_p .$$

where

$$\alpha = -\frac{\ln(1-p)}{\xi_p}.$$

Proof. Since $\ln \bar{F}(t)$ is concave (convex) and $\bar{F}(t)$ is decreasing, then $\frac{\ln \bar{F}(t) - \ln \bar{F}(0^-)}{t - 0}$ is decreasing (increasing). Consequently

$\{\bar{F}(t)\}^{1/t}$ is decreasing (increasing). Thus $\bar{F}(t)^{1/t} \underset{(\leq)}{\geq} (1-p)^{1/\xi_p}$ for

$t \leq \xi_p$, and $\bar{F}(t)^{1/t} \underset{(\geq)}{\leq} (1-p)^{1/\xi_p}$ for $t \geq \xi_p$. ||

Thus if we know a percentile of an IFR (DFR) distribution we immediately have one-sided bounds on the survival probability corresponding to each instant of time. What if we know the mean of an IFR distribution? Then Theorem 2.2 provides a lower bound on survival probability.

Theorem 2.2. If F is IFR with mean μ_1, then

(2.2)

$$\bar{F}(t) \geq \begin{cases} e^{-t/\mu_1} & t < \mu_1 \\ 0 & t \geq \mu_1. \end{cases}$$

The inequality is sharp.

See Barlow and Marshall (1963a) or Barlow and Proschan (1964), Chapter II, for the proof. Note that the exponential distribution with mean μ_1 attains the lower bound for $t < \mu_1$ while the degenerate distribution concentrating at μ_1 attains the lower bound for $t \geq \mu_1$.

An obvious application of Theorem 2.2 is to a system consisting of n independent components in series with distributions F_i(IFR) with means $\mu_i (i = 1, 2, \ldots, n)$. Then using (2.2) the system survival probability is

$$\prod_{i=1}^{n} \bar{F}_i(t) \geq \begin{cases} \exp(-t \sum_{i=1}^{n} 1/\mu_i) & t < \min(\mu_1, \ldots, \mu_n) \\ 0 & \text{elsewhere.} \end{cases}$$

The bound is sharp. This indicates why system reliability is often better than predicted on a parts count basis. Likewise for a parallel system with component distributions F_i with means $\mu_i (i = 1, 2, \ldots, n)$,

$$\prod_{i=1}^{n} F_i(t) \leq \begin{cases} \prod_{i=1}^{n} (1 - e^{-t/\mu_i}) & t < \min(\mu_1, \ldots, \mu_n) \\ 1 & \text{elsewhere.} \end{cases}$$

In fact, the application of Theorem 2.2 can be extended to so-called monotonic structures. (See Barlow and Proschan (1964), Chapter VII, Theorem 3.1.) To define a monotonic structure we let x_i represent the state of the i^{th} component of the structure, with $x_i = 1$ if the i^{th} component is functioning and $x_i = 0$ if the i^{th} component is not functioning, $i = 1, 2, \ldots, n$. Let $\phi(x_1, x_2, \ldots, x_n)$ represent the corresponding state of the structure, with $\phi = 1$ if the structure is functioning and $\phi = 0$ if the structure is not functioning. The structure is said to be monotonic if $\phi(x_1, x_2, \ldots, x_n)$ is monotonic increasing in each argument with $\phi(0, 0, \ldots, 0) = 0$ and $\phi(1, 1, \ldots, 1) = 1$. Intuitively, a monotonic structure is one that performs at least as well if failed components are replaced by functioning components. Most reasonable structures occurring in practice will, of course, be monotonic. See Birnbaum, Esary, and Saunders (1961) for a discussion of monotonic structures (called by them "coherent").

Next suppose stochastically independent components with the reliability of the i^{th} component p_i, $i = 1, 2, \ldots, n$. Then the reliability of the system will be represented by $h(p_1, p_2, \ldots, p_n)$. In the special case $p_1 = p_2 = \ldots = p_n = p$, we will write for convenience $h(p)$. Barlow and Proschan (1964), Chapter VII, show that for monotonic structures $h(p_1, p_2, \ldots, p_n)$ is monotonic in each argument.

Using this fact we may obtain the following lower bound on system reliability for monotonic structures.

<u>Theorem 2.3.</u> <u>Let</u> F_i, IFR <u>with mean</u> μ_i, $i = 1, 2, \ldots, n$, <u>be the failure distributions of the components of a monotonic structure. Then for</u> $t < \min(\mu_1, \ldots, \mu_n)$, <u>the system reliability</u> $h(\bar{F}_1(t), \ldots, \bar{F}_n(t)) \geq h(e^{-t/\mu_1}, \ldots, e^{-t/\mu_n})$, <u>the corresponding system reliability when the components are exponentially distributed with corresponding means.</u>

<u>Proof.</u> Since $h(p_1, \ldots, p_n)$ is monotonic increasing in each argument and by Theorem 2.2 for $t < \min(\mu_1, \ldots, \mu_n)$, $\bar{F}_i(t) \geq e^{-t/\mu_i}$, $i = 1, \ldots, n$, then it follows that for $t < \min(\mu_1, \ldots, \mu_n)$, $h(\bar{F}_1(t), \ldots, \bar{F}_n(t)) \geq h(e^{-t/\mu_1}, \ldots, e^{-t/\mu_n})$. ||

We may apply Theorem 2.2 to the n-fold convolution $F^{(n)}(t)$, where F is IFR with mean μ_1, to obtain

$$F^{(n)}(t) \leq 1 - \sum_{j=0}^{n-1} \frac{(t/\mu_1)^j}{j!} \exp(-t/\mu_1)$$

for $t < \mu_1$. Therefore, if $N(t)$ is the number of renewals in $[0, t]$ of a renewal process based on F,

$$P[N(t) \geq n] \leq \sum_{j=n}^{\infty} \frac{(t/\mu_1)^j}{j!} \exp(-t/\mu_1)$$

for $t < \mu_1$. Thus we have the elementary but important result that under the IFR assumption the Poisson distribution provides a conservative estimate of the probability of n or more failures in $[0, t]$ for t less than the mean life of a single component.

This result has application in the following spare parts situation. Assume that we have one type of tube in n sockets which is replaced immediately upon failure. Let $N_j(t_j)$ denote the number of failures occurring in the j^{th} socket before time t_j, the time the j^{th} socket is to remain in operation. If the sockets are stochastically independent and the life distributions are exponential with parameter λ, then $N_1(t_1) + \ldots + N_n(t_n)$ is a Poisson random variable with parameter $\theta = \lambda \sum_{j=1}^{n} t_j$, and

$$P\left[N_1(t_1) + \ldots + N_n(t_n) \leq N\right] = \sum_{j=0}^{N} \frac{\theta^j e^{-\theta}}{j!} .$$

Cramer (1946), page 203. If all we can assume is that F is IFR with mean $1/\lambda$, and each $t_j < 1/\lambda$, then

$$P\left[N_1(t_1) + \ldots + N_n(t_n) \leq N\right] \geq \sum_{j=0}^{N} \frac{\theta^j}{j!} e^{-\theta}.$$

Using this bound, N, the number of spares to be stocked, can be chosen so that we will be protected with high probability against a shortage of spares.

The best upper bound on $\bar{F}(t)$ when F is IFR is given by the following theorem.

__Theorem 2.4.__ *If* F *is IFR with mean* μ_1 *then*

$$\overline{F}(t) \leq \begin{cases} 1 & t \leq \mu_1 \\ e^{-\omega t} & t > \mu_1, \end{cases}$$

where ω, *depending on* t, *satisfies* $1 - \omega\mu_1 = e^{-\omega t}$.

__Proof.__ Let

$$\overline{G}(x) = \begin{cases} e^{-\omega x} & x < t \\ 0 & x \geq t \end{cases}$$

Then $\ln \overline{G}(x)$ (linear on $[0,t)$) crosses $\ln \overline{F}(x)$ (concave on $[0,t)$) at most once for $0 \leq x < t$. Thus $\overline{G}(x)$ crosses $\overline{F}(x)$ at most once for $0 \leq x < t$. See Figure 2.1.

If $t > \mu_1$, we can always determine ω so that $\int_0^t e^{-\omega x} dx = \mu_1$. For this choice of ω, $\overline{F}(x)$ necessarily crosses $e^{-\omega x}$ exactly once from above for $0 \leq x < t$. Therefore $\overline{F}(t) < e^{-\omega t}$ unless F coincides identically with G. Since G is IFR, the bound is sharp.

For $t \leq \mu_1$, the inequality is obvious. The degenerate distribution concentrating at μ_1 provides the upper bound for $t \leq \mu_1$. ||

Note that as $t \to \infty$, $\omega \to 1/\mu_1$. It can be shown that the upper bound in Theorem 2.4 is asymptotic to e^{-t/μ_1}.

Figure 2.2 illustrates the best upper and lower bounds on $\overline{F}(t)$ when F is IFR and $\mu_1 = 1$. Table I tabulates the upper bound.

We can also obtain bounds on percentiles in terms of the mean and vice versa:

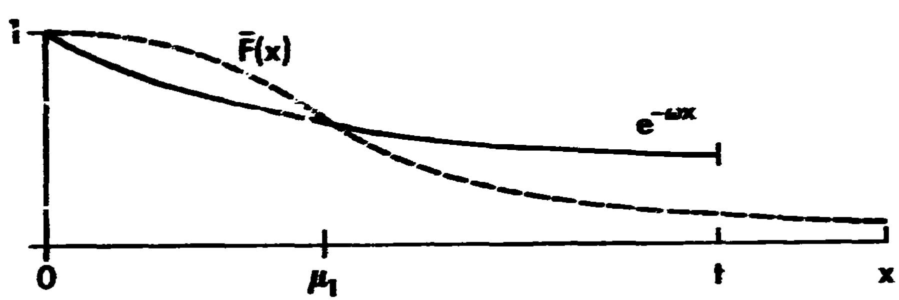


Figure 2.1

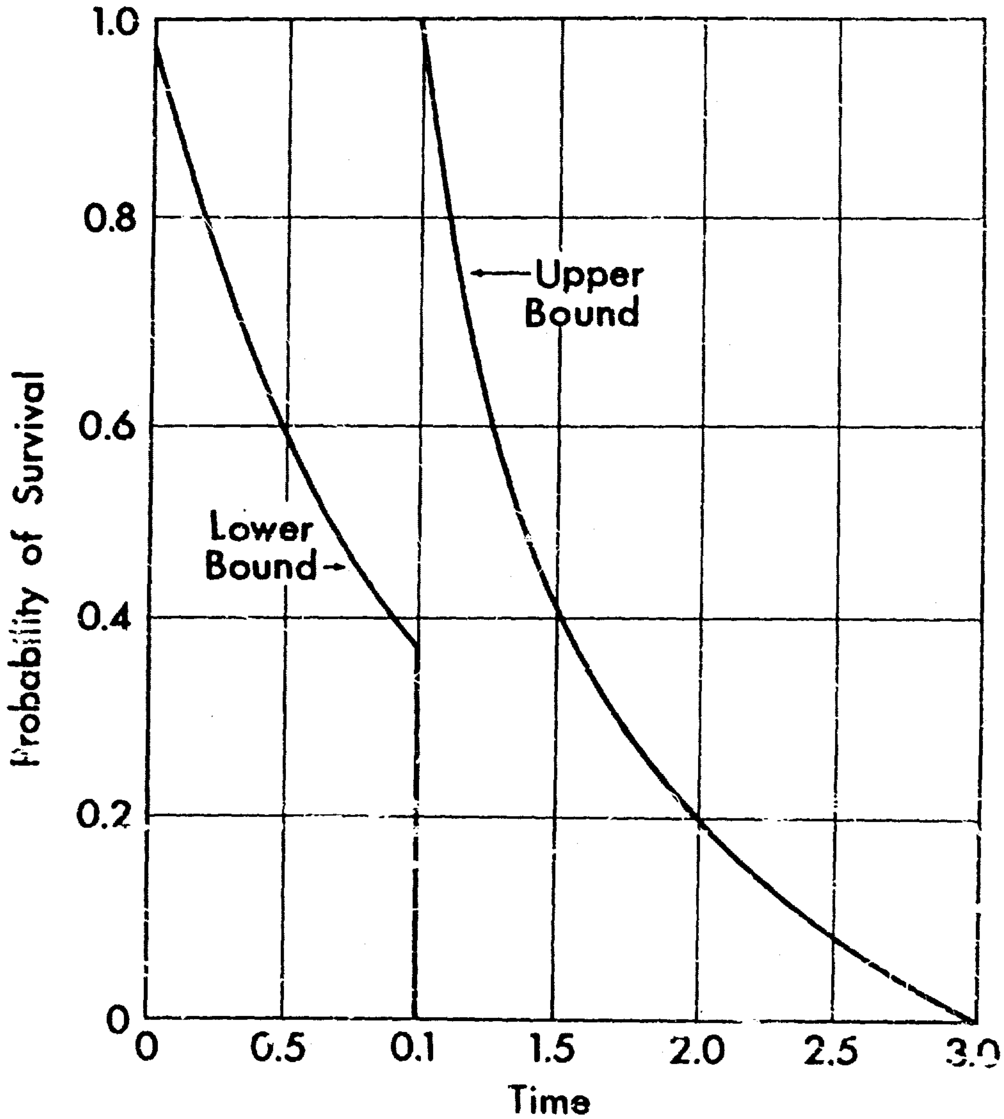


Figure 2.2 Bounds on survival probability for IFR distributions ($\mu = 1$).

Preceding Page Blank

TABLE I

UPPER BOUNDS ON $1 - F(t)$

(F is IFR, $\mu_1 = \int_0^\infty t\,dF(t) = 1$)

t	IFR Bound	Markov Bound (1/t)	t	IFR Bound	Markov Bound (1/t)
1.0	1.000	1.000			
1.1	0.820	0.909	3.1	0.053	0.323
1.2	0.637	0.833	3.2	0.047	0.313
1.3	0.577	0.769	3.3	0.042	0.303
1.4	0.488	0.714	3.4	0.038	0.294
1.5	0.416	0.667	3.5	0.034	0.286
1.6	0.358	0.625	3.6	0.030	0.278
1.7	0.308	0.588	3.7	0.027	0.270
1.8	0.268	0.555	3.8	0.024	0.263
1.9	0.233	0.526	3.9	0.022	0.256
2.0	0.203	0.500	4.0	0.020	0.250
2.1	0.178	0.476	4.1	0.018	0.244
2.2	0.156	0.455	4.2	0.016	0.238
2.3	0.138	0.435	4.3	0.014	0.233
2.4	0.121	0.417	4.4	0.013	0.227
2.5	0.107	0.400	4.5	0.012	0.222
2.6	0.095	0.385	4.6	0.011	0.217
2.7	0.084	0.370	4.7	0.010	0.213
2.8	0.075	0.357	4.8	0.009	0.208
2.9	0.067	0.345	4.9	0.008	0.204
3.0	0.059	0.[illegible]	5.0	0.007	0.200

Preceding Page Blank

Theorem 2.5. Assume F is IFR. For $p \leq 1 - e^{-1}$

$$\left[-\ln(1-p)\right]\mu_1 \leq \xi_p \leq \left[-\frac{\ln(1-p)}{p}\right]\mu_1;$$

for $p \geq 1 - e^{-1}$

$$\mu_1 \leq \xi_p < \left[-\frac{\ln(1-p)}{p}\right]\mu_1,$$

where

$$\xi_p = \sup\{t : F(t) \leq p\}.$$

The inequalities are sharp.

See Barlow and Marshall (1963a) or Barlow and Proschan (1964), Chapter II, for a proof.

In life test samples, some items may not fail at all during the course of the test. Therefore the usual sample average cannot be used. However, some percentile estimates will always be available. Note that using Theorem 2.5 we can obtain bounds on the mean of IFR distributions in terms of percentiles. For example if M is the median, then

$$\frac{M}{2\ln 2} \leq \mu_1 \leq \frac{M}{\ln 2}$$

For F DFR upper bounds on $\overline{F}(t)$ can be given in terms of the mean as shown in:

Theorem 2.6. If F is DFR with mean μ_1, then

$$\overline{F}(t) \leq \begin{cases} e^{-t/\mu_1} & t \leq \mu_1 \\ \frac{\mu_1}{t} e^{-1} & t \geq \mu_1 \end{cases}$$

Preceding Page Blank

The inequality is sharp.

See Barlow and Marshall(1963a) or Barlow and Proschan(1964), ChapterII, for a proof.

The following lemma will provide us with useful moment inequalities and further comparisons with the exponential.

Lemma 2.7. If

(a) F is IFR with mean μ_1 ... $(x) = e^{-x/\mu_1}$,

(b) $\phi(x)$ is increasing (decreasing),

then

$$\int_0^\infty \phi(x)\overline{F}(x)\,dx \underset{(\geq)}{\leq} \int_0^\infty \phi(x)\overline{G}(x)\,dx.$$

Proof. Suppose ϕ is increasing and F is not identically equal to G. Since F is IFR and G is the exponential distribution with the same mean, $\overline{F}$ crosses $\overline{G}$ exactly once from above at, say, t_0; $\overline{F}(t_0) = \overline{G}(t_0)$. Then

$$\int_0^\infty \phi(x)\overline{F}(x)\,dx - \int_0^\infty \phi(x)\overline{G}(x)\,dx = \int_0^\infty \left[\phi(x) - \phi(t_0)\right] \left[\overline{F}(x) - \overline{G}(x)\right] dx \leq 0.$$

To obtain the conclusion for ϕ decreasing, replace ϕ by $-\phi$. ||

Note that a similar lemma is true for DFR distributions with all inequalities reversed. From Lemma 2.7 we obtain an immediate comparison between the moments of an IFR distribution and the corresponding moments of an exponential distribution with the same mean.

Theorem 2.8. If F is IFR(DFR) with r^{th} moment μ_r, then

$$\mu_r \begin{cases} \underset{(\geq)}{\leq} \Gamma(r+1)\,\mu_1^r & r \geq 1 \\ \\ \underset{(\leq)}{\geq} \Gamma(r+1)\,\mu_1^r & 0 \leq r \leq 1. \end{cases}$$

Proof. Choose $\phi(x) = x^{r-1}$ in Lemma 2.7. Then

$$\mu_r = r\int_0^{\infty} x^{r-1}\overline{F}(x)\,dx \leq r\int_0^{\infty} x^{r-1}\overline{G}(x)\,dx = \Gamma(r+1)\mu_1^r ,$$

for $r \geq 1$. For $0 \leq r \leq 1$, $\phi(x) = x^{r-1}$ is decreasing, so that the inequality is reversed.

In particular, for an IFR distribution $\mu_2 \leq 2\mu_1^2$ so that the variance $\sigma^2 \leq \mu_1^2$, and so the coefficient of variation $\sigma/\mu_1 \leq 1$. The inequalities are reversed for DFR distribution(

Lemma 2. 7 can also be used to show that the mean life of a series system with IFR components whose means are $\mu_i (i = 1, 2, \ldots, n)$ exceeds the mean life of a series system with exponential components and means μ_i $(i = 1, 2, \ldots, n)$. Just the reverse is true for a parallel system.

Theorem 2. 9. If $F_i(x)$ is IFR(DFR) with mean μ_i and $\overline{G}_i(x) = e^{-x/\mu_i}$ $(i = 1, 2, \ldots, n)$, then

(a) $$\int_0^{\infty}\prod_{i=1}^{n}\overline{F}_i(x)\,dx \underset{(\leq)}{\geq} \int_0^{\infty}\prod_{i=1}^{n}\overline{G}_i(x)\,dx = 1/\sum_{i=1}^{n} 1/\mu_i ,$$

(b) $$\int_0^{\infty}\left\{1 - \prod_{i=1}^{n} F_i(x)\,dx\right\} \underset{(\geq)}{\leq} \int_0^{\infty}\left\{1 - \prod_{i=1}^{n} G_i(x)\right\} dx.$$

Proof (a) By Lemma 2. 7

$$\int_0^{\infty}\left\{\prod_{j=1}^{i-1}\overline{F}_j(x)\prod_{j=i+1}^{n}\overline{G}_j(x)\right\}\overline{F}_i(x)\,dx \underset{(\leq)}{\geq} \int_0^{\infty}\left\{\prod_{j=1}^{i-1}\overline{F}_j(x)\prod_{j=i+1}^{n}\overline{G}_j(x)\right\}\overline{G}_i(x)\,dx$$

for $1 \leq i \leq n$. By recursion we obtain

$$\int_0^\infty \prod_{j=1}^n \overline{F}_j(x)dx \underset{(\leq)}{\geq} \int_0^\infty \prod_{j=1}^n \overline{G}_j(x)dx = 1/\sum_{i=1}^n 1/\mu_i.$$

(b) The proof is similar.

Upper and lower bounds on $\overline{F}(t)$ when F is IFR and $\mu_1 = 1$, μ_2 are given have been tabulated by Barlow and Marshall (1963b). Short tables are reproduced here as Tables II and III respectively. Table IV, reproduced from Barlow and Marshall (1963b), tabulates the lower bound on $\overline{F}(t)$ when F is DFR and $\mu_1 = 1$, $\mu_2 = 2(.1)4$.

Many additional bounds have been obtained on the failure rate $r(x)$ itself and on the density $f(x)$ by Barlow and Marshall and will be presented in a forthcoming report by them.

We close this section by presenting bounds concerning renewal processes based on underlying IFR(DFR) distributions. First we give a lower bound on $M(t)$, the expected number of renewals in $[0, t)$ true for all renewal processes, and an upper (lower) bound on $M(t)$ when the underlying distribution of the renewal process is IFR(DFR). The proof may be found in Barlow and Proschan (1963).

Theorem 2.10. (i) $M(t) \geq t/\int_0^t \overline{F}(x)dx - 1 \geq t/\mu_1 - 1$.

(ii) If F is IFR(DFR), then

$$M(t) \underset{(\geq)}{\leq} tF(t)/\int_0^t \overline{F}(x)dx \underset{(\geq)}{\leq} t/\mu_1$$

for all $t \geq 0$.

TABLE II

UPPER BOUNDS ON 1 - F(t)

$$\left(F \text{ is IFR}, \ \mu_1 = \int_0^\infty t\,dF(t) = 1, \ \mu_2 = \int_0^\infty t^2 dF(t)\right)$$

t \ μ_2	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.953	0.905
0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.965	0.911	0.862	0.819
0.3	1.000	1.000	1.000	1.000	0.994	0.929	0.873	0.824	0.780	0.741
0.4	1.000	1.000	1.000	0.970	0.901	0.842	0.791	0.746	0.706	0.671
0.5	1.000	1.000	0.957	0.881	0.817	0.763	0.716	0.675	0.639	0.607
0.6	1.000	0.959	0.871	0.800	0.741	0.691	0.648	0.611	0.578	0.549
0.7	0.988	0.878	0.794	0.727	0.672	0.626	0.587	0.553	0.523	0.497
0.8	0.914	0.805	0.725	0.661	0.610	0.568	0.532	0.501	0.474	0.450
0.9	0.850	0.740	0.662	0.602	0.554	0.515	0.482	0.453	0.429	0.407
1.0	0.797	0.683	0.605	0.548	0.503	0.467	0.436	0.410	0.388	0.368
1.1	0.756	0.632	0.555	0.499	0.457	0.423	0.395	0.371	0.351	0.333
1.2	0.633	0.589	0.509	0.456	0.415	0.384	0.358	0.336	0.318	0.302
1.3	0.411	0.555	0.469	0.416	0.378	0.348	0.324	0.304	0.288	0.273
1.4	0.259	0.444	0.434	0.381	0.344	0.316	0.294	0.276	0.260	0.247
1.5	0.163	0.321	0.405	0.349	0.313	0.287	0.266	0.250	0.235	0.224
1.6	0.105	0.231	0.332	0.321	0.285	0.260	0.241	0.226	0.213	0.202
1.7	0.068	0.167	0.254	0.296	0.260	0.237	0.219	0.205	0.193	0.183
1.8	0.046	0.122	0.194	0.259	0.238	0.215	0.198	0.185	0.175	0.165
1.9	0.031	0.089	0.149	0.205	0.218	0.196	0.180	0.168	0.159	0.150
2.0	0.021	0.066	0.116	0.163	0.201	0.178	0.163	0.152	0.143	0.136
2.1	0.015	0.050	0.090	0.130	0.169	0.162	0.148	0.138	0.130	0.123
2.2	0.010	0.038	0.070	0.105	0.138	0.148	0.134	0.125	0.117	0.111
2.3	0.077	0.029	0.056	0.084	0.114	0.135	0.122	0.113	0.106	0.101
2.4	0.005	0.022	0.044	0.068	0.094	0.119	0.111	0.102	0.096	0.091
2.5	0.004	0.017	0.035	0.056	0.077	0.099	0.101	0.093	0.087	0.083
2.6	0.003	0.013	0.028	0.046	0.064	0.083	0.092	0.084	0.079	0.075
2.7	0.002	0.010	0.023	0.037	0.053	0.070	0.084	0.076	0.071	0.068
2.8	0.002	0.008	0.018	0.031	0.045	0.059	0.074	0.069	0.065	0.061
2.9	0.001	0.006	0.015	0.026	0.037	0.050	0.063	0.063	0.059	0.055
3.0	0.001	0.005	0.012	0.021	0.031	0.043	0.054	0.057	0.053	0.050

TABLE III

LOWER BOUNDS ON 1 - F(t)

$(F \text{ is IFR}, \mu_1 = \int_0^\infty t\,dF(t) = 1,\ \mu_2 = \int_0^\infty t^2\,dF(t))$

t \ μ_2	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
0.1	0.974	0.955	0.941	0.930	0.922	0.916	0.912	0.908	0.906
0.2	0.949	0.913	0.886	0.866	0.851	0.840	0.831	0.825	0.821
0.3	0.925	0.872	0.834	0.806	0.785	0.770	0.758	0.750	0.744
0.4	0.900	0.833	0.785	0.750	0.724	0.705	0.691	0.681	0.674
0.5	0.868	0.789	0.736	0.698	0.668	0.646	0.630	0.619	0.611
0.6	0.819	0.731	0.677	0.640	0.613	0.592	0.575	0.562	0.553
0.7	0.747	0.655	0.605	0.573	0.551	0.534	0.521	0.510	0.501
0.8	0.640	0.561	0.524	0.501	0.486	0.475	0.466	0.459	0.453
0.9	0.501	0.459	0.441	0.430	0.423	0.418	0.414	0.411	0.408
1.0	0.367	0.367	0.367	0.367	0.367	0.367	0.367	0.367	0.367
1.1	0.269	0.294	0.306	0.314	0.319	0.323	0.326	0.329	0.330
1.2	0.000	0.235	0.255	0.268	0.277	0.284	0.289	0.294	0.297
1.3	0.000	0.188	0.213	0.229	0.240	0.249	0.257	0.263	0.267
1.4	0.000	0.000	0.177	0.195	0.209	0.219	0.228	0.235	0.240
1.5	0.000	0.000	0.145	0.166	0.181	0.193	0.202	0.210	0.215
1.6	0.000	0.000	0.000	0.141	0.157	0.169	0.179	0.188	0.193
1.7	0.000	0.000	0.000	0.117	0.136	0.149	0.159	0.168	0.174
1.8	0.000	0.000	0.000	0.000	0.116	0.130	0.141	0.150	0.156
1.9	0.000	0.000	0.000	0.000	0.098	0.114	0.125	0.134	0.140
2.0	0.000	0.000	0.000	0.000	0.078	0.098	0.110	0.120	0.126
2.1	0.000	0.000	0.000	0.000	0.000	0.084	0.097	0.107	0.113
2.2	0.000	0.000	0.000	0.000	0.000	0.071	0.085	0.095	0.102
2.3	0.000	0.000	0.000	0.000	0.000	0.057	0.074	0.084	0.091
2.4	0.000	0.000	0.000	0.000	0.000	0.000	0.064	0.075	0.082
2.5	0.000	0.000	0.000	0.000	0.000	0.000	0.055	0.065	0.074
2.6	0.000	0.000	0.000	0.000	0.000	0.000	0.046	0.058	0.066
2.7	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.051	0.059
2.8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.053
2.9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.048
3.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.042

TABLE IV

LOWER BOUNDS ON $1 - F(t)$

$$\left(F \text{ is DFR}, \ \mu_1 = \int_0^\infty t\,dF(t) = 1, \ \mu_2 = \int_0^\infty t^2\,dF(t)\right)$$

t \ μ_2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0.0	0.952	0.909	0.869	0.833	0.800	0.769	0.740	0.714	0.689	0.666
0.1	0.861	0.822	0.787	0.754	0.724	0.696	0.671	0.647	0.625	0.604
0.2	0.779	0.744	0.712	0.683	0.656	0.631	0.608	0.586	0.566	0.548
0.3	0.705	0.673	0.644	0.618	0.594	0.571	0.551	0.531	0.514	0.497
0.4	0.638	0.609	0.583	0.559	0.538	0.517	0.499	0.482	0.466	0.451
0.5	0.577	0.551	0.528	0.506	0.487	0.469	0.452	0.437	0.422	0.409
0.6	0.522	0.499	0.478	0.458	0.441	0.425	0.410	0.396	0.383	0.371
0.7	0.473	0.452	0.432	0.415	0.399	0.385	0.371	0.359	0.347	0.336
0.8	0.428	0.409	0.391	0.376	0.362	0.348	0.336	0.325	0.315	0.305
0.9	0.387	0.370	0.354	0.340	0.327	0.316	0.305	0.295	0.286	0.277
1.0	0.350	0.335	0.321	0.308	0.297	0.286	0.276	0.267	0.259	0.251
1.1	0.317	0.303	0.290	0.279	0.269	0.259	0.250	0.242	0.235	0.228
1.2	0.287	0.274	0.263	0.253	0.243	0.235	0.227	0.220	0.213	0.207
1.3	0.259	0.248	0.238	0.229	0.220	0.213	0.206	0.199	0.193	0.188
1.4	0.235	0.224	0.215	0.207	0.200	0.193	0.186	0.181	0.175	0.170
1.5	0.212	0.203	0.195	0.187	0.181	0.175	0.169	0.164	0.159	0.155
1.6	0.192	0.184	0.176	0.170	0.164	0.158	0.153	0.149	0.144	0.140
1.7	0.174	0.166	0.160	0.154	0.148	0.143	0.139	0.135	0.131	0.127
1.8	0.157	0.150	0.144	0.139	0.134	0.130	0.126	0.122	0.119	0.116
1.9	0.142	0.136	0.131	0.126	0.122	0.118	0.114	0.111	0.108	0.105
2.0	0.129	0.123	0.118	0.114	0.110	0.107	0.104	0.101	0.098	0.095
2.1	0.116	0.111	0.107	0.103	0.100	0.097	0.094	0.091	0.089	0.087
2.2	0.105	0.101	0.097	0.094	0.090	0.088	0.085	0.083	0.081	0.079
2.3	0.095	0.091	0.088	0.085	0.082	0.079	0.077	0.075	0.073	0.071
2.4	0.086	0.083	0.079	0.077	0.074	0.072	0.070	0.068	0.066	0.065
2.5	0.078	0.075	0.072	0.069	0.067	0.065	0.063	0.062	0.060	0.059
2.6	0.070	0.068	0.065	0.063	0.061	0.059	0.057	0.056	0.055	0.053
2.7	0.064	0.061	0.059	0.057	0.055	0.054	0.052	0.051	0.050	0.048
2.8	0.058	0.055	0.053	0.052	0.050	0.049	0.047	0.046	0.045	0.044
2.9	0.052	0.050	0.048	0.047	0.045	0.044	0.043	0.042	0.041	0.040
3.0	0.047	0.045	0.044	0.042	0.041	0.040	0.039	0.038	0.037	0.036

TABLE IV (continued)

LOWER BOUNDS ON 1 - F(t)

$\left(F \text{ is DFR}, \ \mu_1 = \int_0^\infty t\,dF(t) = 1, \ \mu_2 = \int_0^\infty t^2 dF(t)\right)$

t \ μ_2	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0.0	0.645	0.625	0.606	0.588	0.571	0.555	0.540	0.526	0.512	0.500
0.1	0.585	0.567	0.550	0.534	0.519	0.504	0.491	0.478	0.466	0.455
0.2	0.531	0.514	0.499	0.485	0.471	0.458	0.446	0.435	0.424	0.414
0.3	0.481	0.467	0.453	0.440	0.428	0.417	0.406	0.396	0.386	0.377
0.4	0.437	0.424	0.411	0.400	0.389	0.379	0.369	0.360	0.351	0.343
0.5	0.396	0.385	0.374	0.363	0.354	0.344	0.336	0.327	0.320	0.312
0.6	0.360	0.349	0.339	0.330	0.321	0.313	0.305	0.298	0.291	0.284
0.7	0.326	0.317	0.308	0.300	0.292	0.285	0.278	0.271	0.265	0.259
0.8	0.296	0.288	0.280	0.272	0.265	0.259	0.253	0.247	0.241	0.236
0.9	0.269	0.261	0.254	0.248	0.241	0.235	0.230	0.224	0.219	0.215
1.0	0.244	0.237	0.231	0.225	0.219	0.214	0.209	0.204	0.200	0.195
1.1	0.222	0.215	0.210	0.204	0.199	0.195	0.190	0.186	0.182	0.178
1.2	0.201	0.196	0.191	0.186	0.181	0.177	0.173	0.169	0.166	0.162
1.3	0.183	0.178	0.173	0.169	0.165	0.161	0.157	0.154	0.151	0.148
1.4	0.166	0.161	0.157	0.154	0.150	0.147	0.143	0.140	0.137	0.135
1.5	0.151	0.147	0.143	0.140	0.136	0.133	0.130	0.128	0.125	0.123
1.6	0.137	0.133	0.130	0.127	0.124	0.121	0.119	0.116	0.114	0.112
1.7	0.124	0.121	0.118	0.115	0.113	0.110	0.108	0.106	0.104	0.102
1.8	0.113	0.110	0.107	0.105	0.103	0.100	0.098	0.096	0.095	0.093
1.9	0.102	0.100	0.098	0.095	0.093	0.091	0.090	0.088	0.086	0.085
2.0	0.093	0.091	0.089	0.087	0.085	0.083	0.082	0.080	0.079	0.077
2.1	0.084	0.082	0.081	0.079	0.077	0.076	0.074	0.073	0.072	0.070
2.2	0.077	0.075	0.073	0.072	0.070	0.069	0.068	0.066	0.065	0.064
2.3	0.070	0.068	0.067	0.065	0.064	0.063	0.062	0.060	0.059	0.058
2.4	0.063	0.062	0.061	0.059	0.058	0.057	0.056	0.055	0.054	0.053
2.5	0.057	0.056	0.055	0.054	0.053	0.052	0.051	0.050	0.049	0.048
2.6	0.052	0.051	0.050	0.049	0.048	0.047	0.046	0.046	0.045	0.044
2.7	0.047	0.046	0.045	0.045	0.044	0.043	0.042	0.042	0.041	0.040
2.8	0.043	0.042	0.041	0.041	0.040	0.039	0.038	0.038	0.037	0.037
2.9	0.039	0.038	0.038	0.037	0.036	0.036	0.035	0.034	0.034	0.033
3.0	0.036	0.035	0.034	0.034	0.033	0.032	0.032	0.031	0.031	0.030

Note that as a consequence, for an underlying IFR distribution the expected number of renewals is known within an error of 1/2 for each moment of time. Note also that in the DFR case, the inequality (ii) is an improvement over inequality (i) since

$$tF(t)/\int_0^t \overline{F}(x)\,dx \geq \int_0^t \overline{F}(x)\,dx/\int_0^t \overline{F}(x)\,dx = t/\int_0^t \overline{F}(x)\,dx - 1.$$

Next we compare moments of $N(t)$, the number of renewals in $[0, t]$ when the underlying distribution F is IFR(DFR) with the corresponding moments of a Poisson process.

Theorem 2.11. If F is IFR(DFR) with mean μ_1, then

(a) $$EN^k(t) \underset{(\geq)}{\leq} \sum_{j=0}^{\infty} \frac{j^k (t/\mu_1)^j e^{-t/\mu_1}}{j!}.$$

(b) $$\mathrm{Var}\ N(t) \underset{(\geq)}{\leq} EN(t) \underset{(\geq)}{\leq} t/\mu_1.$$

See Barlow and Proschan (1963) for a proof.

3. QUALITATIVE CONCLUSIONS ASSUMING UNDERLYING DISTRIBUTION IS IFR(DFR). In a number of reliability models, the assumption that the underlying failure distributions are IFR(DFR) yields useful qualitative relationships, and in optimization problems, helpful information about the form of the solution. In this section we shall present a selection of reliability models that illustrate how the assumption of underlying IFR(DFR) distributions simplifies the solution.

3.1 Allocation of Spares under Constraints. A system is required to operate for the period $[0, t_0]$. When a component fails, it is immediately replaced by a spare component of the same type if available. If no spare is available, system failure results. Only the spares originally provided may be used for replacements; i.e., no resupply of spares can occur during $[0, t_0]$. The cost of a single spare of the i^{th} type is c_i.

$i = 1, \ldots, k$; the total cost for spares is $c(\underline{n}) = \sum_{i=1}^{k} c_i n_i$, where $\underline{n} = (n_1, \ldots, n_k)$, and n_i = number of spares of type i initially provides, $i = 1, \ldots, k$.

We now describe the original system before any replacements are made. The system consists of d_i "positions" or "sockets" each filled by a component of type i, $i = 1, 2, \ldots, k$. The various components of a given type may be used at different levels of intensity and may be subject to different environmental stresses, so that for full generality we assume the life of the j^{th} component of type i (occupying position i, j, say) has probability distribution F_{ij}, $j = 1, \ldots, d_i$, $i = 1, \ldots, k$.

Each replacement has the same life distribition as its predecessor; component lives are mutually independent. Position i, j is not required to be in constant operation throughout $[0, t_0]$, but rather is scheduled to operate for a period of duration $t_{ij} \leq t_0$, $j = 1, \ldots, d_i$, $i = 1, \ldots, k$. This model is discussed in detail in Proschan (1960).

We seek a family of undominated spares allocations. A spares allocation $\underline{n}$ is undominated if $R(\underline{n}') > R(\underline{n})$ implies $c(\underline{n}') > c(\underline{n})$ while $R(\underline{n}') = R(\underline{n})$ implies $c(\underline{n}') \geq c(\underline{n})$, where $\underline{n}'$ is any other spares allocation, and $R(\underline{n})$ (called system reliability) is the probability of no system shutdown during $[0, t_0]$ resulting from shortage of spares. That is, a spares allocation is undominated if for its cost it achieves maximum reliability. A family of undominated allocations represents a tremendous reduction of possibilities that the decision maker needs to consider in arriving at an optimal spares allocation. We shall develop a method for generating successively larger undominated spares allocations which depends on $\ln R(\underline{n})$ being concave. We shall see below that $\ln R(\underline{n})$ is concave when each of the F_{ij} is IFR.

First note that since shortage in any position results in system failure

(3.2) $$R(\underline{n}) = \prod_{i=1}^{k} R_i(n_i) .$$

where $R_i(m)$ is the probability of no shortage of units of type i assuming m spares of type i are initially provided. We can convert the objective function $R(\underline{n})$ being maximized into a sum of terms each depending on a single unknown only by taking logarithms:

(3.3)
$$\ln R(\underline{n}) = \sum_{i=1}^{k} \ln R_i(n_i).$$

Note that maximizing $\ln R(\underline{n})$ is equivalent to maximizing $R(\underline{n})$ since $\ln x$ is a monotone increasing function of x.

Next we shall describe a procedure for generating a family of undominated allocations which is intuitively quite reasonable. The underlying idea is that we will construct successively larger spares allocations by adding one spare at a time; the spare we add will be the one which provides greatest improvement in system reliability per dollar spent.

Procedure. Start with the cheapest allocation $(0, 0, \ldots, 0)$. Obtain successively more expensive allocations as follows. If the present allocation is $\underline{n}$, determine the index, call it i_0, for which

$$\frac{1}{c_i}\left\{\ln R_i(n_i + 1) - \ln R_i(n_i)\right\}$$

is maximum over $i = 1, 2, \ldots, k$. Then add a single unit of the i_0^{th} type to $\underline{n}$ to obtain the next larger allocation $(n_1, \ldots, n_{i_0-1}, n_{i_0} + 1, n_{i_0+1}, \ldots, n_k)$.

Note that adding the most to $\ln R(\underline{n})$ per dollar spent is equivalent to multiplying $R(\underline{n})$ by the largest factor possible per dollar spent. We shall see in Theorem 3.1 below that the Procedure generates only undominated allocations if $\ln R(\underline{n})$ is concave.

Theorem 3.1. If $\ln R(\underline{n})$ is concave, then each spares allocation generated by the Procedure is undominated.

Proof. Let $\underline{n}^*$ be generated by the Procedure, i_0 denote the index of the last component type added in arriving at $\underline{n}^*$ by the Procedure, and

$$\lambda = \frac{1}{c_{i_0}}\left\{\ln R_{i_0}(n^*_{i_0}) - \ln R_{i_0}(n^*_{i_0} - 1)\right\} .$$

Let $\underline{n}$ be any other allocation such that $R(\underline{n}) > R(\underline{n}^*)$. Designate the set of indices for which $n_i > n_i^*$ by I_1 and the set of indices for which $n_i < n_i^*$ by I_2. Then

$$0 < \ln R(\underline{n}) - \ln R(\underline{n}^*) = \sum_{i \in I_1}\left\{\ln R_i(n_i) - \ln R_i(n_i^*)\right\} - \sum_{i \in I_2}\left\{\ln R_i(n_i^*) - \ln R_i(n_i)\right\}$$

$$\leq \sum_{i \in I_1} \lambda c_i (n_i - n_i^*) - \sum_{i \in I_2} \lambda c_i (n_i^* - n_i),$$

since for i in I_1 and $h = 1, 2, \ldots,$ $\ln R_i(n_i^* + h) - \ln R_i(n_i^* + h - 1) \leq \lambda c_i$, while for i in I_2 and $h = 0, 1, 2, \ldots,$ $\ln R_i(n_i^* - h) - \ln R_i(n_i^* - h - 1) \geq \lambda c_i$. These are consequences of $\ln R(\underline{n})$ being concave, i.e., of $\Delta \ln R_i(n) = \ln R_i(n+1) - \ln R_i(n)$ decreasing in n. Thus

$$0 < \sum_{i=1}^{k} c_i n_i - \sum_{i=1}^{k} c_i n_i^* .$$

In a similar fashion, assuming $R(\underline{n}) = R(\underline{n}^*)$ we may prove that

$$\sum_{i=1}^{k} c_i n_i \geq \sum_{i=1}^{k} c_i n_i^* .$$

Thus $\underline{n}^*$ is undominated. ||

Note that Theorem 3.1 requires that $\ln R(\underline{n})$ be concave. It turns out that $\ln R(\underline{n})$ is not concave for all component failure distributions F_{ij}, but as shown in Theorem 3.2 below, is concave if each F_{ij} is IFR. (As usual, the very natural IFR assumption simplifies the solution considerably.)

First we must establish the relationship between system reliability $R(\underline{n})$ and the F_{ij}. The lives of the component in position i, j and its successive replacements constitute a renewal process (as long as spares are available). Let N_{ij} = number of failures in position i, j during $[0, t_{ij}]$. Then $P[N_{ij} = n] = F_{ij}^{(n)}(t_{ij}) - F_{ij}^{(n+1)}(t_{ij})$. Also $R_i(n)$, the probability that no shortage of spares of type i occurs during $[0, t_0]$ assuming n spares of type i are stocked, is given by

$$R_i(n) = P[N_{i1} + N_{i2} + \cdots + N_{id_i} \leq n],$$

or explicitly, by

(3.4)
$$R_i(n) = \sum_{n_1 + \cdots + n_{d_i} \leq n} \prod_{j=1}^{d_i} P[N_{ij} = n_j].$$

Now we may state

<u>Theorem 3.2.</u> <u>If each F_{ij} is IFR, then $\ell n\, R(\underline{n})$ is concave.</u> The proof may be found in Barlow and Proschan (1964), Chapter VI.

<u>Exponential Failure Distributions.</u> Suppose each failure density is exponential; specifically, suppose

$$f_{ij}(t) = \begin{cases} \dfrac{1}{\mu_{ij}} e^{-t/\mu_{ij}} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases}$$

It follows that

$$P[N_{ij} = n] = e^{-t_{ij}/\mu_{ij}} \left(\frac{t_{ij}}{\mu_{ij}}\right)^n / n!\,,$$

a Poisson frequency function with parameter $\dfrac{t_{ij}}{\mu_{ij}}$ (Arrow, Karlin, Scarf (1958), page 272). Hence

(3.5) $$R_i(n) = e^{-\mu_i} \sum_{j=0}^{n} \mu_i^j / j!\,,$$

where

$$\mu_i = \sum_{j=1}^{d_i} \frac{t_{ij}}{\mu_{ij}} \quad ,$$

since the convolution of Poisson frequency functions is a Poisson frequency function with parameter given by the sum of the separate parameters (Cramér (1946), page 205.).

Using (3.5), it is a relatively simple matter to apply the Procedure above for generating undominated allocations in the present case of underlying exponential failure densities. See Proschan (1960) or Barlow and Proschan (1964), Chapter VI, for worked examples.

3.2 Comparison of Age and Block Replacement. Among the most useful replacement policies currently in popular use are the age replacement policy and the block replacement policy. Under an age replacement policy a unit is replaced upon failure or at age T, a specified positive constant, whichever comes first. Under a block replacement policy a unit is replaced upon failure and at times $T, 2T, 3T, \ldots$. We assume that units fail permanently, independently, and that the time required to perform replacement is negligibly small. Block replacement is easier to administer since the planned replacements occur at regular intervals and so are readily scheduled. This type of policy is commonly used with digital computers and other complex electronic systems. On the other hand, age replacement seems more flexible since under this policy planned replacement takes into account the age of the unit. It is therefore of some interest to compare these two policies with respect to the number of failures, number of planned replacements, and number of removals. ("Removal" refers to both failure replacement and planned replacement.)

Block replacement policies have been investigated by E. L. Welker (1959), R. F. Drenick (1960), and B. J. Flehinger (1962). Age replacement policies have been studied by G. Weiss (1956) and Barlow and Proschan (1962) among others. The results presented below are based on Barlow and Proschan (1963).

We shall compare block replacement with age replacement, both using replacement interval T. For example, block replacement is more wasteful since, as we shall show, more unfailed components are removed than under a policy based on age. Likewise, the total number of removals for both failed and unfailed components is greater. However, as one would suspect, under the IFR assumption the expected number of failures will be less under block replacement. Finally, exactly $[t/T]$ planned replacements will be made in $[0, T]$ under block replacement, while no more than $[t/T]$ can be made under age replacement.

We shall denote the number of removals in $[0, T]$ under a block policy by $N_B(t)$ and the number of removals in $[0, T]$ under an age policy by $N_A(t)$. As we show in Theorem 3.3, $N_B(t)$ is stochastically larger than $N_A(t)$.

<u>Theorem 3.3.</u> $P[N_A(t) \geq n] \leq P[N_B(t) \geq n]$ <u>for</u> $n = 0, 1, 2, \ldots$.

<u>Proof.</u> Let $\{X_k\}_{k=1}^{\infty}$ denote a realization of the lives of successive components. We shall compute what would have occurred under an age and under a block replacement policy. Let $T_A^n (T_B^n)$ denote the time of the n^{th} removal under an age (block) replacement policy. Then

$$T_A^n = \min (T_A^{n-1} + T, T_A^{n-1} + X_n)$$

$$T_B^n = \min (T_B^{n-1} + \alpha, T_B^{n-1} + X_n)$$

where $\alpha (0 \leq \alpha \leq T)$ is the remaining life to a scheduled replacement. Since initially $T_A^1 = T_B^1$, we have by induction $T_A^n \geq T_B^n$. Thus for any realization $\{X_k\}$, $N_A(t)$ is smaller than $N_B(t)$. ||

Next we shall use Theorem 3.3 to establish a lower bound on the renewal function quite independent of replacement. Let $N_A^*(t)$ ($N_B^*(t)$) denote the number of failures in $[0, t]$ under age (block) replacement at interval T. Theorem 3.4 below shows that the number of failures per unit of time under block replacement at interval T is, in the limit, $\frac{M(T)}{T}$, where $M(T)$ is the renewal function $\sum_{k=1}^{\infty} F^{(k)}(T)$.

(3.6) **Theorem 3.4.** $\lim_{t\to\infty} \frac{N_B^*(t)}{t} = \frac{M(T)}{T}$ a.e. .

Proof. Let $N_{Bi}^*(t)$ denote the number of failures in $[(i-1)T, iT]$. Clearly the random variables $N_{Bi}^*(T)$ are independent, identically distributed, and for $kT \leq t < (k+1)T$,

(3.7)
$$\sum_{1}^{k} \frac{N_{Bi}^*(T)}{kT\,\frac{k+1}{k}} \leq \frac{N_B^*(t)}{t} \leq \sum_{1}^{k+1} \frac{N_{Bi}^*(T)}{(k+1)T\,\frac{k}{k+1}} .$$

Letting $t\to\infty$, we have $k\to\infty$, and

$$\lim_{t\to\infty} \frac{N_B^*(t)}{t} = \frac{M(T)}{T}$$

by the strong law of large numbers. ||

From (3.7) we see that also

(3.8)
$$\lim_{t\to\infty} \frac{EN_B^*(t)}{t} = \frac{M(T)}{T} .$$

Since the number of removals with block replacement is stochastically greater than the number of removals with age replacement by Theorem 3.3, we see that

(3.9) $$\lim_{t\to\infty} \frac{EN_B(t)}{t} \geq \lim_{t\to\infty} \frac{EN_A(t)}{t} .$$

But using (3.8)

$$\lim_{t\to\infty} \frac{EN_B(t)}{t} = \frac{M(T)}{T} + \frac{1}{T} ,$$

and

$$\lim_{t \to \infty} \frac{EN_A(t)}{t} = \frac{1}{\int_0^T \bar{F}(x)dx}$$

by the elementary renewal theorem, since for age replacement the times between removals constitute a renewal process. Substituting in (3. 9) we have

$$\frac{M(T)}{T} + \frac{1}{T} \geq \frac{1}{\int_0^T \bar{F}(x)dx}$$

or

(3. 10)
$$M(T) \geq \frac{T}{\int_0^T \bar{F}(x)\,dx} - 1.$$

We thus have the following lower bound on the renewal function, regardless of the underlying distribution.

<u>Theorem 3. 5.</u> $M(t) \geq \frac{t}{\int_0^T \bar{F}(x)\,dx} - 1 \geq \frac{t}{\mu} - 1$ <u>for</u> $0 \leq t \leq \infty$. (3. 11)

<u>Proof.</u> The first inequality has already been established in (3. 10). The second follows from $\int_0^t \bar{F}(x)\,dx \leq \int_0^\infty \bar{F}(x)dx = \mu$. ||

Next we shall compare the number of failures under the two policies and as a consequence obtain an upper bound on the renewal function valid when the underlying failure distribution is IFR. First we obtain the long run average time between failures under a block replacement policy.

Let $\{Y_i\}$ denote the successive times between failures under a block replacement policy having replacement interval T.

<u>Theorem 3.6.</u> $\lim_{t\to\infty} \frac{Y_1+Y_2+\ldots+Y_{N_B^*(t)}}{N_B^*(t)} = \frac{T}{M(T)}$ a.s.

<u>Proof</u>. Note that

$$\frac{Y_1+Y_2+\ldots+Y_{N_B^*(t)}}{N_B^*(t)} \leq \frac{t}{N_B^*(t)} \leq \frac{Y_1+Y_2+\ldots+Y_{N_B^*(t)+1}}{\frac{N_B^*(t)}{N_B^*(t)+1}\{N_B^*(t)+1\}} .$$

Letting $t\to\infty$ and applying Theorem 3.4 we have the desired result. ||

Theorem 3.7 below shows that the number of failures under an age replacement policy is stochastically greater than the number of failures under a block replacement policy.

Theorem 3.7. <u>If</u> F <u>is</u> IFR, <u>then</u>

$$P[N_A^*(t) \geq n] \geq P[N_B^*(t) \geq n].$$

The proof may be found in Barlow and Proschan (1963).

Using Theorem 3.7 we obtain the following upper bounds for the renewal function $M(t)$ when the underlying failure distribution is IFR.

Theorem 3.8. <u>If</u> F <u>is</u> IFR,

$$M(t) \leq \frac{tF(t)}{\int_0^t \overline{F}(x)\,dx} \leq \frac{t}{\mu} \quad \text{for } 0 \leq t < \infty .$$

Proof. By Theorem 3.7

(3.12)
$$\lim_{t\to\infty} \frac{EN_B^*(t)}{t} \leq \lim_{t\to\infty} \frac{EN_A^*(t)}{t} .$$

By (3.8)

(3.13)
$$\lim_{t\to\infty} \frac{EN_B^*(t)}{t} = \frac{M(T)}{T} .$$

It is readily verified that the mean time to an in-service failure under an age replacement policy with replacement interval T is

$$\frac{1}{F(T)} \int_0^T \overline{F}(x)dx .$$

(See Barlow and Proschan (1964), Chapter III.) It follows by the elementary renewal theorem that

(3.14)
$$\lim_{t\to\infty} \frac{EN_A^*(t)}{t} = \frac{F(T)}{\int_0^T \overline{F}(x)dx} .$$

Combining (3.12), (3.13), and (3.14), we conclude

$$M(t) \leq \frac{tF(t)}{\int_0^t \overline{F}(x)dx}$$

the first of the desired conclusions. For F IFR, it is readily verified that the mean time to an in-service failure under an age replacement policy with replacement interval T, $\frac{1}{F(T)} \int_0^T \overline{F}(x)dx$, is a decreasing function of T. (See Barlow and Proschan (1964), Chapter III.) Hence

$$\frac{tF(t)}{\int_0^t \overline{F}(x)\,dx} \leq \frac{tF(\infty)}{\int_0^t \overline{F}(x)\,dx} = \frac{t}{\mu},$$

the second of the desired conclusions. ||

Combining Theorems 3.5 and 3.8, we obtain very close bounds on M(t) when the underlying failure distribution is IFR:

(3.15)
$$\frac{t}{\mu} - 1 \leq M(t) \leq \frac{t}{\mu}, \quad 0 \leq t < \infty.$$

Thus for all non-negative values of t, $M(t)$ may be approximated with an error of at most 1/2.

4. PRESERVATION OF MONOTONE FAILURE RATE. Next we consider operations under which a monotone failure rate is preserved. For example, what structures have the monotone failure rate property when their individual components have this property? Is a monotone failure rate preserved under convolution or under mixture of distributions? The results presented in this section are based on Esary and Proschan (1963) and Barlow, Marshall, and Proschan (1963).

Theorem 4.1. If F_1 and F_2 are IFR, then their convolution H, given by

$$H(t) = \int_{-\infty}^{\infty} F_1(t - x)\,dF_2(x),$$

is also IFR.

Proof. Assume F_1 has density f_1, F_2 has density f_2. For $t_1 < t_2$, $u_1 < u_2$ form

$$D = \left| H(t_i - u_j) \right|_{i,j=1,2} = \left| \int \overline{F}_1(t_i - s) f_2(s - u_j)\, ds \right| = \int_{s_1} \int_{< s_2} \left| \overline{F}_1(t_i - s_k) \right| \left| f_2(s_k - u_j) \right| ds_2\, ds_1$$

by problem 66, page 48, Pólya and Szegö (1925). Integrating the inner integral by parts, we obtain

$$D = \iint_{s_1 < s_2} \begin{vmatrix} \overline{F}_1(t_1 - s_1) & f_1(t_1 - s_2) \\ \overline{F}_1(t_2 - s_1) & f_1(t_2 - s_2) \end{vmatrix} \begin{vmatrix} f_2(s_1 - u_1) & f_1(s_1 - u_2) \\ \overline{F}_2(s_2 - u_1) & \overline{F}_2(s_2 - u_2) \end{vmatrix} ds_2\, ds_1 \, .$$

The sign of the first determinant is the same as that of

$$\frac{f_1(t_2 - s_2)\, \overline{F}_1(t_2 - s_2)}{\overline{F}_1(t_2 - s_2)\, \overline{F}_1(t_1 - s_1)} - \frac{f_1(t_1 - s_2)\, \overline{F}_1(t_1 - s_2)}{\overline{F}_1(t_1 - s_2)\, \overline{F}_1(t_1 - s_1)}$$

assuming non-zero denominators. But

$$\frac{f_1(t_2 - s_2)}{\overline{F}_1(t_2 - s_2)} \geq \frac{f_1(t_1 - s_2)}{\overline{F}_1(t_1 - s_2)}$$

by hypothesis, while

$$\frac{\overline{F}_1(t_2 - s_2)}{\overline{F}_1(t_2 - s_1)} \geq \frac{\overline{F}_1(t_1 - s_2)}{\overline{F}_1(t_1 - s_1)}$$

since, as pointed out in Section 1, $\overline{F}_1$ is logarithmically concave. Thus the first determinant is non-negative. A similar argument holds for the second determinant, so that $D \geq 0$. But this implies $\overline{H}$ is logarithmically concave, and therefore H is IFR.

If F and/or G do not have densities, the theorem may be proved in a similar fashion using limiting arguments. ||

It is of interest to note that the DFR property is not preserved under convolution. A counterexample is obtained if we convolute densities

$$f_1(x) = f_2(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \qquad \text{for } x \geq 0 ,$$

with $1/2 < \alpha < 1$. However, it is true that a mixture of DFR distributions is also DFR, as shown in

<u>Theorem 4.2</u>. <u>If</u> $F_i(t)$ <u>is a</u> DFR <u>distribution in</u> t <u>for each</u> $i = 1, 2, \ldots, a_i \geq 0, i = 1, 2, \ldots,$ <u>and</u> $\sum_{i=1}^{\infty} a_i = 1$, <u>then</u>

$$G(t) = \sum_{i=1}^{\infty} a_i F_i(t)$$

<u>is a</u> DFR <u>distribution</u>.

<u>Proof</u>. First suppose that $F_i(t)$ has a differentiable density $f_i(t)$ Since the density of any DFR distribution must be a decreasing function, we have by Schwarz's inequality that

$$\sum_{i=1}^{\infty} a_i \overline{F}_i(t) \sum_{i=1}^{\infty} \left[-a_i f_i'(t) \right] \geq \left\{ \sum_{i=1}^{\infty} a_i \left[\overline{F}_i(t) \right]^{1/2} \left[-f_i'(t) \right]^{1/2} \right\}^2 .$$

Since $f_i(t)/\overline{F}_i(t)$ is decreasing in t, we must have

$$\overline{F}_i(t) f_i'(t) \leq -\left[f_i(t) \right]^2 .$$

Hence,

$$\sum_{i=1}^{\infty} a_i \bar{F}_i(t) \sum_{i=1}^{\infty}\left[-a_i f_i'(t)\right] \geq \left[\sum_{i=1}^{\infty} a_i f_i(t)\right]^2 ,$$

that is

$$\bar{G}(t)\, g'(t) \leq -\left[g(t)\right]^2$$

where g is the density of G, so that G is DFR.

If each F_i does not have a differentiable density, the same result may be obtained by limiting arguments. ||

Mixtures of IFR distributions are not necessarily IFR. For example, a mixture of two distinct exponentials is not IFR since it is not exponential, and by the above theorem it is DFR. Theorem 4.2 together with the test for DFR distributions in Section 5 may be used to pick up differences in the parameters of pooled samples each coming from an exponentially distributed population. See Proschan (1963).

Theorem 4.1 proves that a system consisting of one unit and a spare has increasing failure rate if the components do. As one would expect, the failure rate of the system is everywhere less than the failure rate of either component if both components have IFR failure distributions.

__Theorem 4.3.__ _If_ F_1 _and_ F_2 _are IFR distributions with failure rates_ $r_1(t)$ _and_ $r_2(t)$ _respectively and_ F _denotes their convolution with failure rate_ $r(t)$, _then_

$$r(t) \leq \min\left[r_1(t),\ r_2(t)\right].$$

__Proof.__ By definition

$$r(t) = \frac{\int_0^t f_1(t-x)\, f_2(x)\,dx}{\bar{F}(t)} .$$

Thus

$$r(t) \le r_1(t)\ \frac{\int_0^t \overline{F}_1(t-x) f_2(x)\,dx}{\overline{F}(t)} = r_1(t) \quad ,$$

the equality being clear from

$$\overline{F}(t) = \int_0^\infty \overline{F}_1(t-x)\,dF_2(x) \quad .$$

Similarly, $r(t) \le r_2(t)$. ||

Next we shall obtain conditions under which an increasing failure rate of like components implies an increasing failure rate for the system. Assume then that the system consists of independent like components, with each component life distributed according to the common probability distribution $F(t)$. At a given instant of time t each component has reliability $p = \overline{F}(t)$; the corresponding system reliability will be designated by $h(p)$. Then we may prove

Theorem 4.4. Assume a structure with reliability function $h(p)$, with each compo[nen]t life independently distributed according to distribution F having dens[it]y f. Then

(a) $$\frac{R(t)}{r(t)} = \frac{p h'(p)}{h(p)}\Bigg|_{p = \overline{F}(t)}$$

where $r(t) = \frac{f(t)}{\overline{F}(t)}$ = component failure rate at time t, and $R(t)$ = system failure rate at time t;

(b) $\frac{R(t)}{r(t)}$ is an increasing function of t if and only if $\frac{p h'(p)}{h(p)}$ is a decreasing function of p;

(c) if $r(t)$ is an increasing function of t and $\frac{p h'(p)}{h(p)}$ is a decreasing function of p, then $R(t)$ is an increasing function of t.

Result (c) gives a simple sufficient condition on a structure which will preserve a monotone failure rate when a structure is constructed out of independent like components. We shall present an important class of structures which satisfy this sufficient condition.

To prove (a), let $S(t)$ represent the probability of structure survival past time t; i.e., $S(t) = h(\overline{F}(t))$. By definition

$$R(t) = \frac{-S'(t)}{S(t)} = \frac{h'(p)}{h(p)}\Bigg|_{p=\overline{F}(t)} \cdot f(t) = \frac{ph'(p)}{h(p)}\Bigg|_{p=\overline{F}(t)} \cdot \frac{f(t)}{\overline{F}(t)} ,$$

so that

$$\frac{R(t)}{r(t)} = \frac{ph'(p)}{h(p)}\Bigg|_{p=\overline{F}(t)} ,$$

establishing (a).

To prove (b), simply note that $p = \overline{F}(t)$ is a decreasing function of t.

Finally, (c) is an immediate consequence of (b) and the fact that p is a decreasing function of t. ||

An important class of structures for which the condition $\frac{ph'(p)}{h(p)}$ is a decreasing function of p are the so-called k out of n structures. A k out of n structure is one that functions if and only if at least k components function. To prove that a k out of n structure consisting of n independent components has a ratio $\frac{ph'(p)}{h(p)}$ decreasing in p, write

$$\frac{h(p)}{ph'(p)} = \frac{1}{p}\int_0^p \left(\frac{t}{p}\right)^{k-1} \left(\frac{1-t}{1-p}\right)^{n-k} dt ,$$

since

$$h(p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(k-1)!(n-k)!} \int_0^p x^{k-1}(1-x)^{n-k} dx ,$$

Mood (1950), page 235. Letting $u = \frac{t}{p}$, we have

$$\frac{h(p)}{ph'(p)} = \int_0^1 u^{k-1} \left(\frac{1-up}{1-p}\right)^{n-k} du .$$

Since $\frac{1-up}{1-p}$ is increasing in p, so is $\frac{h(p)}{ph'(p)}$. Thus if a k out of n structure is composed of independent like components having an increasing failure rate, then the structure itself has an increasing failure rate.

If we note that the time of failure of a k out of n system corresponds to the k^{th} largest in a sample of n observations, then an alternate statement of this result is the following:

Corollary. Suppose $X_1 < X_2 < \ldots < X_n$ are a sample of order statistics based on independent observations from a distribution having increasing failure rate. Then the distribution of X_i has an increasing failure rate, $i = 1, 2, \ldots, n$.

Actually we can generate new structures which have the property that $\frac{ph'(p)}{h(p)}$ is a decreasing function, by composition of structures having this property. Under composition we form a superstructure each element of which consists of copies of a given structure. If $h = f(g)$ with $g'(p) \geq 0$, then since

$$\frac{ph'(p)}{h(p)} = \frac{gf'(g)}{f(g)} \cdot \frac{pg'(p)}{g(p)} ,$$

the property is closed under composition.

The following example shows that a structure consisting of independent like components each having an exponential failure distribution need not have an increasing failure distribution.

Example. Consider a structure composed of two substructures in parallel, the first having k components in series, the second consisting of a single component. Assuming independent components each having exponential distribution for failure

$$F(t) = 1 - e^{-t} \quad ,$$

we compute the probability $S(t)$ of structure survival past time t to be

$$S(t) = 1 - (1-e^{-t})(1-e^{-kt}) \; .$$

Thus

$$R(t) = -\frac{S'(t)}{S(t)} = \frac{e^{-t} + ke^{-kt} - (k+1)e^{-(k+1)t}}{e^{-t} + e^{-kt} - e^{-(k+1)t}} \quad ,$$

and so

$$\text{sgn } R'(t) = \text{sgn } \left[-(k-1)^2 + k^2 e^{-t} + e^{-kt} \right] \; .$$

Note that for $k > 1$, for $t = 0$ the sign is positive, while for $t = \infty$ the sign is negative. Thus the structure failure rate $R(t)$ is not monotonic for $k > 1$.

This is also a counterexample to the conjecture that k out of n structures with unlike IFR components are themselves necessarily IFR. (Simply consider the series substructure as a single component.)

5. STATISTICAL TEST FOR MONOTONE FAILURE RATE. We have seen the important role that distributions with monotone failure rate play in reliability theory. Therefore it would be of great value to have a test to determine whether a sample (or set of samples) comes from a population with monotone failure rate. The test presented in this section is based on Proschan and Pyke.

Let $X_1, X_2, \ldots, X_n$ be a sample of independent observations from the common distribution F with density f, where $f(t) = 0$ for $t < 0$, and failure rate $r(t)$. We wish to choose between the following:

Null Hypothesis, H_0: r is constant.

Alternative Hypothesis, H_1: r is non-decreasing but not constant.

The test statistic is computed as follows. Let $T_1 < T_2 < \ldots < T_n$ be the ordered observations, $D_1 = T_1$, $D_2 = T_2 - T_1$, . . . , $D_n = T_n - T_{n-1}$ the spacings, and $\bar{D}_1 = nD_1$, $\bar{D}_2 = (n-1)D_2$, . . . , $\bar{D}_n = D_n$, the normalized spacings. For i, j = 1, 2, . . ., n let $V_{ij} = 1$ if $\bar{D}_i > \bar{D}_j$, 0 otherwise. The test statistic is

$$V_n = \sum_{\substack{i;\, j=1 \\ i<j}}^{n} V_{ij} \quad .$$

We reject the null hypothesis at the α level of significance if $V_n > v_{n,\alpha}$ where $v_{n,\alpha}$ is determined such that $P\left[V_n > v_{n,\alpha} \mid H_0\right] = \alpha$. (It is obvious how to modify the V_n test if the alternative hypothesis in that r is non-increasing rather than non-decreasing.)

Heuristically we may justify the test as follows. Under the null hypothesis it may be readily verified that $\bar{D}_1$, $\bar{D}_2$, . . . , $\bar{D}_n$ are independent exponential random variables with common parameter, say λ. Hence $P\left[V_{ij} = 1\right] = 1/2$ for i, j = 1, 2, . . . , n, $i \neq j$. However, as shown below, under the alternative hypothesis, $P\left[V_{ij} = 1\right] > 1/2$ for $i < j$, i, j = 1, 2, . . . , n. In fact, each V_{ij} and V_n tend to be larger under the alternative hypothesis, so that rejection of the null hypothesis occurs for large values of V_n. Since under the null hypothesis the distribution of V_n is known, we have available $v_{n,\alpha}$.

<u>Distribution under the Null Hypothesis.</u> Since under H_0, $\bar{D}_1$, $\bar{D}_2$, . . . $\bar{D}_n$ are independently distributed, each having density $\lambda e^{-\lambda t}$, all orderings of $\bar{D}_1$, $\bar{D}_2$, . . . , $\bar{D}_n$ are equally likely. Using this property, Kendall (1938) provides tables for $P\left[V_n \leq k \mid H_0\right]$, $n \leq 10$; more convenient tables are available in Mann(1945). Mann shows that μ_n and σ_n^2, the mean and variance of V_n under H_0, are given by

$$\mu_n = \frac{n(n-1)}{4} \qquad \sigma_n^2 = \frac{(2n+5)(n-1)n}{72},$$

and that V_n is asymptotically normal under H_0.

Unbiasedness of Test. We now show that V_n is unbiased, i.e., that

$$P\left[V_n \geq v_{n,\alpha} \mid H_1\right] \geq \alpha \quad \text{for } 0 < \alpha \leq 1, \ n = 2, 3, \ldots .$$

Make the transformation

$$X_i' = -\ln \bar{F}(X_i) .$$

It follows that

$$P\left[X_i' > u\right] = e^{-u}.$$

Thus each X_i' is distributed according to the exponential distribution with unit mean. Moreover, since the $X_1, \ldots, X_n$ are independent, so are the $X_1', \ldots, X_n'$. Next let $T_1' < T_2' < \ldots < T_n'$ represent the ranked $X_1', X_2', \ldots, X_n'$ so that $T_i' = -\ln \bar{F}(T_i)$, $i = 1, 2, \ldots, n$. Further, let

$$D_i' = (n-i+1)(T_i' - T_{i-1}'), \quad i = 1, 2, \ldots, n,$$

where $T_0' = 0$ by definition. It is easy to verify that the D_i' are independently, identically distributed according to the exponential with unit mean. See for example Epstein and Sobel (1953) or Rényi (1953).

Note that T_i' is an increasing function of T_i. Moreover T_i' is a convex function of T_i as shown in Theorem 4.1, Chapter II of

Barlow and Proschan (1964). It follows that $\bar{D}'_i \geq \bar{D}'_j$ implies $\bar{D}_i \geq \bar{D}_j$ for $i < j$. Thus $V_{ij} \geq V'_{ij}$ where $V'_{ij} = 1$ if $\bar{D}_i \geq \bar{D}_j$. Hence $V_n \geq V'_n$, where $V'_n = \sum_{i<j} V'_{ij}$, so that $P[V_n \geq v_{n,\alpha} | H_1] \geq \alpha$ for $0 < \alpha \leq 1$, $n = 2, 3, \ldots$, implying that V_n is unbiased, as claimed.

Asymptotic Distribution under the Alternative Hypothesis. Under the alternative hypothesis, the asymptotic distribution of V_n is normal under mild restrictions. It follows that V_n is a consistent test.

The asymptotic relative efficiency of the V_n test compared with various possible competing tests is studied in Proschan and Pyke. In general, the V_n test compares quite favorably.

BIBLIOGRAPHY

1. K. J. Arrow, S. Karlin, and H. Scarf (1958), Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, California.

2. R. E. Barlow and A. W. Marshall (1963a), "Bounds for distributions with monotone hazard rate," Boeing Scientific Research Laboratories D1-82-0247.

3. R. E. Barlow and A. W. Marshall (1963b), "Tables of bounds for distributions with monotone hazard rate," Boeing Scientific Research Laboratories D1-82-0249.

4. R. E. Barlow, A. W. Marshall, and F. Proschan (1963), "Properties of probability distributions with monotone hazard rate," Annals of Mathematical Statistics, Vol. 34, No. 2, pp. 375-389.

5. R. E. Barlow and F. Proschan (1962), "Planned replacement," appearing in Studies in Applied Probability and Management Science, edited by K. J. Arrow, S. Karlin, and H. Scarf, Stanford University Press, pp. 63-87.

6. R. E. Barlow and F. Proschan (1963), "Comparison of replacement policies, and renewal theory implications," Boeing Scientific Research Laboratories D1-82-0237, to appear in the Annals of Mathematical Statistics.

7. R. E. Barlow and F. Proschan (1964), Mathematical Theory of Reliability, John Wiley and Sons, New York, in preparation.

8. Z. W. Birnbaum, J. D. Esary, and S. C. Saunders (1961), "Multi-component systems and structures and their reliability," Technometrics, Vol. 3, No. 1, pp. 55-77.

9. H. Cramer (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton.

10. R. F. Drenick (1960), "Mathematical aspects of the reliability problem," Journal of the Society for Industrial and Applied Mathematics, Vol. 8, pages 125-149.

11. B. Epstein (1960), "Testing for the validity of the assumption that the underlying distribution of life is exponential, Part I," Technometrics, Vol. 2, pp. 83-101.

12. B. Epstein and M. Sobel (1953), "Life testing," Journal of the American Statistical Association, Vol. 48, No. 263, pp. 486-502.

13. J. D. Esary and F. Proschan (1963), "Relationship between system failure rate and component failure rates, "Technometrics, Vol. 5, No. 2, pp. 183-189.

14. B. J. Flehinger (1962), "A general model for the reliability analysis of systems under various preventive maintenance policies," Annals of Mathematical Statistics," Vol. 33, pp. 137-156.

15. M. G. Kendall (1938), "A new measure of rank correlation," Biometrika, Vol. 30, pp. 81-93.

16. H. B. Mann (1945). "Nonparametric tests against trend," Econometrica, Vol. 13, pp. 245-259.

17. A. W. Mood (1950), Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc.

18. G. Polya and G. Szego (1925), Aufgaben und Lehrsätze aus der Analysis, Vol. 1, Springer-Verlag, Berlin.

19. F. Proschan (1960), "Optimal system supply," Naval Research Logistics Quarterly, Vol. 7, No. 4, pp. 609-646.

20. F. Proschan (1963), "Theoretical explanation of observed decreasing failure rate," Technometrics, Vol. 5, No. 3, pp. 375-383.

21. F. Proschan and R. Pyke, "Tests for monotone failure rate," in preparation.

22. A. Renyi (1953), "On the theory of order statistics," Acta Math. Acad. Science Hungar., Vol. 4, pp. 191-231.

23. G. Weiss (1956), "On the theory of replacement of machinery with a random failure time," Naval Research Logistics Quarterly, Vol. 3, pp. 279-293.

24. E. L. Welker (1959), "Relationship between equipment reliability, preventive maintenance policy, and operating costs," ARINC Monograph No. 7, Aeronautical Radio Inc., Washington, D. C.

25. M. Zelen and M. C. Dannemiller (1961), "The robustness of life testing procedures derived from the exponential distribution," Technometrics, Vol. 3, pp. 29-49.

UNBIASED ESTIMATES OF RELIABILITY WHEN TESTING AT ONLY ONE EXTREME STRESS LEVEL

A. Bulfinch
Reliability Engineering Branch,
Nuclear Reliability Division
Quality Assurance Directorate
Picatinny Arsenal, Dover, New Jersey

ABSTRACT. Based on the stress-strength concept of reliability for "one-shot" items, it is assumed that an item cannot fail until the stress equals or exceeds the strength. From this premise and the following additional assumptions, methods are given for calculating unbiased estimated of non-time dependent reliability:

1. The relation between the stress and strength standard deviations are known approximately.

2. A single stress level is applied during testing at approximately three standard deviations from the average stress level.

3. The stress and strength distributions are normal.

Calculations are included to show the effect of errors in the assumptions concerning the standard deviations, applied stress level, and rounding-off errors.

This approach further reduces the sample size required to demonstrate high non-time dependent reliability in laboratory testing. It has the added advantage of obtaining unbiased estimates of reliability with the simplest of testing methods.

I. INTRODUCTION. The pressure of time and money in reliability testing requires a never ending quest for simpler methods and smaller sample sizes. Recent work at Picatinny Arsenal has suggested another contribution to this effort.

The usual interpretation of sample results for the determination of non-time dependent reliability, when only attribute type data can be obtained, is based on the binomial distribution. The usual laboratory method of testing is to apply a single level of stress to the sample. Under these conditions very large sample sizes are required to demonstrate reasonably high reliability values that may exist. In addition, this approach results in data very insensitive to changes in reliability. Both of these characteristics are costly shortcomings. However, the simple method of testing is an asset.

The purpose of this paper is to describe a procedure that retains the simple testing method but requires only small sample sizes for any reliability level and produces data that is sensitive to small changes in reliability. This is accomplished by changing the interpretation of the data and supplementing this, in a quantitative way, with knowledge grained from the experience of working with an item over a period of time.

However, the method presented here is limited to the laboratory determination of non-time dependent reliability when only success-failure type of data can be obtained. This type of reliability is based on stress-strength concept presented in an earlier paper (Reference 1).

The procedures proposed are an out-growth of recent work on the evaluation of laboratory methods by means of Monte Carlo sampling techniques. This work showed that when only attribute data can be obtained that:

1. The observed proportion of successes in a sample obtained at a single stress level is a biased estimate of the non-time dependent reliability defined by the stress-strength concept.

2. A sample obtained at a single stress level cannot measure the average or standard deviation of the strength distribution.

3. The observed failure rate, obtained at a single stress level measures the area of the tail of the strength curve to the left of (below) the applied stress ordinate.

From the above, it was realized that sample results obtained at a single stress level furnished information about the strength distribution. This suggested the possiblity of making use of this fact for obtaining unbiased estimates of reliability, with the very simple method of testing at a single stress level, by changing the use made of sample results.

II. METHOD OF CALCULATION. The method of calculation described below is based on the normal deviate:

$$T = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{s_1^2 + s_2^2}}$$

Where: $\bar{X}_1$ = Average stress expected in use

$\bar{X}_2$ = Average strength

s_1^2 = Variance of the stress distribution

s_2^2 = Variance of the strength distribution .

The previous work referred to above shows that precise and unbiased estimates of the true non-time dependent reliability can be obtained by entering a table of areas under the standard normal curve with this calculated T-value. This is true of course only when the stress and strength distributions are normally distributed. The sensitivity of this function to deviations from normality is yet to be demonstrated.

Since testing at a single stress level cannot measure the average and standard deviation of the strength distribution, the above formula was transposed to an equivalent function as follows:

Let X = Any applied stress level in testing. Then:

$$T_1 = \frac{X - \bar{X}_1}{s_1}$$

$$T_1 s_1 = X - \bar{X}_1 \text{ (for the stress distribution)}$$

$$T_2 = \frac{\bar{X}_2 - X}{s_2}$$

$$T_2 s_2 = \bar{X}_2 - X \text{ (for the strength distribution)}$$

$$\bar{X}_2 - \bar{X}_1 = (X - \bar{X}_1) + (\bar{X}_2 - X)$$

$$= T_1 s_1 + T_2 s_2 \text{ (by substitution).}$$

When: $m s_1 = s_2$

$$\bar{X}_2 - \bar{X}_1 = T_1 s_1$$

$$T = \frac{T_1 s_1 + m T_2 s_1}{\sqrt{s_1^2 + (m s_1)^2}}$$

$$= \frac{s_1 (T_1 + m T_2)}{s_1 \sqrt{1 + m^2}}$$

$$= \frac{T_1 + m T_2}{\sqrt{1 + m^2}}$$

III. ASSUMPTIONS. The last formula can be used under the following assumptions:

1. The stress and strength distributions are normal
2. Where $m s_1 = s_2$, m is known
3. The testing is done at a stress of $(\bar{X}_1 + T_1 s_1)$, where T_1 is known approximately.

IV. DISCUSSION OF ASSUMPTIONS.

1. If there is reason to question the assumption of normality, appropriate distribution free methods can be used. However, the form of the distribution should be determined where possible.

2. Experience has shown that m is approximately two. The examples given below show that the value of m can vary widely before seriously affecting the accuracy of the resulting reliability value.

3. $(\bar{X} + T_1 s_1)$ can be defined as the maximum stress expected in use. This level of stress is usually known by the development engineer or is specified in the Military Characteristics. Such a maximum stress can be defined statistically as the stress occurring only once in a thousand or once in ten thousand times. As such, $T_1 = 3.09$ or $T_1 = 3.72$ respectively. The examples given below show that T_1 can also vary widely before seriously affecting the accuracy of the resultant reliability value.

V. USE OF MODIFIED T - FORMULA.

In the above formula, T_7 is measured by the observed failure rate of the sample tested at a single stress level (X). Its numerical value can be obtained by entering a table of areas under the standard normal curve with the proportion of failures in the sample. With this value determined and the values of T_1 and m known or assigned, the above formula can be used without knowing X_2 or s_2 the average and standard deviation of the strength dististribution.

The average and standard deviation of the stress distribution must be separately determined. If this information is not available and cannot be determined, the determination of a numerical value for reliability is impossible.

VI. ACCURACY AND SENSITIVITY

The incentive for using the proposed method of calculating reliability is that it can furnish considerably more information about the existing reliability than the usual way of using sample success-failure results. The examples given in Table I show this quite well. Obtaining 50% sample failures in this method is not as bad as it might seem. If the 50% point of the strength curve is at the three (3.09) sigma point of the stress curve, the reliability equals .9162 (when m equals 2) - not 50%, the proportion of successes in the sample.

TABLE I

ACCURACY AND SENSITIVITY

Sample Size:	$n = 22$
Number of Sample Failures:	$= b$
Standard Deviations:	$2S_1 = S_2$
Testing Level:	$T_1 = 3.09$ (P = .001)

$$T = \frac{3.09 + 2T_2}{\sqrt{1 + (2)^2}}$$

SAMPLE			T-FORMULA		(1-b/n)	
b	b/n	T_2	T	Reliability	Reliability	Difference*
11	.5000	0.00	1.38	.9162	.5000	+.4162
10	.4545	0.11	1.48	.9306	.5455	+.3951
9	.4092	0.23	1.59	.9441	.5908	+.3533
8	.3636	0.35	1.69	.9545	.6364	+.3181
7	.3182	0.47	1.80	.9641	.6818	+.2823
6	.2728	0.60	1.92	.9726	.7272	+.2454
5	.2272	0.75	2.05	.9798	.7728	+.2070
4	.1818	0.91	2.19	.9857	.8182	+.1675
3	.1364	1.10	2.36	.9909	.8636	+.1273
2	.0909	1.34	2.58	.9951	.9091	+.0860
1	.0454	1.69	2.90	.9981	.9546	+.0435

* T-Formula reliability minus the observed proportion of successes in the sample.

The results in Table I show the sensitivity of the proposed method to changes in reliability values. A decrease of .03 in the reliability at the upper end of the scale increases the number of failures in the sample of 22, from zero to six. This is a significant difference at the 95% confidence level.

The above sensitivity is to be compared with the insensitivity of the method of using the observed proportion of successes in the sample as the point estimate of "reliability". In this method, where the binomial distribution pertains, the success probability ("reliability") must decrease approximately 0.23 (1.00 - .77) before the observed number of failures in the sample increases a significant (0 to 5) amount at the 95% level of confidence.

The above comparison of sensitivity shows that the proposed method is sensitive to changes in reliability. That is, the proposed method can detect relatively small changes in reliability with small sample sizes. This is an important property for a laboratory method. It means that small differences between design modifications and small changes occurring during storage can be readily detected.

VII. ERRORS DUE TO ASSUMPTIONS. The relative accuracies of the two methods for determining reliability are shown by the "differences" given in Table I. These differences are to be compared with the errors, resulting from incorrect assumptions shown in Table II. The assumption errors made here are the maximum expected in practice due to total ignorance about the system concerned. Any knowledge gained about a component or a system through experience will improve the accuracy of the assumptions and thereby reduce the resultant errors. This kind of knowledge, from experience, is always available and can be effectively used in the proposed method of calculation.

TABLE II

EFFECT OF ASSUMPTIONS AS FAILURE RATE INCREASES

Test Level: $U_1 + T_1 S_1$

Standard Deviation: $mS_1 = S_2$

FAILURE RATE (b/n)	m	T_1	P_1	T_2	T	POINT ESTIMATE
.02	1	2.33	.0100	2.05	3.10	.99903
.02	1	3.72	.0001	2.05	4.08	.99998
.02	2	3.09	.0010	2.05	3.21	.99934
.02	3	2.33	.0100	2.05	2.68	.99632
.02	3	3.72	.0001	2.05	3.12	.99910

Maximum Error:

Assuming the most favorable (highest reliability) condition when in fact the most unfavorable condition actually exists: .99998 - .99632 = +.00366.

Median Error:

Assuming the median ($m = 2$; $T_1 = 3.09$) condition when the most favorable (1) and Unfavorable (2) conditions exist: (1) .99934 - .99998 = -.00064 (2) .99934 - .99632 = +.00302

Effect of Assumptions (continued):

Test Level: $U_1 + T_1 S_1$

Standard Deviation: $mS_1 = S_2$

FAILURE RATE (b/n)	m	T_1	P_1	T_2	T	POINT ESTIMATE
.05	1	2.33	.0100	1.65	2.81	.99752
.05	1	3.72	.0001	1.65	3.80	.99993
.05	2	3.09	.0010	1.65	2.85	.99781
.05	3	2.33	.0100	1.65	2.30	.98927
.05	3	3.72	.0001	1.65	2.74	.99693

Maximum Error:

Assuming the most favorable (highest reliability) condition when in fact the most unfavorable condition actually exists:

$$.99993 - .98927 = +.01066$$

Median Errors:

Assuming the median ($m = 2$; $T_1 = 3.09$) condition when the most favorable (1) and unfavorable (2) conditions exist:

$$(1)\quad .99781 - .99993 = .00212$$

$$(2)\quad .99781 - .98927 = +.00854$$

Effect of Assumptions (continued):

Test Level: $U_1 + T_1 S_1$

Standard Deviation: $mS_1 = S_2$

FAILURE RATE (b/n)	m	T_1	P_1	T_2	T	POINT ESTIMATE
.10	1	2.33	.0100	1.28	2.55	.99461
.10	1	3.72	.0001	1.28	3.53	.99979
.10	2	3.09	.0010	1.28	2.52	.99413
.10	3	2.33	.0100	1.28	1.95	.97441
.10	3	3.72	.0001	1.28	2.39	.99157

Maximum Error:

Assuming the most favorable (highest reliability) condition when in fact the most unfavorable condition actually exists:

.99979 - .97441 = +.02538

Median Errors:

Assuming the median (m = 2; T_1 = 3.09) condition when the most favorable (1) and unfavorable (2) conditions exists:

(1) .99413 - .99979 = -.00566

(2) .99413 - .97441 = +.01972

Effect of Assumptions (continued):

Test Level: $U_1 + T_1 S_1$

Standard Deviation: $mS_1 = S_2$

FAILURE RATE (b/h)	m	T_1	P_1	T_2	T	POINT ESTIMATE
.20	1	2.33	.0100	0.84	2.24	.98745
.20	1	3.72	.0001	0.84	3.22	.99936
.20	2	3.09	.0010	0.84	2.13	.98341
.20	3	2.33	.0100	0.84	1.53	.93700
.20	3	3.72	.0001	0.84	1.97	97558

Maximum Error:

Assuming the most favorable (highest reliability) condition when in fact the most unfavorable condition actually exists:

$$.99936 - .93700 = .06236$$

TABLE III

SUMMARY OF ERRORS

Failure Rate	Errors Due to Using Sample Proportion of Successes as Point Estimate	Errors Due to Assumptions Maximum	Errors Due to Assumptions Median
.02	+.019	+.004	+.003
.05	+.048	+.011	+.008
.10	+.094	+.025	+.020
.20	+.183	+.062	+.046

The sample errors in Table III were obtained by subtracting (1 - b/n) from the point estimates (in Table II) for M = 2 and T_1 = 3.09 - the median conditions. The assumption errors in Table III were obtained by rounding off the corresponding errors in Table II.

The data in Table III show that both types of errors increase as the observed proportion of failures (failure rate) increases. However, in each case the assumption errors are less than the sampling errors. The magnitude of the assumption errors up through a failure rate of 0.10 is not great enough to seriously affect the reliability value. Some knowledge of T_1 or m will greatly reduce these errors in the calculated reliability.

VIII. EFFECT OF ROUNDING OFF ERRORS. When sample sizes are small, rounding off errors may be important. Their effects at various failure rates are shown in Table IV and Table V.

TABLE IV

EFFECT OF ROUNDING OFF ERRORS

(SAMPLE CALCULATIONS)

Test Level: $U_1 + 3.09\ S_1$

Standard Deviation: $2S_1 = S_2$

b	n	b/n	T_1	P_1	T_2	T	Point Estimate
		.45	3.09	.001	+.13	1.50	.9332
5	10	.50	3.09	.001	.00	1.38	.9162
		.55	3.09	.001	-.13	1.27	.8979

b	n	b/n	T_1	P_1	T_2	T	Point Estimate
		.20	3.09	.001	.84	2.13	[illegible]
5	20	.25	3.09	.001	.68	1.99	.9767
		.30	3.09	.001	.53	1.85	.9678

b	n	b/n	T_1	P_1	T_2	T	Point Estimate
		.05	3.09	.001	1.65	2.85	.9978
5	50	.10	3.09	.001	1.29	2.54	.9945
		.15	3.09	.001	1.04	2.31	.9896

Median Errors:

Assuming the median ($m = 2$; $T_1 = 3.09$) condition when the most favorable (1) and unfavorable (2) conditions exist:

(1) .98341 - .99936 = -.01595

(2) .98341 - .93700 = +.04641

TABLE V

SUMMARY OF EFFECT OF ROUNDING OFF ERRORS

Test Level: $U_1 + 3.09\ S_1$

Standard Deviation: $2S_1 = S_2$

FAILURE RATE	MAXIMUM ERROR
.05	.0006
.10	.0082
.20	.0128
.30	.0185
.40	.0250
.50	.0353

The errors shown in Table V are the differences between the maximum and minimum reliability values for each failure rate (b/n). The method of calculating the maximum and minimum values is based on the assumption of rounding off errors of $\pm$ 0.05 in the failure rate as shown in Table IV.

Although the assumed rounding off error is the maximum expected, its magnitude is not excessive below a failure rate of 0.30. As shown in Table V, this type of error also increases with the failure rate.

IX. USE OF CHEBYSHEV'S INEQUALITY. There is little or no information available on the form of strength distributions of most missiles and missile components. Furthermore, it is costly to obtain. It would be helpful if a distribution free procedure such as Chebyshev's irrequality could be used. As shown in Table VI, the use of Chebyshev's inequality in the modified T-formula resulted in ridiculous values.

TABLE VI

CHEBYSHEV'S INEQUALITY

Test Level: $U_1 + T_1 S_1$

Standard Deviation: $2S_1 = S_2$

Failure Rate (b/n)	T_1	P_1	T_2	T
0.50	31.62	.001	2	15.4

The T-value of 15.4 shown in Table VI is to be compared with the T-value of 1.38 shown in Table I for a failure rate of 0.50. From this, it is concluded that Chevyshev's inequality cannot be used in this application.

X. EXAMPLES. Previous work (Reference 1) has shown that the true non-time dependent reliability of the set of conditions used in these examples can be obtained by means of the following formula:

$$Z \gtrless \frac{U_2 - U_1}{\sqrt{S_1^2 + S_2^2}}$$

Where:

U_1 = True mean of the stress distribution

S_1^2 = True variance of the stress distribution

U_2 = True mean of the strength distribution

S_2^2 = True variance of the strength distribution.

The reliability value obtained by means of the above formula was used to determine the accuracy of the following two methods of using attribute data obtained from the application of a single stress level:

1. Using the observed proportion of successes in the sample as the reliability point estimate

2. Using the observed proportion of failures in the sample as a measure of the area of the strength distribution, below the applied stress, to obtain T_2 in the T-formula.

The errors associated with the two methods of using sample data are to be compared to show the practical value of the method proposed here.

The conditions used in this example are:

Stress	Strength
$U_1 = 10$	$U_2 = 42$
$S_1 = 5$	$S_2 = 10$

The true non-time dependent reliability for this set of conditions can be calculated as follows:

$$Z = \frac{42 - 10}{\sqrt{(10)^2 + (5)^2}} = \frac{32}{125} = 2.86 .$$

The true reliability associated with this Z-value is 0.9979.

1. First Method

Using the observed proportion of successes as the point estimate:

If it is assumed that the testing is done at $U_1 + 3S_1$, then the applied stress will be equal to 25 units. For the set of conditions described above, the portion of the strength distribution below 25 units can be found as follows:

$$Z_2 = \frac{42 - 25}{1} = 1.70.$$

Entering a table of areas under the standard normal curve with this Z_2 value, the following value is obtained:

$$P = .0446$$

The earlier work referred to above shows that this latter value is the expected failure rate of the single-stress-level method. The complement of this value (.9554) would be taken as the "true" mean reliability of this method. The difference between 0.9979 and 0.9554 (0.0425) is considered the expected error of the single-stress-level method when the proportion of successes in the sample is taken as the point estimate.

2. Second Method

Using the observed proportion of successes as a measure of the area in the tail of the strength curve:

The practical value of the method proposed here can best be demonstrated by calculating the magnitude of the errors due to the assumptions made concerning m and Z_1. Using the set of conditions described above, the variations in m and Z_1 used below are the maximum considered likely in practice. Therefore, the errors in the reliability values caused by these variations are the maximum expected.

TABLE VII

VARIATIONS DUE TO ASSUMPTION ERRORS

Failure Failure Rate	T_2	m	Z_1	P_1	T	Point Estimate
1/22	1.70	1	2.33	.0100	2.85	.9978
1/22	1.70	1	3.72	.0001	3.83	.9999
1/22	1.70	3	2.33	.0100	2.35	.9906
1/22	1.70	3	3.72	.0001	2.79	.9974

The following errors were obtained by calculating the differences between the true value and the point estimates shown in Table VII:

ERRORS

M	Z_1	Differences
1	2.33	.0001
1	3.72	.0020
3	2.33	.0073
3	3.72	.0003

These errors are to be compared with 0.0425, the error obtained when the sample result was used as the point estimate of non-time dependent reliability.

XI. CONCLUSIONS.

1. The proposed use of attribute data to estimate non-time dependent reliability by the single-stress-level method is more accurate at all levels of reliability then the usual method of using the proportion of successes in the sample as the reliability point estimate.

2. The proposed method is more sensitive to changes in reliability than the usual method.

3. The proposed method permits the knowledge gained through the experience of working with an item to be used in a quantitative way and thereby reduce the sample size required to obtain an unbiased estimate of reliability.

4. When the true reliability of an item is in fact as high as 0.995 (the usual value of Military Characteristics requirements) and the stress is applied at the three-sigma level, the expected error in the proposed method is less than 1.0%.

REFERENCES

1. Bulfinch A., "Evaluation of Various Laboratory Methods for Determining Reliability", Proceeding Eighth Conference on the Design of Experiments in Army Research, Development and Testing, October 1962.

2. Lloyd, David K., and Lipow, Myron, "Reliability Management, Methods, and Mathematics", Prentice-Hall, Inc., Englewood Cliffs, N.Y., 1962.

COMMENTS ON THE STRENGTH OF REAL STRUCTURAL MATERIALS

Kenneth H. Abbott
Army Materials Research Agency
Watertown Arsenal, Watertown, New York

[Remarks:] Following the talk by A. Bulfinch there was a lively discussion. In the light of the questions posed the following comments should be of interest to those attending Session I.

The most common structural materials for military equipment are metals. As a result, the following comments are specifically related to structural metals and alloys, although they are applicable in a generic sense to other structural materials.

The strength of an engineering alloy, or more generally, the strength distribution for a specific alloy and processing treatment, is normally determined by tensile testing. During a tensile test to determine the maximum stress that can be tolerated before failure occurs, the general behavior of the alloy is illustrated schematically in Figure 1, a typical stress-strain curve. The engineering yield strength is defined as the stress (load per initial area) after a predetermined amount of plastic flow has taken place. The ultimate tensile strength is defined as the stress at the maximum load, and normally occurs after an additional amount of plastic flow. After reaching the ultimate tensile strength, the applied load generally decreases as local necking develops, additional plastic flow occurs, and finally the specimen fractures. Strength, then, is normally determined by the plastic properties of the metal and, per se, is not related to fracture.

Ideally, structures should behave in a similar fashion, i.e., if a sufficient load-carrying cross section is available, the structure should yield (undergo plastic flow) before it fractures. This simple approach, however, is complicated by the geometry of the structure and associated stress field, as well as by fabrication and processing defects such as small voids, inclusions, weld cracks, etc. In an actual structure, the propagation of a crack from an existing defect can result in fracture of the metal before any appreciable general plastic flow takes place. As indicated schematically in Figure 2, fracture can occur at a stress level below the engineering yield strength. This phenomenon of the occurrence of fracture without any general plastic flow is defined as "brittle fracture."

Hence, the assumption that component failure always occurs when a particular strength is exceeded (normally the yield strength) is invalid. Numerous examples are reported in technical literature of both military and civilian structures that failed by fracture at nominal stress levels well below the yield strength. Some instances are reported where such fracture failures occurred at stress levels of only 20% of the yield strength.

Minute defects from which brittle fractures originate cannot be eliminated by non-destructive testing techniques for two reasons. The first is that some such defects are below the resolution limit of the non-destructive testing equipment. The second is that many defects can originate subsequent to component manufacture by various time dependent processes such as corrosion, hydrogen embrittlement, etc. Hence, the design of structures which fail through plastic deformation rather than by fracture is a very difficult and complex process.

The following references are suggested for additional detail:

1. American Society for Testing Materials, "Symposium on Effect of Temperature on the Brittle Behavior of Metals," ASTM No. 158, June 1953.

2. "Conference on Brittle Fracture in Steel," West of Scotland Iron and Steel Institute, May 1953.

3. "Symposium on Prevention of Brittle and Low Temperature Failures of Steel Components," Watertown Arsenal Laboratories, May 1949.

4. Cottrell, A.H., "Theory of Brittle Fracture in Steel and Similar Metals," Transactions AIME, V 212, 1958.

5. "Fracture of Structural Metals," WAL MS-48, June 1962.

TYPICAL STRESS-STRAIN CURVE

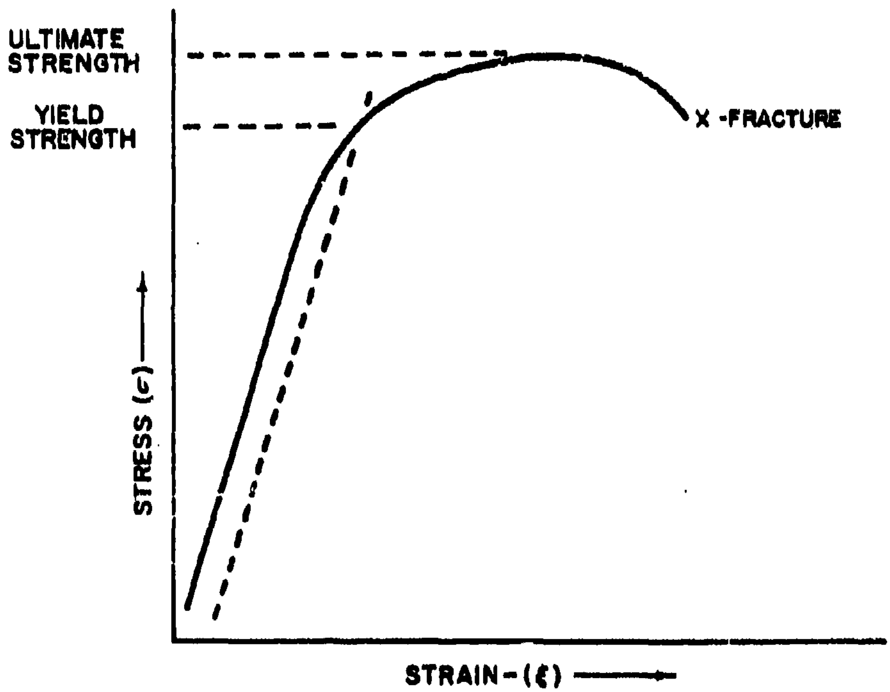


FIGURE 1

19-066-698/ORD-62

PREMATURE FRACTURE

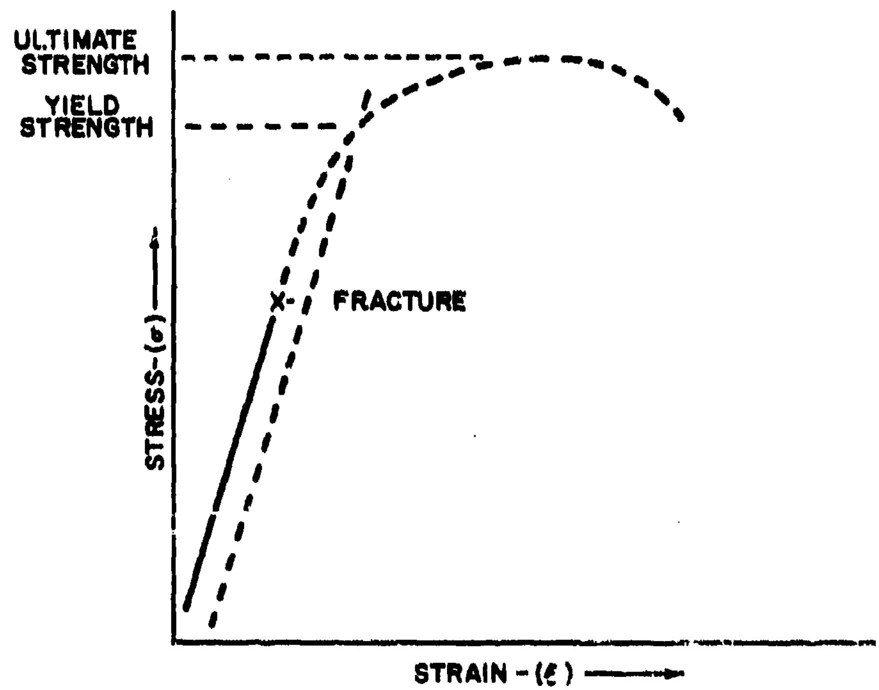


FIGURE 2

19-008-697/ORD-62

THE COLLECTION, PROCESSING AND STORAGE OF DATA ON SERIALIZED NW ITEMS

Manfred W. Krimmer*
Statistics and Analysis Section, Reliability Branch, OAD
Picatinny Arsenal, Dover, New Jersey

The existence of a requirement for a standardized, usable means of data collection and processing for Serialized Nuclear Weapons Items necessitated the writing of a Quality Assurance Instruction (QAI-PA-1) and two major computer programs, namely, an Error Program and a Summary Program. This paper will outline the application of QAI-PA-1 to Serialized NW Items, the fundamentals of the Error and Summary Computer Programs, and the results obtained therefrom.

1. The Quality Assurance Instruction standardizes the method of data collection for those serialized nuclear munitions produced under the jurisdiction of the U. S. Army Munitions Command. The automatic data processing system includes the standardization of forms, recording methods, and conversion of data to punch cards for transmission to Picatinny Arsenal. Computation, analysis and interpretation of the data for quality and reliability evaluation then takes place. The information collected for each item includes acceptance data and performance results as well as stockpile test information.

2. The QAI has three primary purposes.

   a. The Reporting of Identification Information and Test Results.

   b. The Reporting of Drawing Change Documents and Associated Serial Numbers.

   c. The Reporting of Related Serial Numbers.

Following is a brief explanation of the above reports covering The Reporting of Identification Information and Test Results last, it being the primary report of this system, and a discussion of the computer programs and what is done with the results.

3. The reporting of Drawing Change Documents such as engineering orders and technical data change requests and the serialized items pro-

---

*At the present time the author is with the Ammunition Procurement and Supply Agency, Joliet, Illinois.

duced under each, was requested by the reliability engineers at Picatinny Arsenal to aid them in more accurately evaluating the stockpile reliability. This is done on reporting forms SMUPA NS-900 and NS-900A. The first form (Fig. 1) SMUPA NS-900 is a record of all Drawing Numbers, revisions and/or EO's or TDCR numbers. At the start of production or production under a new contract, the contractor will list all the appropriate documents under which production will begin. The second form (Fig. 2) SMUPA NS-900A lists all the serial numbers of the components produced under the drawing, revisions and/or EO's or TDCR's recorded on the first form.

Upon receipt of one or more EO's (or TDCR's), another primary card form, SMUPA Form 900, is filled out listing these additional EO's (or TDCR's). A listing of the component serial numbers for units produced under the new EO's (or TDCR's) is then recorded on a new secondary card, SMUPA Form 900A. When additional EO's (or TDCR's) are received, another primary card form is completed listing only these additions. Similarly, a listing of the component serial number, for units produced under these EO's (or TDCR's) is recorded on a new secondary card.

This procedure is continued until three months have elapsed. At a logical point in production (i.e., prior to incorporation of a new EO (or TDCR)) all the preceding primary cards and their associated secondary cards are forwarded to Picatinny Arsenal in punch card format. The above procedure is then repeated, beginning with an initial primary card listing all the drawing numbers with their up-to-date revisions, and Engineering Order Numbers (or Technical Data Change Requests) not yet associated with a drawing change, under which production will be continued. This process is repeated continuously every three months or until the end of production. The information is then put on punched cards and submitted by the producer.

4. The Reporting of Related Serial Numbers is necessary to maintain an accurate up-to-date record of what assembly an item is a part of throughout its life cycle and what items make up a specific assembly. This information is required by MUCOM through direction of the Army Materiel Command. The information is reported on SMUPA NS-901 (Fig. 3) and lists the serial numbers of all the components making up a specific assembly. This information is submitted by the contractor assembling the item and by field personnel whenever a component is removed or replaced.

5. The last report, but most important, in the system, is the reporting of Identification Information and Test Results. This information is reported on SMUPA NS-902 (Fig. 4). This is the basic document for recording the results of all the inspections for which information is required. It contains all information necessary to identify the item such as, contractor, contract number, spec. number and date of inspection. It also contains the variables test results as called for in the purchase description, it identifies the defect, and gives the disposition of the item. This information is recorded by the item contractor and submitted on punched cards.

6. To simplify the identification of the tests called for in the purchase description, a six digit "test identification code" was developed through numerous conferences with engineers concerned. This code is contained in the PD and is made up as follows. Since all the tests in the PD are called out in paragraph 3 and are numbered, for example 3.10.1.3, the first digit signifying paragraph 3 is omitted and the code begins with the next two digits signifying the first subparagraph. The next subparagraph is then designated by a single digit and the last three digits of the code are individual tests in numerical sequence. For example, 10/1/002 identifies the above paragraph and the second test within that paragraph. The codes are unique within PD's but not among the various PD's.

7. During the gathering of this data by the item contractor, monthly shipments of punched cards containing Identification Information and Test Results come into our office. These cards must be checked for various errors in recording and/or key punching. This is accomplished by means of an Error Program. The information on the cards is put on magnetic tape and fed into the computer and the Error Program checks each card, column by column, for information which does not belong. If all the cards are correct, analyzation of the data takes place. If there are errors in the cards, the computer will print out what the errors are and where to find them. The cards are then corrected by us if the volume of errors is not too great or they are sent back to the contractor for correction.

8. If the data is good, it is analysed. This is done using our Summary Program which takes the data and computes the following:

a. Lists all readings which are out of specification limits.

b. Computes:

(1) Mean

(2) Standard Deviation

(3) (±) 1σ, 2σ and 3σ limits.

c. Prints out the max. and min. values and the range.

d. Lists all readings outside of the sigma limits.

e. Prints out a frequency distribution based on 3 sigma limits.

f. Indicates skewness, which tests the normality of the curve.

g. and indicates kurtosis which measures the flatness of the curve.

This summary program is run for each lot of data received on a monthly basis.

9. A second type of Search and Summary Program, providing the same output as the one above is run for the entire production or for any special cases where specific information is required.

10. The results obtained are then studied and interpreted and a report is written. The reports are issued and a follow up action is maintained until replies to the reports recommendations are received indicating some action being taken.

The data is then retained on master tapes and stored for future use.

It is felt that this system is a good one and it is intended to increase its capabilities and scope as future developments arise.

Requests for copies of QAI-PA-1 or any further information about this system should be directed to:

Commanding Officer
Picatinny Arsenal
ATTN: SMUPA-NR2
Dover, New Jersey

DRAWING CHANGE DOCUMENTS REPORT

PAGE 1 OF 5

ITEM XM 99 AK   CONTRACTOR Widget Corp.   PREPARED BY J. Smith

PRIMARY CARD

C ASSOCIATION NO. 0 0 0 8 | W D C | F | a. Total number of cards b. Card number | 1 0 2 1

1	2	3	4	5	6	7	8	9	10	11	12	13
C	0	0	0	8	W	D	C	F	1	0	2	1

DRAWING NUMBERS, REVISIONS, AND/OR E.O. (OR TDCR) NUMBERS

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
3	6	1	1	9	0	4	*	3	6	[illegible]	1	9	0	7	2	*	2	5	3	8	6	1	7	*	4	8

41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
3	3	1	6	2	1	*	2	5	7	[illegible]	2	6	3	*	1	0	7	4	3	5	2	*	6	2	1	4

68	69	70	71	72	73	74	75	76	77	78	79	80
7	8	5	*	2	2	3	*					

C ASSOCIATION NO. 0 0 0 8 | W D C | F | a. Total number of cards b. Card number | 0 0 0 3

1	2	3	4	5	6	7	8	9	10	11	12	13
C	0	0	0	8	W	D	C	F	0	0	0	3

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
2	5	3	8	6	1	7	*	0	1	3	*	5	4	2	8											

41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67

68	69	70	71	72	73	74	75	76	77	78	79	80

C ASSOCIATION NO. ____ | ___ | __ | a. Total number of cards b. Card number | ____

1	2	3	4	5	6	7	8	9	10	11	12	13
C												

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40

41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67

68	69	70	71	72	73	74	75	76	77	78	79	80

SMUPA-FORM NS-900 MAY 63

APPENDIX B

ASSOCIATED SERIAL NUMBERS REPORT

PAGE 2 OF 5

ITEM XM 99 AK

CONTRACTOR Widget Corp.

PREPARED BY J. Smith

SECONDARY CARD

D[1] ASSOCIATION NUMBER 0 0 0 8 (2–5) SECONDARY CARD NUMBER 0 0 0 2 (6–9) F (10)

COMPONENT SERIAL NUMBER

11 X	12 E	13 C	14 K	15 U	16 M	17	18 R	19 A	20 S	21 S	22 E	23 T	24	25 B	26 E	27 M	28 Z	29 I	30 X	31	32 Y	33 O	34 L	35 Z	36 E	37 S	38
39 N	40 O	41 P	42 D	43 U	44 F	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
67	68	69	70	71	72	73	74	75	76	77	78	79	80														

D[1] ASSOCIATION NUMBER 0 0 0 8 (2–5) SECONDARY CARD NUMBER 0 0 0 4 (6–9) F (10)

COMPONENT SERIAL NUMBER

11 L	12 E	13 R	14 N	15 O	16 Q	17	18 B	19 E	20 B	21 B	22 O	23 C	24	25 Y	26 I	27 T	28 D	29 A	30 R	31	32 K	33 A	34 Z	35 F	36 U	37 N	38
39 D	40 O	41 L	42 Z	43 E	44 K	45	46 T	47 U	48 S	49 G	50 A	51 P	52	53 F	54 O	55 R	56 Y	57 O	58 S	59	60 P	61 A	62 X	63 S	64 I	65 T	66
67 W	68 U	69 F	70 H	71 O	72 P	73	74 D	75 E	76 N	77 T	78 I	79 S	80														

D[1] ASSOCIATION NUMBER 0 0 0 8 (2–5) SECONDARY CARD NUMBER 0 0 0 5 (6–9) F (10)

COMPONENT SERIAL NUMBER

11 R	12 I	13 L	14 T	15 U	16 Z	17	18 B	19 A	20 F	21 F	22 E	23 L	24	25 W	26 I	27 G	28 K	29 A	30 P	31	32	33	34	35	36	37	38
39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
67	68	69	70	71	72	73	74	75	76	77	78	79	80														

PAGE 1 OF 3

RELATED SERIAL NUMBERS REPORT
(Component to Assembly, etc.)

ASSEMBLY REPORTED XM 99 AK   CONTRACTOR Widget Corp.   PREPARED BY R. Jones

PRIMARY CARD

Card	Association Number		Quantity Secondary Cards	
E	4 2 7 1	W D C F	0 1 9	C H 0 0 3 3 3 8

Report Date	Assembly Model No.	Assembly Serial No.
3 2 4 8	0 0 9 9 X 0	X E C K U M

SECONDARY CARD

F ASSOCIATION NUMBER 4 2 7 1 SECONDARY CARD NUMBER 0 0 1

COMPONENT Model Number						COMPONENT Serial Number						COMPONENT Model Number						COMPONENT Serial Number					
3	3	5	9	E	1	R	A	M	K	E	P	0	0	5	7	E	2	M	A	P	R	U	S
3	3	5	9	E	1	Q	I	L	K	A	N	0	0	5	7	E	2	Y	I	N	Q	A	S
0	0	2	2	X	0	S	U	Y	R	I	T	0	0	5	7	E	2	F	O	B	L	A	Y

F ASSOCIATION NUMBER 4 2 7 1 SECONDARY CARD NUMBER 0 0 2

COMPONENT Model Number						COMPONENT Serial Number						COMPONENT Model Number						COMPONENT Serial Number					
0	0	5	7	E	2	R	A	V	H	O	T												

SMUPA FORM NS-901 MAY 63

IDENTIFICATION INFORMATION AND TEST RESULTS REPORT

PAGE 1 OF 15

ITEM XM99 AK CONTRACTOR Widget Corp TESTER A. Wilson INSPECTOR L. Kelly GOV'T INSPECTOR J. North

PRIMARY CARD

IDENTIFICATION INFORMATION

A ASSOCIATION NO. 0 0 0 2 | W D C | F QUANTITY SECONDARY CARDS 0 1 5 | C H 0 0 3 3 3 8

P 0 1 3 1 6 0 3 0 5 DATE OF INSPECTION (TEST) 3 1 9 7 | 0 0 9 9 X 0

COMPONENT SERIAL NO. X E C K U M SUB 1 LOT NO. 0 0 0 0 LOT SIZE 0 0 0 SAMPLE SIZE 0 0

DISP. P | 0 | D ASSEMBLY MODEL NO. 0 0 0 0 0 0 ASSEMBLY SERIAL NO. 0 0 0 0 0 0

SECONDARY CARD

TEST RESULTS

B ASSOCIATION NO. 0 0 0 2 SECONDARY CARD NO. 0 0 1 COMPONENT SERIAL NO. X E C K U M

TYPE TEST	TEST IDENTIFICATION CODE	READING	UNITS OF MEASURE	DEFECT IDENTITY	DISP.
3	0 9 3 0 0 1	0 0 2 2 . 6	0 A M P	2	P

REMARKS:

3	0 9 3 0 0 2	0 0 2 2 . 8	0 A M P	2	P

REMARKS:

3	1 1 6 0 0 1	- 2 0 2 . 5	M E G 0	1	R

REMARKS:

SMUPA-FORM NS-902 MAY 63

APPENDIX A

PROBABILITY THAT STRESS IS LESS THAN STRENGTH AT PRESCRIBED CONFIDENCE LEVELS, FOR NORMALLY DISTRIBUTED DATA

E. L. Bombara
Engine Projects Office, Marshall Space Flight Center

SUMMARY. Under cognizance of Marshall Space Flight Center, the stress vs. strength concept is often employed in small-sample testing during design of engine subsystems.

Under the assumption of normally distributed data taken from small samples of expensive items, this paper presents methods for estimating probability that stress does not exceed strength, with a prescribed level of confidence. Tables are provided to simplify calculations of such probabilities.

Input data are the mean and standard deviation of operating stress from n_{ss} tests, and the mean and standard deviation of strength from n_{th} tests. Precision of these methods, based on approximate results by Welch and Aspin, is indicated vs. sample sizes and ratios of variances. By examination of Tables I and II, sample sizes required for stress and strength tests can be determined. As one might expect, if $\sigma^2_{th}/\sigma^2_{ss}>1$, it is best to select $n_{th} \geq n_{ss}$ in order to achieve highest precision, and conversely for $\sigma^2_{ss}/\sigma^2_{th}>1$ (see Figure III). Comparison of analytical results with computer simulations (Table I) have been made to illustrate the amount of bias of the method when the number of tests is very small; say, 5.

ACKNOWLEDGMENTS. The author would like to express his sincere appreciation to Messrs. Raymond Heathcock and Dale L. Burrows for their findings in Monte-Carlo simulations of the stress-strength problem before the writing of this paper. Their results, contained in these proceedings, were also presented by Mr. Burrows at the Ninth Conference on the Design of Experiments in Army Research, Development and Testing. Gratitude is expressed to Mr. Dennis Nickle, Florida Research and Development Center of Pratt & Whitney Aircraft for Table III (taken from reference 8) and to Dr. George J. Resnikoff for Table IV (taken from reference 7). The author would also like to express his indebtedness to Mr. William Moore for Monte-Carlo programming, Miss Deanna DeBerry for calculations and charts, and Miss Linda Dillard for typing.

INTRODUCTION. Under the concept promoted by Robert Lusser of the Army Ordnance Missile Command (reference 1), a measure of subsystem reliability can be estimated through measurement of safety margins when data is normally distributed. For example, suppose that in development tests of a space vehicle the yield point (strength) of several pressure vessels of the same design (e. g., fuel tanks, thrust chambers, etc.) is measured and that the operating environmental levels of these (stress) are measured in actual flights or in static tests. Certainly it would be comforting to know that all measurements of the latter type are less than all of the yield points, as shown in figure I. The bell-shaped curve indicates the assumption that stress and strength values are normally distributed.

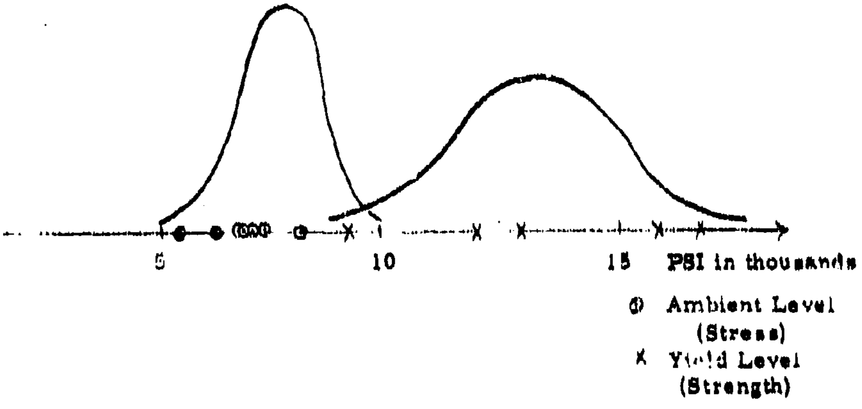


Figure I

Measurements of Stress and Strength

We might be contented with this situation, or, dependent on the vehicle mission, we might still fear the not-too-remote chance that stress exceeds strength on the next equipment to be tested. Certainly, if we observe overlap of the stresses and strengths we should be concerned. On the other hand, if all strengths exceed all stresses by a large amount, the design might be relaxed (reference 2). We are therefore interested in the probability that stress is less than strength in a single mission. We should also be concerned with confidence level, due to the necessarily small sample sizes.

Let us define the observed scatter of stresses and strengths in terms of their respective sample standard deviations, s_{ss} and s_{th}, with corresponding observed means, $\bar{x}_{ss}$ and $\bar{x}_{th}$. The probability that stress is less than strength may be demonstrated at a given confidence level under the assumption of normally distributed stresses and strengths, as a function of $\bar{x}_{th}$, $\bar{x}_{ss}$, s_{th}, s_{ss}, and the respective sample sizes, n_{th} and n_{ss}. For brevity's sake, define probability that stress is less than strength as "stress-strength reliability." As indicated in the appendix, stress-strength reliability, R, can be demonstrated (or not) at a prescribed confidence level, C, in terms of n_{th}, n_{ss}, and the statistic

(1)
$$\hat{K} = \frac{\bar{x}_{th} - \bar{x}_{ss}}{\sqrt{s_{th}^2 + s_{ss}^2}}$$

relative to K. If $\hat{K} \geq K$, R is demonstrated at confidence level C.

This paper deals primarily with methods of finding K as a function of R, C, sample sizes, and variances. σ_{th}^2 is the population variance of strength, estimated by s_{th}^2 from n_{th} observations; σ_{ss}^2 is the population variance of stress estimated by s_{ss}^2 from n_{ss} observations.

ILLUSTRATIVE EXAMPLES.

Case I: $n_{th} = n_{ss} = n$; $b = \sigma_{th}^2 / \sigma_{ss}^2$ Unknown.

If the number of stress and strength measurements are the same, the procedure is relatively simple:

It is desired that R = .99 be demonstrated at 90% confidence. Data is as follows:

$\bar{x}_{ss} = 7.22 \qquad \bar{x}_{th} = 13.28$

$s_{ss} = 1.040 \qquad s_{th} = 3.120$

$n_{ss} = n_{th} = n = 7.$

From (1), $\hat{K} = \dfrac{13.28 - 7.22}{\sqrt{3.120^2 + 1.040^2}} = 1.843.$

Next calculate

Effective Sample Size $\approx f + 1$

$$(2) \qquad \approx \frac{(n-1)(s_{th}^2 + s_{ss}^2)^2}{s_{th}^4 + s_{ss}^4} + 1$$

$$\approx \frac{6(3.120^2 + 1.040^2)^2}{3.120^4 + 1.040^4} + 1 = 8.32.$$

Enter table III using confidence = 90%, R = .99, and conservatively, effective sample size ≈8 to find K = 3.78.

Since $\hat{K} < K$, .99 stress-strength reliability is not demonstrated at the 90% confidence level. Reading to the right in table III, we see that for K = 1.83, R = .84. In this case $\hat{K} > K$, and .84 stress-strength reliability is demonstrated at 90% confidence. The 50% confidence K-value for effective sample size 8 in table III of 1.83 gives R at 50% confidence = .96.

Case 2: $n_{th} \neq n_{ss}$; $b = \sigma_{th}^2 / \sigma_{ss}^2$ Unknown.

If the number of stress measurements, n_{ss}, and strength measurements, n_{th}, are not equal, $\hat{K}$ is defined as in case 1, but the procedure for finding K is more complex:

Using the data of figure I,

$$\bar{x}_{ss} = 7.22 \qquad \bar{x}_{th} = 13.28$$

$$s_{ss} = 1.040 \qquad s_{th} = 3.120$$

$$n_{ss} = 9 \qquad n_{th} = 5$$

From (1), $\hat{R} = \dfrac{13.28 - 7.22}{\sqrt{3.120^2 + 1.040^2}} = 1.843.$

(Again, suppose we wish to determine whether $R = .99$ @ $C = .90$).
From a table of the cumulative normal distribution (e.g., reference 9) find the one-sided value of $Z_R = Z_{.99} = 2.326$

Calculate the parameter of non-centrality of the non-central t statistic:

(3)

$$\delta' \approx z_R \sqrt{\frac{s_{th}^2 + s_{ss}^2}{\dfrac{s_{th}^2}{n_{th}} + \dfrac{s_{ss}^2}{n_{ss}}}} \approx 2.326 \sqrt{\frac{3.120^2 + 1.040^2}{\dfrac{3.120^2}{5} + \dfrac{1.040^2}{9}}} \approx 5.320.$$

Calculate effective degrees of freedom:

(4)

$$f \approx \frac{(n_{th} s_{th}^2 + n_{ss} s_{ss}^2)^2}{\dfrac{n_{th}^2 s_{th}^4}{n_{th}-1} + \dfrac{n_{ss}^2 s_{ss}^4}{n_{ss}-1}} \approx 5.647$$

Calculate

(5) $$\eta = \frac{\delta/\sqrt{2f}}{\sqrt{1+\delta^2/2f}}$$

$$= \frac{5.320/\sqrt{11.29}}{\sqrt{1+\frac{5.320}{11.29}}} = 0.8452\ .$$

Find λ in table IV for $\epsilon = 1 - C = .10$ as a function of η and f. Interpolate linearly on η, then interpolate on $12/\sqrt{f}$, unless $f<9$, in which case linear interpolation on f is sufficient (see reference 7). In this example, for f = 5.643, two-way linear interpolation is sufficient, giving $\lambda = 1.362$.

Then

(6) $$K \approx \frac{\delta + \lambda\sqrt{1 + \delta^2/2f - \lambda^2/2f}}{(1 - \lambda^2/2f)\,\delta/Z_R}\ .$$

Note that δ/Z_R is the radical in equation (3).

$$K \approx \frac{5.320 + 1.362\sqrt{1 + 5.3201^2/11.29 - 1.362^2/11.29}}{(1 - 1.362^2/11.29)\ 5.320/2.326} = 4.19.$$

Since $\hat{R} < K$, .99 stress-strength reliability is not demonstrated at the 90% confidence level. By trial and error we may reduce Z_R enough that R is demonstrated at 90% confidence; e.g. let $Z_R = .8779$, corresponding to R = 0.81 $\delta = 5.320\ (.8779)/2.326 = 2.008$. Then

$$\eta = (2.008/\sqrt{11.29})/\sqrt{1 + (2.008)^2/11.29} = .5130$$

For $\epsilon = 0.10$, table IV gives $\lambda = 1.357$; K = 1.824. Since $\hat{R} > K$, 0.81 stress-strength reliability is demonstrated at the 90% confidence level. Similar calculations show that at the 50% confidence level (use $\epsilon = .50$, table IV), the demonstrated R is 0.96.

In the event that $f > 9$, the interpolation procedure for λ may be accomplished by interpolating on η first, then on $12/\sqrt{f}$. For example, suppose $f = 15.0$, $\eta_L = 0.8251$. For 90% confidence, find $\epsilon = 0.10$ in table IV. Then interpolate in table IV as follows:

$12/\sqrt{f}$	4	3.873	3
$\eta_L \backslash f$	9	15.0	16
0.8	1.3526		1.3380
0.8251	1.3538	1.351	1.3395
0.9	1.3540		1.3398

After linear interpolation based on η_L, the horizontal interpolation is based on $12/\sqrt{f}$, resulting in $\lambda = 1.351$.

<u>Case 3</u>: Equal or Unequal sample Sizes, $b = \sigma_{th}^2/\sigma_{ss}^2$ Known.

If the ratio of variances, $\sigma_{th}^2/\sigma_{ss}^2$ is known, but the level of each variance is unknown, the procedure for finding $\hat{K}$ is the same as in cases 1 & 2. In other words, equation (1) and the same sample values of $\bar{x}_{ss}$, $\bar{x}_{th}$, s_{ss}, s_{th}, n_{th}, and n_{ss} give $\hat{K} = 1.843$ as before.

If we also assume, however, that the ratio of population variances, $b = \sigma_{th}^2/\sigma_{os}^2 = 2.0$, K may be determined* as follows:

$$f \approx \frac{(bn_{th} + n_{ss})^2}{\frac{(bn_{th})^2}{n_{th}-1} + \frac{n_{ss}^2}{n_{ss}-1}} = \frac{(10+9)^2}{\frac{10^2}{4} + \frac{9}{8}} = 13.82 \, . \tag{7}$$

Then

---

* If $n_{th} = n_{ss} = n$, K is more easily found by solving equation (2), substituting σ_{th} for s_{th} and σ_{ss} for s_{ss}, then entering table III with effective sample size as explained in case 1.

(8)

$$\delta \simeq Z_R \sqrt{\frac{b^2 + 1}{\frac{b^2}{n_{th}} + \frac{1}{n_{ss}}}} \; .$$

For $R = .99$, $Z_R = 2.326$,

$$\delta \approx 2.326 \sqrt{\frac{4+1}{\frac{4}{5} + \frac{1}{9}}} = 5.450 \, .$$

Using equation (5),

$$\eta = \frac{5.450/\sqrt{27.64}}{\sqrt{1 + \frac{5.450^2}{27.64}}} = .7201 \, .$$

Linear interpolation on η with interpolation on $12/\sqrt{f}$ (since $f > 9$) in table IV ($\epsilon = 0.10$) gives:

$$K \simeq 4.65$$

using equation (6). As before $\hat{R} < K$, and $R = .99$ is not demonstrated. By trial and error we may reduce R and the corresponding Z_R sufficiently to find that for $R = .85$, $Z_R = 1.036$ and $K = 1.766$. Since $\hat{R} > K$, $R = .85$ is demonstrated at the 90% confidence level. Similarly, $R = .96$ is demonstrated at the 50% confidence level.

<u>Case 4</u>. Equal or Unequal Sample Sizes, Both σ_{th}^2 and σ_{ss}^2 Known.

If the level of each population variance, σ_{th}^2 and σ_{ss}^2 is known separately, sampling variation is due only to $\bar{x}_{th}$ and $\bar{x}_{ss}$, the sample means. Then

(9)

$$\hat{K} = \frac{\bar{x}_{th} - \bar{x}_{ss}}{\sqrt{\sigma_{th}^2 + \sigma_{ss}^2}} .$$

Assuming that $\sigma_{th}^2 = 4.00$, $\sigma_{ss}^2 = 2.00$, $\bar{x}_{th} = 13.28$, $\bar{x}_{ss} = 7.22$,

$$\hat{K} = \frac{13.28 - 7.22}{\sqrt{4.00 + 2.00}} = 2.47.$$

Compute K from:

(10) $$K = Z_A + Z_C \sqrt{\frac{\frac{\sigma_{th}^2}{n_{th}} + \frac{\sigma_{ss}^2}{n_{ss}}}{\sigma_{th}^2 + \sigma_{ss}^2}}$$

(obtained by solving equation (15) after substituting σ_{th}^2 and σ_{ss}^2 for s_{th}^2 and s_{ss}^2, respectively).

Where Z_R and Z_C are critical values of the standard normal deviate exceeded by 1 - R and 1 - C, respectively, found in a table of the cumulative normal distribution. Assuming R = .99 and C = .90, then $Z_R = 2.326$ and $Z_C = 1.282$ (found in a normal distribution table such as reference 9).

Then

$$K = 2.326 + 1.282 \sqrt{\frac{\frac{4}{5} + \frac{2}{9}}{4 + 2}} = 2.510.$$

Since $\hat{K} < K$, R = .99 is not demonstrated at the 90% confidence level. Accordingly Z_R (and therefore K) may be reduced to 1.925 (R = .973), using equation (10) such that $K = \hat{K} = 2.46$. Then R = .973 is demonstrated at the 90% confidence level. In similar manner, R = .993 is demonstrated at the 50% confidence level.

Appendix A

PRECISION AND ACCURACY OF K. Precision in the critical value of K is dependent on whether the ratio of variances (or standard deviations) of stress and strength is known. If this ratio is unknown, as in cases 1 and 2 above, the precision is largely dependent upon the sample sizes used in estimating the variances of stress and strength, measured by the effective degrees of freedom, f. If the ratio is known, precision is as good as the error in tables III and IV. Accuracy of the critical value of K is a function of effective degrees of freedom, f, of the approximated χ^2; i.e., bias decreases as f increases. It should be noted that K is a function of the ratio of $\sigma_{th}^2/\sigma_{ss}^2$ = b, not their levels.

If the levels of each population variance are also known, equation (10), case 4 above shows that K is precisely and accurately determined (no error, within limits of the tabulated cumulative normal distribution). However, such knowledge is rare in practice. The remainder of this discussion will therefore be limited to cases 1, 2 (unknown ratios of variances), and case 3 (known ratio of variances) above.

ACCURACY FOR KNOWN RATIO OF VARIANCES (OR STANDARD DEVIATIONS). When the ratio $b = \sigma_{th}^2/\sigma_{ss}^2$ is known, bias in the critical value of K, due to bias in f and χ^2, may be found by comparing critical values of K as obtained in case 3 above with the more accurate value of K obtained by Monte-Carlo simulation of equation (1) (see reference 2). In illustration, examine table I:

TABLE I
ACCURACY OF K, KNOWN RATIO OF VARIANCES
(OR STANDARD DEVIATIONS)^2
R = .99

$b = \sigma_{th}^2/\sigma_{ss}^2$	n_{th}	n_{ss}	f	K C = .90	K C = .50	K, Monte-Carlo 2000 Simulations C = .90 C = .50
1	5	5	8.0	3.64	2.421	3.79, 2.40
1	7	7	12.0	3.31	2.389	3.45, 2.40
1	9	9	16	3.14	2.372	3.22, 2.34
1	20	20	38.0	2.83	2.35	2.85, 2.35
1	5	9	12.0	3.47	2.40	3.48, 2.406
1	9	5	12.0	3.47	2.40	3.47, 2.405
3	5	9	8.7*	3.67*	2.44*	3.78, 2.44
3	7	7	9.6	3.53	2.41	{3.50, 2.43; 3.54, 2.44} (rep.)
3	9	5	10.5	3.39	2.40	3.41, 2.42
3	20	20	30.4	2.87	2.33	2.89, 2.33
9	5	9	5.6	4.21	2.52	4.22, 2.53
9	7	7	7.5	3.67	2.45	3.69, 2.45
1	50	50	98.0	2.64	2.32	2.64, 2.32

Comparisons with Monte-Carlo solutions indicates that when $b \geq 9$ sample sizes for which $f \geq 5$ give very high accuracy; when $b \geq 1$, $f \geq 20$ gives very high accuracy of the analytical solutions;

* These results are based on data of figure I if s_i^2 should happen to equal σ_i^2.

<u>PRECISION OF K FOR KNOWN RATIO OF VARIANCES (OR STANDARD DEVIATIONS).</u> When the ratio $b = \sigma^2_{th}/\sigma_{ss}^2$ is known, f, δ, η, and K in case 3 are uniquely determined, within the precision of table III and IV. This is readily seen by examination of equations (7), (8), (5), and (6).

<u>PRECISION OF K FOR UNKNOWN OR ASSUMED RATIO OF VARIANCES (OR STANDARD DEVIATIONS).</u> In the absence of knowledge of b in the relationship $\sigma_{th}^2 = b\sigma_{ss}^2$, lack of precision in K is due to the use of s_i^2, the sample estimates of σ_i^2 in equation (2) in case 1, and (3) and (4) in case 2. K was also obtained through Monte-Carlo simulation of equations (2) through (6), obtaining 2000 estimates of K as a function of 2000 random values of both s_{th}^2 and s_{ss}^2, where x_{th} and x_{ss} were sampled randomly from normal distributions with variances σ_{th}^2 and σ_{ss}^2, respectively. Table II below illustrates the over-all precision of the method for the case R = .99, C = .90. It is of interest to examine equation (7), Table II or Figure III regarding f vs. K. For $n_{th} + n_{ss}$ constant and the assumption $b > 1$, f increases as n_{th}/n_{ss} decreases. Conversely, under the assumption $b < 1$, f increases as n_{th}/n_{ss} increases. Further, under all of these conditions that increase f, K decreases, permitting demonstration of reliability at a higher confidence level (or higher reliability at fixed confidence).

TABLE II
PRECISION FOR UNKNOWN VARIANCE RATIOS

σ_{th}^2	σ_{ss}^2	n_{th}	n_{ss}	f*	Central 90 Percentiles of f (5%, 95%)		Central 90 Percentiles of K (5%, 95%)
9	1	5	9	5.6	4.44,11.16		-**, 3.96
9	1	7	7	7.5	6.32,10.55		3.46,3.91
3	1	5	9	8.679	5.26, 11.94		-**, 3.96
3	1	7	7	9.600	6.98,11.94		3.39,3.80
					6.90,11.95	(rep)	3.39,3.81
3	1	9	5	10.516	8.52,11.98		3.40,3.59
3	1	20	20	30.401	24.47,37.06		2.85,2.96
3	1	15	5	16.568	14.48,17.95		3.17,3.20
3	1	25	15	32.342	27.84,37.18		2.85,2.91
3	1	30	10	34.433	31.18,37.64		2.87,2.88
1	1	5	9	11.973	7.87,12.00		3.39,3.70
1	1	7	7	12.000	7.96,12.00		3.38,3.68
1	1	9	5	11.973	8.00,12.00		3.39,3.67
1	1	20	20	38.000	32.22,38.00		2.85,2.88
1	1	15	5	17.921	12.17,18.00		3.17,3.[illegible]
1	1	25	15	37.996	31.79,38.00		2.85,2.90
1	1	30	10	37.969	31.93,38.00		2.87,2.9[illegible]
1	1	50	50	98.000	90.96,98.00		2.63,2.64

* For Known $\sigma_{th}^2 / \sigma_{ss}^2$

**Fifth percentile of f below lower limit of f = 6 in Table IV

Table II indicates that in the absence of knowledge of $b = \sigma_{th}^2 / \sigma_{ss}^2$, a minimum value of f of approximately 20 is required for reasonable precision of the 90% confidence estimate of R.

Figure III indicates what ratio of sample sizes, $a = n_{th}/n_{ss}$ might be selected to obtain large f and therefore high precision of K as a function of $b = \sigma_{th}^2 / \sigma_{ss}^2$ and n_{ss}. The curves in figure III were plotted by putting equation (4) in the form:

$$f \simeq \frac{(ab+1)^2}{\frac{(ab)^2}{an_{ss}-1} + \frac{1}{n_{ss}-1}} \qquad (11)$$

Examination of this equation shows the following limiting characteristics when $n_{ss} > 1$:

As

$$b \to 0, \quad f \to n_{ss} - 1$$
$$b \to \infty, \quad f \to n_{th} - 1$$
$$a \to 0, \quad f \to n_{ss} - 1$$
$$a \to \infty, \quad f \to \infty\,.$$

Also, for $a = 1$, i.e., $n_{th} = n_{ss} = n$, $b = 1$ gives a maximum value of $f = 2(n-1)$.

It is of interest to note that in figure III, if we can assume only that $b = \sigma_{th}^2 / \sigma_{ss}^2 > 1$ then selection of $n_{th}/n_{ss} \geq 1$ gives fairly consistent levels of f, and therefore consistent precision for specified values of n_{th} and n_{ss} due to near-flatness of the curve in that region. Conversely if $\sigma_{ss}^2 / \sigma_{th}^2 > 1$, it is advantageous to select $n_{ss}^2/n_{th}^2 \geq 1$.

Table II shows the 5th and 95th percentiles of K for various combinations of σ_{th}^2, σ_{ss}^2, n_{th} and n_{ss}, with corresponding f. For a ratio s_{th}^2/s_{ss}^2, combinations of n_{th} and n_{ss} may be selected on the basis of small differences of the 5th and 95th percentiles.

If such sample sizes are prohibitive, a possible solution to the problem would be to determine ratios of variances of strength and stress on the basis of experience. Determination of levels of each variance is unnecessary if the ratios are known precisely. Then dependent upon f, n_{th}, n_{ss}, and b, table I would indicate whether or not to use the Monte-Carlo or analytical solution of K, as a function of the permissible amount of bias in the analytical method. As mentioned in reference (3), this bias vanishes if one of the variances is overwhelmingly larger than the other.

Appendix B

DERIVATION OF THE METHOD. Denote x_{th} as strength, and x_{ss} as stress. For brevity's sake, define the probability that stress is less than strength as "stress-strength reliability," R. Then

(12)
$$R = P(x_{ss} < x_{th})$$
$$= P(x_{th} - x_{ss} > 0).$$

Assuming that x_{th} is distributed normally with mean μ_{th} and standard deviation σ_{th} that x_{ss} is distributed normally with mean μ_{ss} and standard deviation σ_{ss}, and that x_{th} and x_{ss} are independent, equation (12) may be written as (reference 6):

(13)
$$R = \frac{1}{\sqrt{2\pi(\sigma_{th}^2 + \sigma_{ss}^2)}} \int_0^{\infty} e^{-\frac{\left[x_{th} - x_{ss} - (\mu_{th} - \mu_{ss})\right]^2}{2(\sigma_{th}^2 + \sigma_{ss}^2)}} \cdot d(x_{th} - x_{ss}).$$

This integral is shown in figure IV:

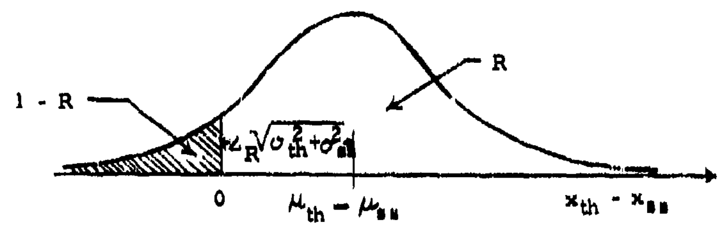


Figure IV
Stress-Strength Reliability

In other words equation (13) states that $x_{th} - x_{ss}$ is normally distributed with mean $\mu_{th} - \mu_{ss}$ and standard deviation $\sqrt{\sigma_{th}^2 + \sigma_{ss}^2}$. Actual stress-strength reliability is a function of the number, Z_R, of such standard deviations between 0 and the mean $(\mu_{th} - \mu_{ss})$. Clearly, the more standard deviations by which $\mu_{th} - \mu_{ss}$ exceeds 0, the higher the stress-strength reliability. When the population means and standard deviations are known, stress-strength reliability is determined (100% confidence) from

$$(14) \qquad Z_R = \frac{\mu_{th} - \mu_{ss}}{\sqrt{\sigma_{th}^2 + \sigma_{ss}^2}}$$

in which case R is obtained from Z_R (one-sided) using a table of the normal distribution

Unfortunately, we seldom know the population means and standard deviations for use in equation (14). In such cases, we can only estimate from samples (but not less than, say, 5) of $\bar{x}_{th} - \bar{x}_{ss}$, the sample estimate of the numerator of equation (14), and $\sqrt{s_{th}^2 + s_{ss}^2}$, the sample estimate of the denominator ($\bar{x}$, the sample mean, denotes estimate of μ; s, the sample standard deviation, denotes estimate of σ). Using such sample data, a procedure for demonstrating stress-strength reliability at a given confidence level is as follows: Find a constant, K, large enough that $\bar{x}_{th} - \bar{x}_{ss} - K\sqrt{s_{th}^2 + s_{ss}^2}$ is less than the proportion R of the population of strength-minus-stress values 100C% of the time. This may be stated mathematically as

$$P(\bar{x}_{th} - \bar{x}_{ss} - K\sqrt{s_{th}^2 + s_{ss}^2} \leq \mu_{th} - \mu_{ss} - Z_R\sqrt{\sigma_{th}^2 + \sigma_{ss}^2}) = C. \quad (15)$$

Then, if $\bar{x}_{th} - \bar{x}_{ss} - K\sqrt{s_{th}^2 + s_{ss}^2} \geq 0$, reliability of at least R is demonstrated at confidence C (figure V). Or, if $\hat{K} = (\bar{x}_{th} - \bar{x}_{ss})/\sqrt{s_{th}^2 + s_{ss}^2} \geq K$, at least R is demonstrated at confidence C.

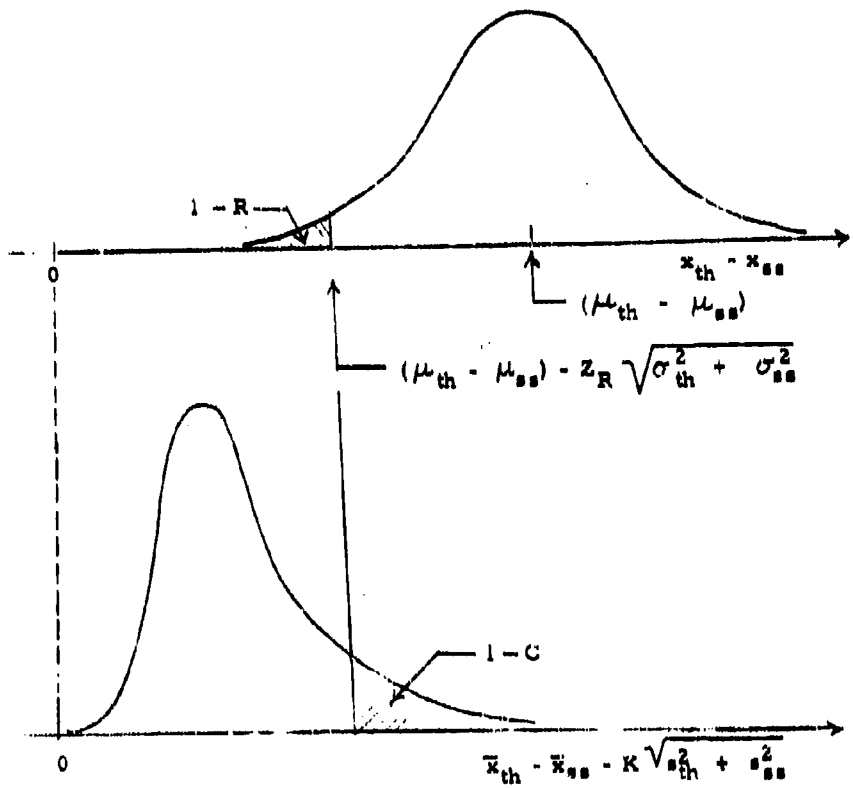


Figure V
Reliability and Confidence

The approach given herein for finding K is an application of theory developed by B. L. Welch and also investigated by Alice A. Aspin. Mr. Welch developed a statistic approximately distributed as χ^2 (reference 3, p. 31); which may be applied as follows in the bivariate case:

(16)
Let

$$s = \lambda_1 s_1^2 + \lambda_2 s_2^2 ,$$

(17)

$$g = \frac{\lambda_1 \sigma_1^4/f_1 + \lambda_2 \sigma_2^4/f_2}{\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2} .$$

Then s/g is approximately distributed as χ^2 with f degrees of freedom, where λ_1 and λ_2 are constants and

(18)

$$f \simeq \frac{(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)^2}{\lambda_1^2 \sigma_1^4/f_1 + \lambda_2^2 \sigma_2^4/f_2} .$$

(19)
Thus

$$\frac{s/g}{f} \simeq \frac{\lambda_1 s_1^2 + \lambda_2 s_2^2}{\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2}$$

is approximately distributed as χ^2/f. In a later paper (reference 4, p. 245), Mr. Welch indicates that in the absence of knowledge of σ_i, f may be estimated by replacing each σ_i in equation (18) with its estimate s_i.

The approximate non-central t statistic utilized herein is defined in the usual manner, substituting χ^2/f as defined above.

SOLUTION OF K, EQUAL SAMPLE SIZES, KNOWN OR ESTIMATED RATIOS (b) OF VARIANCES. If the number of stress measurements, n, is equal to the number of strength measurements, equation (15) may be written as:

$$(20) \quad P\left\{ \frac{\dfrac{\bar{x}_{th} - \bar{x}_{ss} - (\mu_{th} - \mu_{ss})}{\sqrt{(\sigma_{th}^2 + \sigma_{ss}^2)/n}} + Z_R\sqrt{n}}{\sqrt{\dfrac{s_{th}^2 + s_{ss}^2}{\sigma_{th}^2 + \sigma_{ss}^2}}} \leq K\sqrt{n} \right\} = C,$$

By the method of reference 3, equations (23) - (26), the left side of the inequality is approximately the non-central t statistic with parameter of non-centrality $Z_R\sqrt{n}$, which is the same parameter used in calculating tables of one-sided tolerance factors for normal distributions (bold-face values, table II of reference 5). Also, because the right side of the inequality is the same, K is obtained directly from such a table, (table III) as a function of R, C, and effective sample size. (Table III herein is the same as the bold-face values of table II, reference 5, but more complete). Substituting $\lambda_i = 1$ and $f_i = n - 1$ in equation (18) above:

$$(21) \quad f \simeq \frac{(n-1)(\sigma_{th}^2 + \sigma_{ss}^2)^2}{\sigma_{th}^4 + \sigma_{ss}^4} .$$

Which may be estimated as described in the narrative following equation (19) as:

$$(22) \quad f \approx \frac{(n-1)(s_{th}^2 + s_{ss}^2)^2}{s_{th}^4 + s_{ss}^4} .$$

In the absence of knowledge of $b = \sigma_{th}^2 / \sigma_{ss}^2$, the effective sample size for entry in table III (taken by permission from reference 8) is $f + 1$:

$$\text{(23)} \quad \text{Effective sample size} \approx \frac{(n-1)(s_{th}^2 + s_{ss}^2)^2}{s_{th}^4 + s_{ss}^4} + 1 .$$

It may be noted that for the case $\sigma_{th}^2 = b\ \sigma_{ss}^2$, the effective sample size is independent of the standard deviation. In that case, Effective Sample Size $\cong 1 + (n-1)(b+1)^2/2$.

SOLUTION OF K, UNEQUAL SAMPLE SIZES, KNOWN OR ESTIMATED RATIO (b) OF VARIANCES. If the number of stress measurements, n_{ss}, does not equal the number of strength measurements n_{th}, equation (15) may be written as:

$$\text{(24)} \quad P \left\{ \frac{\bar{x}_{th} - \bar{x}_{ss} - (\mu_{th} - \mu_{ss}) + z_R\sqrt{\sigma_{th}^2 + \sigma_{ss}^2}}{\sqrt{n_{th}\left(\frac{s_{th}^2}{n_{th}}\right) + n_{ss}\left(\frac{s_{ss}^2}{n_{ss}}\right)}} \leq K \right\} = C$$

or

(25)

$$P\left\{ \frac{\dfrac{\bar{x}_{th} - \bar{x}_{ss} - (\mu_{th} - \mu_{ss})}{\sqrt{\dfrac{\sigma_{th}^2}{n_{th}} + \dfrac{\sigma_{ss}^2}{n_{ss}}}} + z_R \sqrt{\dfrac{\sigma_{th}^2 + \sigma_{ss}^2}{\dfrac{\sigma_{th}^2}{n_{th}} + \dfrac{\sigma_{ss}^2}{n_{ss}}}}}{\sqrt{\dfrac{n_{th}\left(\dfrac{s_{th}^2}{n_{th}}\right) + n_{ss}\left(\dfrac{s_{ss}^2}{n_{ss}}\right)}{n_{th}\left(\dfrac{\sigma_{th}^2}{n_{th}}\right) + n_{ss}\left(\dfrac{\sigma_{ss}^2}{n_{ss}}\right)}}} \le K \sqrt{\frac{\sigma_{th}^2 + \sigma_{ss}^2}{\dfrac{\sigma_{th}^2}{n_{th}} + \dfrac{\sigma_{ss}^2}{n_{ss}}}} \right\} = C.$$

By definition of non-central t, the quantity in brackets on the left side of the inequality of equation (25) is approximately the non-central t statistic with parameter of non-centrality:

(26)

$$\delta = z_R \sqrt{\frac{\sigma_{th}^2 + \sigma_{ss}^2}{\dfrac{\sigma_{th}^2}{n_{th}} + \dfrac{\sigma_{ss}^2}{n_{ss}}}} .$$

For $R > 0.50$, z_R is the positive one-sided standard normal deviate corresponding to R. If $\sigma_{th}^2 = b\sigma_{ss}^2$, equation (26) becomes:

(26A)

$$\delta = z_R \sqrt{\frac{b + 1}{\dfrac{b}{n_{th}} + \dfrac{1}{n_{ss}}}} .$$

The effective degrees of freedom, f, may also be obtained in the manner of equation (18), where $\lambda_i = n_i$:

(27) $$f \simeq \frac{(n_{th}\,\sigma_{th}^2 + n_{ss}\,\sigma_{ss}^2)^2}{n_{th}^2\,\sigma_{th}^4/(n_{th}-1) + n_{ss}^2\,\sigma_{ss}^4/(n_{ss}-1)}$$

which may be estimated by:

(28) $$f \approx \frac{(n_{th}s_{th}^2 + n_{ss}s_{ss}^2)^2}{n_{th}^2 s_{th}^4/(n_{th}-1) + n_{ss}^2 s_{ss}^4/(n_{ss}-1)}\,.$$

As n_{th} and n_{ss} become large the approximation to the non-central t distribution improves.

In the event that $\sigma_{th}^2 = b\,\sigma_{ss}^2$, equation (27) becomes independent of the standard deviation:

(28A) $$f \simeq \frac{(bn_{th} + n_{ss})^2}{\dfrac{(bn_{th})^2}{n_{th}-1} + \dfrac{n_{ss}^2}{n_{ss}-1}}\,.$$

The critical non central t value is now obtained from table IV, as a function of δ of equation (26) and f of equation (28), by the method of reference 7 as follows:

First, compute

(29) $$\eta = \frac{\delta/\sqrt{2f}}{\sqrt{1 + \delta^2/2f}}\,.$$

Second, find λ in table II as a function of η, f, and confidence $C = 1 - \epsilon$. Turn to the page corresponding to the appropriate value of ϵ. Then interpolate linearly between values of η, and linearly between values of f when $f < 9$. When $f > 9$, interpolate on $12/\sqrt{f}$ between values of f. For confidence levels between .50 and .95 and reliability levels of .50 or more, use only positive values of λ (see reference 7). The critical non-central t exceeded by $\epsilon = 1 - C$ is:

$$(30) \qquad t_{1-\epsilon} \simeq \frac{\delta + \lambda\sqrt{1 + \delta^2/2f - \lambda^2/2f}}{1 - \lambda^2/2f}$$

And, as indicated by equations (25) and (26), the critical value of K exceeded by 1 - C is:

$$(31) \qquad K \simeq \frac{t_{1-\epsilon}}{\delta / z_R} .$$

(Note that δ / z_R is the radical in equation (26), which has been computed in the process of calculation).

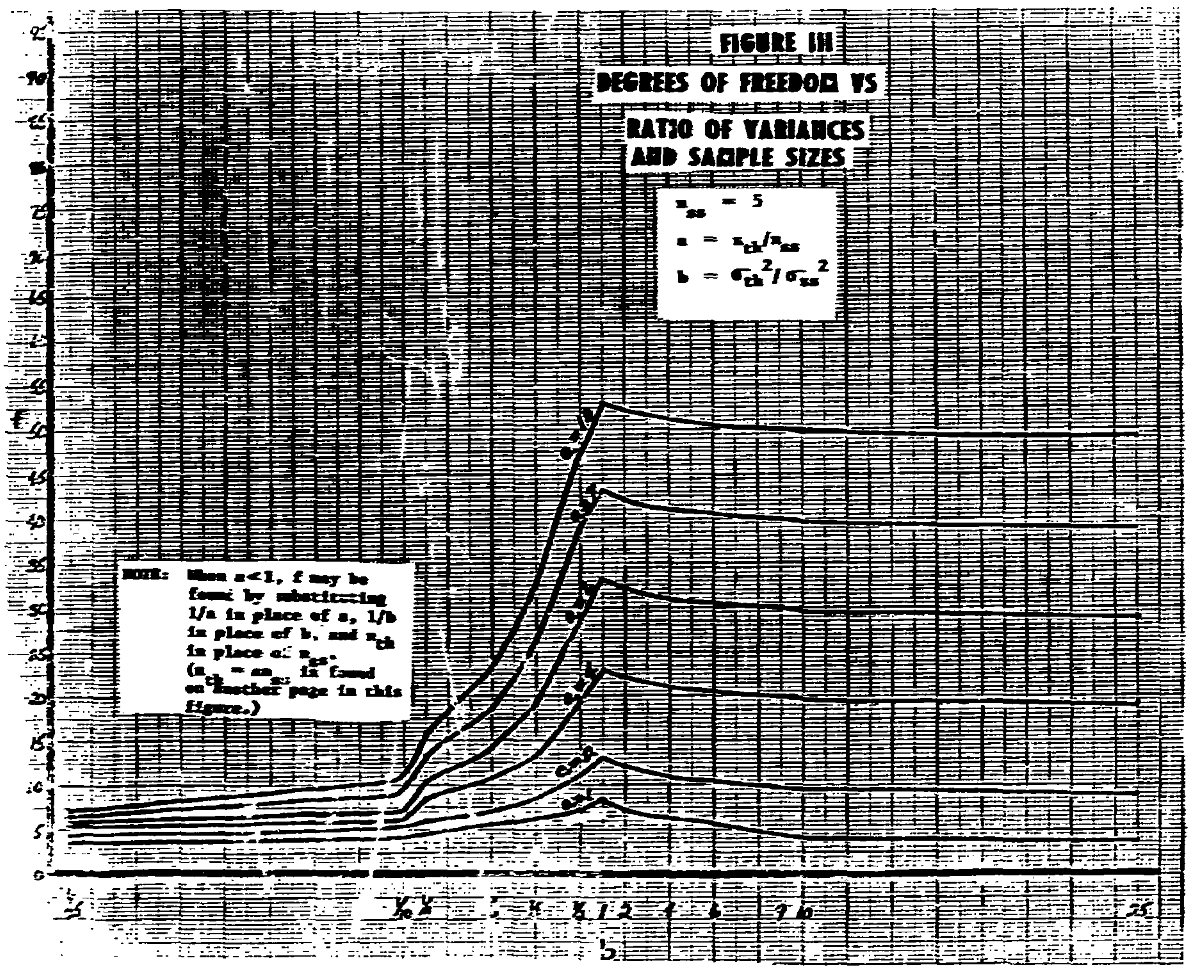
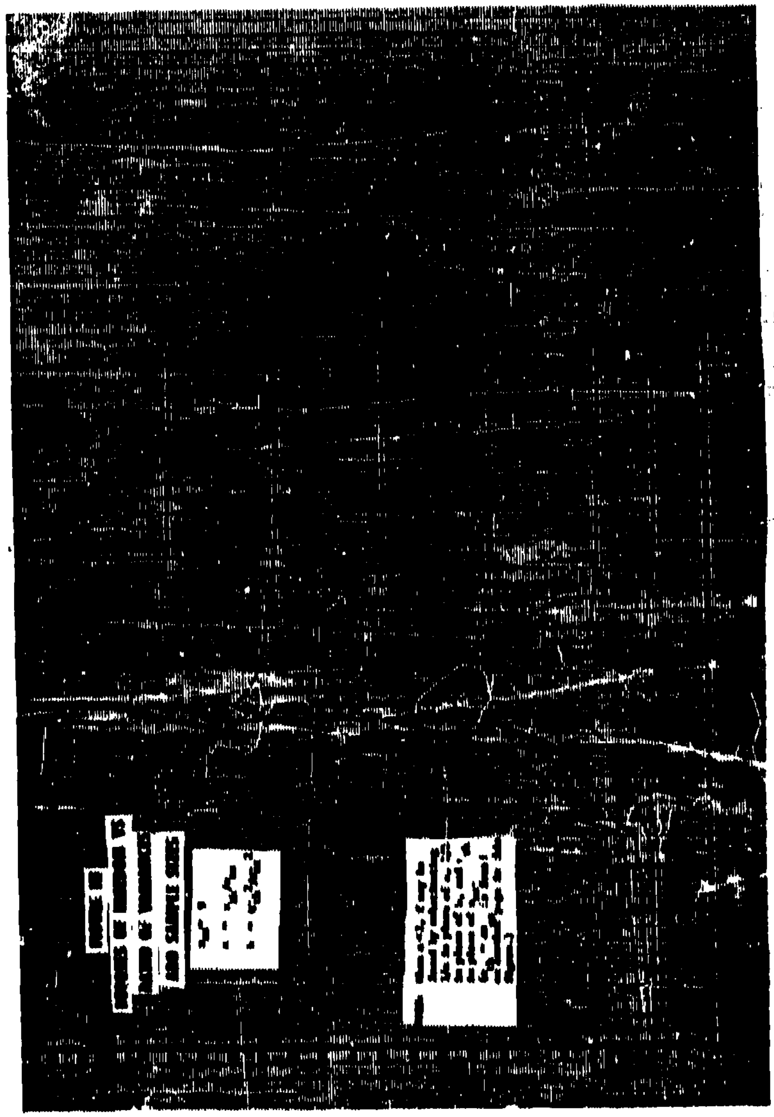
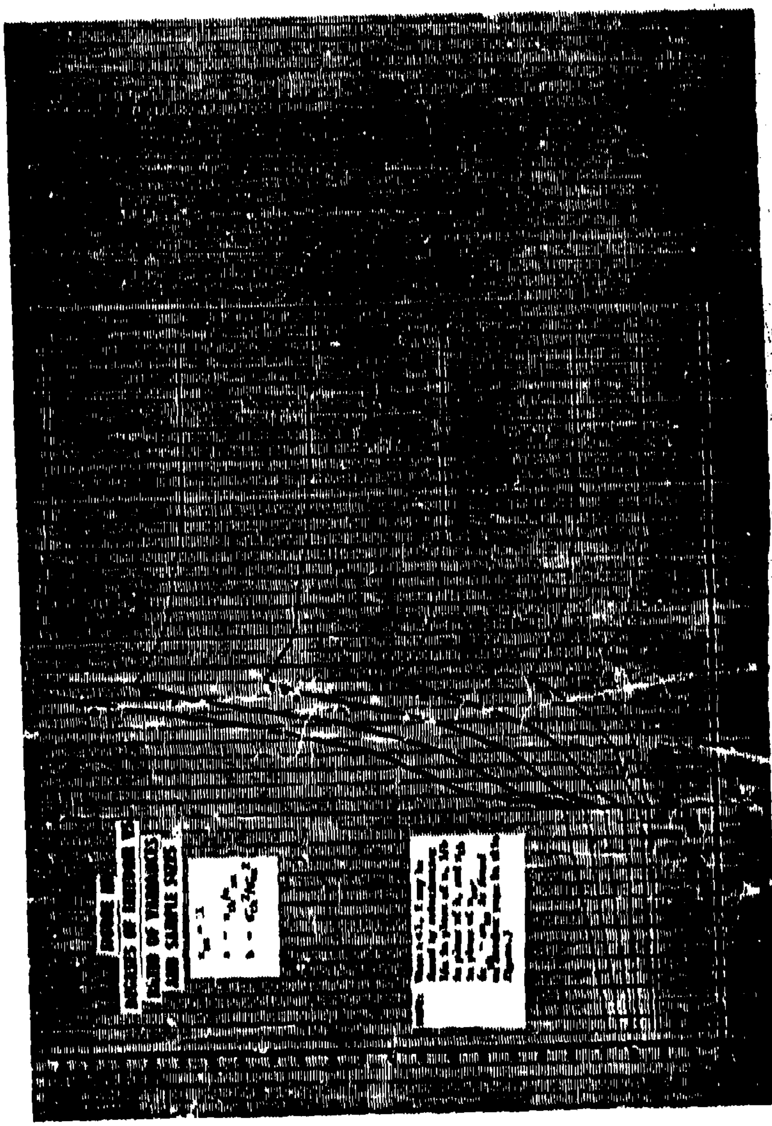


FIGURE III
DEGREES OF FREEDOM VS
RATIO OF VARIANCES
AND SAMPLE SIZES
$n_{ss} = 5$
$a = n_{th}/n_{ss}$
$b = \sigma_{th}^2/\sigma_{ss}^2$
NOTE: When a<1, f may be found by substituting 1/a in place of a, 1/b in place of b, and n_{th} in place of n_{ss}. ($n_{th} = an_{ss}$; is found on another page in this figure.)
f
b





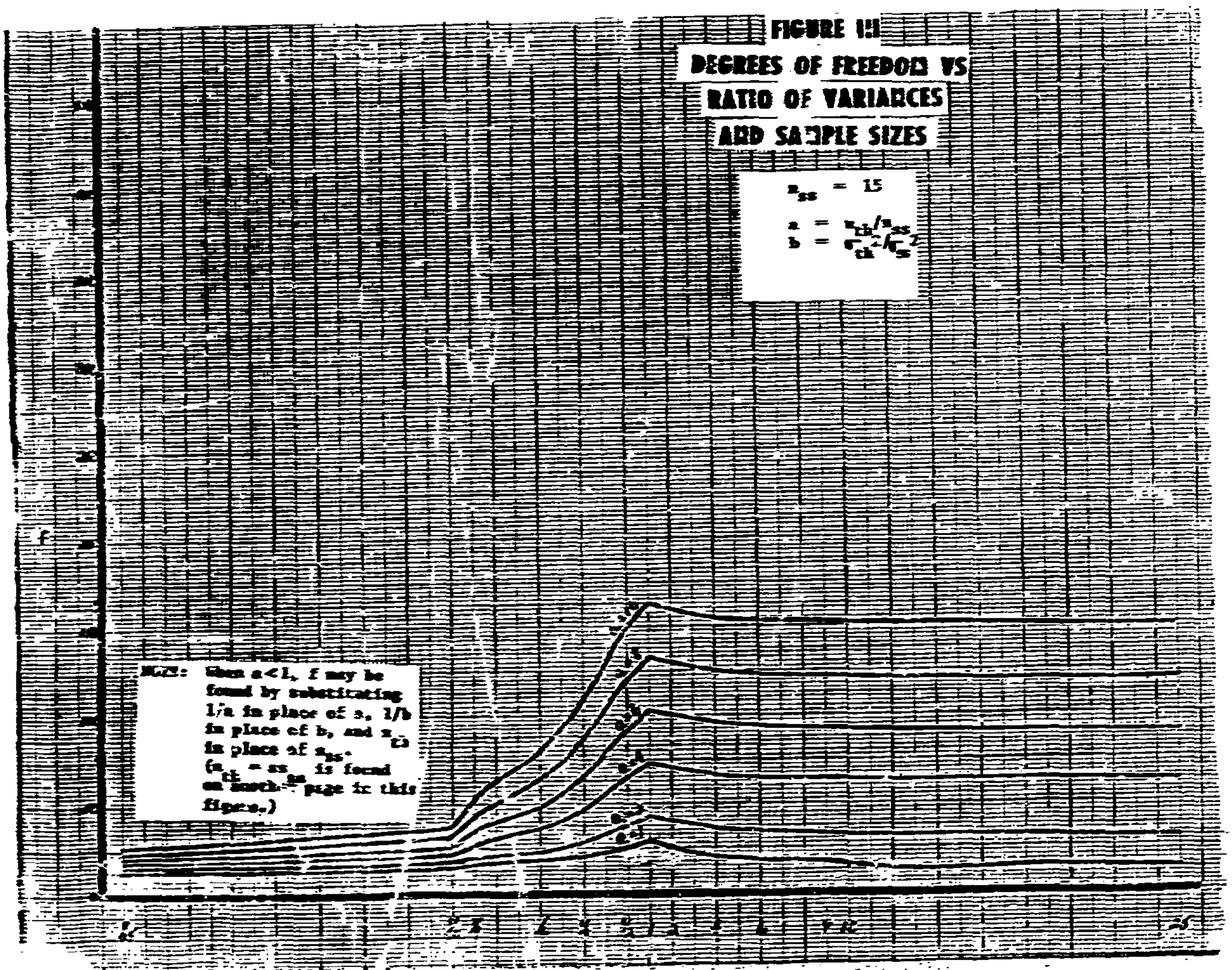


FIGURE
DEGREES OF FREEDOM VS
RATIO OF VARIANCES
AND SAMPLE SIZES
$n_{ss} = 15$
$a = n_{th}/n_{ss}$
$b = \sigma_{th}^2/\sigma_{ss}^2$
NOTE: When a < 1, f may be found by substituting 1/a in place of a, 1/b in place of b, and n_{th} in place of n_{ss}.
$(n_{th} = n_{ss}$ is found on another page in this figure.)

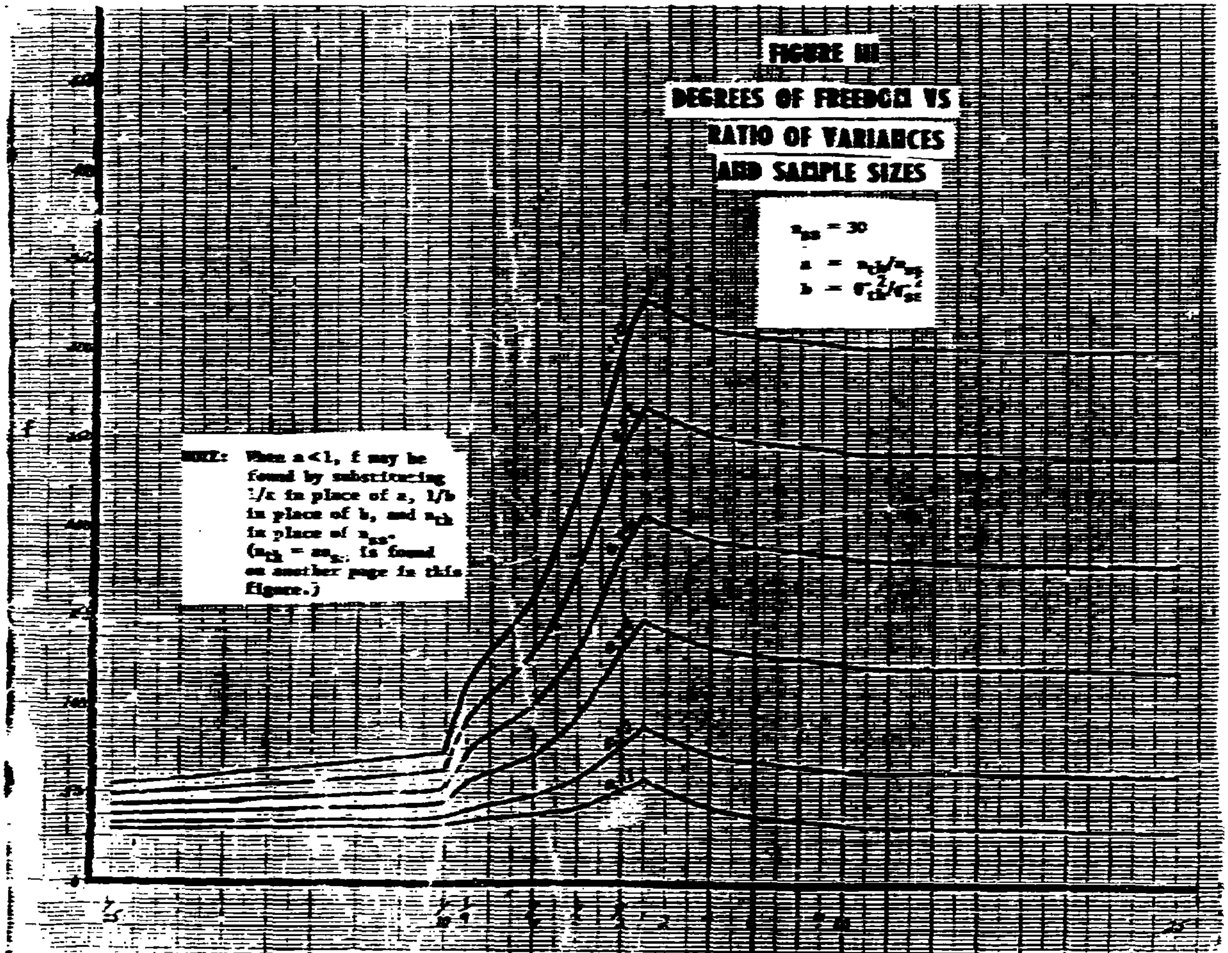
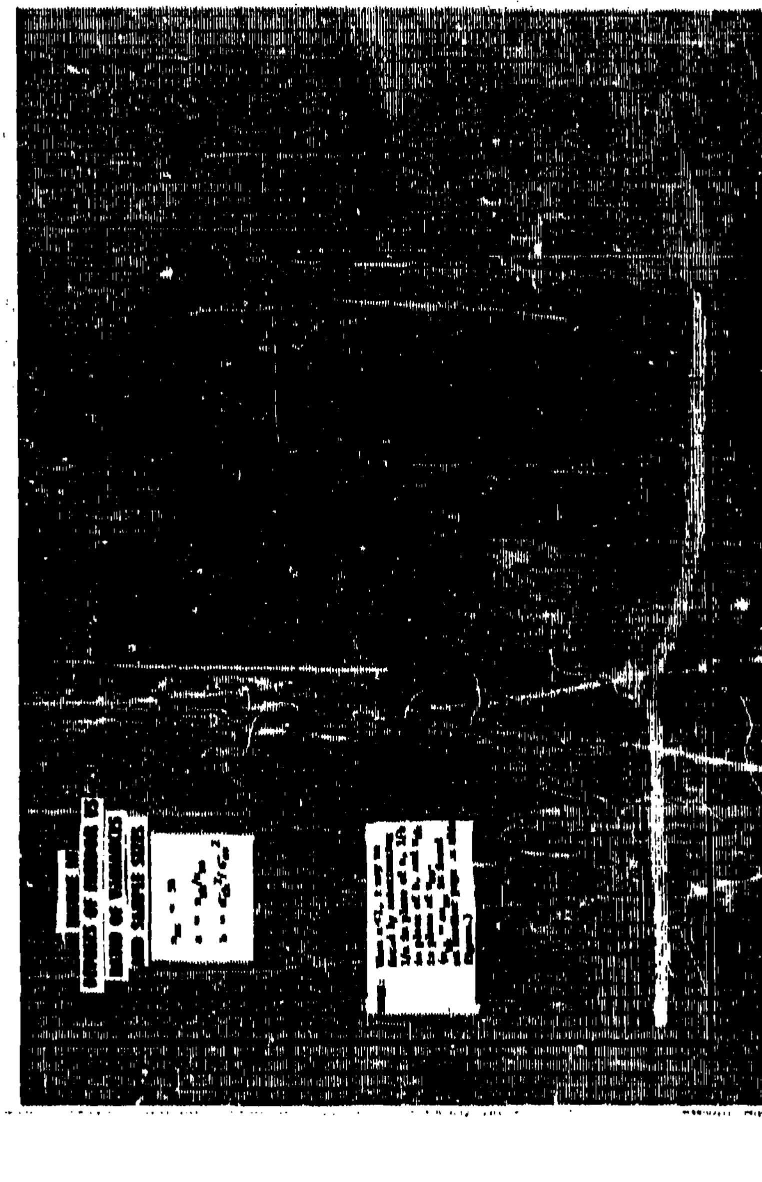


FIGURE III
DEGREES OF FREEDOM VS RATIO OF VARIANCES AND SAMPLE SIZES
$n_{se} = 30$
$a = n_{th}/n_{se}$
$b = \sigma_{th}^2/\sigma_{se}^2$
NOTE: When a < 1, f may be found by substituting 1/a in place of a, 1/b in place of b, and n_{th} in place of n_{se}. (n_{th} = [illegible] is found on another page in this figure.)
f



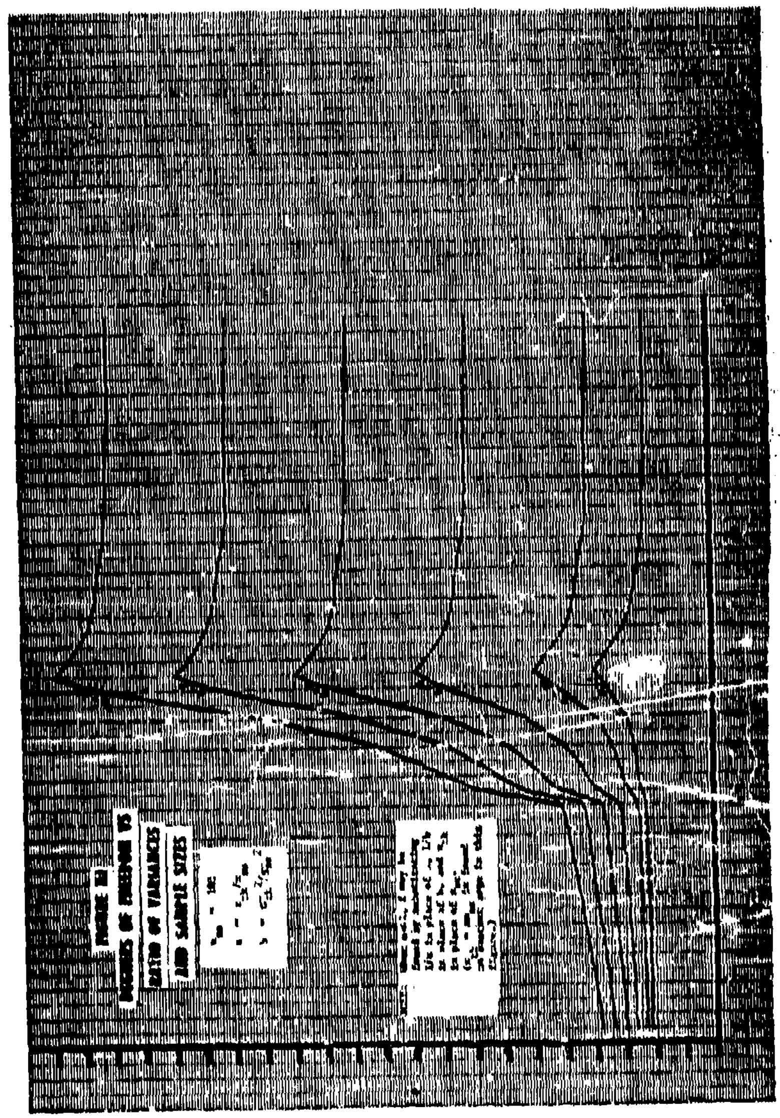

DEGREES OF FREEDOM VS
RATIO OF VARIANCES
AND SAMPLE SIZES

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.5000

SAMPLE SIZE	**P** 0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
5.	0.	0.2695	0.5603	0.7[illegible]	0.9028	1.1144	1.2386	1.3817	1.4463	1.5166	1.5938
6.	0.	0.2666	0.5527	0.[illegible]	0.8898	1.0977	1.2198	1.3605	1.4241	1.4931	1.5691
7.	0.	0.2643	0.5483	0.7058	0.8814	1.0868	1.2075	1.3466	1.4093	1.4774	1.5523
8.	0.	0.2628	0.5444	0.70[illegible]0	0.8758	1.0799	1.1993	1.3369	1.3992	1.4668	1.5411
9.	0.	0.2615	0.5420	0.69[illegible]7	0.8712	1.0739	1.1929	1.3299	1.3917	1.4588	1.5327
10.	0.	0.2606	0.5399	0.6950	0.8678	1.0697	1.1879	1.3241	1.3856	1.4525	1.5260
11.	0.	0.2598	0.5383	0.6929	0.8652	1.0662	1.1841	1.3197	1.3810	1.4476	1.5208
12.	0.	0.2592	0.5371	0.6915	0.8630	1.0635	1.1809	1.3162	1.3772	1.4436	1.5166
13.	0.	0.2588	0.5360	0.6893	0.8612	1.0611	1.1783	1.3132	1.3741	1.4403	1.5131
14.	0.	0.2583	0.5351	0.6880	0.8596	1.0592	1.1761	1.3107	1.3714	1.4375	1.5101
15.	0.	0.2580	0.5343	0.6873	0.8583	1.0575	1.1742	1.3085	1.3692	1.4351	1.5076
16.	0.	0.2577	0.5336	0.6867	0.8572	1.0561	1.1726	1.3067	1.3672	1.4330	1.5054
17.	0.	0.2574	0.5331	0.6859	0.8562	1.0548	1.1711	1.3051	1.3655	1.4312	1.5035
18.	0.	0.2572	0.5325	0.6852	0.8553	1.0537	1.1699	1.3037	1.3640	1.4297	1.5018
19.	0.	0.2569	0.5321	0.6846	0.8545	1.0528	1.1688	1.3024	1.3627	1.4283	1.5004
20.	0.	0.2568	0.5317	0.6841	0.8539	1.0519	1.1678	1.3013	1.3615	1.4270	1.4991
21.	0.	0.2566	0.5313	0.6836	0.8532	1.0511	1.1669	1.3003	1.3605	1.4259	1.4979
22.	0.	0.2564	0.5310	0.6831	0.8527	1.0504	1.1661	1.2994	1.3595	1.4249	1.4968
23.	0.	0.2563	0.5307	0.6827	0.8522	1.0498	1.1654	1.2986	1.3587	1.4240	1.4958
24.	0.	0.2562	0.5304	0.6824	0.8517	1.0492	1.1647	1.2978	1.3579	1.4232	1.4949
25.	0.	0.2560	0.5302	0.6821	0.8513	1.0486	1.1641	1.2971	1.3572	1.4224	1.4941
30.	0.	0.2556	0.5292	0.6807	0.8496	1.0465	1.1617	1.2944	1.3543	1.4194	1.4909
35.	0.	0.2552	0.5284	0.6798	0.8484	1.0450	1.1600	1.2925	1.3523	1.4172	1.4887
40.	0.	0.2550	0.5279	0.6791	0.8475	1.0439	1.1588	1.2911	1.3508	1.4157	1.4870
45.	0.	0.2548	0.5275	0.6786	0.8469	1.0431	1.1578	1.2900	1.3496	1.4145	1.4857
50.	0.	0.2547	0.5272	0.67[illegible]2	0.8463	1.0424	1.1571	1.2891	1.3487	1.4135	1.4847
60.	0.	0.2544	0.5267	0.6775	0.8455	1.0414	1.1559	1.2878	1.3474	1.4121	1.4832
70.	0.	0.2543	0.5264	0.67[illegible]	0.8449	1.0406	1.1551	1.2869	1.3464	1.4110	1.4921
80.	0.	0.2542	0.5261	0.6768	0.8445	1.0401	1.1545	1.2862	1.3457	1.4103	1.4813
90.	0.	0.2541	0.5259	0.676[illegible]	0.8442	1.0397	1.1540	1.2857	1.3451	1.4097	1.4807
100.	0.	0.2540	0.5258	0.676[illegible]	0.8439	1.0394	1.1537	1.2853	1.3447	1.4092	1.4802
150.	0.	0.2538	0.5253	0.6757	0.8432	1.0384	1.1525	1.2840	1.3434	1.4078	1.4787
200.	0.	0.2537	0.5251	0.6754	0.8428	1.0379	1.1520	1.2834	1.3427	1.4071	1.4780
250.	0.	0.2536	0.5249	0.6752	0.8425	1.0376	1.1517	1.2830	1.3423	1.4067	1.4775
300.	0.	0.2536	0.5249	0.6751	0.8424	1.0374	1.1514	1.2828	1.3421	1.4064	1.4772
350.	0.	0.2535	0.5248	0.6750	0.8423	1.0373	1.1513	1.2826	1.3419	1.4062	1.4770
400.	0.	0.2535	0.5247	0.6749	0.8422	1.0372	1.1512	1.2825	1.3417	1.4061	1.4769
450.	0.	0.2535	0.5247	0.6749	0.8421	1.0371	1.1511	1.2824	1.3416	1.4060	1.4768
500.	0.	0.2535	0.5247	0.6748	0.8421	1.0370	1.1510	1.2823	1.3415	1.4059	1.4767
1000.	0.	0.2534	0.5245	0.6747	0.8413	1.0367	1.1507	1.2819	1.3411	1.4055	1.4762

This table is reproduced with permission from tables by Pratt & Whitney Aircraft Corporation

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.5000

P

SAMPLE SIZE	0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
5.	1.6800	1.7784	1.8943	2.0370	2.2266	2.5254	2.7989	2.9874	3.3626	3.5842	4.0535
6.	1.6539	1.7505	1.8542	2.0039	2.1897	2.4824	2.7503	2.9352	3.3036	3.5211	3.9818
7.	1.6360	1.7314	1.8435	1.9817	2.1654	2.4547	2.7197	2.9023	3.2663	3.4812	3.9364
8.	1.6242	1.7188	1.8301	1.9669	2.1488	2.4354	2.6978	2.8787	3.2395	3.4525	3.9037
9.	1.6151	1.7091	1.8196	1.9555	2.1363	2.4212	2.6820	2.8617	3.2200	3.4317	3.8799
10.	1.6080	1.7016	1.8116	1.9469	2.1268	2.4103	2.6699	2.8488	3.2053	3.4159	3.8620
11.	1.6025	1.6958	1.8053	1.9401	2.1193	2.4016	2.6602	2.8383	3.1935	3.4032	3.8475
12.	1.5981	1.6910	1.8002	1.9345	2.1131	2.3945	2.6522	2.8299	3.1838	3.3929	3.8357
13.	1.5943	1.6870	1.7959	1.9299	2.1080	2.3887	2.6457	2.8228	3.1758	3.3844	3.8260
14.	1.5912	1.6837	1.7924	1.9260	2.1037	2.3837	2.6402	2.8169	3.1691	3.3771	3.8178
15.	1.5885	1.6808	1.7893	1.9227	2.1000	2.3795	2.6355	2.8118	3.1633	3.3710	3.8108
16.	1.5862	1.6784	1.7867	1.9198	2.0969	2.3759	2.6314	2.8075	3.1584	3.3657	3.8047
17.	1.5842	1.6762	1.7844	1.9173	2.0941	2.3727	2.6278	2.8037	3.1541	3.3610	3.7995
18.	1.5824	1.6744	1.7823	1.9151	2.0917	2.3599	2.6247	2.8003	3.1503	3.3570	3.7948
19.	1.5809	1.6727	1.7805	1.9132	2.0895	2.3675	2.6220	2.7974	3.1469	3.3534	3.7907
20.	1.5795	1.6712	1.7790	1.9114	2.0876	2.3652	2.6195	2.7947	3.1439	3.3502	3.7870
21.	1.5782	1.6699	1.7775	1.9099	2.0859	2.3633	2.6173	2.7924	3.1412	3.3473	3.7838
22.	1.5771	1.6686	1.7762	1.9085	2.0843	2.3615	2.6153	2.7902	3.1387	3.3447	3.7808
23.	1.5761	1.6675	1.7750	1.9072	2.0829	2.3599	2.6135	2.7883	3.1365	3.3423	3.7781
24.	1.5751	1.6665	1.7740	1.9060	2.0816	2.3584	2.6118	2.7865	3.1345	3.3401	3.7756
25.	1.5743	1.6656	1.7730	1.9050	2.0805	2.3570	2.6103	2.7849	3.1327	3.3381	3.7734
30.	1.5709	1.6620	1.7691	1.9007	2.0758	2.3517	2.6043	2.7784	3.1254	3.3303	3.7645
35.	1.5685	1.6594	1.7663	1.8978	2.0725	2.3479	2.6001	2.7739	3.1202	3.3248	3.7582
40.	1.5667	1.6576	1.7643	1.8956	2.0701	2.3451	2.5970	2.7706	3.1164	3.3208	3.7536
45.	1.5653	1.6561	1.7628	1.8939	2.0682	2.3430	2.5946	2.7680	3.1135	3.3176	3.7500
50.	1.5642	1.6550	1.7615	1.8925	2.0667	2.3412	[illegible]	2.7660	3.1112	3.3151	3.7471
60.	1.5626	1.6532	1.7597	1.8905	2.0645	2.3387	[illegible]	2.7629	3.1077	3.3114	3.7429
70.	1.5615	1.6520	1.7584	1.8891	2.0629	2.3369	2.5878	2.7607	3.1053	3.3088	3.7399
80.	1.5606	1.6511	1.7574	1.8881	2.0618	2.3356	2.5863	2.7591	3.1034	3.3069	3.7377
90.	1.5600	1.6504	1.7566	1.9872	2.0609	2.3345	2.5851	2.7579	3.1020	3.3053	3.7360
100.	1.5594	1.6478	1.7560	1.8856	2.0602	2.3337	2.5842	2.7569	3.1009	3.3041	3.7346
150.	1.5579	1.6432	1.7542	1.8846	2.0580	2.3312	2.5815	2.7539	3.0975	3.3005	3.7305
200.	1.5571	1.6473	1.7533	1.8837	2.0569	2.3300	2.5801	2.7525	3.0959	3.2967	3.7285
250.	1.5566	1.6468	1.7528	1.8831	2.0563	2.3293	2.5793	2.7516	3.0949	3.2977	3.7272
300.	1.5563	1.6465	1.7525	1.8827	2.0559	2.3288	2.5787	2.7510	3.0942	3.2970	3.7264
350.	1.5561	1.6463	1.7522	1.8624	2.0556	2.3284	2.5783	2.7506	3.0937	3.2965	3.7259
400.	1.5559	1.6461	1.7520	1.8822	2.0553	2.3282	2.5780	2.7503	3.0934	3.2961	3.7254
450.	1.5558	1.6459	[illegible]	1.8821	2.0552	2.3280	2.5778	2.7500	[illegible]	3.2958	3.7251
500.	1.5557	1.6458	1.7517	1.8819	2.0550	2.3278	2.5776	2.7498	[illegible]	3.2956	3.7248
1000.	1.5552	1.6453	1.7512	1.8814	2.0544	2.3271	2.5768	2.7490	[illegible]	[illegible]	3.7236

THIS TABLE IS REPRODUCED WITH PERMISSION FROM

TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: X-FACTORS

CONFIDENCE LEVEL =0.6000

SAMPLE SIZE	0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
5.	0.1208	0.3960	0.6979	0.8686	1.0603	1.2863	1.4198	1.5740	1.6438	1.7198	1.8034
6.	0.1090	0.3797	0.6759	0.8420	1.0289	1.2492	1.3790	1.5289	1.5968	1.6706	1.7518
7.	0.1001	0.3678	0.6597	0.8236	1.0076	1.2237	1.3510	1.4980	1.5646	1.6370	1.7168
8.	0.0929	0.3590	0.6481	0.8101	0.9918	1.2052	1.3307	1.4756	1.5412	1.6126	1.6913
9.	0.0872	0.3517	0.6387	0.7993	0.9794	1.1907	1.3149	1.4583	1.5233	1.5940	1.6718
10.	0.0825	0.3457	0.6310	0.7907	0.9695	1.1792	1.3025	1.4449	1.5093	1.5794	1.6565
11.	0.0785	0.3407	0.6248	0.7836	0.9614	1.1698	1.2924	1.4339	1.4978	1.5674	1.6440
12.	0.0749	0.3364	0.6196	0.7775	0.9546	1.1620	1.2839	1.4246	1.4082	1.5574	1.6335
13.	0.0719	0.3327	0.6148	0.7724	0.9487	1.1552	1.2766	1.4157	1.4800	1.5489	1.6247
14.	0.0691	0.3294	0.6108	0.7679	0.9436	1.1494	1.2703	1.4099	1.4729	1.5415	1.6170
15.	0.0667	0.3265	0.6073	0.7639	0.9391	1.1443	1.2648	1.4039	1.4667	1.5351	1.6103
16.	0.0645	0.3239	0.6041	0.7604	0.9352	1.1398	1.2600	1.3987	1.4613	1.5294	1.6044
17.	0.0625	0.3215	0.6012	0.7572	0.9316	1.1358	1.2557	1.3940	1.4564	1.5244	1.5992
18.	0.0607	0.3194	0.5987	0.7544	0.9285	1.1322	1.2518	1.3898	1.4521	1.5199	1.5945
19.	0.0590	0.3174	0.5964	0.7518	0.9256	1.1289	1.2483	1.3860	1.4482	1.5158	1.5903
20.	0.0575	0.3157	0.5942	0.7494	0.9230	1.1260	1.2451	1.3825	1.4446	1.5121	1.5864
21.	0.0560	0.3140	0.5923	0.7473	0.9205	1.1232	1.2422	1.3794	1.4414	1.5088	1.5829
22.	0.0547	0.3125	0.5905	0.7453	0.9183	1.1207	1.2395	1.3765	1.4384	1.5057	1.5797
23.	0.0535	0.3111	0.5888	0.7434	0.9163	1.1184	1.2370	1.3738	1.4356	1.5028	1.5767
24.	0.0523	0.3098	0.5872	0.7417	0.9144	1.1163	1.2347	1.3714	1.4331	1.5002	1.5740
25.	0.0513	0.3085	0.5858	0.7401	0.9126	1.1143	1.2326	1.3691	1.4307	1.4977	1.5715
30.	0.0467	0.3033	0.5797	0.7335	0.9053	1.1061	1.2238	1.3596	1.4210	1.4876	1.5610
35.	0.0432	0.2994	0.5751	0.7285	0.8998	1.0999	1.2173	1.3526	1.4137	1.4801	1.5532
40.	0.0403	0.2962	0.5715	0.7246	0.8954	1.0951	1.2121	1.3471	1.4080	1.4742	1.5471
45.	0.0380	0.2936	0.5685	0.7213	0.8919	1.0912	1.2080	1.3426	1.4034	1.4695	1.5422
50.	0.0360	0.2915	0.5661	0.7187	0.8890	1.0880	1.2045	1.3389	1.3996	1.4656	1.5382
60.	0.0329	0.2880	0.5621	0.7144	0.8844	1.0828	1.1991	1.3331	1.3937	1.4594	1.5318
70.	0.0304	0.2853	0.5591	0.7112	0.8808	1.0790	1.1950	1.3287	1.3891	1.4548	1.5270
80.	0.0284	0.2832	0.5567	0.7086	0.8780	1.0759	1.1917	1.3253	1.3856	1.4511	1.5232
90.	0.0268	0.2814	0.5547	0.7065	0.8758	1.0734	1.1891	1.3225	1.3827	1.4481	1.5201
100.	0.0254	0.2799	0.5531	0.7047	0.8738	1.0713	1.1869	1.3201	1.3802	1.4456	1.5175
150.	0.0207	0.2749	0.5476	0.6988	0.8675	1.0643	1.1795	1.3123	1.3723	1.4374	1.5090
200.	0.0179	0.2719	0.5443	0.6954	0.8638	1.0603	1.1753	1.3079	1.3677	1.4327	1.5042
250.	0.0160	0.2699	0.5421	0.6931	0.8613	1.0576	1.1725	1.3049	1.3646	1.4295	1.5009
300.	0.0146	0.2685	0.5405	0.6914	0.8595	1.0557	1.1705	1.3027	1.3624	1.4272	1.4986
350.	0.0136	0.2673	0.5393	0.6901	0.8581	1.0542	1.1689	1.3010	1.3607	1.4255	1.4968
400.	0.0127	0.2664	0.5383	0.6890	0.8570	1.0530	1.1676	1.2997	1.3593	1.4241	1.4953
450.	0.0120	0.2657	0.5373	0.6882	0.8561	1.0520	1.1665	1.2986	1.3582	1.4229	1.4941
500.	0.0113	0.2650	0.5369	0.6875	0.8553	1.0511	1.1657	1.2977	1.3572	1.4220	1.4931
1000.	0.0080	0.2616	0.5331	0.6833	0.8512	1.0467	1.1610	1.2928	1.3522	1.4168	1.4879

THIS TABLE IS REPRODUCED WITH PERMISSION FROM TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.8000

SAMPLE SIZE	P 0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
5.	1.8962	2.0037	2.1297	2.2850	2.4918	2.8183	3.1177	3.3243	3.7361	3.9795	4.4954
6.	1.8426	1.9463	2.0684	2.2190	2.4195	2.7359	3.0261	3.2264	3.6256	3.8619	4.3622
7.	1.8059	1.9077	2.0274	2.1749	2.3713	2.6812	2.9655	3.1617	3.5531	3.7844	4.2747
8.	1.7792	1.8795	1.9976	2.1430	2.3366	2.6421	2.9223	3.1156	3.5011	3.7290	4.2119
9.	1.7587	1.8580	1.9748	2.1186	2.3101	2.6122	2.8892	3.0804	3.4615	3.6868	4.1643
10.	1.7427	1.8411	1.9568	2.0993	2.2891	2.5886	2.8632	3.0526	3.4304	3.6537	4.1269
11.	1.7296	1.8273	1.9422	2.0837	2.2722	2.5695	2.8421	3.0302	3.4052	3.6268	4.0966
12.	1.7186	1.8158	1.9301	2.0708	2.2581	2.5536	2.8246	3.0116	3.3843	3.6046	4.0715
13.	1.7094	1.8061	1.9198	2.0598	2.2461	2.5402	2.8098	2.9958	3.3667	3.5859	4.0504
14.	1.7014	1.7977	1.9109	2.0503	2.2359	2.5287	2.7972	2.9823	3.3515	3.5698	4.0323
15.	1.6944	1.7903	1.9032	2.0421	2.2270	2.5187	2.7861	2.9705	3.3384	3.5558	4.0165
16.	1.6882	1.7839	1.8964	2.0348	2.2191	2.5099	2.7764	2.9603	3.3269	3.5436	4.0027
17.	1.6828	1.7782	1.8903	2.0284	2.2121	2.5021	2.7679	2.9512	3.3167	3.5327	3.9905
18.	1.6779	1.7730	1.8849	2.0226	2.2059	2.4951	2.7602	2.9430	3.3075	3.5230	3.9796
19.	1.6735	1.7684	1.8801	2.0174	2.2003	2.4888	2.7533	2.9357	3.2994	3.5143	3.9698
20.	1.6695	1.7642	1.8756	2.0127	2.1952	2.4831	2.7470	2.9291	3.2919	3.5064	3.9609
21.	1.6658	1.7604	1.8716	2.0084	2.1906	2.4779	2.7414	2.9230	3.2852	3.4992	3.9528
22.	1.6625	1.7569	1.8679	2.0045	2.1864	2.4732	2.7361	2.9175	3.2790	3.4927	3.9454
23.	1.6594	1.7537	1.8645	2.0009	2.1825	2.4688	2.7314	2.9124	3.2733	3.4866	3.9387
24.	1.6565	1.7507	1.8614	1.9976	2.1788	2.4648	2.7269	2.9077	3.2681	3.4811	3.9324
25.	1.6539	1.7479	1.8584	1.9945	2.1755	2.4611	2.7228	2.9034	3.2632	3.4759	3.9266
30.	1.6430	1.7365	1.8465	1.9818	2.1618	2.4458	2.7061	2.8856	3.2434	3.4549	3.9030
35.	1.6349	1.7280	1.8376	1.9723	2.1517	2.4345	2.6937	2.8725	3.2287	3.4393	3.8855
40.	1.6285	1.7214	1.8307	1.9650	2.1438	2.4257	2.6841	2.8623	3.2174	3.4273	3.8720
45.	1.6234	1.7161	1.8251	1.9591	2.1374	2.4187	2.6764	2.8541	3.2083	3.4176	3.8612
50.	1.6192	1.7118	1.8205	1.9542	2.1322	2.4128	2.6700	2.8474	3.2008	3.4097	3.8523
60.	1.6126	1.7049	1.8135	1.9466	2.1240	2.4038	2.6601	2.8369	3.1891	3.3973	3.8384
70.	1.6076	1.6997	1.8078	1.9409	2.1179	2.3969	2.6526	2.8290	3.1803	3.3880	3.8280
80.	1.6037	1.6956	1.8036	1.9364	2.1130	2.3915	2.6468	2.8228	3.1735	3.3807	3.8198
90.	1.6005	1.6922	1.8001	1.9327	2.1091	2.3872	2.6420	2.8177	3.1679	3.3748	3.8132
100.	1.5978	1.6895	1.7972	1.9296	2.1056	2.3836	2.6381	2.8136	3.1633	3.3699	3.8077
150.	1.5891	1.6804	1.7877	1.9196	2.0951	2.3717	2.6252	2.7999	3.1482	3.3539	3.7898
200.	1.5840	1.6752	1.7822	1.9[illegible]39	2.0890	2.3650	2.6178	2.7922	3.1396	3.3448	3.7797
250.	1.5807	1.6717	1.7786	1.9101	2.0849	2.3605	2.6130	2.7871	3.1339	3.3388	3.7729
300.	1.5783	1.6692	1.7760	1.9073	2.0819	2.3572	2.6095	2.7833	3.1298	3.3345	3.7681
350.	1.5764	1.6672	1.7739	1.9052	2.0797	2.3548	2.6068	2.7805	3.1266	3.3311	3.7644
400.	1.5749	1.6657	1.7723	1.9035	2.0779	2.3528	2.6046	2.7782	3.1241	3.3285	3.7614
450.	1.5737	1.6644	1.7710	1.9021	2.0764	2.3511	2.6028	2.7764	3.1221	3.3263	3.7590
500.	1.5727	1.6633	1.7699	1.9009	2.0752	2.3498	2.6014	2.7748	3.1203	3.3245	3.7570
1000.	1.5672	1.6577	1.7640	1.8948	2.0686	2.3426	2.5936	2.7666	3.1113	3.3149	3.7463

THIS TABLE IS REPRODUCED WITH PERMISSION FROM TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.7000

P

SAMPLE SIZE	0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
5.	0.2545	0.5405	0.8609	1.0438	1.2512	1.4973	1.6431	1.8122	1.8886	1.9724	2.0644
6.	0.2284	0.5082	0.8155	0.9948	1.1947	1.4315	1.5712	1.7331	1.8066	1.8857	1.9749
7.	0.2090	0.4845	0.7887	0.9610	1.1558	1.3861	1.5222	1.6799	1.7512	1.8289	1.9145
8.	0.1940	0.4664	0.7661	0.9359	1.1274	1.3531	1.4860	1.6400	1.7098	1.7859	1.8698
9.	0.1820	0.4516	0.7484	0.9158	1.1045	1.3270	1.4583	1.6103	1.6791	1.7540	1.8366
10.	0.1719	0.4400	0.7341	0.8998	1.0865	1.3065	1.4361	1.5862	1.6542	1.7282	1.8097
11.	0.1633	0.4300	0.7221	0.8866	1.0716	1.2895	1.4180	1.5666	1.6338	1.7071	1.7878
12.	0.1559	0.4215	0.7119	0.8753	1.0591	1.2753	1.4027	1.5501	1.6168	1.6894	1.7694
13.	0.1494	0.4140	0.7031	0.8656	1.0483	1.2631	1.3896	1.5360	1.6022	1.6743	1.7538
14.	0.1437	0.4075	0.6954	0.8571	1.0389	1.2525	1.3783	1.5238	1.5896	1.6613	1.7402
15.	0.1386	0.4017	0.6886	0.8497	1.0306	1.2432	1.3684	1.5131	1.5786	1.6499	1.7284
16.	0.1339	0.3964	0.6825	0.8430	1.0232	1.2349	1.3596	1.5037	1.5688	1.6398	1.7179
17.	0.1298	0.3917	0.6770	0.8370	1.0167	1.2276	1.3517	1.4952	1.5601	1.6308	1.7086
18.	0.1260	0.3875	0.6721	0.8317	1.0107	1.2209	1.3447	1.4876	1.5523	1.6227	1.7002
19.	0.1225	0.3836	0.6676	0.8268	1.0053	1.2149	1.3383	1.4808	1.5452	1.6154	1.6926
20.	0.1193	0.3800	0.6635	0.8223	1.0004	1.2095	1.3324	1.4745	1.5388	1.6087	1.6857
21.	0.1163	0.3767	0.6597	0.8182	0.9959	1.2044	1.3271	1.4688	1.5329	1.6026	1.6794
22.	0.1136	0.3737	0.6562	0.8144	0.9917	1.1998	1.3222	1.4635	1.5274	1.5970	1.6736
23.	0.1110	0.3708	0.6530	0.8109	0.9879	1.1955	1.3176	1.4586	1.5224	1.5918	1.6682
24.	0.1086	0.3682	0.6499	0.8076	0.9843	1.1915	1.3134	1.4541	1.5178	1.5370	1.6632
25.	0.1063	0.3657	0.647[illegible]	0.8045	0.9810	1.1870	1.3095	1.4499	1.5134	1.5325	1.6586
30.	0.0968	0.3553	0.6353	0.7918	0.9671	1.1725	1.2932	1.4326	1.4956	1.5641	1.6396
35.	0.0895	0.3473	0.6263	0.7821	0.9566	1.1609	1.2809	1.4195	1.4822	1.5503	1.6253
40.	0.0836	0.3409	0.6192	0.7745	0.9483	1.1518	1.2713	1.4093	1.4716	1.5394	1.6141
45.	0.0788	0.3357	0.6134	0.7683	0.9415	1.1444	1.2635	1.4010	1.4631	1.5306	1.6050
50.	0.0747	0.3313	0.6085	0.7630	0.9359	1.1383	1.2570	1.3940	1.4560	1.5233	1.5974
60.	0.0681	0.3242	0.6007	0.7547	0.9270	1.1285	1.2467	1.3831	1.4447	1.5118	1.5855
70.	0.0630	0.3186	0.5947	0.7484	0.9201	1.1210	1.2388	1.3748	1.4362	1.5030	1.5764
80.	0.0589	0.3144	0.5899	0.7433	0.9147	1.1151	1.2326	1.3682	1.4294	1.4960	1.5692
90.	0.0555	0.3107	0.5860	0.7391	0.9102	1.1102	1.2275	1.3628	1.4239	1.4903	1.5634
100.	0.0526	0.3078	0.5827	0.7356	0.9064	1.1062	1.2232	1.3582	1.4192	1.4855	1.5585
150.	0.0429	0.2976	0.57[illegible]	0.7239	0.8939	1.0926	1.2090	1.3432	1.4038	1.4697	1.5422
200.	0.0371	0.2916	0.5651	0.7170	0.8866	1.0847	1.2007	1.3345	1.3949	1.4605	1.5327
250.	0.0332	0.2875	0.5607	0.7124	0.8817	1.0794	1.1951	1.3286	1.3888	1.4543	1.5264
300.	0.0303	0.2845	0.5576	0.7090	0.8781	1.0755	1.1911	1.3243	1.3845	1.4498	1.5217
350.	0.0281	0.2821	0.5549	0.7064	0.8753	1.0725	1.1879	1.3210	1.3811	1.4464	1.51[illegible]
400.	0.0263	0.2803	0.5529	0.7043	0.8730	1.0701	1.1854	1.3183	1.3784	1.4436	1.515
450.	0.0247	0.2787	0.55[illegible]	0.7025	0.8712	1.0681	1.1833	1.3162	1.3761	1.4413	1.5130
500.	0.0235	0.2774	0.54[illegible]	0.7010	0.8696	1.0664	1.1816	1.3143	1.3742	1.4394	1.5110
1000.	0.0166	0.2703	0.5423	0.6931	0.8613	1.0574	1.1722	1.3045	1.3642	1.4291	1.5004

THIS TABLE IS REPRODUCED WITH PERMISSION FROM

TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.7000

SAMPLE SIZE	0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
5.	2.1676	2.2855	2.4245	2.5958	2.8243	[illegible]	[illegible]	3.7475	4.2060	4.4773	[illegible]
6.	2.0737	2.1867	2.3197	2.4838	2.7025	3.0483	[illegible]	3.5856	4.0238	4.2831	[illegible]
7.	2.0103	2.1198	2.2488	2.4082	2.6207	2.9566	3.2651	3.4781	3.9033	4.1548	[illegible]
8.	1.9637	2.0710	2.1974	2.3533	2.5611	2.8895	3.1912	3.3995	3.8154	4.0614	[illegible]
9.	1.9289	2.0345	2.1588	2.3121	2.5164	2.8393	3.1359	3.3407	3.7494	3.9912	[illegible]
10.	1.9009	2.0052	2.1279	2.2792	2.4808	2.7995	3.0921	3.2941	3.6974	3.9359	[illegible]
11.	1.8781	1.9812	2.1027	2.2523	2.4518	2.7670	3.0564	3.2562	3.6549	3.8908	[illegible]
12.	1.8589	1.9612	2.0815	2.2299	2.4275	2.7399	3.0266	3.2246	3.6196	3.8533	4.3481
13.	1.8426	1.9441	2.0636	2.2108	2.4069	2.7169	3.0014	3.1978	3.5897	3.8215	4.3130
14.	1.8285	1.9294	2.0481	[illegible]	2.3892	2.6971	2.9797	3.1748	[illegible]	3.7942	4.2824
15.	1.8162	1.9165	2.0345	[illegible]	2.3737	2.6798	2.9608	3.1547	[illegible]	3.7705	4.2557
16.	1.8053	1.9051	2.0226	2.1673	2.3601	2.6646	2.9441	3.1370	3.5219	3.7495	4.2322
17.	1.7956	1.8950	2.0119	2.1560	2.3479	2.6510	2.9293	3.1213	3.5044	3.7310	4.2114
18.	1.7869	1.8859	2.0024	2.1458	2.3370	2.6389	2.9160	3.1072	3.4887	3.7143	4.1927
19.	1.7790	1.8776	1.9937	2.1367	2.3272	2.6280	2.9040	3.0945	3.4745	3.6993	4.1758
20.	1.7718	1.8702	1.9859	2.1284	2.3182	2.6180	2.8931	3.0830	3.4617	3.6857	4.1605
21.	1.7652	1.8633	1.9787	2.1208	2.3100	2.6089	2.8832	3.0725	3.4500	3.6732	4.1466
22.	1.7592	1.8570	1.9721	2.1138	2.3025	2.6006	2.8741	3.0628	3.4392	3.6619	4.1338
23.	1.7537	1.8512	1.9660	2.1074	2.2956	2.5929	2.8657	3.0539	3.4293	3.6514	4.1220
24.	1.7485	1.8458	1.9604	2.1014	2.2892	2.5858	2.8579	3.0457	3.4202	3.6417	4.1112
25.	1.7437	1.8408	1.9551	2.0958	2.2833	2.5792	2.8507	3.0380	3.4117	3.6327	4.1011
30.	1.7239	1.8203	1.9336	2.0731	2.2588	2.5521	2.8211	3.0067	3.3769	3.5957	4.0597
35.	1.7091	1.8048	1.9174	2.0560	2.2405	2.5318	2.7990	2.9833	3.3509	3.5683	4.0290
40.	1.6975	1.7927	1.9048	2.0427	2.2262	2.5160	2.7817	2.9651	3.3307	3.5468	4.0050
45.	1.6881	1.7830	1.8945	2.0319	2.2147	2.5032	2.7678	2.9504	3.3144	3.5296	3.9857
50.	1.6803	1.7748	1.8860	2.0229	2.2051	2.4926	2.7563	2.9382	3.3008	3.5152	3.9697
60.	1.6679	1.7621	1.8727	2.0088	2.1900	2.4760	2.7382	2.9190	3.2796	3.4928	3.9445
70.	1.6586	1.7523	1.8625	1.9981	2.1786	2.4633	2.7244	2.9045	3.2636	3.4758	3.9256
80.	1.6511	1.7446	1.8544	1.9896	2.1695	2.4534	2.7136	2.8931	3.2509	3.4624	3.9106
90.	1.6451	1.7383	1.8479	1.9827	2.1622	2.4452	2.7048	2.8838	3.2406	3.4515	3.8984
100.	1.6400	1.7330	1.8424	1.9770	2.1560	2.4385	2.6974	2.8760	3.2320	3.4424	3.8883
150.	1.6232	1.7156	1.8242	1.9578	2.1356	2.4161	2.6731	2.8503	3.2036	3.4124	3.8549
200.	1.6134	1.7055	1.8137	1.9468	2.1239	2.4031	2.6591	2.8356	3.1873	3.3952	3.8356
250.	1.6069	1.6987	1.8066	1.9394	2.1160	2.3945	2.6497	2.8257	3.1764	3.3836	3.8228
300.	1.6021	1.6938	1.8015	1.9340	2.1103	2.3882	2.6429	2.8185	3.1685	3.3753	3.8134
350.	1.5984	1.6899	1.7975	1.9299	2.1058	2.3833	2.6376	2.8130	3.1624	3.3688	3.8063
400.	1.5955	1.6869	1.7944	1.9265	2.1023	2.3795	2.6334	2.8085	3.1575	3.3637	3.8005
450.	1.5930	1.6844	1.7911	1.9238	2.0994	2.3763	2.6300	2.8049	3.1535	3.3594	3.7958
500.	1.5910	1.6823	1.7896	1.9215	2.0970	2.3736	2.6271	2.8019	3.1501	3.3559	3.7919
1000.	1.5801	1.6719	1.7778	1.9092	2.0839	2.3593	2.6116	2.7856	3.1322	3.3370	3.7708

THIS TABLE IS REPRODUCED WITH PERMISSION FROM

TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.8000

SAMPLE SIZE	0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
	P										
5.	0.4204	0.7284	1.0795	1.2823	1.5140	1.7903	1.9545	2.1456	2.2322	2.3267	2.4310
6.	0.3754	0.6717	1.0059	1.1984	1.4173	1.6776	1.8324	2.0122	2.0937	2.1826	2.2807
7.	0.3422	0.6313	0.9549	1.1403	1.3514	1.6021	1.7505	1.9228	2.0011	2.0864	2.1805
8.	0.3166	0.6003	0.9171	1.0978	1.3030	1.5464	1.6906	1.8579	1.9339	2.0167	2.1080
9.	0.2964	0.5760	0.8879	1.0651	1.2661	1.5045	1.6456	1.8093	1.8835	1.9643	2.0535
10.	0.2793	0.5563	0.8639	1.0387	1.2365	1.4708	1.6094	1.7701	1.8430	1.9224	2.0099
11.	0.2651	0.5397	0.8441	1.0168	1.2120	1.4432	1.5797	1.7381	1.8099	1.8881	1.9744
12.	0.2528	0.5256	0.8273	0.9983	1.1915	1.4200	1.5549	1.7114	1.7823	1.8595	1.9447
13.	0.2421	0.5133	0.8129	0.9824	1.1739	1.4002	1.5338	1.6886	1.7588	1.8352	1.9195
14.	0.2326	0.5025	0.8003	0.9686	1.1586	1.3830	1.5154	1.6689	1.7385	1.8142	1.8977
15.	0.2242	0.4930	0.7891	0.9564	1.1452	1.3680	1.4994	1.6517	1.7207	1.7959	1.8787
16.	0.2166	0.4844	0.7791	0.9456	1.1332	1.3546	1.4852	1.6365	1.7051	1.7797	1.8619
17.	0.2097	0.4767	0.7703	0.9359	1.1226	1.3427	1.4726	1.6230	1.6911	1.7652	1.8470
18.	0.2035	0.4698	0.7622	0.9272	1.1130	1.3321	1.4612	1.6108	1.6785	1.7523	1.8336
19.	0.1978	0.4634	0.7549	0.9192	1.1043	1.3224	1.4509	1.5998	1.6672	1.7406	1.8214
20.	0.1925	0.4576	0.7483	0.9120	1.0963	1.3135	1.4416	1.5897	1.6568	1.7299	1.8104
21.	0.1877	0.4522	0.7421	0.9053	1.0891	1.3055	1.4330	1.5806	1.6474	1.7202	1.8003
22.	0.1832	0.4472	0.7364	0.8992	1.0823	1.2980	1.4251	1.5721	1.6387	1.7112	1.7911
23.	0.1790	0.4426	0.7312	0.8935	1.0761	1.2911	1.4178	1.5643	1.6307	1.7029	1.7825
24.	0.1751	0.4383	0.7262	0.8882	1.0704	1.2847	1.4110	1.5571	1.6233	1.6953	1.7746
25.	0.1714	0.4343	0.7217	0.8833	1.0650	1.2783	1.4047	1.5504	1.6163	1.6881	1.7672
30.	0.1560	0.4174	0.7026	0.8628	1.0427	1.2542	1.3737	1.5227	1.5879	1.6587	1.7369
35.	0.1441	0.4044	0.6881	0.8472	1.0258	1.2357	1.3571	1.5018	1.5664	1.6367	1.7141
40.	0.1346	0.3941	0.6766	0.8349	1.0125	1.2211	1.3437	1.4855	1.5496	1.6194	1.6962
45.	0.1267	0.3857	0.6672	0.8249	1.0017	1.2092	1.3312	1.4722	1.5360	1.6054	1.6818
50.	0.1201	0.3786	0.6594	0.8165	0.9927	1.1993	1.3208	1.4612	1.5247	1.5937	1.6698
60.	0.1095	0.3673	0.6458	0.8032	0.9703	1.1837	1.3044	1.4437	1.5068	1.5753	1.6508
70.	0.1013	0.3585	0.6372	0.7929	0.9673	1.1718	1.2918	1.4305	1.4931	1.5613	1.6364
80.	0.0946	0.3515	0.6295	0.7848	0.9586	1.1623	1.2818	1.4199	1.4823	1.5502	1.6249
90.	0.0892	0.3457	0.6232	0.7781	0.9515	1.1545	1.2737	1.4113	1.4735	1.5411	1.6156
100.	0.0846	0.3408	0.6179	0.7725	0.9455	1.1480	1.2669	1.4041	1.4661	1.5335	1.6078
150.	0.0689	0.3244	0.6001	0.7537	0.9254	1.1263	1.2441	1.3801	1.4415	1.5083	1.5818
200.	0.0597	0.3147	0.5897	0.7427	0.9137	1.1137	1.2305	1.3662	1.4273	1.4937	1.5668
250.	0.0533	0.3082	0.5826	0.7353	0.9058	1.1052	1.22[illegible]	1.3569	1.4177	1.4839	1.5567
300.	0.0487	0.3033	0.5774	0.7298	0.9000	1.0990	1.2155	1.3500	1.4107	1.4767	1.5493
350.	0.0451	0.2995	0.5734	0.7256	0.8956	1.0942	1.2105	1.3447	1.4053	1.4711	1.5436
400.	0.0421	0.2965	0.5702	0.7222	0.8920	1.0903	1.2065	1.3404	1.4009	1.4667	1.5390
450.	0.0397	0.2941	0.5675	0.7194	0.8890	1.0871	1.2032	1.3370	1.3974	1.4630	1.5353
500.	0.0377	0.2919	0.5652	0.7171	0.8865	1.0844	1.2004	1.3340	1.3944	1.4600	1.5321
1000.	0.0266	0.2806	0.5531	0.7044	0.8731	1.0701	1.1855	1.3183	1.3783	1.4435	1.5152

THIS TABLE IS REPRODUCED WITH PERMISSION FROM TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE VI: K-FACTORS

CONFIDENCE LEVEL =0.8000

SAMPLE SIZE	0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
5.	2.5480	2.6820	2.8402	3.0357	3.2968	3.7104	4.0911	4.3543	4.8801	5.1914	5.8523
6.	2.3907	2.5166	2.6651	2.8484	3.0932	3.4809	3.8378	4.0845	4.5777	4.8697	5.4895
7.	2.2860	2.4067	2.5490	2.7247	2.9592	3.3307	3.6724	3.9087	4.3809	4.6604	5.2538
8.	2.2103	2.3274	2.4654	2.6357	2.8630	3.2229	3.5540	3.7829	4.2402	4.5110	5.0856
9.	2.1533	2.2676	2.4023	2.5685	2.7905	3.1419	3.4650	3.6864	4.1344	4.3984	4.9588
10.	2.1080	2.2202	2.3524	2.5155	2.7332	3.0776	3.3944	3.6133	4.0506	4.3095	4.8588
11.	2.0710	2.1814	2.3116	2.4722	2.6864	3.0255	3.3372	3.5526	3.9828	4.2375	4.7779
12.	2.0401	2.1491	2.2777	2.4362	2.6476	2.9822	3.2897	3.5023	3.9267	4.1779	4.7110
13.	2.0138	2.1217	2.2489	2.4056	2.6147	2.9455	3.2496	3.4597	3.8792	4.1276	4.6545
14.	1.9912	2.0981	2.2240	2.3793	2.5864	2.9140	3.2151	3.4232	3.8385	4.0844	4.6060
15.	1.9714	2.0775	2.2024	2.3564	2.5617	2.8866	3.1851	3.3914	3.8031	4.0468	4.5639
16.	1.9540	2.0593	2.1833	2.3362	2.5400	2.8624	3.1587	3.3634	3.7720	4.0138	4.5269
17.	1.9385	2.0431	2.1663	2.3182	2.5207	2.8410	3.1353	3.3386	3.7444	3.9845	4.4941
18.	1.9246	2.0286	2.1511	2.3021	2.5034	2.8217	3.1143	3.3163	3.7196	3.9583	4.4646
19.	1.9120	2.0154	2.1373	2.2875	2.4877	2.8044	3.0953	3.2962	3.6973	3.9346	4.4380
20.	1.9005	2.0035	2.1248	2.2743	2.4735	2.7886	3.0781	3.2780	3.6770	3.9131	4.4140
21.	1.8900	1.9926	2.1134	2.2622	2.4606	2.7742	3.0624	3.2614	3.6585	3.8935	4.3921
22.	1.8804	1.9826	2.1029	2.2511	2.4487	2.7610	3.0480	3.2461	3.6416	3.8756	4.3720
23.	1.8716	1.9733	2.0932	2.2409	2.4377	2.7489	3.0347	3.2321	3.6260	3.8591	4.3535
24.	1.8633	1.9648	2.0842	2.2314	2.4276	2.7376	3.0224	3.2191	3.6116	3.8438	4.3364
25.	1.8557	1.9568	2.0759	2.2226	2.4181	2.7272	3.0111	3.2071	3.5982	3.8297	4.3205
30.	1.8243	1.9241	2.0417	2.1865	2.3794	2.6844	2.9644	3.1577	3.5434	3.7716	4.2556
35.	1.8007	1.8976	2.0160	2.1594	2.3505	2.6524	2.9296	3.1209	3.5026	3.7284	4.2073
40.	1.7822	1.8804	1.9960	2.1383	2.3279	2.6274	2.9024	3.0922	3.4708	3.6947	4.1696
45.	1.7672	1.8[illegible]49	1.9797	2.1212	2.3096	2.6072	2.8804	3.0689	3.4451	3.6675	4.1392
50.	1.7548	1.8519	1.9662	2.1070	2.2944	2.5905	2.8622	3.0497	3.4238	3.6450	4.1140
60.	1.7352	1.8316	1.9450	2.0846	2.2705	2.5641	2.8336	3.0195	3.3903	3.6096	4.0745
70.	1.7203	1.81[illegible]1	1.9289	2.0676	2.2524	2.5442	2.8119	2.9966	3.3650	3.5828	4.0447
80.	1.7085	1.8059	1.9161	2.0542	2.2381	2.5284	2.7948	2.9785	3.3450	3.5617	4.0211
90.	1.6988	1.79[illegible]9	1.9057	2.0432	2.2264	2.5156	2.7808	2.9638	3.3288	3.5445	4.0019
100	1.6908	1.7855	1.8969	2.0[illegible]41	2.2161	2.5048	2.7692	2.9515	3.3152	3.5302	3.9859
150.	1.6640	1.7578	1.8680	2.[illegible]037	2.184[illegible]	2.4693	2.7307	2.9110	3.2704	3.4828	3.9351
200.	1.6485	1.7417	1.8513	1.[illegible]962	2.1657	2.4488	2.7084	2.8875	3.2445	3.4555	3.9027
250.	1.6380	1.7309	1.8401	1.9744	2.1531	2.4351	2.6936	2.8718	3.2272	3.4372	3.8824
300.	1.6304	1.7230	1.8319	1.9[illegible]58	2.1440	2.4251	2.6827	2.8604	3.2146	3.4240	3.8676
350.	1.6246	1.7170	1.825[illegible]	1.9592	2.1370	2.4174	2.6744	2.8516	3.2049	3.4137	3.8562
400.	1.6199	1.7121	1.820[illegible]	1.9539	2.1313	2.4112	2.6677	2.8446	3.1972	3.4056	3.8471
450.	1.6160	1.7081	1.8164	1.9496	2.1267	2.4061	2.6622	2.8388	3.1908	3.3988	3.8396
500.	1.6127	[illegible].7047	1.81[illegible]9	1.9459	2.1228	2.4019	2.6576	2.8340	3.1855	3.3932	3.8333
1000.	1.5953	1.6867	1.79[illegible]2	1.9263	2.1020	2.3791	2.6330	2.8080	3.1569	3.3630	3.7998

THIS TABLE IS REPRODUCED WITH PERMISSION FROM

TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.9000

SAMPLE SIZE	P 0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
5.	0.6857	1.0409	1.3554	1.6974	1.9760	2.3109	2.5109	2.7440	2.8499	2.9656	3.0934
6.	0.6023	0.9345	1.3172	1.5393	1.7943	2.0999	2.2817	2.4937	2.5902	2.6955	2.8119
7.	0.5439	0.8621	1.2247	1.4346	1.6753	1.9629	2.1336	2.3324	2.4229	2.5216	2.6307
8.	0.5001	0.8077	1.1580	1.3598	1.5899	1.8647	2.0284	2.2191	2.3056	2.3998	2.5039
9.	0.4657	0.7662	1.1064	1.3021	1.5254	1.7914	1.9492	2.1329	2.2164	2.3075	2.4081
10.	0.4373	0.7325	1.0654	1.2565	1.4742	1.7332	1.8868	2.0656	2.1468	2.2353	2.3331
11.	0.4138	0.7046	1.0317	1.2192	1.4324	1.6859	1.8362	2.0110	2.0904	2.1770	2.2725
12.	0.3937	0.6811	1.0036	1.1880	1.3977	1.6467	1.7943	1.9659	2.0438	2.1287	2.2225
13.	0.3763	0.6608	0.9795	1.1515	1.3681	1.6135	1.7589	1.9278	2.0045	2.0880	2.1803
14.	0.3609	0.6431	0.9585	1.1385	1.3427	1.5850	1.7284	1.8951	1.9707	2.0532	2.1441
15.	0.3473	0.6276	0.9403	1.[illegible]185	1.3205	1.5601	1.7019	1.8666	1.9414	2.0229	2.1127
16.	0.3352	0.6137	0.9240	1.1007	1.3009	1.5382	1.6786	1.8416	1.9156	1.9962	2.0851
17.	0.3242	0.6012	0.9095	1.0848	1.2834	1.5187	1.6579	1.8194	1.8927	1.9726	2.0607
18.	0.3143	0.5900	0.8964	1.0706	1.2677	1.5012	1.6392	1.7996	1.8721	1.9513	2.0387
19.	0.3053	0.5798	0.8846	1.0576	1.2535	1.4853	1.6224	1.7814	1.8536	1.9321	2.0188
20.	0.2970	0.5704	0.8737	1.0458	1.2405	1.4709	1.6071	1.7651	1.8367	1.9148	2.0008
21.	0.2893	0.5618	0.8638	1.0351	1.2287	1.4578	1.5931	1.7501	1.8214	1.8989	1.9845
22.	0.2822	0.5539	0.8546	1.0251	1.2178	1.4457	1.5803	1.7365	1.8073	1.8844	1.9694
23.	0.2756	0.5465	0.8462	1.0160	1.2078	1.4346	1.5685	1.7239	1.7943	1.8710	1.9556
24.	0.2694	0.5396	0.8383	1.0074	1.1985	1.4243	1.5576	1.7122	1.7823	1.8586	1.9428
25.	0.2637	0.5332	0.8310	0.9995	1.1898	1.4147	1.5475	1.7014	1.7712	1.8472	1.9310
30.	0.2395	0.5065	0.8006	0.9668	1.1541	1.3753	1.5058	1.6570	1.7255	1.8001	1.8823
35.	0.2209	0.4861	0.7776	0.9421	1.1273	1.3458	1.4746	1.6238	1.6914	1.7650	1.8461
40.	0.2061	0.4700	0.7595	0.9226	1.1063	1.3227	1.4502	1.5978	1.6647	1.7375	1.8177
45.	0.1940	0.4568	0.7447	0.9068	1.0892	1.3039	1.4304	1.5768	1.6431	1.7153	1.7948
50.	0.1838	0.4458	0.7324	0.8936	1.0749	1.2883	1.4140	1.5594	1.6253	1.6969	1.7759
60.	0.1674	0.4281	0.7128	0.8727	1.0524	1.2637	1.3881	1.5320	1.5971	1.6680	1.7460
70.	0.1547	0.4115	0.6978	0.8568	1.0355	1.2451	1.3685	1.5112	1.57[illegible]	1.6460	1.7234
80.	0.1445	0.4037	0.6859	0.8441	1.0217	1.2302	1.3529	1.4947	1.52[illegible]	1.6287	1.7056
90.	0.1361	0.3948	0.6761	0.8337	1.0105	1.2181	1.3402	1.4813	1.5451	1.6145	1.6910
100.	0.1291	0.3872	0.6679	0.8250	1.0012	1.2080	1.3296	1.4700	1.5336	1.6027	1.6788
150.	0.1051	0.3620	0.6403	0.[illegible]759	0.9701	1.1744	1.2943	1.4328	1.4955	1.5636	1.6386
200.	0.0909	0.3471	0.6243	0.[illegible]789	0.9520	1.1548	1.2739	1.4113	1.4734	1.5409	1.6153
250.	0.0813	0.3370	0.6134	0.7675	0.9399	1.1417	1.2601	1.3968	1.4586	1.5258	1.5997
300.	0.0742	0.3296	0.6054	0.7591	0.9310	1.1321	1.2501	1.3863	1.4478	1.5147	1.5883
350.	0.0687	0.3239	0.5997	0.7526	0.9241	1.1247	1.2424	1.3781	1.4395	1.5062	1.5795
400.	0.0642	0.3193	0.594[illegible]	0.7474	0.9186	1.1188	1.2362	1.3716	1.4328	1.4993	1.5725
450.	0.0605	0.3154	0.5902	0.7432	0.9141	1.1139	1.2311	1.3663	1.4273	1.4937	1.5668
500.	0.0574	0.3122	0.58[illegible]8	0.7395	0.9102	1.1098	1.2268	1.3618	1.4227	1.4890	1.5619
1000.	0.0406	0.2948	0.56[illegible]2	0.7202	0.8897	1.0878	1.2039	1.3376	1.3981	1.4637	1.5360

THIS TABLE IS REPRODUCED WITH PERMISSION FROM
TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.9000

SAMPLE SIZE	0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
					P						
5.	3.2368	3.4013	3.5958	3.8364	4.1583	4.6690	5.1398	5.4653	6.1160	6.5017	7.3210
6.	2.9425	3.0922	3.2690	3.4875	3.7796	4.2433	4.6707	4.9665	5.5579	5.9084	6.6530
7.	2.7531	2.8933	3.0589	3.2636	3.5372	3.9713	4.3714	4.6482	5.2018	5.5299	6.2268
8.	2.6207	2.7546	2.9125	3.1377	3.3686	3.7825	4.1639	4.4278	4.9554	5.2681	5.9323
9.	2.5210	2.6502	2.8028	2.9911	3.2427	3.6417	4.0093	4.2636	4.7721	5.0734	5.7133
10.	2.4427	2.5683	2.7165	2.8795	3.1440	3.5316	3.8887	4.1357	4.6294	4.9220	5.5434
11.	2.3796	2.5023	2.6471	2.8258	3.0645	3.4430	3.7916	4.0327	4.5146	4.8001	5.4065
12.	2.3276	2.4479	2.5899	2.7651	2.9992	3.3702	3.7118	3.9487	4.4203	4.7001	5.2942
13.	2.2831	2.4021	2.5417	2.7141	2.9442	3.3090	3.6448	3.8771	4.3413	4.6162	5.2001
14.	2.2461	2.3629	2.5005	2.6706	2.8973	3.2568	3.5877	3.8165	4.2738	4.5447	5.1199
15.	2.2135	2.3288	2.4647	2.6325	2.8566	3.2115	3.5382	3.7641	4.2155	4.4829	5.0506
16.	2.1848	2.2989	2.4334	2.5933	2.8209	3.1719	3.4949	3.7182	4.1645	4.4288	4.9899
17.	2.1594	2.2724	2.4056	2.5679	2.7893	3.1368	3.4566	3.6777	4.1194	4.3810	4.9363
18.	2.1365	2.2485	2.3805	2.5435	2.7609	3.1053	3.4221	3.6412	4.0788	4.3380	4.8882
19.	2.1159	2.2271	2.3581	2.5197	2.7354	3.0770	3.3912	3.6085	4.0425	4.2995	4.8450
20.	2.0973	2.2076	2.3377	2.4982	2.7123	3.0513	3.3632	3.5789	4.0096	4.2646	4.8060
21.	2.0803	2.1899	2.3191	2.4785	2.6912	3.0280	3.3378	3.5519	3.9776	4.2329	4.7705
22.	2.0647	2.1737	2.3021	2.4606	2.6720	3.0067	3.3145	3.5273	3.9523	4.2039	4.7381
23.	2.0503	2.1587	2.2865	2.4440	2.6542	2.9870	3.2931	3.5047	3.9272	4.1773	4.7083
24.	2.0371	2.1449	2.2720	2.4288	2.6379	2.9689	3.2734	3.4838	3.9010	4.1528	4.6808
25.	2.0248	2.1321	2.2586	2.4146	2.6227	2.9521	3.2551	3.4644	3.8825	4.1300	4.6553
30.	1.9744	2.0797	2.2038	2.3567	2.5608	2.8836	3.1804	3.3854	3.7949	4.0375	4.5516
35.	1.9369	2.0407	2.1630	2.3137	2.5147	2.8327	3.1250	3.3269	3.7300	3.9686	4.4749
40.	1.9075	2.0102	2.1311	2.2802	2.4789	2.7931	3.0819	3.2814	3.6796	3.9153	4.4152
45.	1.8839	1.9856	2.1054	2.2531	2.4499	2.7612	3.0472	3.2447	3.6390	3.8723	4.3672
50.	1.8642	1.9652	2.0841	2.2307	2.4260	2.7348	3.0185	3.2144	3.6054	3.8368	4.3276
60.	1.8334	1.9332	2.0507	2.1955	2.3885	2.6934	2.9735	3.1670	3.5529	3.7813	4.2657
70.	1.8100	1.9090	2.0254	2.1689	2.3601	2.6622	2.9396	3.1312	3.5134	3.7395	4.2190
80.	1.7916	1.8898	2.0055	2.1479	2.3377	2.6376	2.9129	3.1030	3.4822	3.7065	4.1822
90.	1.7765	1.8743	1.9892	2.1309	2.3195	2.6176	2.8912	3.0801	3.4569	3.6797	4.1524
100.	1.7640	1.8612	1.9757	2.1166	2.3043	2.6009	2.8731	3.0610	3.4358	3.6574	4.1275
150.	1.7224	1.8181	1.9308	2.0695	2.2542	2.5458	2.8134	2.9980	3.3663	3.5840	4.0457
200.	1.6984	1.7933	1.9049	2.0424	2.2253	2.5141	2.7790	2.9618	3.3263	3.5418	3.9987
250.	1.6823	1.7767	1.8876	2.0273	2.2060	2.4929	2.7561	2.9376	3.2996	3.5137	3.9673
300.	1.6706	1.7645	1.8750	2.0110	2.1919	2.4775	2.7394	2.9200	3.2802	3.4932	3.9445
350.	1.6616	1.7552	1.8653	2.0068	2.1811	2.4656	2.7265	2.9065	3.2653	3.4775	3.9270
400.	1.6543	1.7477	1.8575	1.9920	2.1724	2.4561	2.7163	2.8957	3.2534	3.4649	3.9130
450.	1.6484	1.7416	1.8511	1.9859	2.1653	2.4483	2.7078	2.8868	3.2436	3.4546	3.9016
500.	1.6434	1.7364	1.8458	1.9803	2.1593	2.4418	2.7008	2.8794	3.2354	3.4459	3.8919
1000.	1.6167	1.7089	1.8171	1.9503	2.1274	2.4069	2.6630	2.8396	3.1916	3.3997	3.8405

THIS TABLE IS REPRODUCED WITH PERMISSION FROM
TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.9500

SAMPLE SIZE	**P** 0.5000	0.6000	0.7000	0.7500	0.8000	0.8500	0.8750	0.9000	0.9100	0.9200	0.9300
5.	0.9539	1.3693	1.8615	2.1519	2.4865	2.8894	3.1308	3.4127	3.5411	3.6813	3.8364
6.	0.8224	1.1981	1.6377	1.8950	2.1907	2.5463	2.7592	3.0076	3.1205	3.2438	3.3801
7.	0.7340	1.0863	1.4943	1.7319	2.0050	2.3329	2.5281	2.7561	2.8599	2.9732	3.0985
8.	0.6697	1.0062	1.3924	1.6173	1.8752	2.1635	2.3674	2.5820	2.6797	2.7864	2.9042
9.	0.6197	0.9448	1.3168	1.5321	1.7783	2.0732	2.2490	2.4541	2.5473	2.6489	2.7613
10.	0.5797	0.8957	1.2566	1.4650	1.7033	1.9880	2.1573	2.3547	2.4444	2.5423	2.6505
11.	0.5465	0.8558	1.2077	1.4106	1.6423	1.9189	2.0833	2.2748	2.3619	2.4568	2.5617
12.	0.5185	0.8224	1.1672	1.3657	1.5922	1.8622	2.0226	2.2094	2.2944	2.3870	2.4892
13.	0.4943	0.7939	1.1329	1.3279	1.5500	1.8147	1.9718	2.1547	2.2379	2.3286	2.4287
14.	0.4733	0.7693	1.1035	1.2954	1.5139	1.7741	1.9285	2.1082	2.1899	2.2789	2.3772
15.	0.4548	0.7477	1.0779	1.2672	1.4826	1.7390	1.8910	2.0679	2.1483	2.2360	2.3327
16.	0.4383	0.7286	1.0553	1.2424	1.4552	1.7082	1.8582	2.0327	2.1121	2.1985	2.2938
17.	0.4234	0.7115	1.0352	1.2203	1.4308	1.6810	1.8292	2.0016	2.0800	2.1653	2.2595
18.	0.4101	0.6961	1.0170	1.2004	1.4089	1.6564	1.8031	1.9736	2.0511	2.1355	2.2286
19.	0.3979	0.6822	1.0006	1.1825	1.3891	1.6343	1.7796	1.9484	2.0251	2.1087	2.2009
20.	0.3867	0.6694	0.9857	1.1662	1.3711	1.6143	1.7583	1.9257	2.0017	2.0845	2.1759
21.	0.3764	0.6577	0.9721	1.1514	1.3548	1.5961	1.7390	1.9050	1.9804	2.0625	2.1531
22.	0.3669	0.6470	0.9595	1.1377	1.3398	1.5794	1.7213	1.8861	1.9609	2.0424	2.1323
23.	0.3581	0.6370	0.9480	1.1251	1.3260	1.5641	1.7050	1.8687	1.9430	2.0239	2.1132
24.	0.3499	0.6277	0.9373	1.1135	1.3132	1.5500	1.6900	1.8526	1.9265	2.0069	2.0956
25.	0.3423	0.6191	0.9273	1.1027	1.3014	1.5368	1.6761	1.8378	1.9112	1.9911	2.0793
30.	0.3103	0.5834	0.8863	1.0583	1.2529	1.4832	1.6192	1.7771	1.8488	1.9268	2.0128
35.	0.2858	0.5563	0.8555	1.0251	1.2167	1.4432	1.5770	1.7322	1.8025	1.8792	1.9637
40.	0.2664	0.5349	0.8313	0.9990	1.1883	1.4120	1.5440	1.6971	1.7665	1.8421	1.9254
45.	0.2505	0.5175	0.8116	0.9778	1.1654	1.3868	1.5174	1.6688	1.7374	1.8122	1.8945
50.	0.2372	0.5029	0.7952	0.9603	1.1464	1.3660	1.4954	1.6454	1.7134	1.7875	1.8690
60.	0.2158	0.4797	0.7693	0.9325	1.1164	1.3332	1.4609	1.6088	1.6758	1.7487	1.8291
70.	0.1993	0.4619	0.7495	0.9115	1.0937	1.3083	1.4348	1.5811	1.6474	1.7195	1.7990
80.	0.1861	0.4477	0.7338	0.8947	1.0757	1.2887	1.4141	1.5593	1.6250	1.6965	1.7753
90.	0.1753	0.4360	0.7210	0.8811	1.0610	1.2727	1.3973	1.5415	1.6068	1.6778	1.7560
100.	0.1661	0.4262	0.7102	0.8696	1.0487	1.2594	1.3833	1.5267	1.5916	1.6622	1.7400
150.	0.1351	0.3934	0.6743	0.8316	1.0080	1.2152	1.3370	1.4778	1.5474	1.6107	1.6870
200.	0.1169	0.3741	0.6533	0.8094	0.9844	1.1897	1.3102	1.4495	1.5125	1.5810	1.6565
250.	0.1045	0.3611	0.6392	0.7945	0.9686	1.1725	1.2923	1.4306	1.4932	1.5612	1.6361
300.	0.0953	0.3515	0.6288	0.7837	0.9570	1.1601	1.2793	1.4169	1.4791	1.5468	1.6213
350.	0.0882	0.3441	0.6209	0.7753	0.9481	1.1504	1.2692	1.4063	1.4683	1.5357	1.6098
400.	0.0825	0.3381	0.6144	0.7685	0.9409	1.1428	1.2612	1.3979	1.4596	1.5268	1.6007
450.	0.0777	0.3332	0.6097	0.7635	0.9350	1.1364	1.2545	1.3909	1.4525	1.5195	1.5932
500.	0.0737	0.3291	0.6047	0.7583	0.9301	1.1311	1.2490	1.3850	1.4465	1.5134	1.5869
1000.	0.0521	0.3067	0.5803	0.7333	0.9035	1.1026	1.2193	1.3538	1.4146	1.4806	1.5533

THIS TABLE IS REPRODUCED WITH PERMISSION FROM

TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE III: K-FACTORS

CONFIDENCE LEVEL =0.9500

SAMPLE SIZE	P 0.9400	0.9500	0.9600	0.9700	0.9800	0.9900	0.9950	0.9970	0.9990	0.9995	0.9999
5.	4.0105	4.2101	4.4461	4.7379	5.1284	5.7487	6.3210	6.7172	7.5101	7.9802	8.9794
6.	3.5331	3.7087	3.9162	4.1733	4.5174	5.0639	5.5680	5.9170	6.6147	7.0285	7.9079
7.	3.2392	3.4005	3.5912	3.8269	4.1424	4.6434	5.1056	5.4256	6.0654	6.4448	7.2512
8.	3.0366	3.1883	3.3676	3.5889	3.8850	4.3553	4.7891	5.0874	5.6903	6.0467	6.8039
9.	2.8874	3.0320	3.2027	3.4139	3.6963	4.1448	4.5584	4.8448	5.4167	5.7559	6.4767
10.	2.7719	2.9110	3.0754	3.2785	3.5501	3.9813	4.3790	4.6542	5.2047	5.5310	6.2246
11.	2.6795	2.8144	2.9737	3.1706	3.4338	3.8517	4.2369	4.5035	5.0367	5.3528	6.0245
12.	2.6040	2.7355	2.8908	3.0826	3.3391	3.7461	4.1213	4.3810	4.9002	5.2080	5.8619
13.	2.5410	2.6697	2.8216	3.0093	3.2602	3.6582	4.0251	4.2790	4.7867	5.0875	5.7268
14.	2.4874	2.6137	2.7629	2.9470	3.1932	3.5837	3.9436	4.1926	4.6905	4.9855	5.6124
15.	2.4412	2.5655	2.7122	2.8934	3.1355	3.5196	3.8734	4.1183	4.6078	4.8979	5.5141
16.	2.4008	2.5234	2.6680	2.8466	3.0852	3.4636	3.8123	4.0535	4.5357	4.8215	5.4285
17.	2.3652	2.4862	2.6290	2.8053	3.0408	3.4144	3.7584	3.9965	4.4723	4.7543	5.3532
18.	2.3331	2.4527	2.5939	2.7682	3.0010	3.3701	3.7101	3.9453	4.4151	4.6940	5.2856
19.	2.3043	2.4227	2.5624	2.7349	2.9652	3.3305	3.6668	3.8995	4.3645	4.6400	5.2252
20.	2.2783	2.3956	2.5340	2.7049	2.9330	3.2947	3.6278	3.8582	4.3186	4.5914	5.1707
21.	2.2547	2.3710	2.5082	2.6776	2.9038	3.2623	3.5924	3.8207	4.2770	4.5473	5.1214
22.	2.2331	2.3485	2.4847	2.6527	2.8771	3.2327	3.5601	3.7866	4.2391	4.5071	5.0764
23.	2.2133	2.3279	2.4631	2.6299	2.8526	3.2056	3.5305	3.7552	4.2043	4.4703	5.0352
24.	2.1951	2.3089	2.4431	2.6088	2.8300	3.1806	3.5033	3.7264	4.1723	4.4364	4.9973
25.	2.1782	2.2913	2.4247	2.5894	2.8092	3.1575	3.4781	3.6998	4.1427	4.4051	4.9622
30.	2.1093	2.2196	2.3497	2.5102	2.7243	3.0636	3.3758	3.5916	4.0227	4.2780	4.8201
35.	2.0583	2.1666	2.2943	2.4517	2.6618	2.9945	3.3005	3.5120	3.9345	4.1847	4.7157
40.	2.0187	2.1254	2.2512	2.4063	2.6132	2.9408	3.2421	3.4503	3.8662	4.1124	4.6349
45.	1.9867	2.0922	2.2165	2.3698	2.5742	2.8977	3.1952	3.4008	3.8113	4.0543	4.5701
50.	1.9604	2.0648	2.1879	2.3397	2.5420	2.8612	3.1566	3.3601	3.7662	4.0066	4.5168
60.	1.9191	2.0220	2.1432	2.2926	2.4918	2.8068	3.0964	3.2965	3.6959	3.9323	4.4338
70.	1.8880	1.9897	2.1095	2.2572	2.4540	2.7652	3.0512	3.2488	3.6431	3.8764	4.3715
80.	1.8635	1.9643	2.0830	2.2293	2.4243	2.7325	3.0157	3.2113	3.6017	3.8326	4.3227
90.	1.8436	1.9437	2.0615	2.2067	2.4001	2.7059	2.9869	3.1809	3.5681	3.7971	4.2831
100.	1.8270	1.9265	2.0435	2.1875	2.3800	2.6839	2.9629	3.1556	3.5401	3.7676	4.2502
150.	1.7723	1.8698	1.9845	2.1258	2.3140	2.6113	2.8843	3.0727	3.4486	3.6709	4.1424
200.	1.7408	1.8372	1.9506	2.0902	2.2761	2.5697	2.8392	3.0251	3.3961	3.6155	4.0806
250.	1.7198	1.8154	1.9279	2.0664	2.2508	2.5420	2.8091	2.9935	3.3612	3.5786	4.0396
300.	1.7045	1.7996	1.9115	2.0491	2.2324	2.5218	2.7873	2.9705	3.3359	3.5519	4.0098
350.	1.6728	1.7874	1.8988	2.0359	2.2183	2.5064	2.7706	2.9529	3.3164	3.5314	3.9870
400.	1.6634	1.7777	1.8887	2.0253	2.2071	2.4941	2.7573	2.9389	3.3009	3.5150	3.9688
450.	1.6 57	1.7697	1.8814	2.0166	2.1978	2.4839	2.7465	2.9273	3.2882	3.5016	3.9538
500.	1.6492	1.7630	1.8734	2.0093	2.1901	2.4754	2.7371	2.9176	3.2775	3.4903	3.9413
1000.	1.6146	1.7272	1.8362	1.9703	2.1487	2.4301	2.6883	2.8661	3.2207	3.4304	3.8746

THIS TABLE IS REPRODUCED WITH PERMISSION FROM TABLES BY PRATT & WHITNEY AIRCRAFT CORPORATION

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\epsilon = .05$

$12/\sqrt{f}$				4	3	2	1	0
η \ f	6	7	8	9	16	36	144	∞
-1.0	1.554	1.563	1.569	1.5744	1.5952	1.6141	1.6307	1.6449
-.9	1.553	1.561	1.567	1.5721	1.5923	1.6116	1.6292	1.6449
-.8	1.565	1.570	1.574	1.5787	1.5950	1.6121	1.6289	1.6449
-.7	1.584	1.587	1.589	1.5910	1.6017	1.6150	1.6297	1.6449
-.6	1.608	1.606	1.607	1.6065	1.6109	1.6195	1.6313	1.6449
-.5	1.629	1.627	1.624	1.6230	1.6212	1.6251	1.6333	1.6449
-.4	1.651	1.645	1.642	1.6391	1.6322	1.6313	1.6359	1.6449
-.3	1.668	1.662	1.657	1.6535	1.6428	1.6379	1.6387	1.6449
-.2	1.680	1.674	1.670	1.6657	1.6525	1.6443	1.6417	1.6449
-.1	1.690	1.684	1.679	1.6755	1.6610	1.6503	1.6448	1.6449
.0	1.695	1.690	1.686	1.6828	1.6683	1.6558	1.6477	1.6449
.1	1.697	1.693	1.690	1.6875	1.6739	1.6606	1.6505	1.6449
.2	1.697	1.695	1.693	1.6901	1.6780	1.6646	1.6529	1.6449
.3	1.694	1.693	1.693	1.6908	1.6806	1.6677	1.6550	1.6449
.4	1.690	1.691	1.691	1.6897	1.6819	1.6700	1.6568	1.6449
.5	1.685	1.687	1.687	1.6873	1.6818	1.6712	1.6581	1.6449
.6	1.680	1.682	1.683	1.6839	1.6805	1.6714	1.6588	1.6449
.7	1.674	1.677	1.678	1.6796	1.6782	1.6707	1.6590	1.6449
.8	1.668	1.672	1.674	1.6747	1.6750	1.6691	1.6586	1.6449
.9	1.662	1.665	1.668	1.6694	1.6710	1.6667	1.6576	1.6449
1.0	1.655	1.660	1.662	1.6638	1.6665	1.6634	1.6559	1.6449

$$\eta = \frac{\delta}{\sqrt{2f}} \Big/ (1 + \delta^2/2f)^{1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$t_{1-\epsilon}(\delta) = t_\epsilon(-\delta)$$

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\epsilon = .10$

$12/\sqrt{f}$				4	3	2	1	0
η \ f	6	7	8	9	16	36	144	∞
-1.0	1.150	1.161	1.169	1.1765	1.2049	1.2319	1.2575	1.2816
-.9	1.166	1.175	1.182	1.1877	1.2120	1.2358	1.2591	1.2816
-.8	1.186	1.192	1.198	1.2018	1.2209	1.2408	1.2611	1.2816
-.7	1.208	1.212	1.214	1.2175	1.2310	1.2465	1.2635	1.2816
-.6	1.231	1.231	1.233	1.2341	1.2418	1.2526	1.2661	1.2816
-.5	1.252	1.251	1.250	1.2503	1.2526	1.2591	1.2688	1.2816
-.4	1.272	1.269	1.267	1.2657	1.2634	1.2656	1.2718	1.2816
-.3	1.290	1.286	1.283	1.2801	1.2738	1.2720	1.2748	1.2816
-.2	1.305	1.299	1.296	1.2931	1.2836	1.2784	1.2778	1.2816
-.1	1.318	1.313	1.308	1.3048	1.2926	1.2846	1.2809	1.2816
.0	1.329	1.323	1.319	1.3150	1.3009	1.2903	1.2838	1.2816
.1	1.339	1.332	1.328	1.3236	1.3084	1.2957	1.2866	1.2816
.2	1.346	1.340	1.335	1.3309	1.3149	1.3007	1.2893	1.2816
.3	1.351	1.346	1.341	1.3369	1.3207	1.3053	1.2919	1.2816
.4	1.354	1.350	1.345	1.3417	1.3256	1.3093	1.2943	1.2816
.5	1.357	1.353	1.349	1.3455	1.3297	1.3129	1.2965	1.2816
.6	1.360	1.355	1.352	1.3485	1.3331	1.3159	1.2984	1.2816
.7	1.361	1.358	1.354	1.3509	1.3358	1.3186	1.3001	1.2816
.8	1.362	1.359	1.355	1.3526	1.3390	1.3207	1.3016	1.2816
.9	1.363	1.360	1.357	1.3540	1.3398	1.3224	1.3028	1.2816
1.0	1.364	1.361	1.358	1.3554	1.3413	1.3240	1.3038	1.2816

$$\eta = \frac{\delta}{\sqrt{2f}} \Big/ (1 + \delta^2/2f)^{1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$t_{1-\epsilon}(\delta) = t_{\epsilon}(-\delta)$$

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\epsilon = .20$

$12/\sqrt{f}$				4	3	2	1	0
η \ f	6	7	8	9	16	36	144	∞
−1.0	.673	.686	.697	.7055	.7408	.7753	.8089	.8416
−.9	.699	.711	.718	.7261	.7551	.7841	.8129	.8416
−.8	.726	.734	.741	.7460	.7690	.7926	.8168	.8416
−.7	.750	.757	.761	.7649	.7821	.8007	.8206	.8416
−.6	.773	.777	.780	.7824	.7944	.8083	.8242	.8416
−.5	.794	.796	.797	.7987	.8059	.8156	.8276	.8416
−.4	.814	.814	.813	.8138	.8165	.8225	.8308	.8416
−.3	.831	.829	.828	.8278	.8271	.8291	.8340	.8416
−.2	.848	.845	.843	.8409	.8367	.8354	.8371	.8416
−.1	.863	.859	.856	.8532	.8459	.8415	.8401	.8416
.0	.877	.871	.868	.8648	.8547	.8475	.8431	.8416
.1	.890	.884	.880	.8759	.8632	.8532	.8460	.8416
.2	.902	.895	.890	.8864	.8715	.8589	.8489	.8416
.3	.913	.907	.901	.8964	.8795	.8645	.8519	[illegible]
.4	.924	.916	.911	.9061	.8873	.8701	.8548	[illegible]
.5	.934	.927	.920	.9156	.8950	.8756	.8576	.8416
.6	.945	.937	.930	.9849	.9026	.8811	.8506	[illegible]
.7	.955	.947	.940	.9342	.9103	.8867	.8637	.8416
.8	.965	.957	.950	.9433	.9180	.8924	.8668	.8416
.9	.976	.966	.959	.9526	.9260	.8983	.8700	.8416
1.0	.986	.977	.969	.9623	.9341	.9044	.8734	[illegible]

$$\eta = \frac{\delta}{\sqrt{2f}} \Big/ (1 + \delta^2/2f)^{1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$t_{1-\epsilon}(\delta) = t_\epsilon(-\delta)$$

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\epsilon = .30$

$12/\sqrt{f}$				4	3	2	1	0
η \ f	6	7	8	9	16	36	144	∞
-1.0	.339	.353	.365	.3739	.4124	.4503	.4876	.5244
-.9	.368	.380	.389	.3971	.4291	.4610	.4927	.5244
-.8	.396	.405	.412	.4188	.4446	.4709	.4975	.5244
-.7	.421	.428	.434	.4386	.4589	.4799	.5018	.5244
-.6	.443	.449	.453	.4566	.4718	.4882	.5057	.5244
-.5	.465	.467	.471	.4730	.4838	.4959	.5095	.5244
-.4	.483	.485	.487	.4880	.4947	.5030	.5129	.5244
-.3	.501	.501	.501	.5019	.5049	.5097	.5162	.5244
-.2	.516	.515	.515	.5148	.5145	.5160	.5193	.5244
-.1	.531	.529	.528	.5272	.5237	.5221	.5223	.5244
.0	.546	.543	.541	.5391	.5326	.5281	.5253	.5244
.1	.561	.556	.553	.5507	.5414	.5340	.5282	.5244
.2	.575	.570	.565	.5624	.5502	.5398	.5312	.5244
.3	.588	.582	.578	.5741	.5591	.5458	.5343	.5244
.4	.603	.595	.590	.5859	.5682	.5520	.5375	.5244
.5	.618	.610	.603	.5980	.5776	.5584	.5407	.5244
.6	.632	.624	.617	.6106	.5874	.5652	.5442	.5244
.7	.648	.638	.630	.6237	.5976	.5723	.5479	.5244
.8	.663	.652	.645	.6374	.6084	.5799	.5518	.5244
.9	.680	.668	.660	.6517	.6199	.5879	.5561	.5244
1.0	.697	.685	.675	.6668	.6320	.5965	.5607	.5244

$$\eta = \frac{\delta}{\sqrt{2f}} \Big/ (1 + \delta^2/2f)^{1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$t_{1-\epsilon}(\delta) = t_\epsilon(-\delta)$$

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\epsilon = .40$

$12/\sqrt{f}$				4	3	2	1	0
η \ f	6	7	8	9	16	36	144	∞
-1.0	.060	.075	.087	.0964	.1362	.1755	.2146	.2533
-.9	.089	.101	.110	.1197	.1534	.1869	.2202	.2533
-.8	.116	.126	.135	.1411	.1692	.1972	.2252	.2533
-.7	.141	.149	.155	.1607	.1836	.2065	.2298	.2533
-.6	.163	.169	.174	.1786	.1967	.2151	.2339	.2533
-.5	.183	.187	.192	.1948	.2086	.2229	.2378	.2533
-.4	.201	.204	.207	.2096	.2195	.2301	.2413	.2533
-.3	.219	.220	.222	.2233	.2297	.2366	.2446	.2533
-.2	.234	.235	.236	.2362	.2392	.2431	.2478	.2533
-.1	.249	.249	.248	.2464	.2484	.2491	.2508	.2533
.0	.264	.262	.262	.2604	.2573	.2551	.2537	.2533
.1	.279	.276	.274	.2723	.2662	.2610	.2567	.2533
.2	.293	.289	.287	.2844	.2753	.2670	.2597	.2533
.3	.309	.304	.300	.2968	.2845	.2733	.2629	.2533
.4	.324	.318	.314	.3097	.2943	.2798	.2661	.2533
.5	.341	.334	.328	.3233	.3045	.2866	.2696	.2533
.6	.359	.351	.343	.3378	.3155	.2940	.2733	.2533
.7	.378	.368	.360	.3533	.3273	.3020	.2773	.2533
.8	.397	.386	.377	.3700	.3400	.3106	.2818	.2533
.9	.418	.406	.396	.3877	.3538	.3201	.2866	.2533
1.0	.441	.427	.416	.4067	.3686	.3303	.2919	.2533

$$\eta = \frac{\delta}{\sqrt{2f}} \Big/ (1 + \delta^2/2f)^{1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$t_{1-\epsilon}(\delta) = t_\epsilon(-\delta)$$

TABLE IV

TABLES TO FACILITATE CALCULATION OF K

$\sigma = .50$

$12/\sqrt{f}$				4	3	2	1	0
η \\ f	6	7	8	9	16	36	144	∞
.0	.000	.000	.000	.0000	.0000	.0000	.0000	.0000
.1	.015	.013	.012	.0119	.0089	.0059	.0030	.0000
.2	.030	.028	.026	.0242	.0181	.0120	.0060	.0000
.3	.046	.042	.039	.0369	.0275	.0182	.0091	.0000
.4	.063	.057	.[illegible]54	.0504	.0376	.0250	.0124	.0000
.5	.080	.074	[illegible]69	.0648	.0483	.0320	.0159	.0000
.6	.099	.092	.085	.0804	.0599	.0397	.0168	.0000
.7	.120	.111	.104	.0974	.0726	.0482	.0238	.0000
.8	.143	.132	.123	.1159	.0865	.0574	.0286	.0000
.9	.167	.155	.145	.1361	.1017	.0676	.0336	.0000
1.0	.194	.179	.167	.1578	.1182	.0787	.0393	.0000

$$\eta = (\delta/\sqrt{2f})\ (1 + \delta^2/2f)^{-1/2} \qquad t = \frac{\delta + \lambda(1 + \delta^2/2f - \lambda^2/2f)^{1/2}}{1 - \lambda^2/2f}$$

$$\lambda(-\delta) = \lambda(\delta)$$

REFERENCES

1. Lusser, R. Reliability Through Safety Margins, U. S. Army Ordnance Missile Command, October 1958.

2. Heathcock, R., and Burrows, D. L., Monte Carlo Application for Developing a Design Reliability Goal Compatible with Small Sample Requirements, a paper presented at the Ninth Conference on the Design of Experiments in Army Research, Development and Testing, October 23, 1963.

3. Welch, B. L., The Generalization of "Student's" Problem When Several Different Population Variances are Involved, Biometrika, Vol. 34, 1947, pp. 28-35.

4. Aspin, Alice A., Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated, Biometrika, Vol. 36, 1949, Appendix by Welch, pp 293-296.

5. Bombara, E. L., ARGMA TR 2B1R, Reliability of Compliance with One-sided Specification Limits when Data is Normally Distributed, September 15, 1961. Also contained in ARODR 62-2, Proceedings of the Seventh Conference on the Design of Experiments in Army Research, Development and Testing, August, 1962.

6. Norman, J. E., Estimating Reliability as a Function of Stress/Strength Data, Army Rocket and Guided Missile Agency, Redstone Arsenal, Alabama, November 1961.

7. Resnikoff, George J., Technical Report #55, Tables to Facilitate the Computation of Percentage Points of the Non-Central t-Distribution, Applied Mathematics and Statistics Laboratories, Stanford University, California, October 11, 1961.

8. Nickle, Dennis E., PWA FR 781, One-Sided Statistical Tolerance Factors, Pratt & Whitney Aircraft Corporation, West Palm Beach, Florida, October 1963.

9. Tables of Normal Probability Functions, National Bureau of Standards, U. S. Department of Commerce.

STATISTICAL TESTING TECHNIQUES USED IN THE DEVELOPMENT OF THE PRATT & WHITNEY AIRCRAFT RL10 ROCKET ENGINE FOR THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Harold J. Tiedemann
Pratt & Whitney Aircraft
Florida Research and Development Center

INTRODUCTION. Pratt & Whitney Aircraft's experience in the development of many different kinds of propulsion systems has shown clearly the necessity of a comprehensive development program of three to five years to bring a new power plant from preliminary design to operational capability. Reliable operation of such a system is achieved by integrated and experienced engineering design and development teams, proven manufacturing methods, rigid quality and vendor control, and a supporting field service. Emphasis must be placed on a comprehensive test program involving perhaps thousands of tests of both individual components and full-scale engines. There is no means to realize the full potential and reliable operation of a power plant or system except by a comprehensive development program which involves many cycles of design, testing, corrective design, and retesting. The element of time cannot be factored out of this development cycle.

During the course of most such development programs a number of complex problems arise where the time involved in test evaluation can be minimized and more complete information obtained by the use of statistical testing techniques. Several representative examples of these testing techniques have been used in the RL10 Rocket Engine Development Program being conducted for the NASA - Marshall Space Flight Center. These include Full Factorials, Latin Squares, and Random and Multiple Balance techniques. The purpose of this paper is to discuss these techniques in detail, pointing out their best areas of application as well as their advantages and disadvantages, and to describe several RL10 engine development problems which were successfully solved by the use of these methods.

These include a Full Factorial to determine thrust control setting repeatability, and a partial Full Factorial to provide additional fuel pump stall margin. Also included is the description of a Latin Square test program designed to determine the principal factors in controlling transient thrust overshoot. Finally, a Random Balance test is described which was used to determine methods to control thrust control hysteresis and non-linearity.

STATISTICAL TESTING TECHNIQUES.

A. THE FULL FACTORIAL TEST.

In the full Factorial testing technique, tests are made with all possible combinations of the independent variables; therefore, all interactions between parameters will be identified. Thus, a Full Factorial might be described as a fully expanded version of the step-by-step approach. A Full Factorial experiment is illustrated as follows:

Suppose that there are three independent variables, each at two levels, of which either the main effect of a single variable or the interaction of two, or even three, of the variables has an effect on the result. The three independent variables are denoted by A, B, and C and the levels by the subscripts 1 and 2. The number of possible combinations is $2^3 = 8$. A matrix, or block test-plan representation of the Full Factorial test is given as follows:

	B_1	B_2	B_1	B_2
A_1	C_1	C_1	C_2	C_2
A_2	C_1	C_1	C_2	C_2

There are several advantages in running a Full-Factorial-designed test program. These include complete information on individual main effects as well as complete information on all interactions. Furthermore, for the same number of tests, the Full Factorial approach gives a more precise estimate of the main effects than the step-by-step approach. The Full Factorial is the only type of designed experiment that will positively identify all interactions among the variables in the experiment, none of which may be identified by the step-by-step approach.

The tests must be run in a random or near-random manner to avoid distorting the results. An effect might be falsely identified as significant if runs at the low level of a variable were made at one time and the runs at the high level were made later. A shift in the measured output from the first to the second time period could be caused by any factor changing with time. Occasionally it may appear impossible to finish a Full Factorial, or some suspicion may exist that a main effect of a single parameter rather than the interaction of several parameters is the true reason for a given result. If so, it may be more judicious to run the first few tests in a predetermined sequence to gain the greatest preliminary information.

Full-Factorial-designed experiments have been used many times in the development of the RL 10; this report will discuss two types of Full Factorial experiments. The first of these, which is designated Full Factorial A, concerns a test made to determine the rather complex interrelationships that exist between engines and thrust controls. The second, which is designated Full Factorial B, concerns an analog program to determine the important parameters in a fuel pump stall situation. Although not strictly an engine test program, Full Factorial B represents the use of a limited Full Factorial experiment in the analysis of several parameters to evaluate their contribution to fuel pump stall.

1. FULL FACTORIAL A.

The Full Factorial A test program was run to determine the thrust control setting repeatability from engine-to-engine and to show the effect of running a series of preset thrust controls on a given single engine. In this manner individual control and engine characteristics were demonstrated independently. As a further variable, both hot and cold thrust control housings and thrust chambers were run. A representation of the plan is given below. The independent variables were engine number, thrust control number, and thrust control and chamber temperatures.

	Thrust Control A			Thrust Control B		Thrust Control C	
	Trim Hot	Repeat Hot	Repeat Cold	Repeat Hot	Repeat Cold	Repeat Hot	Repeat Cold
Engine I							
Engine II	"	"	"	"	"	"	"

The Full Factorial test program was justified because both individual engine and thrust control effects on engine trim setting repeatability had to be known so that specific steps could be taken to improve trim setting and trim repeatability. Also, it was known that a great deal of corollary information of the effect of thrust control characteristics on other engine performance parameters would be made available.

In this test, both engines were trimmed to rated thrust conditions with Thrust control A on the first run of the series and two repeat runs made with the same thrust control. Thrust Control A was then bench calibrated and Thrust Controls B and C were set to this calibration. The remaining runs of the plan were then made. The effect that each thrust control setting had on individual run-to-run trim repeatability and engine-to-engine repeatability was demonstrated in this manner; assuming that all other influences were known or approximated. From this test, it was shown that custom setting of thrust controls appears to be feasible although transient thrust control performance and steady state repeatability need further definition.

Additional conclusions are that the trim engine propellant mixture ratio has set stabilization time and is apparently independent of start inlet temperature. However, stabilization time for trim thrust is indicated at different values for a chamber cold start than for a chamber hot start. Other important conclusions on the effect of control characteristics on thrust overshoot, rate of acceleration, stability, and repeatability resulted from this test. Thrust control bench techniques required to preset thrust controls for field operation were determined.

2. FULL FACTORIAL B

Early in the RL 10 engine development program it was apparent that fuel pump stall was an occasional phenomenon of the engine; its occurence was dependent on several environmental and operational factors. Test experience and analytical studies indicated that at least 15 of these factors appeared to have a bearing on pump stall. Obviously, even if the number of values per parameter were limited to just two, the number of possible combinations of all of these factors in a Full Factorial would be a large number of test runs (2^{15}) or a large number of analog machine runs (as it was in this case).

The 15 factors that most affected pump stall were:

1. Fuel pump bleed valve area; pump interstage and discharge bleed valves are used on the RL10.

2. Fuel pump bleed valve closing schedule; that is, the points in the start transient where the bleed valve starts to close and where it finishes closing.

3. Fuel pump blade characteristics; that is, the amount of sweep in the first and second stages.

4. The fuel pump inlet pressure at the start signal and throughout the transient.

5. Engine thrust control gain -- usually expressed in terms of psi of chamber pressure per unit of mixture ratio variation.

6. The maximum thrust control bypass area that opens near the top of the start transient to limit engine overshoot by passing fuel around the propellant drive turbine.

7. The Lox pump pressure at which the mixture ratio valve on the Lox side of the engine opens.

8. Lox pump inlet pressure at start and during the start transient.

9. The jacket metal temperature that determines the amount of energy imparted to the turbine drive of the RL10.

10. Venturi area upstream of the fuel turbine.

11. Turbine to venturi area ratio.

12. Fuel pump discharge orifice diameter.

13. Turbine efficiency.

14. Fuel pump efficiency.

15. The amount of time the thrust control is open during the over- shoot control period.

To reduce the number of runs required to complete a Full Factorial test on these parameters, it was decided to concentrate only on those factors that were considered most important for the results of the experiment. Based on available experience, these were (1) fuel pump discharge orifice diameter, (2) fuel-side venturi area, and (3) the ratio of turbine area to venturi area.

A Full Factorial experiment of 24 runs was made as the first part of the analog program with four pump discharge orifice diameter values, three venturi area values, and two values of the ratio of turbine area to venturi area. Because this was an analog program, there was no advantage in running these tests in a random order. Based on the Full Factorial of 24 runs, optimum values of fuel pump discharge orifice diameter, venturi area, and turbine venturi area ratio were chosen to give maximum stall margin during the start transient. Although these were the most important factors as indicated above, they were not necessarily the only ones that could contribute to a stall situation.

The remainder of the analog program consisted of (1) holding the values of fuel pump discharge orifice diameter, venturi area, and the ratio of turbine to venturi area at the values found optimum from the 24-run Full Factorial and (2) independently varying each of the remaining parameters over a two or three-value range. For example, the fuel pump bleed valve closing pressure was varied between 100 and 300 psia, the fuel pump bleed areas were varied between 0 and 0.2 square inches for the fuel pump discharge valve, and between 0.2 and 0.5 square inches for the fuel pump interstage valve. Similar variations were made in the remaining parameters at either two or three levels and the effect of each of these parameters was evaluated. Based on this, a great deal of valuable information was gained on the effect of each of these factors on pump stall.

Some of the parameters are not necessarily controllable. For example, turbine and pump efficiencies are not independent variables but rather are variables that are dependent upon design considerations. Enough information was gained to greatly increase the stall margin of the RL10 engine as a result of this and several other experiments conducted both in the analog program and in the actual test program.

B. THE LATIN SQUARE TEST

If there is evidence that interactions are not significant between the several influencing parameters, and if information of the main effects of the several parameters is what is most desired, a Latin Square test program can be set up in which the parameters are arranged as illustrated in the table on the next page, with three parameters.

	a	b	c	d
I	A	B	C	D
II	B	A	D	C
III	C	D	A	B
IV	D	C	B	A

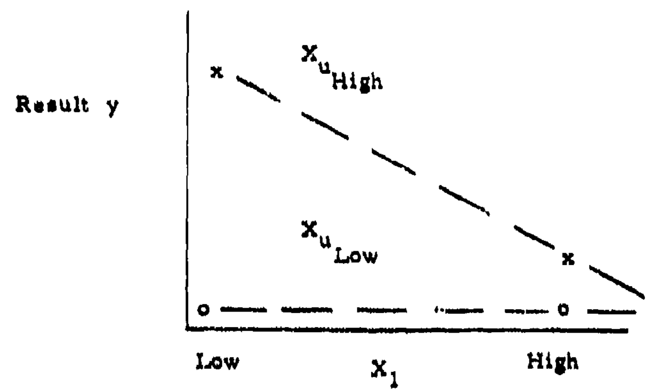
Parameter 1 (a, b, c, d)
Parameter 2 (A, B, C, D)
Parameter 3 (I, II, III, IV)

Here the three parameters, each at four levels, are arranged in a 16-test experiment. Note that each main effect of each of the three parameters is capable of being evaluated separately by the series of 16 tests. (A Full Factorial would require 64 tests.) However, all interactions (if existent) between these several parameters are mixed in with the main effects. The advantage of a Latin Square program is that not only are main effects established quickly but they are also established with a minimum number of runs.

An additional parameter may be identified with the same number of runs as with the Latin Square by adding another variable to the Latin Square. This is known as a Graeco-Latin Square; adding two variables (5 total) would result in a Hyper-Graeco-Latin Square. These additional estimates of main effects are made at the expense of the precision of the estimate of experimental error; i.e., the total degree of freedom is the same for all three designs, and each main effect requires a number of degrees of freedom that is one less than the number of its levels investigated.

It should be pointed out that in engineering problems it is difficult to state before the experiment, a priori, whether or not an interaction may exist. Many phenomena have demonstrated the existence of interactions. Interactions, so easily neglected in conventional engineering analyses, can be quite insidious. Therefore, Full Factorials are usually preferred to Latin Squares, Fractional Factorials, and other "short-cut" experiments.

The following curve demonstrates the danger of not considering interactions. X_1 is the variable being investigated, and X_u is the unknown variable.



While changing X_1 from a low level to a high level, if the unknown variable X_u happens to be at a high level, we get a very significant-appearing main effect of X_1. On the other hand, if, while changing X_1 from a low level to a high level, X_u were constant at its low level, we would discard X_1 as being completely unimportant. Here we have a strong ($X_1 \times X_u$) interaction, with possible moderate main effects of X_u and X_1. An objective search technique called "variation research" is available for finding unknown variables to put into statistically designed experiments. This search technique is not described in this paper.

During the earlier phases of the RL10 development program, thrust overshoot during the engine start transient was higher than desired limits on some runs. Three parameters that appeared to affect overshoot were thrust chamber metal temperature (which determines turbine power on acceleration in the RL10 regenerative cycle), thrust control time constant, and thrust control body or servo pressure. It was necessary to ascertain the main effects of each of these parameters as quickly as possible to determine where maximum development effort should be expended. Therefore, a simple four-run Latin Square program was set up. Since only four runs were involved, the order of tests was not randomized. The program consisted of the following:

Parameter	Levels	
Thrust control time constant, seconds	0.30	0.55
Thrust control body pressure at thrust control actuation, psia	55	70
Thrust chamber metal temperature, °R	300	570

Thrust Control Time Constant

Thrust Chamber Metal Temperature		0.55	0.30
	570°R	55 psia	70 psia
	300°R	70 psia	55 psia

(Cell entries: Thrust Control Body Pressure)

Note that this program required only four runs, while a Full Factorial would have required eight runs (2^3).

From the results of this program it was determined that both thrust control constant and body pressure had important main effects in determining thrust overshoot. The role of thrust chamber metal temperature in determining overshoot was verified.

Subsequent to this program, development was continued on the thrust control to limit thrust overshoot. This development effort required the running of several Full Factorial programs along the lines suggested by the results of the original Latin Square program.

C. THE RANDOM BALANCE TEST

If a large number of factors are believed to be influential in obtaining a certain result, a simple Full Factorial may involve an excessive number of tests. For example, if there are 10 factors each at three levels, a Full Factorial would require a total of 59,049 tests.

A random representative sample of such a Full Factorial as the Random Balance-designed experiment can be used to reveal the stronger main effects and interactions of the 10 variables. The combinations of levels of each variable for each test are then chosen at random. If 30 tests were permitted in this design, then each of the 3 levels of each factor would be tested 10 times. A Random Balance design may be analyzed with McBee edge punch cards, graphical regression analysis, and tests of significance, including analysis of variance.

If there are suspected interactions between several factors, a Multiple Balance-designed experiment would, with almost no loss, more completely evaluate such suspected interactions. For example, if 3 of the 10 factors are expected to be interacting, the 3 variables are laid out in a Full Factorial, 3^3=27 tests, with three of these cells (chosen at random) replicated or repeated for a total of 30 tests. The tests of this factorial are then listed in the random order of running and the remaining seven variables have their levels randomly but equally distributed throughout these 30 runs. The analyses are the same with the exception of the analysis of the factorial. This factorial analysis positively identifies significant interactions. Although in theory a great deal of information about the interactions of all of the parameters is lost, most problems yield to this test approach in practice because the one or two most important factors or combinations of factors are separated from the unimportant majority. This assumes, of course, that the factors considered for the random or multiple balance experiment were selected with good engineering judgment.

During the course of the development of the RL10, it became apparent that thrust control hysteresis or nonlinearity was a problem. This was evidenced by failure of the thrust control to repeat, for successive runs, a given setting of thrust and mixture ratio during steady-state operation. By applying engineering judgment the number of suspected causes was reduced to 10 factors, which are listed below. Factors A, B, C, and D were suspected as being most important, while factors A, B, and C were suspected to interact. The number in parenthesis after each of the factors indicates the levels of each of the factors.

A Bellows Assemblies (4)
B Reference Spring (2)
C Ball or Race Guides (2)
D Bypass Valve Assembly (2)
E Bypass Valve Spring (2)

F Lower Housing (2)
G Feedback Springs (2)
H Reference Spring Guides (3)
I Carriage (bellows) (2)
J Method of Thrust Control Assembly (2)

A Full Factorial of the 10 factors would require 3072 tests ($4x3x2^8$) to obtain complete knowledge of all interactions. Since factors A, B, and C were suspected to interact, a Full Factorial of these factors was built into a Random Balance experiment, thus making it a Multiple Balance experiment. It was pointed out that analysis could begin on about the tenth test.

The 32 tests are shown as follows. It will be noted that the Full Factorial of factors A, B, and C, with two replicates per cell, were actually run with the remaining factors selected at random, thus a Multiple Balance design, as shown on the next page.

MULTIPLE BALANCE DESIGN

Test No.	A	B	C	D	E	F	G	H	I	J
1	1	-	-	-	+	+	+	1	+	+
2	4	+	-	+	+	+	+	1	+	+
3	3	-	+	+	+	+	-	2	-	-
4	4	-	+	+	-	-	-	2	+	+
5	4	-	-	-	-	-	-	2	-	-
6	3	-	+	+	-	-	-	1	+	-
7	3	+	+	-	+	-	-	1	-	+
8	2	-	+	+	+	+	-	3	-	-
9	1	+	-	-	-	+	+	3	-	+
10	2	+	-	-	+	-	-	1	+	+
11	4	+	-	+	+	+	+	3	+	-
12	2	-	+	-	-	-	+	1	+	+
13	1	+	-	-	-	+	+	2	+	+
14	4	+	+	+	-	-	-	2	+	+
15	1	-	+	-	-	-	-	2	-	-
16	3	-	-	+	+	+	-	3	-	+
17	2	-	-	+	+	-	+	1	-	+
18	2	+	-	-	-	-	-	2	+	-
19	2	+	+	+	-	+	+	3	-	-
20	4	-	-	-	+	-	+	2	-	-
21	1	-	+	-	-	-	-	1	+	+
22	4	-	+	-	-	-	+	3	-	+
23	1	-	+	-	-	+	+	3	+	+
24	2	-	-	-	-	+	-	1	-	-
25	3	+	-	+	+	+	+	3	+	-
26	1	-	-	+	+	-	-	3	+	+
27	1	+	+	+	+	+	+	2	+	-
28	3	+	-	+	+	+	+	2	-	-
29	3	+	+	-	-	+	+	3	+	-
30	2	+	+	+	-	-	-	3	-	+
31	4	+	+	+	+	+	+	1	-	-
32	3	-	-	-	+	-	-	2	-	-

Factor A at 4 levels: 1, 2, 3, 4
Factor H at 3 levels: 1, 2, 3
Factors B, C, D, etc. at 2 levels: "+" "-"

The results of these tests showed that Factors I (carriage), J (assembly), A (bellows), D (bypass valve assembly), and F (lower housing) have the greatest main effects. Interestingly, an interaction between Factors A, B, and C was not conclusively demonstrated. The combination of levels of the factors which gave minimum thrust control hysteresis was sufficiently demonstrated by this test to enable a satisfactory solution of the hysteresis problem to be found. Thus, the problem was solved in a 32-run program in a far shorter time than would have been possible with the conventional cut-and-dry testing techniques.

CONCLUSIONS. Great advantages are possible with statistical testing techniques if certain guide lines are observed:

1. The urge to believe that mid-program results have solved the problem is strong and must be resisted. The program must be carried through to its conclusion.

2. The tests must be run in a random fashion even though this is not necessarily the fastest approach. For example, running the first half of a program exclusively on Stand A, and the second half of a program on Stand B, may add a degree of confounding that would not have occurred if random stand changes had been made. If such tests must be run in a nonrandom fashion, or in blocks, the results must be analyzed with this in mind.

3. Statistical testing is not a substitute for sound engineering judgment but simply a method for obtaining the most efficient testing program. There is no conflict between sound engineering judgment and statistical testing techniques.

COMPUTER SIMULATION STUDY OF BRUCETON AND PROBIT METHODS OF SENSITIVITY TESTING

John B. Gayle
National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Huntsville, Alabama

SUMMARY. A computer simulation technique has been used to evaluate the suitability of the Bruceton and Probit methods of sensitivity testing for launch vehicle applications. The results indicated that either method permits satisfactory estimates for the mean value of a normal distribution. However, for launch vehicle applications, estimates of the levels of stimulus corresponding to very high or very low percentage reactions usually are required; neither method permits reliable estimates for such levels when the distribution is non-normal, a condition which occurs with sufficient frequency that assumption of a normal distribution cannot be justified as routine procedure.

INTRODUCTION. The rapid growth of a launch vehicle technology has resulted in the widespread application of a variety of materials which are capable of reacting, either alone or with some other material, in the form of an explosion when subjected to a suitable stimulus. Test methods for determining the extent of the hazard associated with any particular material or material combination usually consist of subjecting samples to predetermined stimuli and noting the frequency of reactions. Because of the all-or-none nature of the results, a series of tests usually is conducted in which replicate samples are subjected to different stimuli and the results subjected to some form of statistical analysis, the details of which may be dictated by the particular test procedure employed. The method most generally accepted in the explosive industry is the Bruceton up-and-down method which specifies both the experimental procedure and the statistical analysis. An equally applicable method which is more generally accepted in the biological sciences is the Probit method which also specifies both the experimental procedure and the statistical analysis.

The purpose of this investigation was to determine the suitability of each of these methods for launch vehicle applications.

PROCEDURE. The amount of information which can be obtained from any laboratory study of a particular material usually is limited to that derived from the selected test method and no absolute basis exists for determining the extent to which this information is actually descriptive of the material being tested. Available information also is severely limited by the number of tests which can be included in any experimental study. This limitation is particularly important for all-or-none type data for which a minimum of 10 to 100 individual tests are required to obtain even a preliminary result.

To circumvent these and similar difficulties, the approach used in this investigation consisted of generating data for "systhetic explosives" of rigorously defined characteristics and comparing the various population parameters with those determined by the Bruceton and Probit methods of analysis.

A. BRUCETON OR UP-AND-DOWN METHOD

This method has been described by several investigators (Ref. 1). In practice, the range of stimuli which can be applied in a given test apparatus is divided into discrete increments of uniform spacing. Thus, for an impact apparatus for which the height of drop can be varied from 0 to 60 inches, one can define 31 levels separated by distances of 2 inches*.

To actually carry out a test, an initial level is selected which is guessed to be close to the mean value and a single sample tested at that level. If an explosion or other reaction is noted, a plus is recorded; if not, a minus. A second sample then is tested at the next lower level if a plus was recorded for the previous sample, or at the next higher level if a minus was recorded for the previous sample. A third sample then is tested in the same manner and so on until the scheduled series of tests is completed.

---

* For some applications, it is advisable to convert all heights to log units and separate the levels by uniform increments measured in log units. However, except for this transformation of units, the test is carried out in the usual manner.

For purposes of calculation, either the pluses or minuses may be used; however, the results for the first few tests are discarded up to but not including the result just prior to the first change in sign (plus to minus or minus to plus). Further details of the calculations have been adequately described and, therefore, are not included.

The method yields a mean value, corresponding to the level at which there is a 50/50 probability of a reaction, a standard deviation which is used to compute the level corresponding to any other probability of a reaction, and a standard deviation of the mean. It should be noted that the up-and-down aspect of the test procedure tends to compress the data about the mean value and, consequently, yields a biased (low) estimate of the standard deviation which is subsequently adjusted to obtain an unbiased estimate.

To carry out a computer sampling study, synthetic explosives were defined for which the probability of a reaction corresponding to any given level could be determined by use of a table of random digits. Consider, for example, a level for which the probability of a reaction is required to be 70 percent. A simulated test consists simply of drawing a pair of random digits and making a test to see whether or not the numerical value of the pair is equal to or less than 69. If so, a plus is recorded; if not, a minus. The level for the next test is selected by adding or subtracting one from the previous level (depending on the outcome of the previous test) and the series continued in the same manner as would be the case in an actual experiment.

As a matter of convenience, all of the test data were used for computations, none of the initial values being discarded. However, all initial tests were made at levels for which the probability of a reaction was in the range $0 < \% < 100$, and the effects of entering the program at random levels and at different fixed levels were considered. Each series of tests, consisting of 100 samples, was repeated 50 times and standard methods of statistical analysis were used to determine the reproducibility of the sample estimates.

The levels corresponding to a 5 percent probability of a reaction were estimated from the averages of the means and standard deviations for each group of 50 series.

B. PROBIT METHOD

The experimental portion of this method differs from the Bruceton method principally in that the individual tests are carried out at levels of stimuli selected by the investigator rather than those dictated by the preceding test result (Ref. 2) . Although both the particular levels selected and the number of tests made at each level may be varied within wide limits, in practice, approximately 5 levels usually are selected which are guessed to fall close to and on either side of the mean, and the number of tests at each level is constant, usually around 20.

The novel feature of this method is the transformation of the non-linear percentage reaction versus stimulus level plots obtained directly from the data, to linear plots (assuming a normal distribution) which is accomplished by expressing the reaction frequencies in standard deviation units referred to as "probits." Subsequent manipulation of the transformed data yields estimates of the mean, standard deviation, and standard deviation of the mean.

The computer sampling study for this method was carried out using the same synthetic explosives or populations that were used with the Bruceton method, the levels for testing being varied to determine the extent to which the particular levels selected influenced the results. Again 100 samples were used for each test series, 20 being taken at each of 5 levels. Also, each test series was repeated a total of 50 times to permit direct estimates of the reproducibility of data obtained by this method. The levels corresponding to 5 percent reaction were again estimated from the average results for each group of 50 series.

C. RESULTS

As indicated above, each series of 100 tests, whether made using the Bruceton or Probit method of analysis, provides estimates of the mean, standard deviation, and standard deviation of the mean. By repeating each series 50 times, average values were obtained for each of these estimates based on the calculations peculiar to each particular method of testing. In addition, application of conventional methods of statistical analysis to each group of 50 results provided estimates of the reproducibility of these values. Population values for the mean, standard deviations, and standard deviations of the mean were calculated from the frequency distributions in the usual manner; the levels corresponding to 5 percent reaction were interpolated from plots of the cumulative frequency distributions for each population.

The results for the Bruceton and Probit methods of analysis are not strictly comparable because of the influence of the point of entry for the Bruceton method, and the particular levels selected for testing for the Probit method. However, general comparisons, such as those given in the following sections, are believed to be indicative of the relative characteristics of the two methods.

The several populations or "synthetic explosives" selected for testing are shown graphically in FIG 1. Also, the probabilities of a reaction at any given level are summarized for each population in Table 1. It should be noted that the extreme ends (1 to 2 percent) of each distribution were truncated for convenience in simulating test data by using pairs of random digits. These populations include both normal and non-normal (peaked, skewed, and bimodal) types.

Application of either the Bruceton or Probit method assumes that the distribution of data is normal. Approximately normal distributions therefore were selected for the first two populations to obtain data under ideal conditions of testing.

Table 2 summarizes the results obtained by the Bruceton method for population No. 1. Table 3 gives similar results for population No. 2 which differs largely in that the 50 percent level was shifted from a population value of 4.00, which coincides with one of the levels selected for testing, to 4.5 which is midway between two of the levels selected for testing.

The estimated mean values are in excellent agreement with the population values. Also, the estimated standard deviations and standard deviations of the means are very close to the population values with some slight bias in the direction of low estimates being evident. This bias results in slightly high estimates of the levels corresponding to 5 percent reactions, but the discrepancy is of doubtful practical significance.

Tables 4 and 5 summarize the results obtained by the Probit method for populations No. 1 and 2. When the levels selected for testing are closely grouped around the mean, all of the estimates are

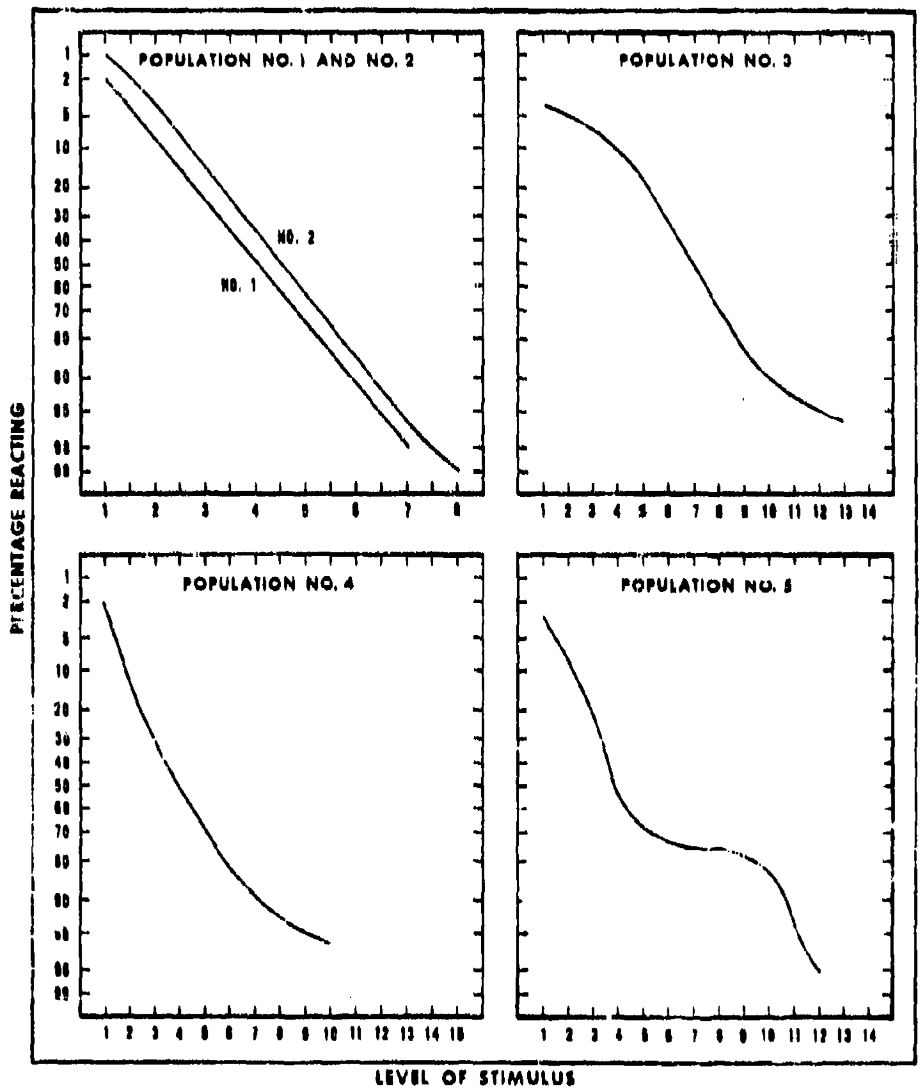


FIGURE 1. CHARACTERISTICS OF SYNTHETIC EXPLOSIVES SELECTED FOR TESTING

Table 1. Cumulative Frequency Distributions for Synthetic Explosives

Values in body of table indicate
reaction frequency at given level, cumulative percent

Population Type	Normal	Normal	Peaked	Skewed	Bimodal
Population No.	1	2	3	4	5
Level					
0	0	0	0	0	0
1	2	1	4	2	3
2	9	4	5	12	8
3	25	15	7	30	23
4	50	37	10	50	53
5	75	63	18	68	68
6	91	85	32	82	73
7	98	96	50	90	76
8	100	99	68	93	76
9		100	82	95	78
10			90	96	83
11			93	100	93
12			95		98
13			96		100
14			100		

Table 2. Summary of Bruceton Method Results for Normal Distribution (Population No. 1)

	Level for First Test in Series	Result	Std. Error of Result
Mean or 50% level			
Population value	-	4.00	-
Sample estimate	4	3.99	.03
do	6	4.05	.03
do	2-6*	4.01	.03
Standard Deviation			
Population value	-	1.49	-
Sample estimate	4	1.37	.04
do	6	1.42	.04
do	2-6*	1.43	.04
Standard Deviation of Mean			
Population value	-	0.21	-
Sample estimate	4	0.19	.00
do	6	0.19	.00
do	2-6*	0.20	.01
5% Level			
Population value	-	1.50	-
Sample estimate**	4	1.74	-
do	6	1.71	-
do	2-6*	1.66	-

* The level for the first test in each series was selected at random from levels 2 through 6.

** ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

Table 3. Summary of Bruceton Method Results for Normal Distribution (Population No. 2)

	Level for First Test in Series	Result	Std. Error of Result
Mean or 50% level			
Population value	-	4.50	-
Sample estimate	2-7*	4.49	0.03
Standard Deviation			
Population value	-	1.48	-
Sample estimate	2-7*	1.38	0.03
Standard Deviation of Mean			
Population value	-	0.21	-
Sample estimate	2-7*	0.19	0.00
5% level	-		
Population value	-	2.08	-
Sample estimate**	2-7*	2.22	-

* The level for the first test in each series was selected at random from levels 2 through 7 except that the probability of selecting levels 2 or 7 was one-half that of selecting 3 through 6.

** ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

Table 4. Summary of Probit Method Results for Normal Distribution (Population No. 1)

	Levels at Which Tests were made	Result	Std. Error of Result
Mean or 50% level			
Population value	-	4.00	-
Sample estimate	2-6	4.00	.18
do	0-4	2.89	.04
Standard Deviation			
Population value	-	1.49	-
Sample estimate	2-6	1.51	.04
do	0-4	1.86	.12
Standard Deviation of Mean			
Population value	-	0.21	-
Sample estimate	2-6	.22	.00
do	0-4	.32	.02
5% level			
Population value	-	1.58	-
Sample estimate*	2-6	1.52	-
do	0-4	-.17	-

* ($M-1.645\,\sigma$) computed from overall average values for mean and standard deviation.

Table 5. Summary of Probit Method Results for Normal Distribution (Population No. 2)

	Levels at Which Tests were made	Results	Std Error of Result
Mean or 50% level			
Population value	-	4.50	-
Sample estimate	2-6	4.24	0.02
do	0-4	3.08	0.05
do	4-8	5.21	0.C
Standard Deviation			
Population value	-	1.48	-
Sample estimate	2-6	1.46	0.04
do	0-4	2.72	0.45
do	4-8	1.44	0.35
Standard Deviation of Mean			
Population value	-	0.21	-
Sample estimate	2-6	0.22	0.00
do	0-4	0.52	0.09
do	4-8	0.23	0.01
5% Level			
Population value	-	2.08	-
Sample estimate*	2-6	1.84	-
do	0-4	-1.39	-
do	4-8	2.84	-

* (M-1.645 σ) computed from overall average values for mean and standard deviation.

Table 6. Summary of Bruceton Method Results for Peaked Distribution (Population No. 3)

	Level for First Test in Series	Result	Std. Error of Result
Mean or 50% level			
Population value	-	7.00	-
Sample estimate	3-11*	7.06	0.04
Standard Deviation			
Population value	-	2.69	-
Sample estimate	3-11*	2.21	0.09
Standard Deviation of Mean			
Population value	-	0.38	-
Sample estimate	3-11*	0.29	0.01
5% level			
Population value	-	2.00	-
Sample estimate**	3-11*	3.42	-

* The level for the first test in each series was selected at random from levels 3 through 11 except that the probability of selection of level 7 was twice that for the other levels.

** ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

Table 7. Summary of Bruceton Method Results for Skewed Distribution (Population No. 4)

	Level for First Test in Series	Result	Std. Error of Result
Mean or 50% level			
Population value	-	4.32	-
Sample estimate	1-10*	4.21	0.04
Standard Deviation			
Population value	-	2.20	-
Sample estimate	1-10*	2.35	0.12
Standard Deviation of Mean			
Population value	-	0.31	-
Sample estimate	1-10*	0.31	0.01
5% Level			
Population value	-	1.45	-
Sample estimate**	1-10*	0.34	-

* The level for the first test in each series was selected at random from levels 1 through 10.

** ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

Table 8. Summary of Probit Method Results for Peaked Distribution (Population No. 3)

	Levels at Which Tests were made	Result	Std. Error of Result
Mean or 50% Level			
Population value	-	7.00	-
Sample estimate	5-9	7.01	0.01
do	0,3,6,9,12	6.92	0.06
do	0-4	2.85	0.08
Standard Deviation			
Population value	-	2.69	-
Sample estimate	5-9	2.18	0.07
do	0,3,6,9,12	6.57	2.75
do	0-4	6.14	1.15
Standard Deviation of Mean			
Population value	-	0.38	-
Sample estimate	5-9	0.30	0.01
do	0,3,6,9,12	1.39	0.62
do	0-4	1.30	0.27
5% Level			
Population value	-	2.00	-
Sample estimate*	5-9	3.42	-
do	0,3,6,9,12	-3.89	-
do	0-4	-7.25	-

* ($M-1.645\ \sigma$) computed from overall average values for mean and standard deviation.

Table 9. Summary of Probit Method Results for Skewed Distribution (Population No. 4)

	Levels at Which Tests were made	Result	Std. Error of Result
Mean or 50% Level			
Population value	-	4.32	-
Sample estimate	3-7	4.72	0.02
do	0,2,4,6,8	4.27	0.04
do	0-4	2.82	0.03
Standard Deviation			
Population value	-	2.20	-
Sample estimate	3-7	2.22	0.07
do	0,2,4,6,8	7.00	4.10
do	0-4	1.81	0.08
Standard Deviation of Mean			
Population value	-	0.31	-
Sample estimate	3-7	0.31	0.01
do	0,2,4,6,8	1.49	0.95
do	0-4	0.3	0.01
5% Level			
Population value	-	1.45	-
Sample estimate*	3-7	1.07	-
do	0,2,4,6,8	-7.25	-
do	0-4	-0.16	-

* ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

Table 10. Summary of Bruceton Method Results for Bimodal Distribution (Population No. 5)

	Level for First Test in Series	Result	Std. Error of Results
Mean or 50% Level			
Population value	-	5.18	-
Sample estimate	2	4.22	0.05
do	7	4.40	0.04
do	11	4.84	0.04
do	2-11*	4.47	0.05
Standard Deviation			
Population value	-	3.23	-
Sample estimate	2	2.10	0.10
do	7	2.53	0.15
do	11	5.41	0.17
do	2-11*	3.24	0.23
Standard Deviation of Mean			
Population value	-	0.46	-
Sample estimate	2	0.28	0.01
do	7	0.32	0.02
do	11	0.44	0.08
do	2-11*	0.33	0.06
5% Level			
Population value	-	1.60	-
Sample estimate**	2	0.77	-
do	7	0.24	-
do	11	-4.06	-
do	2-11*	0.86	-

* The level for the first test was selected at random from levels 2 through 11.

** ($M-1.645\sigma$) computed from overall average values for mean and standard deviation.

in excellent agreement with the population values. However, when the levels selected for testing are distributed solely below the mean, the deviations between the sample estimates and population values for the mean and standard deviations become appreciable, and the sample estimates of the levels corresponding to 5 percent probability of reactions deviate markedly from the population values.

Tables 6 through 9 summarize results for two populations which are either peaked or skewed and thus depart appreciably from normality as shown in FIG 1. For the Bruceton method, the mean and standard deviations for population No. 3 are in good agreement with the population values, whereas the deviation for the level corresponding to 5 percent reactions is somewhat larger. This discrepancy is caused by the decreasing slope of the cumulative distribution curve shown in FIG 1 for values below approximately 15 percent. For population No. 4, the slope of the distribution curve increases for this same range of values, and the agreement between population values and sample estimates for the 5 percent level is slightly better.

Probit method results for population No. 3 are similar to those for the Bruceton method when the levels selected for testing are closely grouped around the mean. However, when the levels were dispersed over the entire range of those available, the agreement deteriorated appreciably. When the levels were limited to the lower half of the distribution, the sample estimate of the level corresponding to 5 percent reactions was markedly in error in the opposite direction to that noted when the samples were closely grouped around the mean. Results for population No. 4 exhibited somewhat different trends with the 5 percent levels estimated from samples closely grouped around the mean and those limited to the lower half of the distribution being in closer agreement with the population values, than that for which the sampling levels were dispersed over the entire range.

Tables 10 and 11 summarize results for a bimodal population, No. 5, (i.e., the distribution resulting from combination of two other distributions) such as might be expected for the mixtures of diverse chemical compounds used for proprietary materials. Bruceton method estimates of the 5 percent levels were particularly sensitive to the level selected for the first test. Inspection of some of the individual test data indicated that when the first test of a series was made in that portion of the combined distribution corresponding to the larger of the two single distributions, there was an appreciable number of instances in which the entire series was confined to that portion of the distribution. For these series, the characteristics of the smaller distribution would have little influence on the results. This situation, did not occur when the first level selected for testing was in that portion of the combined distribution corresponding to the smaller of the two single distributions. However, regardless of the point of entry selected, the sample estimates of the 5 percent levels were not considered to be in satisfactory agreement with the population values.

Probit method results for this distribution also were extremely sensitive to the particular levels selected for testing with agreement ranging from poor, in some instances, to ridiculous in others.

DISCUSSION AND CONCLUSIONS. The results reported herein, as well as similar results for several other populations which have not been included, demonstrate conclusively that tests at levels located close to the mean value of a normal distribution provide excellent estimates of the characteristics of the population regardless of whether the Bruceton or Probit method is used. However, when the distribution is non-normal, and, particularly, when it is bimodal in character, the estimates provide only rough indications of the population parameters, and in particular, the estimates of the levels corresponding to very large or very small percentage reactions become extremely unreliable. These discrepancies are the direct result of linear extrapolations of non-linear data. No relief from this problem is possible with the Bruceton method, which is specifically intended to concentrate the testing at levels close to the mean. However, with the Probit method, the operator is free to select the levels at which tests are carried out and can thus concentrate his efforts on the particular end of the distribution of greatest interest for his application. The results obtained in this investigation, however, do not indicate that even this modification provides reliable estimates of any of the statistical parameters considered since linear extrapolation and interpolation are still used with non-linear data.

Most launch vehicle applications are concerned with either the high or low end of a distribution. In fact, probably no point of the distribution is of less practical significance than the 50 percent point. Thus, an explosive of 50 percent response to a given stimulus would be unacceptable from either a performance or safety point of view. Conversely, levels corresponding to very large or very small percentage reactions, usually 95-5 or 99-1, must be considered. Results given in this report indicate that neither the Bruceton or Probit methods provide reliable estimates of stimuli corresponding to very large or small probabilities of reaction unless the population is normally distributed. For many applications, the requirement of a "normal distribution" may be satisfied if only that portion near the mean approximates a normal distribution. However, for reliability and safety applications, the normality of the distribution must include not only that area around the mean but also the area between the mean and the percentage frequency of interest. Assumption of such normality does not appear justifiable in view of the heterogeneous character of the systems undergoing test. Determination of the normality or lack of normality of such a system would require expenditures of time and effort in excess of those which can be justified. It is recognized, however, that valid comparisons of parameters for related distributions are sometimes possible even when the assumption of normality is appreciably in error.

To obtain additional information regarding the normality of distributions encountered in actual testing, results of LOX impact tests on a number of materials were examined. Only a few of these materials gave results which could be reasonably interpreted as normal, the distributions for the other materials varying widely. FIG 2 presents typical curves for one normally (titanium alloy) and two non-normally distributed sets of data. The curve for pickled titanium is of particular interest in that the slope of the curve changes sign twice within a relatively narrow range of impact energies. Ordinarily, cumulative frequency distributions are obtained by cumulating frequency of occurrence data. The nature of the cumulating process is such that changes in the sign of the slope of the data are precluded. Conversely, for sensitivity test data, each point on the curve is determined directly and no cumulating operation is involved. The changes in sign noted for these data, and also for data reported by other investigators, suggest that the mechanism of the process is complex and varies with the level of stimulus supplied. It, therefore, is debatable as to whether the data should be considered to represent a frequency distribution in any sense of the word.

In view of these considerations, it appears that the best method available at this time is to carry out a systematic search for a sensitivity threshold corresponding to some acceptable frequency of reaction and to utilize graphical and statistical techniques to evaluate the results. Such a procedure is used in LOX impact testing at this Center. It should be noted that data for such a procedure can be subjected to the Probit calculations, in the event that estimates of the mean and/or standard deviations are considered essential.

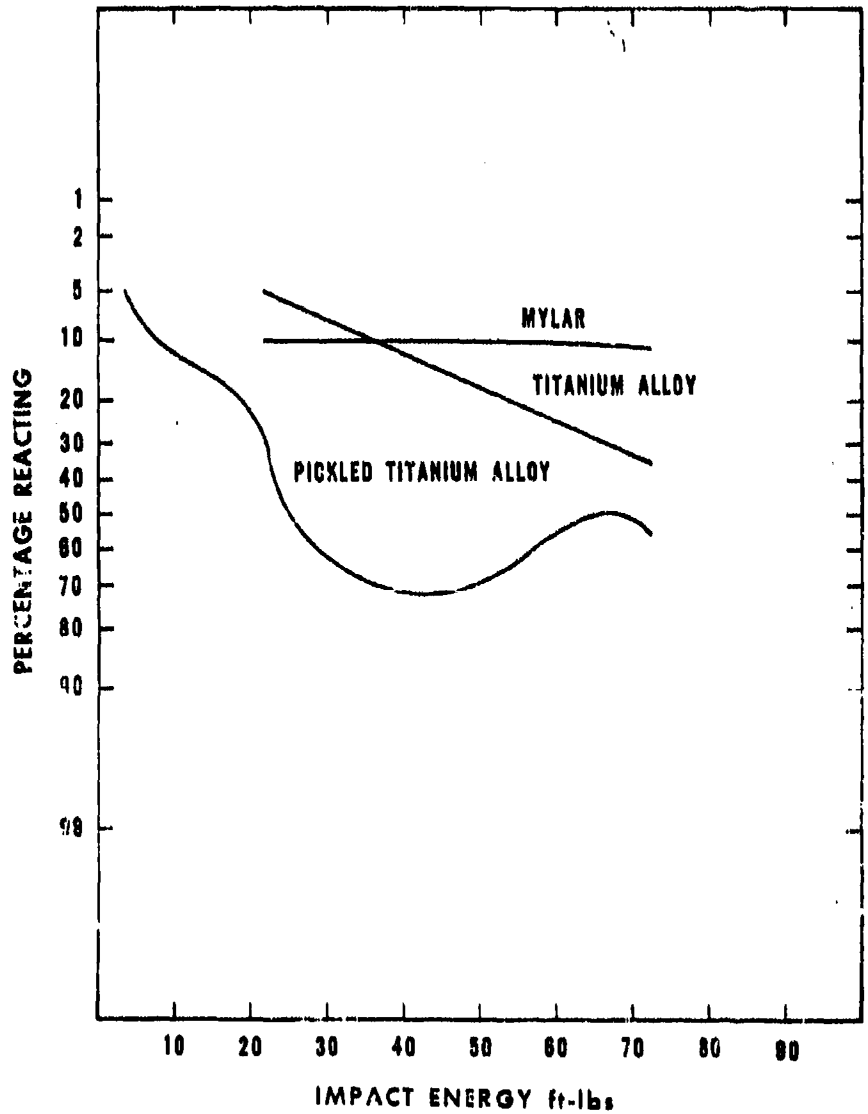


FIGURE 2. TYPICAL RESULTS FOR LOX IMPACT SENSITIVITY TESTS

REFERENCES

1. Statistical Research Group, Princeton University, "Statistical Analysis for a new Procedure in Sensitivity Experiments," OSRD 4040, July 1944.

2. Finney, D. J., "Probit Analysis," Second Edition, Cambridge University Press, 1962.

THE IMPACT OF ADMINISTRATIVE LIMITATIONS ON THE DESIGN OF EXPERIMENTS

Gerhard J. Isaac
U. S. Army Medical Research and Nutrition Laboratory
Denver, Colorado

ABSTRACT. Experiments often must be carried out with less than ideal designs because of limitations having their origin at administrative levels. Designs may be modified to reflect administrative decisions on the utilization of available resources of personnel, funds, and equipment in a research program. The nature and effect of some of these modifications is discussed. An example is presented and some alternatives are explored.

Experimental designs often are influenced by administrative limitations placed upon the experimenter or the research unit. The result may be a frustrated researcher, or at least one whose ingenuity may be sorely tried as he searches for a feasible compromise between what he considers an ideal design for exploring the problem and the practical realities of the situation. This may also place a heavy burden on the statistician-advisor in suggesting a statistical design which will maximize the quantity and quality of information obtained under somewhat less than ideal conditions.

The limitations referred to are those having their origin at administrative levels rather than at the research level. They may be in the nature of specific directives related to particular activities, merely general statements of policy, or even not specifically spelled out at all. The imposition of time limits may be one form of limitation. The necessity for an early administrative decision may lead to a request for a quick answer to some research problem. Perhaps more often the availability of funds or personnel is at issue. Closely related is the question of priorities for various projects. Or even the facilities available are limited and expansion is impossible. In other words, we are concerned with the impact on the design of research projects of administrative decisions on the utilization of available resources of personnel, equipment and funds in accomplishing the mission of the organization.

The thought might be interpolated here that the experimenter himself also makes many similar decisions when he decides how he is to divide his time and effort among various projects or segments of projects. Actually in practice there is a merging of these influences in the final decisions regarding operations.

In his first thinking about a problem an experimenter may feel that an extensive experiment is required because existing knowledge about the problem is very limited. A re-statement of the problem may provide a setting for developing, within the various limitations present, a design which is both feasible and statistically valid. It is, of course, always assumed that relevant findings of others have been examined to provide guidance in delimiting the scope of the project. Repetition of the work of others simply to see if you can duplicate their procedures can be a wasteful process. This is not to say that there may not at times be reasons to repeat experiments of others to see if the conditions can be reproduced and confirmatory results obtained. But, repetition simply for the sake of repetition can be wasteful of resources.

Let us then explore some of the consequences of the imposition of administrative limitations on the conduct of experiments. Adjusting the design to meet the situation can be done in many different ways. It has already been mentioned that a restatement of the problem may provide the basis for modifying the original proposal. Some obvious changes quickly come to mind. A smaller number of subjects might be used or the number of treatment levels reduced. The experimental period might be shortened. Test parameters which are to be evaluated might be restricted. Sometimes a pilot study is an economical way to establish limits within which the final study is run. With careful planning a pilot study can be part of the initial stages of a more comprehensive study. With limits established for the area of primary interest it may become possible to use a fairly simple design with straight forward comparisons. This statement is not intended to imply that it may not be desirable to use factorials or other design forms which increase the information available more rapidly than the cost.

In making adjustments in the design, care must be exercised to insure that the design does not end up "unbalanced" and introduce complications into the statistical analysis and the interpretation of the results. Unbalance is not necessarily fatal, but it could be if the implications to the analysis were not fully anticipated in redesigning the experiment. If not anticipated, the statistician might be faced with a "salvage job" or a great loss of information because an important part of the data became unanalyzable.

How will the statistician meet these problems? Mention has been made of the use of smaller numbers of subjects. Standard statistics text books contain discussions of the effect upon the inferences to be drawn of changes in the numbers of subjects or observations. Somewhere along the line it is always pointed out that the selection of the number of observations to be used depends in part on the variability expected in the various parameters, and on the precision of the measurements. The probability that inferences about a population or universe drawn from sample observations are valid depends upon such factors as these. Important in the problem then is the probability level that is considered appropriate in the case of a particular experiment.

Except in the most unusual situations replicate determinations should be performed. Usually duplicates are enough, though if the particular procedure is known to lack precision it might be well to use triplicates. The decision rests on the variability expected to occur in the individual measurements. The variability expected among the subjects in turn has much to do with the number of subjects to be used. Equally important here is the size of the differences between treatment groups which will be considered important. The sample size must be large enough to detect differences that are important in the light of what is known or estimated about the variation in the population. It could lead to unfortunate consequences if decisions on these matters, for example, rested solely on the availability of personnel, rather than taking into account variations of the types just mentioned. However, to the extent that the presence of administrative limitations forces harder thinking and more careful planning, and thus leads to a "tighter" design, the results may even be beneficial in the long run.

I would now like to explore with you a specific project, the design of which required taking into account some administrative limitations of the types I have been talking about. Briefly, the purpose of this project is to test in the field five different rations designed for non-resupply situations of perhaps 10 days duration for small groups of men. Study parameters include certain biochemical procedures on blood and urine samples, performance tests, and the subjects' evaluation of the rations. Out of this study will come statements regarding the nutritional adequacy and acceptability of the test rations. Also, it is expected that suggestions will come up that might aid in the development of improved rations for use in this particular type of field situation.

The ideal design that immediately comes to mind is to organize 5 patrol groups and set up a 5 x 5 latin square pattern for feeding the rations. In this way each man would be on each ration sometime during the test, and during each cycle each ration would be tested under the same prevailing environmental conditions.

The time required would be the 10 days in the field during each cycle plus a rest period between patrols. A rest period of this kind serves two purposes. One is a recovery from the prior test conditions before being subjected to a new test. The other purpose is to give time for remeasurement of basal study parameters. This would permit reestablishment of normal values for each man immediately prior to entering a new phase of the experiment. Incidentally a morale factor is also involved. It would be difficult to maintain morale among the troops if they were on patrol continuously for a period as long as required by a project of this kind.

It was pointed out that a full schedule on a 5 x 5 latin square pattern would tie up the troops and test personnel involved for nearly 90 days. This was considered excessive and a limit of approximately 55 days was set. In partial mitigation it was determined that 60 men might be counted on, so that 6 test groups of 10 men each could be formed. With regard to the test rations it was noted that numbers 1, 3, and 5 were considered slightly more important than 2 and 4.

With these specifications before us the problem was to set up a workable design for the field work. The plan established provided for 6 groups of 10 men each who were to be sent on 10 day patrols, three different times. The ration to be carried and consumed by each group was assigned by a random procedure for each phase, with the modification that rations 1, 3, and 5 appear four times and 2 and 4 appear three times (Table 1). After the patrol groups are formed they will be assigned to feeding sequences A through F by drawing out of a hat or other random process.

TABLE 1. ASSIGNMENT OF RATIONS TO GROUPS AND PHASES

	Patrol Group					
	A	B	C	D	E	F
Phase 1	3	1	2	5	4	3
Phase 2	5	4	5	2	3	1
Phase 3	1	3	1	4	2	5

How have we fared in setting up our design? We have assured that no patrol group tests the same ration twice. During each phase all rations are tested, with 3, 5 and 1 each duplicated in one of the three phases. The overall time, including preliminary briefing, "before" and "after" studies, 10 day patrols, and 5 day rest periods, will be kept to 55 days. In terms of the administrative limitations imposed we have met the specifications.

What about the analysis and the inferences to be drawn from the results? For any of the test parameters where large differences appear there should be no problem. It is in the area of the more subtle differences that there may be a problem. Will we have to say that we expect considerable variation among the men because of the nature of the test? Any questions dealing with the subjects' evaluation of the rations will have all the subjective elements that occur in all food acceptability tests. The biochemical procedures are reasonably precise and objective, though variation among normal subjects often is great. The performance tests also are essentially objective, but in this area there always is some question about the effect of motivation on test scores and about the physiological validity of the tests as measures of response to the various rations.

In this test our first interest is in the nutritional adequacy of the diet in maintaining a soldier, operating under the prescribed conditions, as an efficient fighting machine. Next in importance is the acceptability of the ration because of the relation of acceptability to adequate intake and to morale generally.

In conclusion then, I suggest that in a test of the kind described, the design should be developed as far as possible in accord with sound statistical practices, but that some deviations from the ideal are not necessarily fatal since the interest is more in gross differences than in those of a more subtle nature.

VERIFICATION OF PRODUCT ACCEPTANCE INSPECTION BY ATTRIBUTES*

Joseph Mandelson
Directorate for Quality Assurance, Chemical Corps Materiel Command
Edgewood Arsenal, Maryland

The framework of reference of this paper is that of Department of Defense Handbook H109, 6 May 1960, "Statistical Procedures for Determining Validity of Suppliers' Attributes Inspection." The purpose of the paper is to reveal the thinking which guided the statistical research performed and which must form the groundwork for the administrative uses to which the Handbook can be put. H109 may be applied to an increasingly popular contractual arrangement between consumer and supplier in industrial and governmental circles. This arrangement now forms part of the procurement policy of the Department of Defense, namely, that the supplier, prior to offering his product for acceptance, will perform the inspections and tests necessary to ascertain that the material meets all quality requirements established by the consumer and made part of the contract. To do this, the supplier binds himself to establish a system of quality inspection over and above his own system of quality control, to cover all those quality characteristics, inspection procedures and tests required by the consumer to satisfy him that the product meets all contractual requirements which define material acceptable to him.

One aspect just mentioned may be a bit confusing. It may not be immediately apparent why the supplier should institute a quality inspection system over and in addition to his quality control system. It is a major objective in quality control to watch the effect of the manufacturing process on component quality characteristics of interest to minimise production of non-conforming items. When such material seems about to be produced, suitable changes are instituted in the process to bring it back into statistical control and to continue production of conforming material.[1] On the other hand, the major objective of the quality inspection system is to determine that the material manufactured meets all quality requirements prescribed in the contract. In a sense, it is a measure of the success of the quality control system.

---

*This paper was presented at the Conference on Reliability Assurance Techniques for Semi-Conductor Specification, and appears in the Proceedings of this conference. We wish to thank the editors of said Proceedings for permission to reprint this paper.

To return to the policy of the Department of Defense and of many industrial consumers, for many practical reasons, they insist that the supplier offer for acceptance only material which he has good and sufficient reason to believe conforms with all contract requirements. The only way to gain such knowledge objectively is to perform conscientiously and competently the various inspections and tests by which the contract defines product acceptability. This, of course, clearly requires that the supplier establish a system of quality inspection for his finished product which parallels the system which the consumer must plan to assure that the quality provisions of the contract are being honored. The supplier will certainly include the cost of his quality inspection system in his over-all item cost; the consumer must expect this. At the same time, since the consumer will eventually pay its cost, he has a special right and interest in the supplier's quality inspection system, one might say a proprietary interest. In a sense, the consumer is buying not only a product but a service -- the quality inspection system. The consumer can use the supplier's system to reduce his own acceptance inspection cost and yet lose little, if anything, in assurance of product quality.

This can be done through procedures collectively called "product verification" based on the following philosophy: If the supplier's quality inspection system is competently established and run so that it yields valid, objective, data, equivalent in all respects to the data the consumer himself would get by inspecting the same lot, then the supplier's data are just as trustworthy as a basis for acceptance as the consumer's. Hence, if the latter is assured that the supplier's quality inspection system yields valid data, he may forthwith accept any material the supplier offers for he may confidently expect it to meet contractual requirements were he himself to inspect it fully. This assurance is obtained by the procedures collectively described as product verification.

The first step in product verification is to determine that the supplier has the organization and physical means in being or available to him to perform all the inspections and tests required by the contract to determine the acceptability of the product. This means not only the several gages and pieces of equipment involved, but the means to calibrate or standardize them periodically as required.

The next step is to determine that the supplier has a definite plan or program for using this equipment as specified and that the personnel assigned to do this work are competent in all technical and administrative phases. The consumer must particularly be satisfied that the supplier has established a reporting system that will make available to both parties complete and timely information as to the technical findings of the quality inspections.

Third, the consumer must be sure that the supplier's quality inspection personnel and his own acceptance personnel use the same inspection and test procedures and standards so that they get the same results on inspecting a given sample. To avoid differences, misunderstandings, and disputes, visual quality standards should be established to the extent required so that, ideally, both parties would always agree as to whether a given item is defective.

Fourth, we enter upon validation, the field of coverage of H109. The consumer must perform his own inspections and tests of the material offered for acceptance. It is not the primary purpose of this activity to determine the acceptability of the material inspected. Rather, the objective is to determine whether the results obtained by the supplier's quality inspection system are valid and essentially the same as those obtained by the consumer and may, therefore, be used to justify acceptance of the material offered. In other words, the supplier offers for acceptance lots of material which his quality inspection system has found, through objective evidence, to meet all contractual requirements. If his results are determined to be valid by the consumer, then the latter should be willing to accept the supplier's inspection data as sufficient evidence of the acceptability of the material. It is assumed that the consumer's inspection results are a standard. If the results generated by the supplier's quality inspection system are essentially the same as the standard, then acceptance may fairly and properly be based upon them.

The essence of the product verification which Department of Defense policy requires of the Government inspector, then, is to determine the existence of the physical means and organization for quality inspection, of the necessary training and know-how on the part of the supplier's personnel using these means, and continued surveillance of the apparatus and procedures employed. Validation is the last step and is a tool for checking the effectiveness of the entire system of quality inspection which is constantly under surveillance by the consumer's inspector. It is plain, from the breadth of responsibilities necessarily associated with this policy, that the consumer's inspector's span of capabilities and competence must be far beyond those of the gage-pushing inspector of yesteryear if he is to be expected to perform competently the duties laid upon him.

For one thing, he must be thoroughly familiar with the concepts laid down in this paper and equally with the new responsibilities he bears. He must recognize that the specific techniques necessary to execute these responsibilities may be expected to vary, and therefore must be developed anew, in each supplier's plant. This program presents a definite challenge to the old-line inspector directed toward the up-grading of his abilities in the modern age.

The purpose of validation is to check the supplier's system of inspection. Hence, it is desired to check results on both accepted and rejected lots for these will enter into computation of the process average which, in MIL-STD 105, controls reduced, normal, and tightened inspection. It is the consumer's purpose, then, to use his validation results to justify the data generated by the supplier's quality inspection system; he is not trying to find whether to accept the material. It is this point which seems most difficult for the newcomer to validation to grasp. As long as the supplier's results are found valid, his data may properly be used to justify acceptance of the material he offers as meeting all requirements. When the consumer's results appear to differ significantly from those of the supplier, as indicated in application of Table I or Table III of H109, then the consumer's representative must review the supplier's quality inspection system from stem to stern to determine the cause of the discrepancies noted. Administratively, there seems to be little point in taxing the supplier with the fact that his data do not jibe with those of the consumer unless he can be told why. In this area, one must keep his criticism constructive but this does not absolve the supplier from the responsibility of reviewing his own activity once he is told the validity of his data is in question.

Department of Defense Handbook H109 was prepared to furnish statistical tools whereby the validity of the supplier's data may be adjudged in comparison with data obtained in validation inspection. Since the Handbook is intended for use by inspectors in the field, it was designed to require a minimum of computation, both in quantity and sophistication, and the procedures were simplified to the degree possible. The purpose was to publish a procedure which would require no previous statistical background on the part of the user.

It is assumed in H109 that the supplier has already inspected and tested the lot in question for all quality characteristics listed in contractual requirements and that the data generated have been made available in detail to the consumer. It does not matter whether the lot was found acceptable or rejectable. In either case, the consumer may sample and test the same lot for the purpose of verifying the validity of the data furnished by the supplier. He takes a sample of such size that the ratio of the supplier's sample to his is r, where r = 1, 2, 3, 5, or 8, as shown in Table I, H109. The consumer inspects the sample and finds a number of defects or defectives, say d_c. He compares d_c with d_s, the number of defects or defectives found by the supplier in his sample. This comparison is made in Table I which furnishes the limit for d_c for any d_s within the given sample ratios, r. If this limit is equalled or exceeded, the consumer may proceed on the theory that the supplier's data are invalid and that, therefore, his quality inspection system should be reviewed to detect the shortcoming responsible. If such shortcoming is found, the consumer should reject the supplier's data for the lot in question and substitute his own as a basis for deciding whether the lot is acceptable or otherwise. The "action" limit of Table I was computed on a 5-percent level of statistical significance so the chance is excellent that an "action" indication will result in finding a true physical discrepancy. However, the possibility still remains that no shortcoming really exists. In this case, there is as much reason to trust one set of data as the other but it is perhaps most fair to request the supplier to perform a reinspection of the questionable lot under the direct surveillance of the consumer's inspector.

H109 also contains procedures to detect small biases or insidious discrepancies which cause relatively small bias in the results and, therefore, would show up only in a sort of historical review of the evidence from a number of consecutive validations. Thus, Table I indicates discrepant results from a single validation, but the tendency for discrepant validation results to accumulate in a sequence of validations is picked up by Tables II and III, H109. Thus, from Table II (which has different sections, one for each value of r) we can obtain a so-called "check rating" for each pair of d_c and d_s observed from each validation performed. The check ratings obtained from Table II are equally valid when different r's are used or for different defects or defect classes. Its omnibus nature makes it quite flexible and valuable. The check ratings are added together and their sum compared with the "median" values and the "warning" and "action" limits given in Table III for the number of lots whose check rating, are summed. The median value is what is expected if the supplier's results are commensurate with those of the consumer. If the summed check ratings reach the "Warning Critical Value" there is reason to seek a discrepancy in the supplier's quality inspection system. As before, this limit was calculated at a 5-percent level of statistical significance and the action to be taken is that already described. However, if the "Action Critical Value" is reached (at the 1-percent level of significance) the consumer is justified in rejecting the supplier's data, whether or not he has been able to locate the source of trouble in the latter's system.

Whenever observed differences have been traced to some shortcoming in the supplier's quality inspection system and the system has been corrected to the consumer's satisfaction, he can start on a new cycle of validation. Until some shortcoming has been found and corrected, there is no reason for starting a new cycle.

Quite aside from the contractual nature of the agreement, the supplier must always recognize that his quality inspection system is being sold as a service to the consumer and, as such, must satisfy him as to the propriety of its structure, procedures, and results, the validity of which is checked by his product verification activities. The supplier must always be willing to modify or improve his system as requested by the consumer so long as the validity of his results is not affected adversely thereby. The supplier's willingness to please the consumer stems from recognition that the latter is paying for the system and uses the data it generates as a basis for acceptance of the product.

ADDITIONAL ANALYSIS OF MISSILE TRAJECTORY MEASURING SYSTEMS

Oliver L. Kingsley and Bernie R. Free
Range Instrumentation Systems Office
White Sands Missile Range
New Mexico

I. INTRODUCTION. Four flight tests have been conducted for the study of accuracy and precision of missile trajectory measuring systems at White Sands Missile Range. Data collected from the first two tests have been analyzed, interim reports have been published, and high lights of the results were presented by Mr. Kingsley at previous conferences on the design of experiments. Data from the third flight test is in the process of being analyzed.

The purpose of this paper is to present a summary of the results from the third flight test and to compare them with results from the first two tests. The four tracking systems to be discussed are the ballistic camera, Askania cinetheodolite, DOVAP, and FPS-16 radar.

The ballistic camera, which is a fixed camera, photographs a flashing light on the missile against a star-trail background. The plate from each camera yields angular data and missile position is computed using the method of over-determination.

The Askania cinetheodolite is a tracking camera. It photographs the missile along with internal dials which show the azimuth and elevation angles of the optical axis for each frame of the film. Position data is computed using the over-determination method.

DOVAP is a continuous wave electronic system which utilizes the doppler principle to determine position data.

Each FPS-16 radar measures range and azimuth and elevation angles. Data from each radar can be used to estimate the missile trajectory, or data from several radars can be combined to give a composite estimate. The radars from which data is combined form a composite radar system.

Although the term "system" is sometimes used in this analysis to refer to one of the four trajectory measuring systems, in a broader sense it means the end-to-end process of collecting, converting, reducing and reporting position data. It is in this broader sense that the estimates of accuracy and precision have meaning. In other words, the data on which analyses were performed contain the effects of final data reduction and reporting.

The analysis for the third test covered three sections of the trajectory as shown in Table I. Data were available from all systems, except DOVAP, for all three trajectory sections. Data were available from the DOVAP system for Section I only. The bias error analysis for each system will be presented first. Then the precision analysis will be discussed. Significant improvements in system performance will be pointed out.

TABLE I

Section	Data Points Sampled	Nominal Time Along Trajectory (seconds)
I	48	34 to 60
II	44	101 to 125
III	55	163 to 192

II. BIAS ERROR ANALYSIS.

A. Ballistic Camera. When properly located the ballistic camera system can yield very good unbiased trajectory data. For this reason the ballistic camera system was chosen as the bias standard for WSMR and bias errors in the system are assumed to be zero. Using data from the ballistic camera system, bias estimates for the other systems were obtained as follows:

A parameter measured by the jth instrumentation system at the ith time may be expressed as:

$$x_{ij} = x_{ti} + b_{ij} + e_{ij}$$

Where x_{ti} = the true value of the parameter at the ith time,

b_{ij} = the bias error of the jth system at the ith time,

and e_{ij} = the random error of the jth system at the ith time.

An error for the jth system at the ith time is written as

$$\Delta x_{ij} = x_{ij} - x_{iBC}$$

where BC represents the ballistic camera system. This can be rewritten as:

$$\Delta x_{ij} = x_{ti} + b_{ij} + e_{ij} - (x_{ti} + b_{iBC} + e_{iBC}) = b_{ij} + e_{ij} - e_{iBC}$$

since it is assumed that $b_{iBC} = 0$. Next, it is assumed that the sum of the n random errors associated with the n measurements made by the jth system goes to zero. This assumption is written as:

$$\sum_{i=1}^{n} e_{ij} = 0. \quad (j = \text{any instrumentation system})$$

Now, if the Δx_{ij} are summed for the n measurements, we have

$$\sum_{i=1}^{n} \Delta x_{ij} = \sum_{i=1}^{n} b_{ij} + \sum_{i=1}^{n} e_{ij} - \sum_{i=1}^{n} e_{iBC}$$

The last two terms on the right go to zero under the above assumption, and the equation becomes

$$\sum_{i=1}^{n} \Delta x_{ij} = \sum_{i=1}^{n} b_{ij}$$

The sum of the left can be computed directly from data from the ith system and corresponding data from the ballistic camera system. It is an estimate of the sum of the n bias errors, b_{ij}, associated with the n measurements made by the jth system. On taking the mean as follows

$$\frac{1}{n} \sum_{i=1}^{n} \Delta x_{ij} = \frac{1}{n} \sum_{i=1}^{n} \Delta b_{ij} = \bar{b}_j$$

the expression on the left yields a bias estimate for the jth system.

B. Askania Cinetheodolite. Table II presents bias estimates, in terms of cartesian components, obtained from the first three tests. These estimates

TABLE II

Flight Test	BIAS ESTIMATES FOR ASKANIA CINETHEODOLITES Askania Component Minus B.C. Component (Estimates in Feet)		
	North	East	Up
1	6	-7	-29
2	3	-1	-30
3	4	-4	-27

TABLE III

Flight Test	BIAS ESTIMATES FOR DOVAP DOVAP Component Minus B.C. Component (Estimates in Feet)		
	North	East	Up
1	46	-20	-80
2	24	7	-68
3 (36.9 mc)	13	-6	12
(36.2 mc)	9	-8	13

were computed using trajectory data from all sections. For each component the bias estimates obtained from the three tests are alike in sign and similar in magnitude. Estimates for the North and East components are reasonably small, whereas they are consistantly larger in the Up component. Apparently, the corrected elevation angles are smaller than the true elevation angles. Perhaps a better approximation of the refraction correction would yield a desired improvement of the negative bias in the Up component.

For Test No. 3 all three component bias estimates were significant at the 5% level compared to an expected value of zero. Also, the analysis of variance of error data revealed a significant shift in the magnitude of system bias over Section I in the North component, Section II in all three components, and Section III in the North and Up components.

C. DOVAP. Test No. 3 was instrumented with two DOVAP systems. The standard system operated at 36.2 mc. The second system, which operated at 36.9 mc, was used to test the Interstate Transponder. DOVAP data were available in Section I only. Table III presents the component bias estimates for the system for the first three tests.

Each test revealed improved system bias. Electronic reading and digitizing equipment was introduced into the reduction process on Test No. 2. For Test No. 3 signal propagation velocities were estimated using a table look-up technique which took into account the missile height above the transmitter. These changes in data reduction greatly reduced the system bias. Further reduction of system bias was investigated through improvements in start point determination and by making adjustments for the relative locations on the missile of the flashing light (ballistic camera system reference point) and the DOVAP antenna.

The bias estimates for Test No. 3 were all significant at the 5% level compared to an expected value of zero. Also, there was a significant shift in the magnitude of system bias over the section in the North and East component

D. FPS-16 Radar. Bias estimates for three of the FPS-16 radars are shown in Table IV. All of the radars operated in the skin track mode. The estimates are, in general, larger than those for the Askania and DOVAP systems.

The bias estimates for R-112 are smaller for Tests No. 1 and 3 than for Test No. 2. They range in magnitude form 2 to 59 feet. For R-114 the estimates are very good, except for the 66 feet in the East component of the Test No. 3. They range in magnitude from 0 to 12 feet. Radars R-112 and R-114 are both located at the southern end of the range and tracked the missile as it moved north.

TABLE IV

BIAS ESTIMATES FOR THE FPS-16 RADARS				
Radar	Flight Test	Radar Component Minus B.C. Component (Estimates in Feet)		
		North	East	Up
R-112	1	-22	24	33
	2	59	30	55
	3	-24	2	20
R-114	1	-2	8	4
	2	1	0	-12
	3	6	66	3
R-122	1	-11	18	-27
	2	2	13	3
	3	2	25	-34

TABLE V

BIAS ESTIMATES FOR COMPOSITE* RADAR SYSTEM			
Flight Test	Composite Component Minus B.C. Component (Estimates in Feet)		
	North	East	Up
1	-12	17	4
2	21	14	15
3	-8	34	-6

*Respective components of R-112, 114, and 122 averaged.

The bias estimates for R-122 are more consistent in size for the three tests, partly because of its location near mid-range. On the average it was closer to the trajectory sections analyzed than either R-112 or R-114. The estimates range in magnitude from 2 to 34 feet.

Table V presents the bias estimates for a composite trajectory obtained by averaging the respective components of all three radars. These estimates are very similar for those of R-122.

With allowances made for the skin track mode most of the radar bias estimates for the North and Up components were significant at the 5% level compared to an expected value of zero. Also, there were significant shifts in the magnitude of system bias in all three components over Section I for R-112 and R-114, and in the Up component over Section III for R-122.

Since their installation at WSMR the radar systems have been used extensively. A series of calibration and evaluation tests have been proposed; these tests have been planned for the near future.

III. <u>PRECISION ANALYSIS</u>. Precision is defined as a measure of variability of a random variable about its mean value. For this analysis it is synonymous with the statistical term, standard deviation. Three methods were used to obtain precision estimates for the four trajectory measuring systems. A brief description of each method follows:

1. Over-determination Point Estimate of Precision - Over-determination, by the method of least squares, of space points on a trajectory yields, as a side product, variance estimates for each space point. When the variance estimates for a sample of space points are pooled the result is a point estimate of precision. A major disadvantage of this method is that it is sensitive to system bias errors in the input data. However, for this analysis there are ways to isolate a biased system for investigation.

2. Multi-Instrument Estimate of Precision - This method, sometimes referred to as the Simon-Grubbs method, requires a simultaneous sample of trajectory space points from each of three or more instrumentation systems. This method yields a precision estimate for each system involved. It is insensitive to constant system bias errors in the input data, but shifting system bias errors will enlarge the estimates produced by this method.

3. Variate Difference Estimate of Precision - Successive differencing of a sample of trajectory space points determined by a single system proceeds until the systematic elements of the data become negligible and the random

TABLE VI

PRECISION ESTIMATES FOR BALLISTIC CAMERAS			
Flight Test	Component Standard Deviation (Estimates in Feet)		
	North	East	Up
1	2	6	10
2	4	6	9
3	6	4	6

TABLE VII

PRECISION ESTIMATES FOR ASKANIA CINETHEODOLITES			
Flight Test	Component Standard Deviation (Estimates in Feet)		
	North	East	Up
1	11	11	8
2	10	15	12
3	8	8	12

element becomes dominant. The precision estimate for the system is based on the residual random element. This method requires a data sample with points equally spaced in time and assumes the data can be approximated by a polynomial. All linear trends, which are first degree elements, are filtered out and thus the precision estimates by this method tend to be smaller than those obtained by the two methods mentioned above.

A. Ballistic Camera. Precision estimates for the ballistic camera system are shown in Table VI. The estimates were obtained by a multi-instrument method where variances were pooled over the three trajectory sections. The precision estimates are ten feet or less in magnitude and similar for the three tests. These estimates correspond to an average ballistic camera angular precision of 10" of arc. The new BC-4 ballistic cameras, of which four have been installed at WSMR, are designed for an angular precision of 1" of arc. This will lower the precision estimates by a factor of ten. Hopefully, the BC-4 system will become the standard for the range in the near future.

B. Askania Cinetheodolite. Table VII presents precision estimates for the Askania cinetheodolite system for the first three tests. These are also multi-instrument estimates with the variances pooled over the sections. In general, these estimates are about twice as large as those obtained for the ballistic camera system. They are similar for the three tests and correspond to a system angular precision of approximately 30" of arc. These precision estimates agree very well with those computed for the system over a number of years of operation.

C. DOVAP. Precision estimates for the DOVAP system are shown in Table VIII. These were obtained by the variate difference technique. Except for Test No. 2 the estimates are less than 0.6 feet. They are comparable to precision estimates expected from the BC-4 ballistic camera system. For all three tests the DOVAP has been the most precise system at WSMR. The multi-instrument method was also used to obtain estimates of precision for this system. However, some of the variances were negative in sign. Since these are less meaningful for the purpose intended and more difficult to interpret, they are not included in this paper.

D. FPS-16 Radars. Precision estimates for three of the FPS-16 radars are shown in Table IX. These are multi-instrument estimates. Variability in the radars is considerably larger than in any of the other systems. Generally, it increased over the three tests, especially for R-112 in the East and Up components. The estimates for R-114 are smaller than those for R-112. The

TABLE VIII

PRECISION ESTIMATES FOR DOVAP			
Flight Test	Component Standard Deviation (Estimates in Feet)		
	North	East	Up
1	0.2	0.4	0.3
2	2.0	0.6	1.0
3 (36.2 mc)	0.2	0.3	0.6
(36.9 mc)	0.2	0.3	0.6

TABLE IX

PRECISION ESTIMATES FOR THE FPS-16 RADARS				
Radar	Flight Test	Component Standard Deviation (Estimates in Feet)		
		North	East	Up
R-112	1	18	46	34
	2	25	68	92
	3	34	134	76
R-114	1	13	30	29
	2	16	63	28
	3	21	44	73
R-122	1	29	29	21
	2	21	18	20
	3	32	44	27

largest change is in the Up component for Test No. 3. For R-122 the precision estimates are more consistant than for the other two radars. Bias estimates for R-122 also had this characteristic.

Radar system cartesian component variability, shown in Table IX, must be viewed in light of the variability of the measured parameters, namely range and azimuth and elevation angles. The computed cartesian components depend, in various degrees, on these measured parameters. For instance, for both R-112 and R-114 the range measurement is essentially a measurement of the North component, whereas the azimuth and elevation angle measurements account mostly for the computed East and Up components respectively. This relationship arises from the fact that both R-112 and R-114 are located at the southern end of the range and the missile moved, in general, North. For R-122 the relationship is not so evident since it is located near mid-range. Table X presents precision estimates in terms of measured parameters for the three radar systems.

A comparison of Tables IX and X shows that the precision estimate of 154 feet in the East component for R-112 on Test No. 3 corresponds to the large precision estimate (0.55 mils) obtained for azimuth measurements made by this radar. Likewise, the estimate of 73 feet in the Up component for R-114 on Test No. 3 corresponds to the estimate obtained for elevation measurements for this radar. The precision estimates of Table X reveal more directly the performance of the radars.

Following the installation of the first FPS-16 radar in 1958 a series of evaluation tests were conducted. In a report[1] covering these tests the general conclusions were that the expected variability of the radars would be on the order of 10 yards in range and 0.2 mils in azimuth and elevation. The range precision estimates of Table X meet this expectation. One might suspect this large angular variability to be caused by glinting since the radars operated in the skin track mode. However, for the three tests being considered the maximum component in the precision estimates attributable to glinting is only about 0.09 mils and is therefore considered negligible. As stated in the section on bias errors a series of calibration and evaluation tests have been planned for the radars. Hopefully, the angular variability will be reduced.

[1] See Reference 6

TABLE X

PRECISION ESTIMATES FOR FPS-16 RADARS				
Radar	Flight Test	Standard Deviation for Measured Parameters		
		Range (yds)	Azimuth (mils)	Elevation (mils)
R-112	1	2.7	0.42	0.28
	2	4.2	0.37	0.52
	3	7.3	0.55	0.31
R-114	1	3.2	0.17	0.15
	2	4.8	0.26	0.21
	3	4.2	0.21	0.33
R-122	1	1.4	0.20	0.23
	2	2.6	0.23	0.15
	3	7.8	0.60	0.26

IV. SUMMARY AND CONCLUSIONS.

A. The DOVAP continues to be the most precise trajectory measuring system at WSMR. With improvements in techniques of start point determination, and with continued refinement of propagation velocity estimates, the DOVAP can also become one of the least biased systems.

B. The present ballistic camera system is the second most precise system at WSMR. The new BC-4 system is expected to be ten times more precise than the present system.

C. The angular precision of the Askania cinetheodolites is estimated at 36" of arc. A large bias still exists in the Up component. Better approximation of the refraction correction will probably improve elevation determination.

D. The radars did not perform as well as was expected. System bias needs to be reduced. Also, improvements are needed in system precision, especially in angular data.

REFERENCES

1. Simon, L. E., "On the Relation of Instrumentation to Quality Control," Instruments, Vol.19, Nov. 1946.

2. Grubbs, F. E., "On Estimating Precision of Measuring Instruments and Product Variability," J.A.S.A., Vol. 43, (1948), pp. 243-264.

3. Davis, R. C., "Techniques for Statistical Analysis of Cine-Theodolite Data," NAVORD Report 1299, China Lake, California, (March 22, 1951).

4. Davis, R. C., "Techniques for Statistical Analysis of Continuous Wave Doppler Data," NAVORD Report 1312, NOTS 383, April 1951.

5. Kendall, M. G., "The Advanced Theory of Statistics," Vol. II, Third Edition, C. Griffin Co., Ltd., London (1951).

6. Pearson, K. E., "Evaluation of the AN/FPS-16 (System No. 1) at White Sands Missile Range," WSMR Technical Memorandum #606, Feb. 1959.

7. "Final Data Report No. 13,569, Ballistic Camera Data for Nike Hercules Flt 85 HE MX 13538 In Support of Accuracy and Precision of Trajectory Measuring Systems Test No. 3(U)." Launched 29 January 1962, IRM-DRD, WSMR, NMex (9 August 1962). Classified Confidential.

8. "Final Data Report No. 14001, Askania Trajectory Data for Nike Hercules Flt 85 HE MX 13538 In Support of Accuracy and Precision of Trajectory Measuring Systems Test No. 3 (U)." Launched 29 January 1962, IRM-DRD, WSMR, NMex (21 August 1962). Classified Confidential.

9. "Final Data Report No. 14163, DOVAP Trajectory Data for Nike Hercules Flt 85 HE MX 13538 In Support of Accuracy and Precision of Trajectory Measuring Systems Test No. 3 (U)." Launched 29 January 1962, IRM-DRD, WSMR, NMex (5 September 1962). Classified Confidential.

10. "Final Data Report no. 14366, Radar Data for Nike Hercules Flt 85 HE MX 13538 In Support of Accuracy and Precision for Trajectory Measuring Systems Test No. 3 (U)." Launched 29 January 1962, IRM-DRD, WSMR, NMex (4 October 1962). Classified Confidential.

11. Weiss, J. E. and Kingsley, O. L., "Study of Accuracy and Precision of Trajectory Measuring Systems," Interim Report of Operation PRECISE Round No. 1, IRM-RIDD, WSMR, NMex (30 June 1961).

12. Kingsley, O. L. and White, J. E., "An Analysis of the Bias and Precision Estimates of Trajectory Measuring Systems," Interim Report on Operation PRECISE Round No. 2, IRM-RIDD, WSMR, NMex (30 April 1962).

ASPECTS TO CONTROL LIQUID PROPELLANT SLOSHING BASED UPON EXISTING THEORY

W. R. Eulitz
Marshall Space Flight Center
Huntsville, Alabama

SUMMARY. Liquid propellants usually carried by a launch vehicle in cylindrical containers, represent one part in a complex of several coupled spring systems (missile bending, bulking, etc.). By external forces, the liquid spring system can be excited to heavy liquid mass motions which can be detrimental to the performance of the launch vehicle. The vehicle control motion may be considered the most influential factor for exciting liquid oscillations.

During the last decade, many attempts have been made to ascribe the response of a liquid to exciting oscillations. Although principally a nonlinear problem, the theory has been confined to the first order terms only, due to mathematical difficulties. This linearized theory has been discerned in satisfactory agreement with many experimental results at least as far as the first liquid resonance is concerned which likewise is the most adverse condition for the vehicle control system.

The interpretation of the existing theory leads to dimensionless parameters which, incorporated in a nomograph, provide quick orientation on liquid behavior under varying oscillatory conditions. Such data help to define critical vehicle flight periods and to predetermine proper design parameters.

The survey of the parameters of oscillatory liquid motion suggests possible means of suppressing liquid sloshing. The pros and cons of several methods are discussed and furthermore, the prospects for proper measurements of liquid surface motion are indicated in the paper.

ASPECTS TO CONTROL LIQUID PROPELLANT SLOSHING BASED UPON EXISTING THEORY

W. R. Eulitz
Marshall Space Flight Center
Huntsville, Alabama

SUMMARY. Liquid propellants usually carried by a launch vehicle in cylindrical containers, represent one part in a complex of several coupled spring systems (missile bending, bulking, etc.). By external forces, the liquid spring system can be excited to heavy liquid mass motions which can be detrimental to the performance of the launch vehicle. The vehicle control motion may be considered the most influential factor for exciting liquid oscillations.

During the last decade, many attempts have been made to ascribe the response of a liquid to exciting oscillations. Although principally a nonlinear problem, the theory has been confined to the first order terms only, due to mathematical difficulties. This linearized theory has been discerned in satisfactory agreement with many experimental results at least as far as the first liquid resonance is concerned which likewise is the most adverse condition for the vehicle control system.

The interpretation of the existing theory leads to dimensionless parameters which, incorporated in a nomograph, provide quick orientation on liquid behavior under varying oscillatory conditions. Such data help to define critical vehicle flight periods and to predetermine proper design parameters.

The survey of the parameters of oscillatory liquid motion suggests possible means of suppressing liquid sloshing. The pros and cons of several methods are discussed and furthermore, the prospects for proper measurements of liquid surface motion are indicated in the paper.

DEFINITION OF SYMBOLS

Symbol	Definition
ζ	Liquid amplitude
ζ_s	Liquid amplitude under splash condition
X_o	Forced amplitude of transverse container motion
ω	Forced frequency in rad/sec
ω_1	First resonant frequency in rad/sec
f	Forced frequency in cps
f_1	First resonant frequency in cps
$\lambda = \frac{\omega^2}{\omega_1^2} = \frac{f^2}{f_1^2}$	
$R_f = \frac{\omega}{\omega_1} = \frac{f}{f_1} = \sqrt{\lambda}$	
a	Radius of container
$d = 2a$	Diameter of container
r, θ	Polar coordinates of points of the liquid surface
g	Longitudinal acceleration of container
$J_1(\zeta)$	Bessel functions (B. F.)
α_n	Zero's of first derivative of B. F. of first order and first kind
$n = \frac{g_n}{g_o}$	Acceleration ratio (g_o = acc. due to gravity)
$K = \tanh 2\alpha \frac{h}{d}$	

Symbol	Definition
h	Filling height of liquid in the container
F	Force exerted by the liquid toward the tank wall
W	Weight of liquid in the container

INTRODUCTION. In recent years, comprehensive studies, experimentally and theoretically, clarified the response of a liquid under forced vibrations to such an extent that the effect of slosh motion is predictable by theory under simplified conditions. Nevertheless, the application of the theory for practical purposes has always been a difficult undertaking. As the consequence of these difficulties, connected with some skepticism in the dependence on theoretical predictions in this particular field, extreme conservatism in vehicle design has sometimes been practiced as the alternative.

This uncertainty stimulated a study on the limitations of the existing theory and on the effectiveness of the parameters involved in order to discern the practical possibilities of suppressing liquid slosh motion and their applicabilities with respect to container configuration and flight condition. The results of this study are presented in this paper.

THE BASIC EXPERIMENTAL AND THEORETICAL FACTS CONCERNING LIQUID SLOSH MOTION. The problem of liquid oscillations is principally a nonlinear problem like all problems of hydrodynamics. Due to mathematical difficulties, the theory has been linearised in order to solve the problem. The linearization of the theory is permissible if the exciting acceleration of the system ($x_0\,\omega^2$) is very small compared to the longitudinal acceleration (g_n) of the liquid container, (see Ref. 1 through 4). This is one limitation of the existing theory.

With these assumptions (small acceleration of excitation and linearization of the theory), it has been possible to deduce the equation of the resonant frequencies of the liquid system in a flat bottomed circular cylindrical container and the equation of the liquid surface under forced lateral sinusoidal oscillations.

The general form of the equation of the liquid surface in motion is (Ref. 1 and 5)

$$\zeta = x_0 \frac{a\,\omega^2}{g} \left[\frac{r}{a} + \sum_{n=0}^{\infty} \frac{2 I_1 (\alpha_n \frac{r}{a})}{(\alpha_n^2 - 1) I_1 (\alpha_n) \left(\frac{\omega_n^2}{\omega^2} - 1\right)} \right] \cos\theta \cos \omega t$$

The surface equation ascribes the elevation (ζ) of any point of the liquid surface with polar coordinates (r, θ) at any instant of the cycle ($\cos \omega t$) above the zero level.

From this theory, as well as from the experiment, it follows rigorously that:

(1) The maximum amplitudes of liquid occur in the plane of motion at the tank wall during the first period of resonance (wave length about 2 tank diameters). At higher resonances the maximum liquid amplitudes develop at the interior of the liquid surface (wave length smaller than one tank diameter).

(2) The largest liquid amplitudes ever possible in an oscillating liquid system also occur during the first resonance period.

(3) The largest liquid amplitudes are reached when the longitudinal acceleration of the liquid ($\zeta \omega^2$) is equal to the counteracting acceleration due to gravity (g_0) or thrust of the vehicle (ng_0). Under this condition the liquid starts violently splashing (slosh or splash condition, see Ref. 1 and 3). This, simultaneously, is the upper limitation of the validity of the surface equation following from the linearized theory (Ref. 1).

(4) The lower limitation of the theory concerning maximum liquid amplitudes is beyond any practical consideration as discussed in detail in Ref. 1 (very large exciting amplitudes).

(5) The theory is strongly valid for flat bottomed cylindrical containers only.

Since the maximum liquid amplitudes ($\cos\omega t = 1$) at the container wall ($r = a = d/2$) in the plane of motion ($\cos\theta = 1$) during the first resonant period (ω_1) are of most practical interest because of the forces exerted toward the tank wall, and if the frequency ratio squared is expressed by the Greek letter $\lambda = \omega^2/\omega_1^2$, ω designating the forced frequency, the liquid surface equation simplifies (Ref. 4)

$$\zeta = x_0 K \alpha \lambda \left[1 + \frac{2\lambda}{(\alpha^2 - 1)(1 - \lambda)} \right]$$

or dimensionless

$$(1) \qquad \frac{\zeta}{x_0 K} = \alpha \lambda \left[1 + \frac{2\lambda}{(\alpha^2 - 1)(1 - \lambda)} \right]$$

The first resonant frequency ω_1 which is one factor of ($\lambda = \dfrac{\omega^2}{\omega_1^2}$) follows from the theory as

$$(2) \qquad \omega_1^2 = \frac{2\alpha\, n g_0}{d} \tanh 2\alpha \frac{h}{d} = 2\alpha g_0 \cdot \frac{nK}{d}$$

where K identifies the hyperbolic tangent term which depends on the filling height - diameter ratio h/d. The term K approaches unity if $h \geq d$.

The liquid starts splashing, if the acceleration of liquid along the tank wall ($\zeta_s \omega^2$) is equal to the acceleration due to gravity (g_0) or thrust ($g_n = n \cdot g_0$), as pointed out earlier, or (Ref. 4):

$$\zeta_s = \frac{n g_o}{\omega^2} = \frac{n g_o}{\omega_1^2 \lambda} = \frac{1}{2\alpha\lambda} \cdot \frac{d}{K}$$

and

$$(3) \qquad \frac{\zeta_s}{x_o K} = \frac{1}{2\alpha\lambda} \cdot \frac{d}{x_o K^2} .$$

The general equation for the net forces on the tank wall which is the integration of the pressure distribution, (Ref. 5 and 6), also simplifies if considered under the most critical conditions (splash condition) and the assumptions above. Then, forces (F) can be expressed in terms of the total propellant weight (W) according to the following equation (see Appendix):

$$(4) \qquad \frac{F}{W} = \frac{1}{2\alpha^2\lambda} \cdot \frac{d}{h} + \frac{x_o K^2}{d} \lambda \left(\frac{2\alpha}{K} - \frac{d}{h} \right).$$

Equations (1) through (4) represent the fundamental formulation of liquid propellant response to sinusoidal tank oscillations under the following conditions of practical importance: Maximum liquid amplitudes ($\cos \omega t = 1$) in the plane of motion ($\cos \psi = 1$) at the tank wall ($r = a$) during the first resonant period (ω_1) under forced vibrations ($x_o \omega^2$) in a cylindrical flat bottomed container (d) under varying conditions (n, K) and acceleration equilibrium ($\zeta\omega^2 = n g_o$). These equations cover all container sizes and all flight conditions.

Equation (1) indicates that any frequency ratio λ is coupled with a special parameter $\zeta / x_o K$ which may be designated the "liquid amplitude coefficient". Knowing this coefficient, the actual liquid amplitude ζ can be concluded by multiplying $\zeta / x_o K$ with the exciting amplitude x_o and the term K which is a function of the filling height.

Equation (3) shows that each parameter pair, $\zeta/x_0\kappa$ and λ is the splash parameter pair for a "design parameter" $d/x_0\ \kappa^2$ which satisfies this equation. This consideration leads to a nomograph (Figure 1) which has been explained in detail in an earlier report (Ref. 4). The corresponding parameters, $\zeta/x_0\kappa$ and λ, according to equation (1), are plotted versus the corresponding design parameters, $d/x_0\ \kappa^2$, following from equation (3).

For a container of $d \approx 10$ m (= 400 in) filled to a height larger than the diameter d (κ = 1) and oscillated with an exciting amplitude $x_0 \approx 10$ cm (= 4 in), the design parameter would be $d/x_0\ \kappa^2 = 100$. The nomograph (Figure 1) indicates for $d/x_0\ \kappa^2 = 100$ an amplitude coefficient of $\zeta/x_0 = 29$. At this condition the liquid starts splashing if the frequency ratio $R_f = f/f_1$ (square root of λ) is about 0.975. The same situation would exist for a tank with $d \approx 50$ cm (= 20 in) operated with an exciting amplitude $x_0 \approx 0.5$ cm (= 0.2 in). In the first case, the actual liquid amplitude would be $\zeta = 29 \times 4 \approx 2.95$ m (= 116 in) in the second case $\zeta = 29 \times 0.2 \approx 14.7$ cm (= 5.8 in). If the frequency ratio in both cases is smaller than 0.975, then the liquid surface swings smoothly; if the frequency ratio is larger than 0.975, the liquid is violently splashing.

In Figure 2 the κ values are tabulated for different filling height ratios h/d and the nomograph gives the first resonant frequencies f_1 for any flight condition (right side double scale) in dependence of the parameter $d/n\ \kappa$ according to equation (2). A straight line from the point indicating the exciting frequency f (left scale) to the first resonant frequency f_1 of the system (right scale) provides the frequency ratio $R_f = f/f_1$ (intersection with central scale). With this value the amplitude coefficient $\zeta/x_0\kappa$ for this particular condition can be found from the nomograph of Figure 1.

In a similar way, the largest forces possible on the tank wall of an oscillating liquid system (splash condition) are plotted in percentage of the momentary propellant weight (F/W) in Figure 3 versus the design parameter $d/x_0\ \kappa^2$ for different filling ratios h/d according to equation (4).

DESIGN CRITERIA FOR LIQUID SLOSH CONTROL FOLLOWING FROM THEORY. With the nomographs discussed above, the response of liquid propellant to any vehicle flight condition can be estimated with fair approximation. From the fundamental equations (1, 2, and 3), as well as from the nomograhp Figure 1, it follows immediately that the liquid amplitude coefficient $\zeta/x_0 \kappa$ increases with increasing design parameter $d/x_0\ \kappa^2$. The design parameter can increase either with the tank diameter d, or with decreasing exciting amplitude x_0 or decreasing filling ratio h/d (κ). However, since the amplitude coefficient curve is almost a straight line at least for the higher design parameters (see nomograph Figure 1), the effect of x_0 and κ on the actual liquid splash amplitudes ζs (highest liquid amplitudes possible in the system) is small because the amplitude coefficient is approximately proportional to the product $x_0 \kappa$. Thus, the most efficient factor on the design parameter is the tank diameter d. The larger d the larger are the liquid amplitudes to be expected in the system, and consequently, the larger the forces on the tank wall.

The nomograph Figure 1 also shows that for design parameters $d/x_0\ \kappa^2 > 70$ the frequency ratios $R_f = f/f_1$ are very close to unity ($R_f \geq 0.965$). The differences in the critical frequency ratios (splash condition) are small among large container design parameters. This means, a slight increase of x_0 which reduces the original $d/x_0\ \kappa^2$, would suddenly create the detrimental splash amplitudes because the critical frequency ratio under the varied condition is smaller than the applied frequency ratio.

This suggests that frequency larger than 0.9 should be avoided in any case. A frequency ratio of $R_f = f/f_1 = 0.85$ may be considered "safe" for all practical cases in order to provide a smooth oscillation of the liquid level. As the nomograph Figure 1 indicates, such a frequency ratio corresponds to a design parameter of $d/x_0\ \kappa^2 = 11.3$. The propellants in a container of $d \approx 10$ m (= 400 in) would start splashing (highest liquid amplitude possible) at this frequency ratio if the exciting amplitude would be $x_0 \approx 90$ cm (≈ 35.4 in); ($d/x_0\ \kappa^2 = 400/35.4 \approx 10{:}0.9 = 11.3$; κ assumed unity). However, such an extreme exciting amplitude is very unlikely in practice. For a 2.5 m (≈ 100 in) container, the exciting amplitude x_0 would be about 22.8 cm (= 9 in) to achieve the splash amplitude under the same extreme conditions.

Fortunately, during vehicle flight, the frequency ratio decreases automatically if the exciting frequency is maintained constant. It follows from equation (2), and the nomograph Figure 2 shows, that the first resonant frequency ω_1 (f_1) increases with increasing vehicle acceleration (n). Thus, the frequency ratio λ decreases and the liquid amplitudes decrease too (Equation 1 and nomograph Fig 1). The decreasing height of the liquid ($\kappa < 1$) during drainage, however, increases λ again so that splashing toward the end of the powered flight is very likely.

On the other hand, the natural frequencies of the liquid system (first resonance) are small for large diameter containers; they are inversely proportional to the spqare root of the diameter d (Equation 2 and nomograph Fig 2). This effect is aggravated if the vehicle acceleration ratio n is small (upper stages). It is easy to understand that small resonant frequencies are easier to approach by any vehicle motion than are larger resonant frequencies. In other words, liquids in large diameter containers are more sensitive to any movement than those in small diameter containers, or, slosh control by vehicle control frequencies is more difficult to maintain if the tank diameter is large and if the vehicle acceleration is low (near zero g).

At higher exciting frequencies the situation changes. Figure 4 shows the pattern of the curve of liquid amplitude coefficients versus higher frequency ratios λ , the peaks indicating the resonances. The other curve intersecting the amplitude curve at the beginning represents the splash condition for a particular parameter $d/x_o \kappa^2$. This graph illustrates that, at higher resonances, splashing of liquid starts before the theoretical amplitudes of liquid are obtained. This means, at higher resonances, there is always splashing and the theory according to Equation (1) is ineffective. No theory exists yet which ascribes the liquid surface under these conditions. The only fact we know is that liquid amplitudes at this stage are limited by the equilibrium of the accelerations acting on the liquid (splash condition).

Another important fact concerning the container design follows from Figure 3. It shows that the force - ratio increases considerably with decreasing h/d - ratio for a particular design parameter $d/x_o \kappa^2$. This effect, of course, is fairly compensated by the decreasing propellant weight during the drainage process. However, if an equal propellant volume is considered in two different tank configurations, first, in a long but small diameter tank, and second, in a short but large diameter tank, it can be concluded from Figure 3 that the forces the tank wall in the first case are much smaller than in the second case. This again suggests the design of long but small diameter containers rather than short but large diameter containers.

PRACTICAL METHODS TO CONTROL LIQUID SLOSH MOTION. The interpretation of the theory of liquid slosh motion during the first resonance period in the previous section already illustrated that the basic container design is of importance for handling this problem. Large but small diameter containers are preferable for two reasons; first, the very low resonant frequency of liquid in large diameter containers, and second, the larger forces toward the tank wall to be expected in large diameter containers. These requirements following from liquid slosh characteristics, however, contradict (in most of the practical cases) design requirements which suggest containers to be built as short as possible in order to avoid bending, or for other structural reasons. To serve both requirements, longitudinally compartmenting the large diameter containers might be considered a fair approach to an optimum. Attempts in this direction have been made by the design of so-called "scallop" or "multicell" tanks. Even though the first resonant frequency of the compartments will not correctly be in agreement with Equation (2) because Equation (2) applies only to cylindrical containers (Ref. 1), it will be higher than in the single tank and thus, be more advantageous to control liquid slosh motion as discussed in the previous section.

The cluster principle as experienced in SATURN I may be considered an incidental modification of the compartment concept. Here, in addition, the effect of long but small diameter tanks is advanced by a combination of liquid systems of different resonances. If the center container with its lower resonant frequency is excited to the extreme splash amplitudes, the liquid surfaces of the outer containers are still smoothly swinging because the frequency ratio λ in the center container is high (low resonant frequency) while the frequency ratio in the outboard containers is relatively small (higher resonant frequency). This combination certainly stablizes one portion of the available liquid propellants while the other portion is unstable and the net forces on the tank wall are reduced accordingly.

Equation (2) shows that the resonant frequencies are essentially dependent on the n/d ratio, and at "shallow water" conditions also on the h/d ratio (K). This means that the first resonance frequency is very low if the vehicle acceleration is small ($n \rightarrow o$), which concerns upper stages after separation. Since such stages usually are designed short (large diameters) for structural reasons, the ratio n/d and thus, the natural frequency of the liquid system is extremely small. Any vehicle motion will immediately create the critical splash condition: the liquid portions thrown up under these conditions, but dropping back to the liquid surface very slowly. The consequence would be an enhanced heat tranafer from the low temperature liquid to the higher temperature ullage gas which could jeopardize the ullage pressure.

This consideration suggests the design of honeycomb containers for upper stages, the cells extending in longitudinal direction with diameters small enough to provide a larger n/d ratio (higher resonance) for each cell in spite of the low vehicle acceleration (n). Such a honeycomb container could be made of very light-weight (small wall thickness) because the pressures toward the cell walls cancel each other except on the outside walls.

It should be noted that from another point of view the honeycomb concept is also advantageous for solving the problems of liquid behavior under low gravity conditions. It is known that, at low g, the influence of the surface tension becomes more and more effective on the shape of the liquid surface. J. T. Neu and R. J. Good, in an interesting study on this particular subject (Ref. 7), also arrive at the conclusion that a honeycomb container, according to their suggestion of conically shaped hexagonal cells, would be the proper solution for controlling liquid propellants under low gravity.

The characteristics discussed so far provide design criteria for preventing violent liquid slosh motion on a natural basis. They are deduced form the existing theory which is in good agreement with the experiment for the most critical conditions formulated earlier in this paper. Although there are many other effects, especially during vehicle flight, which can change the liquid motion considerably (interference by engine vibrations, bending, tank breathing, etc.), the conditions discussed above are always actual.

In the past the procedure almost everytime was to design a container which satisfies all structural and weight requirements, and to consider liquid slosh motion as a secondary problem. This led to the installation of so-called anti-slosh or slosh suppressing devices, which sometimes are based on rather eccentric ideas. Only few of those ideas are somehow related to theoretical facts. One of them is the floating can-type anti-slosh device (Ref. 8) which has been successfully flight tested some years ago (Fig 5). The function of this device is based upon the pressure equation which shows that only the upper portion of the liquid down to a depth of about one-quarter diameter of the container is in motion, which determines the length of the cans, and based on the fact that the spring constant of the liquid system in motion is increased. The friction between the floats and on the tank wall cannot be considered responsible for the damping characteristics as sometimes erroneously anticipated. This anticipation is disproved by the fact that can devices of too light weight where the friction is supposed to be the same, are not feasible. The floating ball which covers the total liquid surface can be considered an extended can device. The floating lid which also covers the total liquid surface actually provides for all liquid levels a "full tank" condition. The liquid is encapsuled drainage and thus, acts like a solid body. These devices, of course, are feasible only for containers which are long compared with the diameter.

For conceivable reasons, all floating devices have many opponents particularly among designers. Emphasis is placed on so-called fixed devices. Some of them are shown in Figure 6. In most cases, these devices are ring-like, mounted at the tank wall. Many tests proved that simple flat rings have the same effect as the other more complicated devices shown in Figure 6. The function of those devices actually has no correlation to the existing theory because the damping characteristic of such baffles is evidently a nonlinear effect, a problem which is not solved yet. As a stop-gap, a linear damping factor has been introduced into the liquid surface equation in a similar way as customarily implemented in equations for mechanical vibrations. The damping factor then, is determined by experiment. It has been found that the ratio between the ring width (w) and the radius of the tank (r) provides a damping factor within a satisfactory margin. The ratio of $w/r = 0.15$ may be considered a practically useful average. This means that the ring width in large diameter containers is considerable, not to mention the necessary structural strength of such rings because they have to consume all the forces exerted by the liquid in motion.

The solid cross device shown in Figure 6 is a compartmentation of the liquid bulk which changes the natural frequency of the liquid within each compartment, as discussed earlier. If the walls of the cross are perforated, the natural frequency of the liquid will not change considerably, but the damping will be satisfactory (Ref. 9). All these fixed devices except compartmentation are comparable with "break-waters" in the ocean. They break the liquid waves which evolve within the tank according to the theory. From this point of view it follows again that no correlation between the linearised theory and the behavior of the liquid due to baffling can exist. The baffles actually disturb what the theory describes.

In recent years, another effect has been applied for controlling liquid slosh motion. If a tube is submerged longitudinally along the tank wall in the plane of motion, the phase between the oscillating liquid surfaces, inside and outside the tube, changes. If the submerging depth of the tube opening is about one tank diameter, the phase shift is approximately 180 degrees. This means, at this stage the liquid level inside the tube oscillates inversely to the level in the container. The phase shift takes place while the opening of the tube passes the space between a depth of one-quarter and three-quarters of the tank diameter below the zero liquid level. At a constant exciting frequency the liquid surface in the tank oscillates with amplitudes according to equation (1), while the liquid level inside the tube is going up and down as in a U-tube but with different phase. The sequence is illustrated in Figure 7. Here, four semi-annular tubes are attached to the container wall, one pair in the plane of motion; the other pair normal to the direction of motion. The longitudinal cross section indicates the differently oscillating levels in the tank and in the tubes. The reason for the phase shift has not been clarified yet, theoretically. Qualitatively considered, this configuration resembles the anti-rolling tank principle which has been applied to large ships by Frahm, in 1902 (Ref. 10). The tubes shown in Figure 7 may be considered the "absorber system" which is attached to the "ship" (vehicle container). In contrast to Frahm's concept, the "sea" in our case is inside the "ship" and excited by the "ship" motion. This principle has been successfully model-tested. It is quite different from the principles applied so far, and should not be confused with compartmenting the tank or with baffling. Practically no forces due to liquid motion are acting on the tube walls The pressures inside and outside the tubes cancel each other. Thus, the tubes can be made of very thin, lightweight material. By this method, the liquid is divided in at least two interfering oscillating systems which do not disturb the coherence of the liquid as baffles do (violent splashing) and which can be applied to the most critical areas of the container.

The principle illustrated by Fig 7 also can be combined with a floating bell device as shown in Fig 8. The double-walled cylindrical portion of the bell is compartmented so that each compartment acts like a tube as illustrated in Figure 7. Due to the phase shift discussed earlier, the liquid flow in the open area within the bell and in the double-wall section of the bell, compensate each other as indicated by the arrows in the sketch Figure 8.

LIQUID SURFACE MEASURING METHODS DURING SLOSH MOTION. Another problem closely connected with damping liquid slosh motion, is the measurement of liquid surface deformation particularly during vehicle flight. The methods developed and utilized so far, however, give only poor, and even misleading information. The most common methods are differential pressure and capacitance measurements. Since these measurements are only point-measurements of the liquid surface, a comprehensive knowledge on the real shape of the surface at any instant, is not possible.

This experience stimulated a study, contracted recently by MSFC to SPACO, INC., Huntsville, on the feasibility of the stereo-television principle for monitoring the shape of the total liquid surface during slosh motion. This study came up with the interesting result that it will be possible to scan the liquid surface with a TV-camera and to monitor and convert the data into a contour mapping of the surface at any instant. Such contour mapping of the liquid surface characterized by contour lines of equal liquid level would provide respectable information on liquid behavior under varying conditions which would help not only to clarify the frequency spectrum acting on the liquid, but also could be utilized as an improved method to solve the problem of exact propellant loading.

APPENDIX

The general equation of the forces toward the tank wall in (Ref. 5 and 6):

$$F = \omega^2 x_0 e^{i\omega t} m \left[\frac{r}{a} + 2 \sum_{n=1}^{\infty} \frac{\tanh 2\alpha_n \frac{h}{d}}{2\alpha_n \frac{h}{d} (\alpha_n^2 - 1)\left(\frac{\omega_n^2}{\omega^2} - 1\right)} \right]$$

For the amplitude of force ($e^{i\omega t} = 1$); $W = mg$; at the tank wall ($r = a$); $\kappa = \tanh 2\alpha_n \frac{h}{d}$, and the first resonance of liquid ($\alpha_n = \alpha$; $\omega_n = \omega_1$).

$$F = \frac{\omega^2 x_0}{g} W \left[1 + \frac{\kappa}{\alpha \frac{h}{d}} \cdot \frac{1}{(\alpha^2 - 1)\left(\frac{\omega_1^2}{\omega^2} - 1\right)} \right]$$

Since $\omega^2 = \lambda \omega_1^2$ and $\omega_1^2 = \frac{2\alpha n g_0 \kappa}{d}$

$$\frac{F}{W} = \frac{2\alpha\lambda\kappa x_0}{d} \left[1 + \frac{\kappa}{\alpha \frac{h}{d}} \cdot \frac{\lambda}{(\alpha^2 - 1)(1 - \lambda)} \right] \qquad (1)$$

On the other hand

$$\zeta = \alpha\kappa\lambda x_0 \left[1 + \frac{2\lambda}{(\alpha^2 - 1)(1 - \lambda)} \right] \qquad (2)$$

and under the splash condition

$$\zeta = \frac{n \cdot g_0}{\omega^2} = \frac{n g_0}{\lambda \omega_1^2} = \frac{1}{2\alpha\lambda} \cdot \frac{d}{\kappa} \qquad (3)$$

Equalizing (2) and (3) and transforming

$$\frac{\lambda}{(\alpha^2 - 1)(1 - \lambda)} = \frac{1}{4\alpha^2 \lambda^2} \cdot \frac{d}{x_0 \kappa^2} - \frac{1}{2} \qquad (4)$$

Substituting in (1) and transforming

$$\frac{F}{W} = \frac{1}{2\alpha^2 \lambda} \cdot \frac{d}{h} + \frac{x_0 \kappa^2}{d} \lambda \left(\frac{2\alpha}{\kappa} - \frac{d}{h} \right) \qquad (5)$$

Equation (5) is the basis for the nomograph Fig 3; λ is a function of $d/x_0\kappa^2$ according to equation (4).

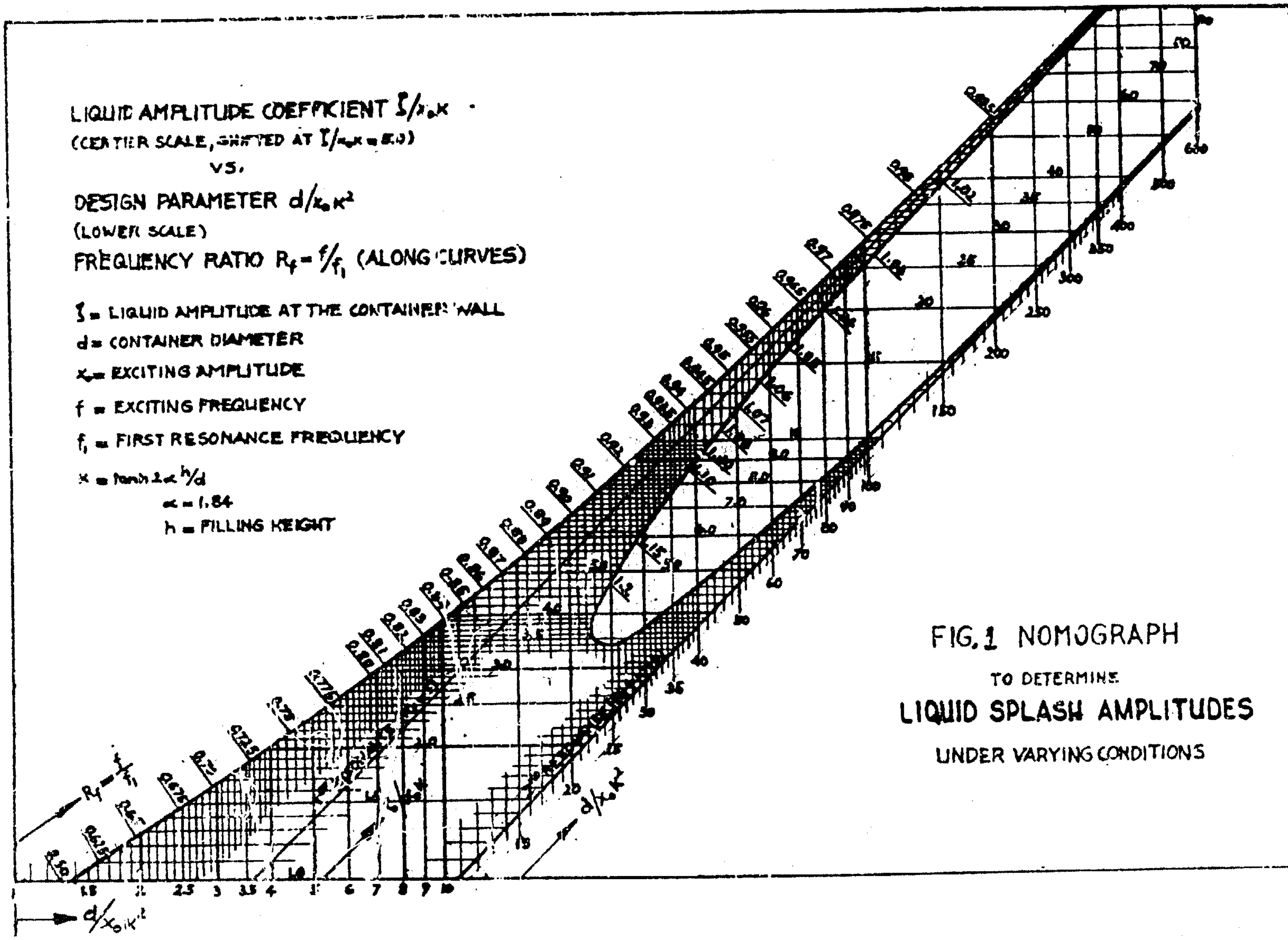


FIG. 1 NOMOGRAPH TO DETERMINE LIQUID SPLASH AMPLITUDES UNDER VARYING CONDITIONS

CORRELATION BETWEEN
h/d ; K ; K^2
(TABULATED)

h/d	K	K^2
1.0	0.999	0.998
0.95	0.998	0.996
0.90	0.997	0.994
0.85	0.996	0.992
0.80	0.995	0.990
0.75	0.992	0.984
0.70	0.988	0.976
0.65	0.983	0.967
0.60	0.976	0.952
0.55	0.965	0.931
0.50	0.951	0.905
0.45	0.929	0.863
0.40	0.900	0.810
0.35	0.859	0.738
0.30	0.802	0.643
0.25	0.726	0.527
0.20	0.627	0.393
0.15	0.502	0.252
0.10	0.351	0.123

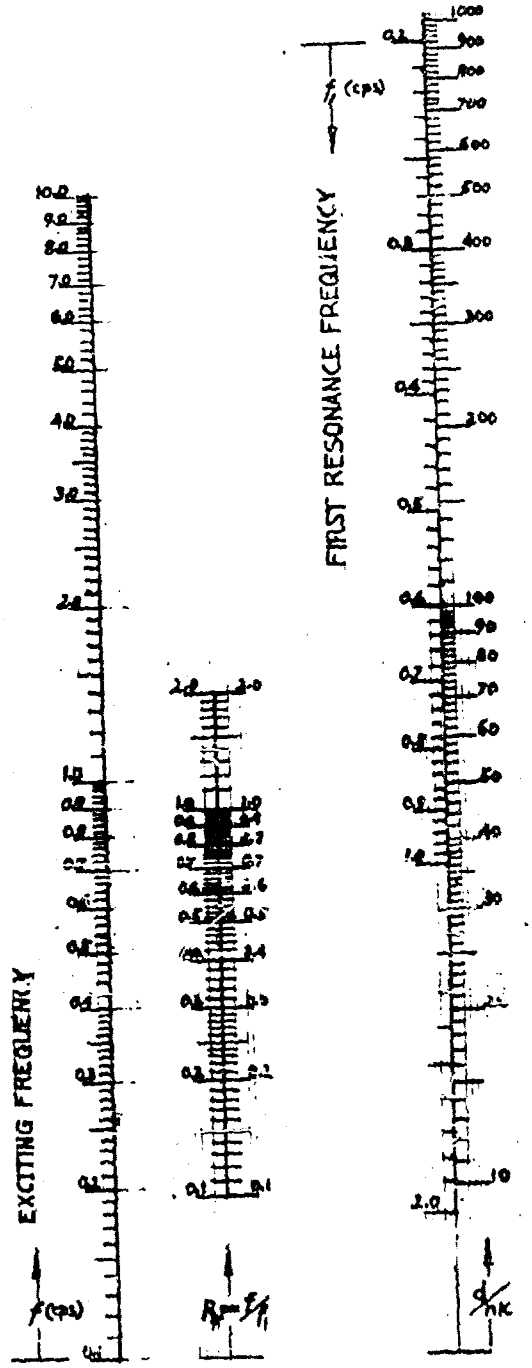


FIG.2 NOMOGRAPH

TO DETERMINE f_1 AND R_f BY d/nK AND TABULATED h/d, K, K^2-VALUES.

f_2 (SECOND RESONANCE FREQUENCY) = 1.703 · f_1
η (ACCELERATION RATIO) = g_n/g_0
g_n = LONGITUDINAL ACCELERATION
g_0 = ACCELERATION DUE TO GRAVITY

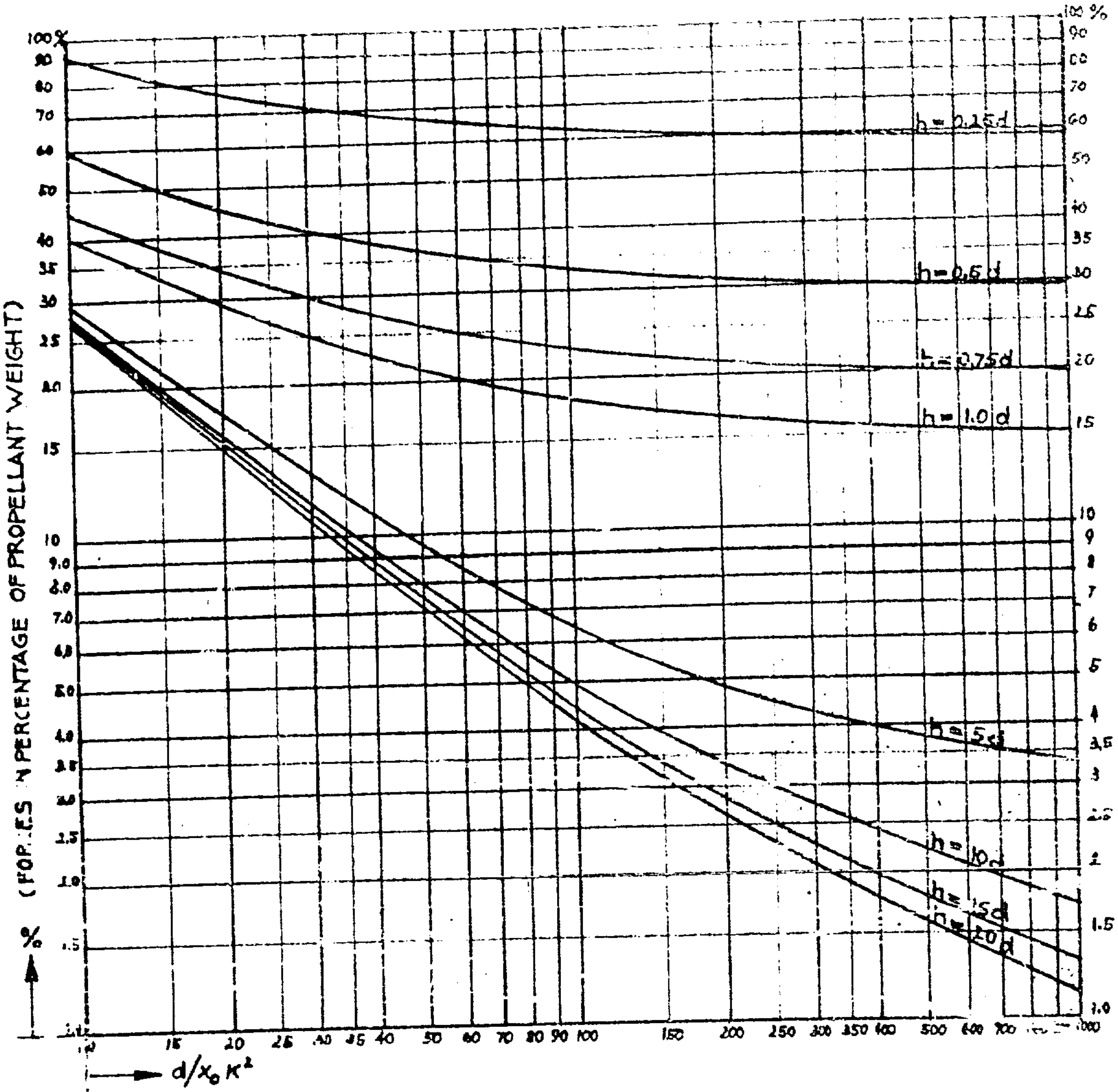


FIG 3 NOMOGRAPH TO DETERMINE
FORCES DUE TO SLOSHING IN PERCENTAGE OF PROPELLANT WEIGHT
VS.
DESIGN PARAMETER $d/x_0 K^2$
FOR DIFFERENT h/d - VALUES

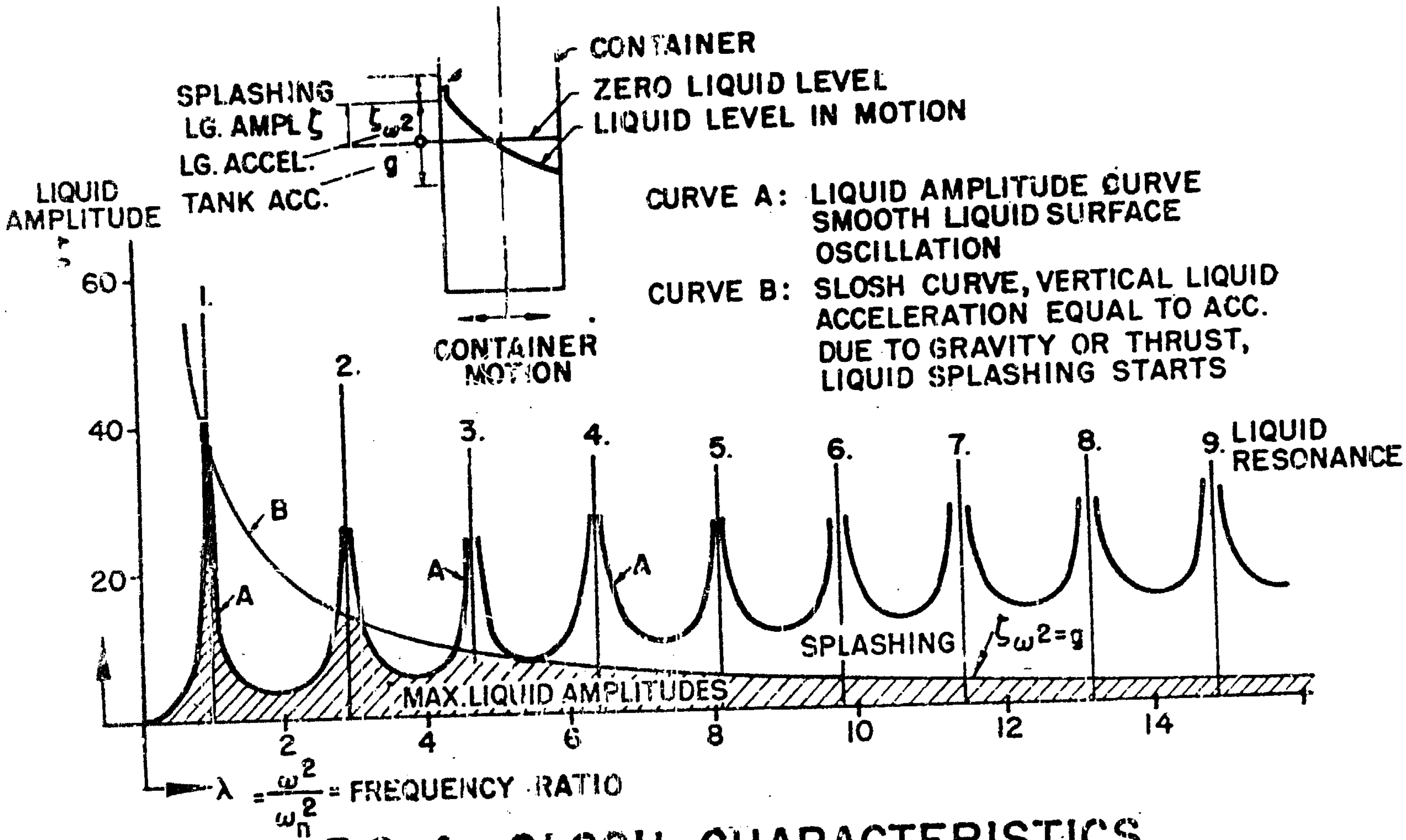


FIG 4 SLOSH CHARACTERISTICS OF HIGHER MODES OF RESONANCE

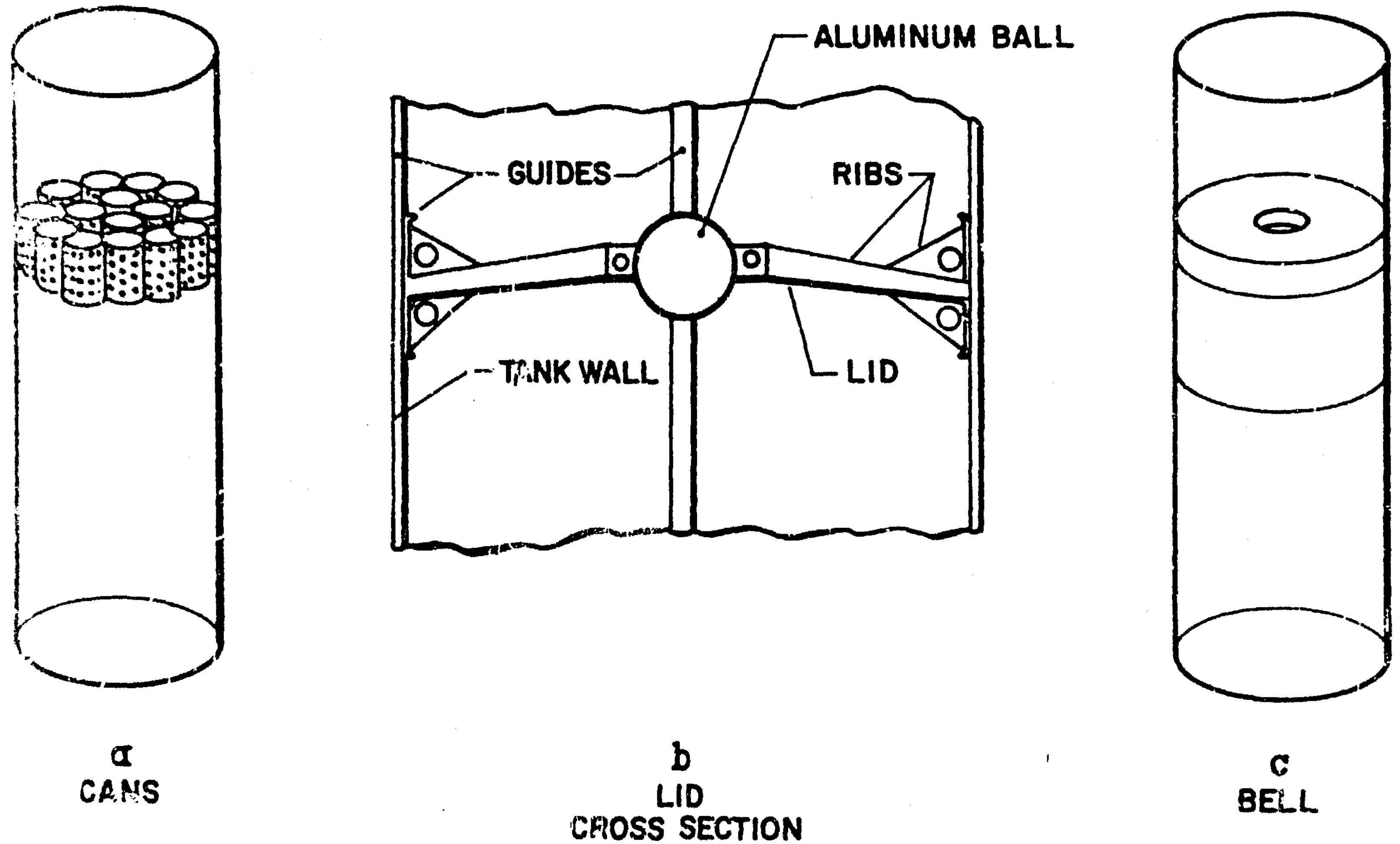


FIG. 5 FLOATING ANTI-SLOSH DEVICES

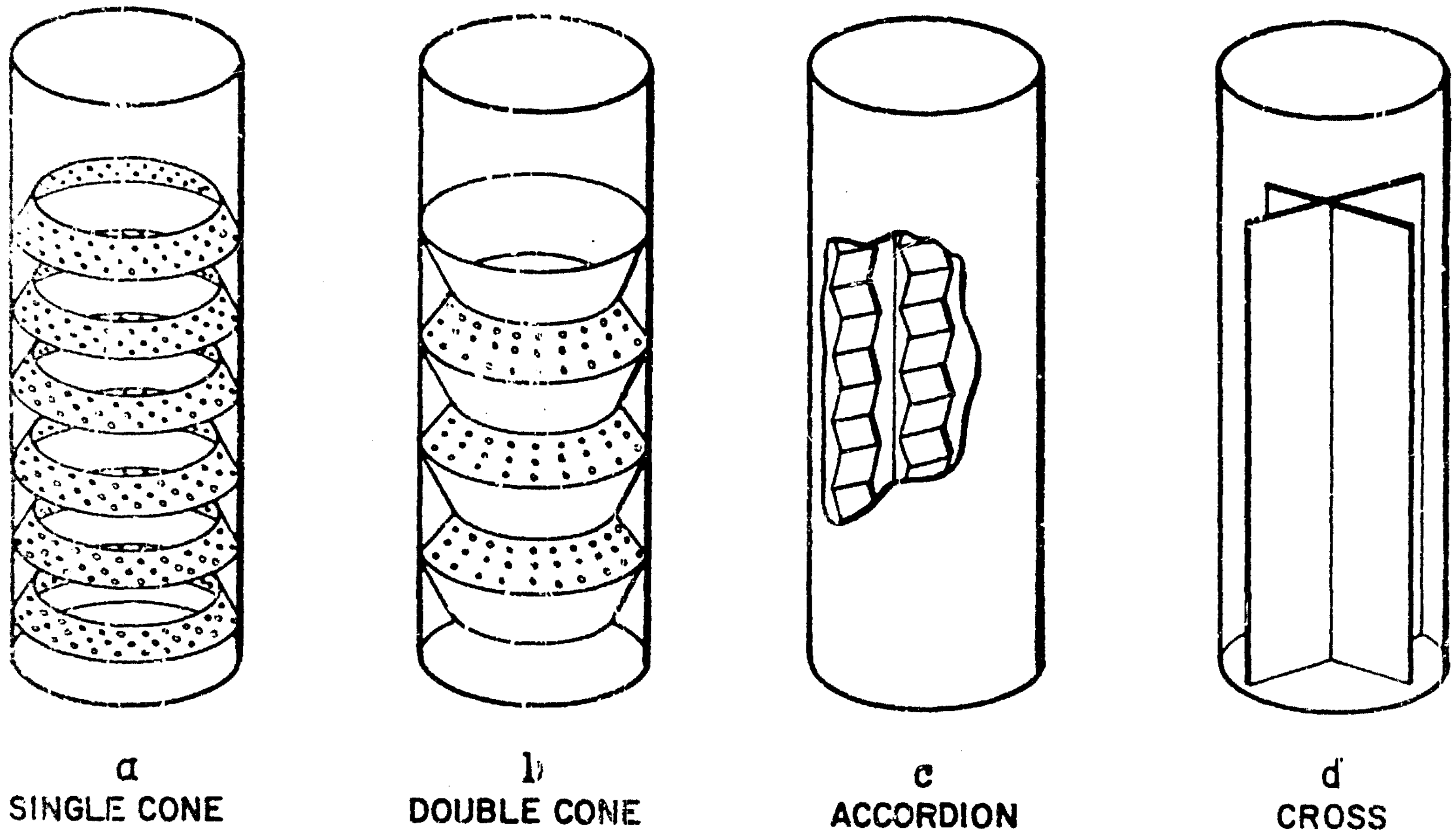


FIG. 6 FIXED ANTI-SLOSH DEVICES

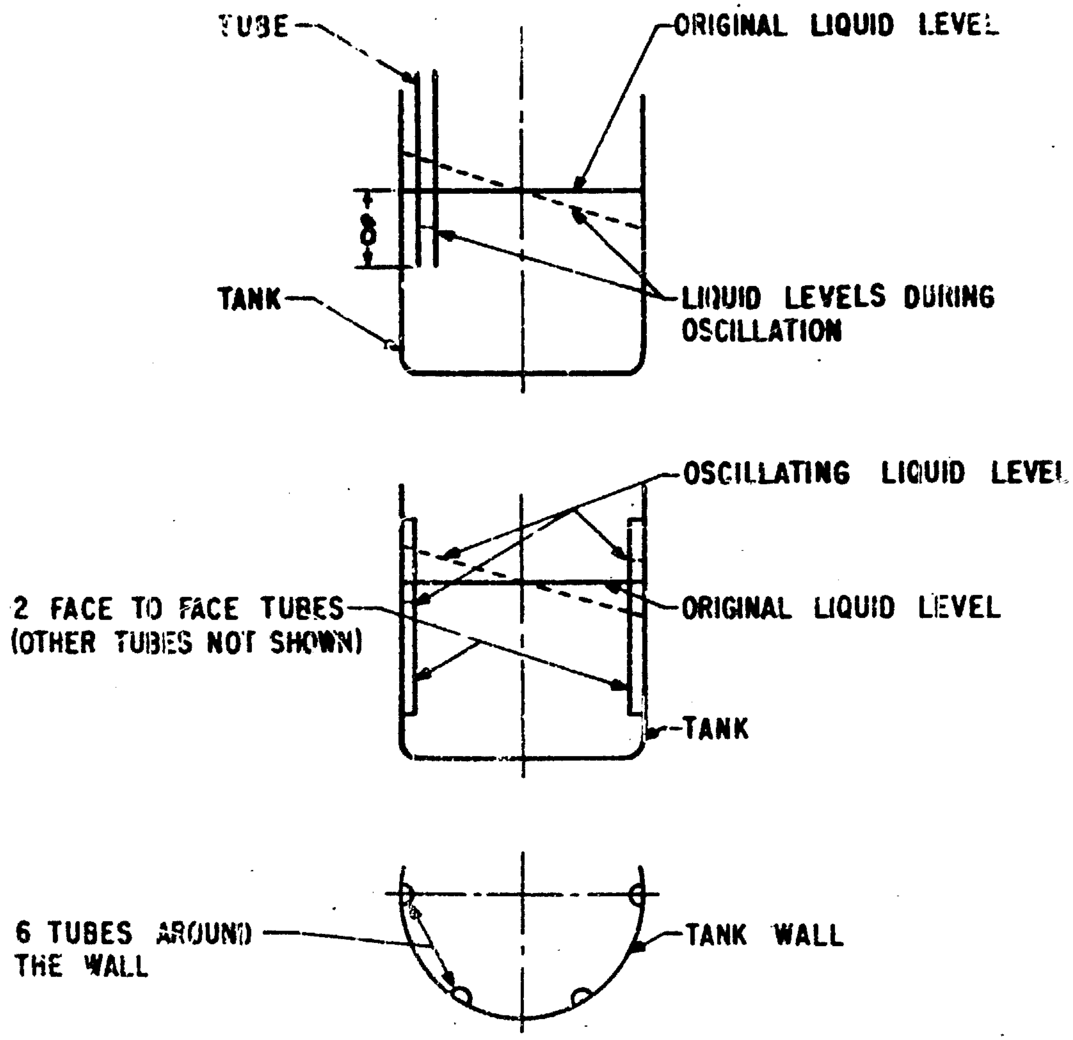

TUBE
ORIGINAL LIQUID LEVEL
δ
TANK
LIQUID LEVELS DURING
OSCILLATION
OSCILLATING LIQUID LEVEL
2 FACE TO FACE TUBES
(OTHER TUBES NOT SHOWN)
ORIGINAL LIQUID LEVEL
TANK
6 TUBES AROUND
THE WALL
TANK WALL

FIG. 7

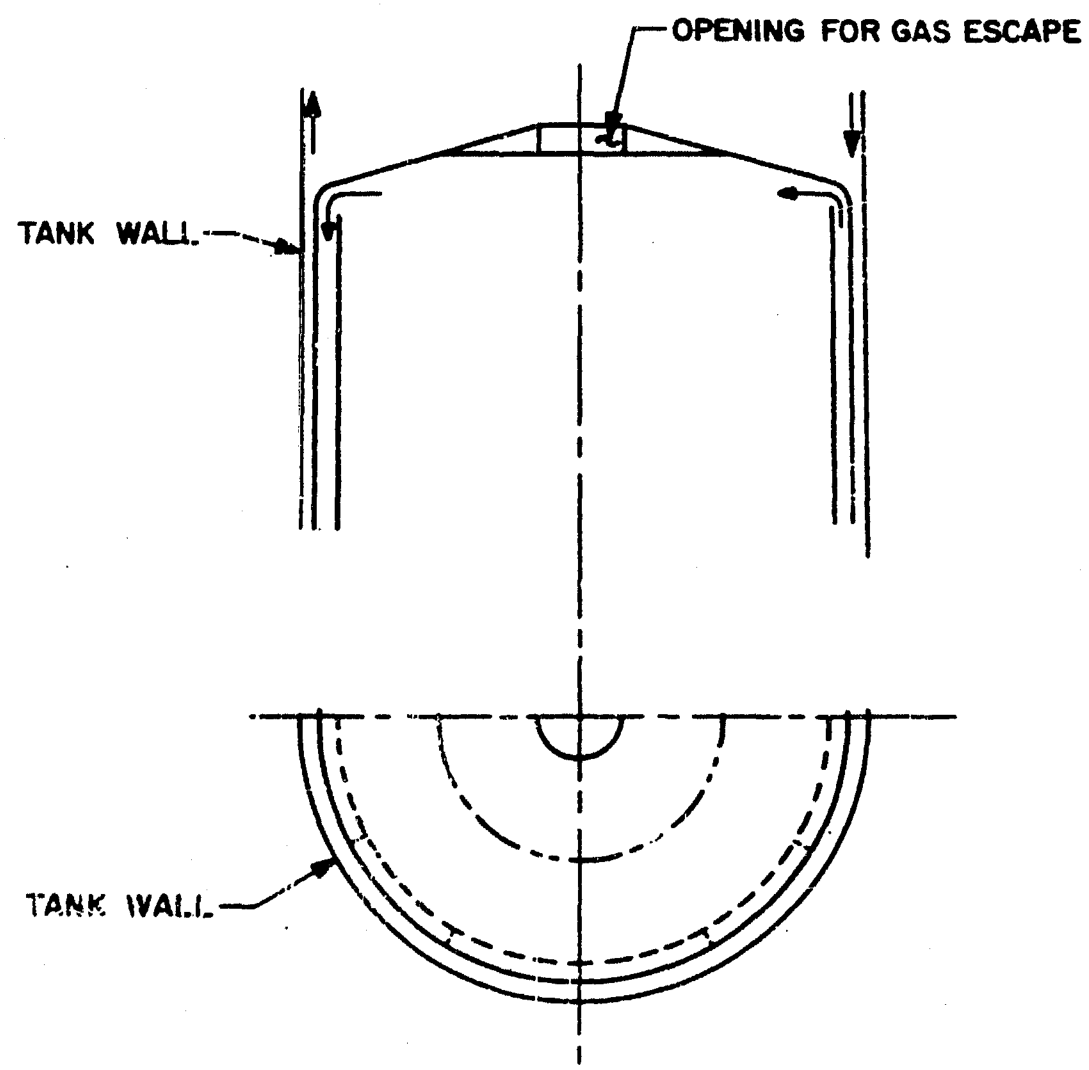


FIG. 8

REFERENCES

1. W. R. Eulitz and R. F. Glaser, Comparative Experimental and Theoretical Considerations on the Mechanism of Fluid Oscillations in Cylindrical Containers, NASA-MTP-P&VE-P-61-11; 1961.

2. W. R. Eulitz, Analysis and Control of Liquid Propellant Sloshing During Missile Flight, NASA-MTP-P&VE-P-61-22; 1961.

3. W. R. Eulitz; Discrepancies with Δ-p Sensors for Monitoring Liquid Propellant Sloshing During Rocket Flight, NASA-MTP-P&VE-P-62-9, 1962.

4. W. R. Eulitz, Practical Consequences of Liquid Propellant Slosh Characteristics Derived by Nomographic Methods, NASA-MTP-P&VE-P-63-7, 1963.

5. H. F. Bauer, Propellant Sloshing ABMA-DA-TR-18-58, 1958.

6. H. F. Bauer, Theory of the Fluid Oscillations in a Circular Cylindrical Ring Tank Partially Filled with Liquid, NASA-TN-D-557, 1960.

7. J. T. Neu and R. J. Good, Equilibrium Behavior of Fluids in Containers at Zero Gravity, AIAA Journal, Vol. 1, No. 4, April 1963.

8. W. R. Eulitz, A Can-Type Anti-Slosh Device Derived from Basic Slosh Studies, ABMA, Rpt. No. DSD-TR-4-58, 1958.

9. H. N. Abramson, W. H. Chu, and L. R. Garza, Liquid Sloshing in 45° Sector Compartmented Cylindrical Tanks. Technical Report No. 3, Contract NAS8-1555, Southwest Res. Inst. Nov. 1962.

10. J. P. Den Hartog, Mechanical Vibration, McGraw Hill, 3rd Ed., New York 1947, Page 136.

BASIC CONSIDERATIONS FOR THE PRELIMINARY DESIGN OF A SHOCK TUBE FOR THE INVESTIGATION OF THE ACTION OF A NUCLEAR EXPLOSION WAVE UPON A MISSILE

Dietrich E. Gudzent*
Structures and Mechanics Laboratory, Army Missile Command

Modern warfare makes it necessary to include all those environmental conditions which are expected to be imposed on all types of missiles in flight and prior to launch during a nuclear attack. The essential effects of a nuclear explosion are:

(1) The blast,
(2) The heat radiation, and
(3) The nuclear radiation.

From the point of view of the Structures and Mechanics Laboratory, Army Missile Command, all three effects are important. The blast and the heat radiation endanger the integrity of the missile structure; the nuclear radiation as well as the heat radiation can change the properties of materials to an unbearable extent. Little is known as yet of the full effects of shock waves and the accompanying extreme temperature gradient, the radiation of the fireball, and the nuclear decay on missile structures. In order to meet the extreme environment of a nuclear explosion, it appears extremely necessary to investigate the fundamental characteristics of response of the missile structure under such circumstances.

Since the Dynamics Analysis Branch of the Structures and Mechanics Laboratory is primarily interested in the dynamical response of missile structures to a nuclear explosion, this paper is restricted to the blast effects.

The problem of predicting the characteristics of dynamic loading on a missile structure, given the peak overpressure and the duration and shape of the positive and negative phases of the blast wave, can be approached in four ways:

(1) Theoretical analysis,
(2) The use of high explosive charges
(3) The full-scale nuclear test, and
(4) The shock tube.

*The author is now at the NASA Marshall Space Flight Center.

It would be like carrying coal to Newcastle to question the usefulness of the theoretical analysis. The theoretical analysis should deliver a mathematical model of the events when a blast wave passes over a missile and should also develop a model for the structure mathematically, based on structural response characteristics. However, the analytical methods are not yet developed enough to relinquish experimental investigations. Among the experimental methods the shock tube is admirably suited to studies of this nature. It provides a laboratory controlled safe method for obtaining shock waves of desired peak pressures in a relatively inexpensive manner; these shock waves may be imposed on structural models of arbitrary shapes, and of course, of limited dimensions. Accurate pressure-time measurements can be made on each face of the missile model. Naturally, the shock tube has not only advantages but also a few disadvantages such as limited control of the shape of the pressure pulse following the initial pressure, and a limited duration of the pressure pulse. However, despite these small disadvantages, the shock tube has been in the past an extremely powerful tool for the design of blast-resistant protective structures and should also usefully serve in the future for the investigation of the effects of shock waves on missile structures.

Please let me give you first a brief survey of the phenomena after a nuclear explosion in the atmosphere. Almost immediately after the detonation occurs, the expansion of the hot gases initiates a pressure wave in the surrounding air. As the pressure wave propagates away form the center of the explosion, the following (or inner) part moves through a region which has been previously compressed and heated by the leading (or outer) parts of the wave. The disturbance moves with the velocity of sound and since this velocity increases with temperature and pressure of the air through which the wave is moving, the inner part of the wave moves more rapidly and catches up with the outer part. The wavefront thus gets steeper and steeper and within a very short period it becomes abrupt. The advancing shock front provides a very sudden increase of pressure from normal atmospheric to the peak shock pressure. The shock front thus behaves like a moving wall of highly compressed air.

As the expansion proceeds, the pressure distribution in the region behind the shock front gradually changes. The overpressure is no longer constant but drops off continuously toward the center. At later times when the shock front has progressed some distance from the center, a rarefaction wave develops at the center causing a drop in pressure below the initial atmospheric value. Thus, a suction phase develops. The shock wave weakens as it progresses outward and its velocity drops toward the velocity of sound in the initial cooler air.

If the bomb is detonated at a distance h above the surface of the earth, the shock wave will have the general configuration indicated in Figure 1 (Figures are at the end on this article) during the brief time interval before it impinges upon the surface.

A short time later the radius of the shock front becomes greater than h, and that protion of the incident shock wave which impinges upon the earth's surface is reflected back forming the reflected shock wave as illustrated in Figure 2. The symbol α represents the angle of incidence of the shock wave with the earth's surface. The reflected shock wave overpressure P_r is a function of the incident shock overpressure P_s and the angle of incidence, but always

$$P_r \geq P_s \quad .$$

The reflected shock front travels through the atmosphere at a higher velocity than the incident shock and gradually overtakes and merges with it to form a single shock front called the Mach stem, as shown in Figure 2. The fused shock front thus formed is normal to and travels parallel to the earth's surface. The Mach stem formation is initiated when the angle of incidence of shock wave becomes greater than approximately 45°. Once formed, the height of the Mach stem gradually increases as the radius of the shock wave becomes greater.

The importance of the Mach stem phenomena is that it causes two shock waves to fuse into a single shock wave of higher overpressure and of greater destructive power to structures located in its path.

The peak overpressure P_s existing in the shock wave adjacent to the ground surface is a function of the distance from the point of burst and the yield of the weapon; its value is plotted in Figure 3 for four weapon sizes, 20 KT, 100 KT, 1 MT, and 20 MT. These curves are weapon bursts at ground surface. The shock front velocity U is a fuction of the peak overpressure P_s (Figure 4). Its value is given through the following equation:

$$U = C_0 \left(\frac{\gamma + 1}{2} + \frac{\gamma + 1}{2} \cdot \frac{P_s}{P_0} \right)^{1/2}$$

where

C_0 = ambient speed of sound

γ = ratio of the specific heat of air

P_0 = ambient pressure (ahead of the shock front) of the atmosphere.

With

$$\frac{U}{C_0} = M_s \quad \text{(Shock Mach Number)}$$

the equation simplifies to

$$M_s = \left(\frac{6}{7} + \frac{6\, P_s}{7\, P_0} \right)^{1/2}$$

where the velocity of the shock front is expressed as a multiple of the ambient sound speed.

After a reflection, the shock wave becomes stronger and the reflected overpressure ratio, P_r/P_s, is plotted in Figure 5 as a function of angle of incidence, α, of the shock front. This figure applies to both an inclined shock front striking the surface of the earth and a shock front striking any surface at an angle of incidence α.

In the second part of my paper I want to discuss the design criteria of a shock tube in order to simulate a nuclear explosion. Please, let me give you first a brief survey of the elementary shock tube theory.

Figure 6 shows the most simple shock tube assembly. The shock tube consists of a rigid cylinder divided into two sections by a gastight diaphragm. One-half of the tube, known as the compression chamber, contains a gas at a pressure p_4, which is in excess of the pressure p_1 of the gas in the other half of the tube known as the expansion chamber. The gases on either side of the diaphragm need not necessarily be of the same chemical type.

When the diaphragm is caused to shatter, a shock wave travels into the expansion chamber and a rarefaction wave travels back into the compression chamber. A flow of gas is behind the shock front and, in order to avoid pressure variations building up, the flow velocity is uniform in the region between the shock front and the tail of the rarefaction wave. This is also a region of constant pressure. The distribution of pressure along the tube before and after the diaphragm has shattered is also shown on the picture.

The dotted line denotes the position occupied by that gas which was originally at the diaphragm. The gas to the right has been compressed and heated by the shock wave but the gas to the left of this line has expanded from the compression chamber and has, therefore, been cooled. At this position there will be, therefore, in general, a change of type, temperature, and density; velocities are the same on both sides. Such a point is known as a contact discontinuity.

Figure 7 shows a plot in the x, t plane (x corresponds to the length of the shock tube with $x = 0$ at the diaphragm and t is the time) of the processes occurring in a shock tube with both ends closed. The shock front position is represented by the line 0A with slope $dt/dx = 1/U$ (U = velocity of the shock wave) and the contact discontinuity by 0B with slope $dx/dt = 1/U'$ (U' = velocity of the contact discontinuity). This meets the reflected shock AB at B, where the shock wave once again undergoes a reflection (not shown in the plot). The slope of the contact surface curve in the x, t plane is then very steep because the flow velocity is low in this region. The head of the rarefaction wave travelling from the diaphragm is represented by the characteristics 0D and the path of the reflected rarefaction wave is denoted by DE.

Figure 8 shows the situation in a shock tube a short time after the diaphragm was removed. The strong line gives the pressure distribution versus distance, the dashed line the original pressure distribution, and the dashed-dotted line the contact surface.

Finally, the next equation will yield the static pressure ratio P_{14} across the diaphragm necessary to give a pressure ratio P_{12} across the shock front:

$$P_{14} = P_{12}\left[1 - A_{14}\frac{(\gamma_4 - 1)(1 - P_{12})}{\left\{2\gamma_1 P_{12}\left[P_{12}(\gamma - 1) + \gamma_1 + 1\right]\right\}}\right]^{\frac{2\gamma_4}{\gamma_4 - 1}}$$

For $\gamma_1 = \gamma_4 = 7/5$ and $A_{14} = a_1/a_4 = 1$ follows

$$P_{14} = P_{12}\left[1 - \frac{1 - P_{12}}{\{7P_{12}(P_{12} + 6)\}^{1/2}}\right]^{7}$$

It is clear from the first equation that the starting pressure ratio P_{14} required to produce a shock of pressure ratio P_{12} is a function of three independent parameters, A_{14}, γ_1 and γ_4. A_{14} is the ratio of the speeds of sound in the region 1 and 4; γ_1, γ_4 is the specific heat in region 1 and 4. The theory may, of course, be formulated in terms of other parameters such as internal energy, for example. The primary condition for producing a shock is the difference in pressure across the diaphragm and, having fixed γ_1 and γ_4 by a choice of gases, one still has a free parameter which becomes fixed when the temperature ratio is specified. In practice it is found very advantageous to use a light gas to drive a heavy one, such as helium on air, because this combination gives a smaller value of A_{14} without change in temperature, and results in a stronger shock without increasing the pressure ratio across the diaphragm.

Figure 9 shows the relation between the shock Mach number M_s and the pressure ratio P_{41} across the diaphragm for several values of A_{41}. For a strong shock the following equation yields the shock Mach number:

$$M_s = \left[\frac{\gamma_1 + 1 + (\gamma_1 + 1)P_{12}}{2\gamma_1}\right]^{1/2}$$

Using a certain substitution yields

$$M_s^2 = \frac{\gamma_1 + 1}{2\gamma_1} + \left[\frac{\gamma_1^2 + 2\gamma_1 + 1}{A_{14}^2(\gamma_4 - 1)^2} + 2\right]$$

Figure 10 shows the variation of the shock pressure ratio, P_{21}, with the diaphragm pressure ratio, P_{41}, and Figure 11 the variation of the shock Mach number, M_s, and particle velocity u_2, with diaphragm pressure ratio, P_{41}, for the gas combination air/air, He/air, and H_2/air.

Before any design data can be given, it is necessary to decide which specific conditions shall be simulated. Any point around a nuclear explosion is characterized by overpressure and duration of flow. Hence, it must be determined for what overpressure and flow duration the shock tube shall be designed, though the flow duration is of minor importance since special instrumentation and theoretical considerations elminate the need to simulate the full duration of flow.

Although present missiles can only endure a very small overpressure, it appears reasonable that future missiles should bear a much higher overpressure in the magnitude of 100 psi. 100 psi is almost equivalent to a pressure ratio of the shock wave of 1:8.

Formerly given graphs have shown immediately that the desired pressure ratio of the shock front $P_{21} = 8$ is obtained by a pressure ratio across the diaphragm

$P_{41} = 200$ for air/air

$P_{41} = 26$ for He/air

$P_{41} = 16$ for H_2/air.

The notation air/air, He/air, and H_2/air means that air, helium, and hydrogen are used as driver gases in the high pressure chamber, respectively, and only air is used as driven gas in the low pressure chamber.

The pressure ratio $P_{21} = 8$ yields a velocity ratio of the shock wave $M_s = 2.6$. If the shock tube is designed to bear an overpressure of about 3000 psi ($P_{41} = 200$), then it is possible to attain shock pressure ratio $P_{21} = 15.5$ with helium and $P_{21} = 23$ with hydrogen instead of air as driver gas which is equivalent to about 230 psi and 345 psi, respectively.

The numbers given above show clearly that is is easily possible to simulate the peak overpressure of a nuclear blast with a shock tube built with moderately strong material.

The length of the shock tube is a very important factor since the duration of the flow depends on it. Figure 12 represents the course of events by a time-distance plot which can be used to determine the length of a shock tube. In the picture the origin is taken at the position of the diaphragm and at the time it breaks. The abscissa $X = x/\ell$ is distance along the channel divided by the length of the pressure chamber ℓ . The ordinate $Y = a_1 t/\ell$ is likewise dimensionless and, hence, the plot is applicable to any shock tube. The velocity of the shock front is considered as constant and has been previously presented in the equations. Its path is represented by the straight line a. The contact surface is represented by the straight line b. The rarefaction wave is designated by d and after reflection on the closed end by e.

At any particular X the length of a verical line within a zone which it crosses represents the duration of the flow in that zone for an ideal shock tube. The verical line through the point B extending from point B to the intersection with the line a at B' represents the longest possible time of flow τ between the shock front and the arrival of the contact surface before other disturbances given by the rarefaction wave may occur.

In order to calculate the very important point B for any given parameters γ_4, a_1, and a_4 the following equations can be used:

$$X_B = (Y_B - Y_C)\left(\frac{3-\gamma_4}{2} \cdot \frac{u_2}{a_1} + A_{41}\right) + X_C$$

$$Y_B = \frac{Y_C\left(\frac{3-\gamma_4}{2} \cdot \frac{u_2}{a_1} + A_{41}\right) - X_C}{\frac{1-\gamma_4}{2} \cdot \frac{u_2}{a_1} + A_{41}}$$

where

$$X_C = Y_C\left(\frac{\gamma_4 + 1}{2} \cdot \frac{u_2}{a_1} - A_{41}\right)$$

$$Y_C = A_{14}\left(1 - \frac{\gamma_4 - 1}{2} \cdot \frac{u_2}{a_1} \cdot A_{14}\right)^{\frac{\gamma_4 + 1}{2(1-\gamma_4)}}$$

$$\frac{u_2}{a_1} = (P_{21} - 1)\left(1 - \frac{\gamma_4 - 1}{\gamma_4 + 1}\right) \cdot \sqrt{\frac{1}{\left(P_{21} + \frac{\gamma_4 - 1}{\gamma_4 + 1}\right)\left(1 + \frac{\gamma_4 - 1}{\gamma_4 + 1}\right)}}$$

The following equation yields the length ℓ of the compression chamber expressed in the same system of units in which a_1 is given:

$$\ell = \frac{a_1 \tau M_s}{Y_B M_s - X_B}$$

Consequently, the overall length L of the whole shock tube is given by:

$$L = \ell X_B + \ell$$

Assuming a pressure ratio $P_{12} = 0.15$, a flow duration $\tau = 3$ sec, and a velocity of sound in air $a_1 = 332$ m/sec, the equations yield the overall length of a shock tube to 5.2 kilometers. Stronger shocks as assumed or use of helium or hydrogen would lead to even longer tubes. From these results it becomes obvious that it is practically impossible to simulate the full duration of the overpressure phase, but as already mentioned, the need for this can be eliminated by special arrangements.

If the equations are solved for τ, the flow duration behind the shock wave in a shock tube of given length of the compression chamber can be computed.

Assuming again $P_{12} = 0.15$, $a_1 = 332$ m/sec, and the length $L = 30$ m, the equation yields $\tau = 0.016$ seconds.

No principal considerations can be given for the diameter of the shock tube since no primary physical factors depend on this parameter. Nevertheless, in order to avoid boundary layer effects and as far as possible to ease the measurements, the diameter of the shock tube should be as large as justifiable from the standpoint of general design and economics.

It has been shown that from the environment of a nuclear explosion the most important part, namely the blast wave, can be simulated with a shock tube. Other parameters as flow duration and the shape of the positive phase of the overpressure are difficult to simulate, yet the need for this can be eliminated by special instrumentation and theoretical considerations.

In order to meet future requirements of missile structures in respect to blast loading it is necessary to produce a pressure ratio across the shock front $P_{21} = 8$ which is equivalent to an overpressure $p_2 = 105$ psi. The simple shock tube theory shows that this requirement can easily be achieved by a pressure ratio $P_{41} = 200$ (air/air) across the diaphragm of the shock tube. Length and diameter of the shock tube should be as large as justifiable from the standpoint of general design and economics.

REFERENCES

1. Bleakney, W. and R. J. Emrich, The Shock Tube in "High Speed Aerodynamics and Jet Propulsion," Vol VIII, High Speed Problems of Aircraft and Experimental Methods, Princeton, N. J., 1961.

2. Bradley, J. N., Shock Waves in Chemistry and Physics; London, New York, 1962.

3. Glasstone, S., The Effects of Nuclear Weapons; Washington, 1962.

4. Gudzent, D. E., Basic Considerations for the Preliminary Design of a Shock Tube for the Investigation of the Action of a Nuclear Blast Upon a Missile; AMC Report No. RS-TN-63-3, 1963.

5. Hansen, R. J., Weapons-Effects Data Applicable to Blast-Resistant Design in "Structural Design for Dynamics Loads"; New York, Toronto, London, 1959

6. Hansen, R. J., Discussion of Laboratory and Theoretical Developments for Improved Blast-Resistant Design in "Structural Design for Dynamic Loads"; New York, Toronto, London, 1959.

7. Martin, W. A., A Review of Shock Tubes and Shock Tunnels, Convair Report No ZR-658-050, 1958.

8. Oertel, H., Experimentelle Hyperschallforschung; Raketentechnik und Raumfahrtforschung, 1959.

9. Wright, J. K., Shock Tubes; London, New York, 1961.

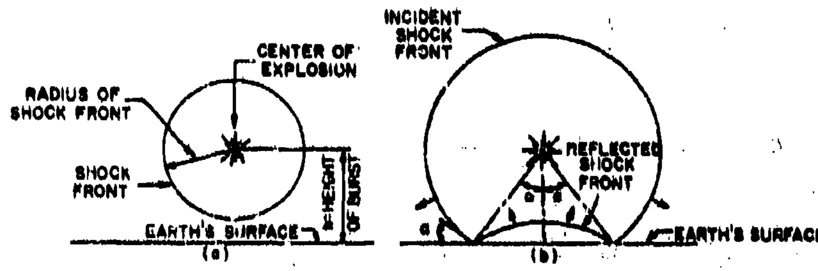


FIGURE 1 SHOCK FRONT FROM ABOVE-GROUND NUCLEAR EXPLOSION BEFORE AND AFTER REFLECTION ON THE SURFACE OF THE EARTH.

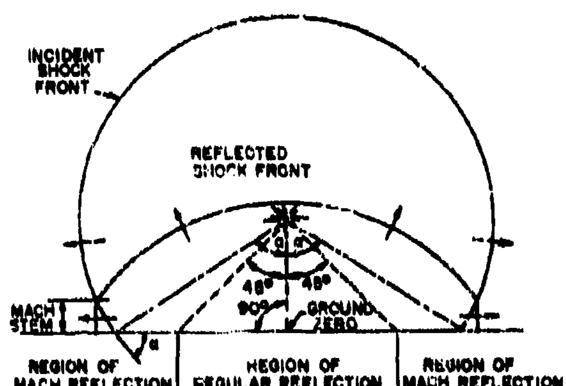


FIGURE 2 SHOCK-REFLECTION PHENOMENA IN REGION WHERE α IS GREATER THAN 45°.

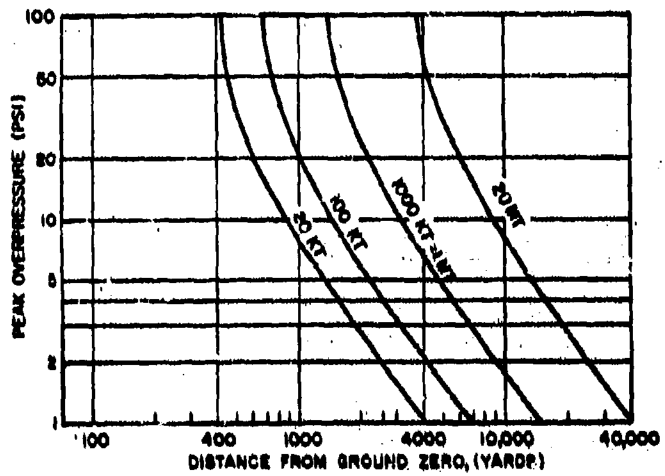


FIGURE 3 PEAK OVERPRESSURE VERSUS DISTANCE FOR SEVERAL YIELDS OF NUCLEAR BOMBS.

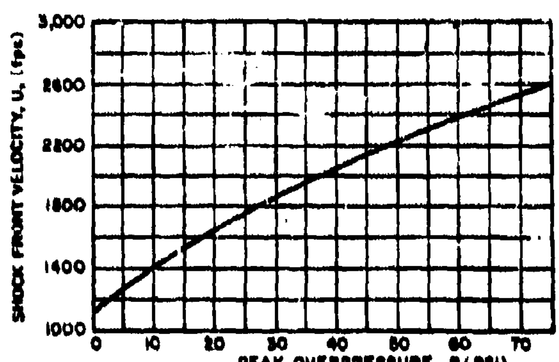


FIGURE 4 SHOCK-FRONT VELOCITY VS PEAK OVERPRESSURE

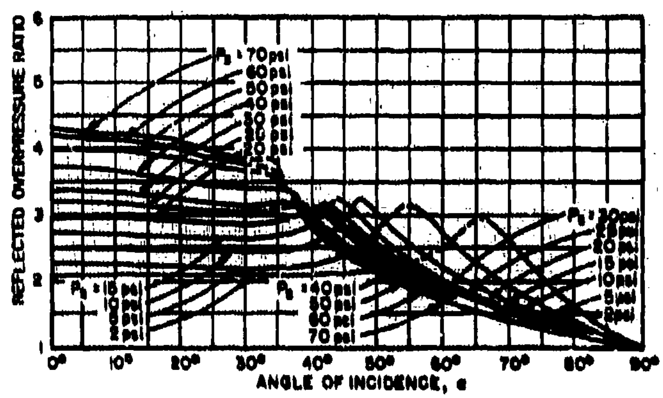


FIGURE 5 REFLECTED OVERPRESSURE RATIO VS ANGLE OF INCIDENCE FOR VARIOUS PEAK OVERPRESSURES.

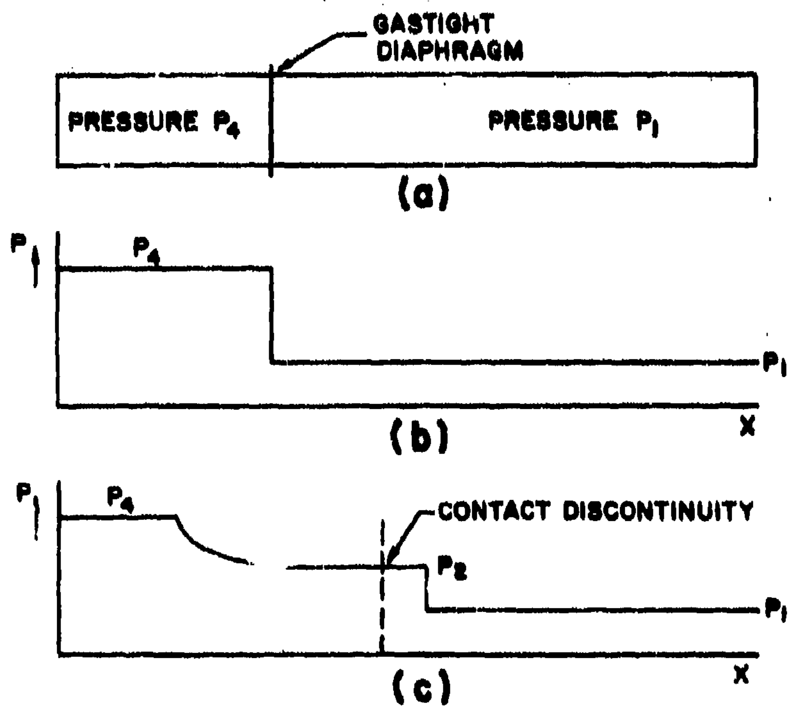


FIGURE 6 PRESSURE WAVES IN A SHOCK TUBE

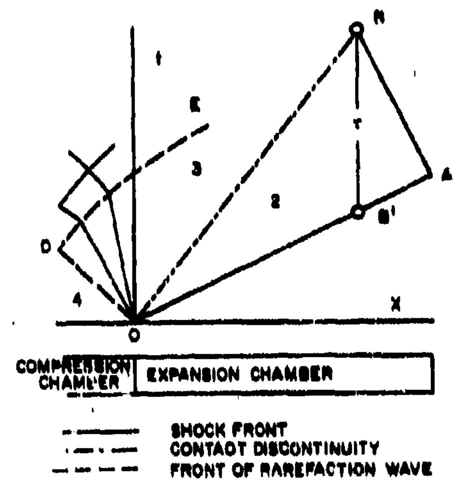


FIGURE 7. DISTANCE-TIME PLOT OF EVENTS IN A SHOCK TUBE.

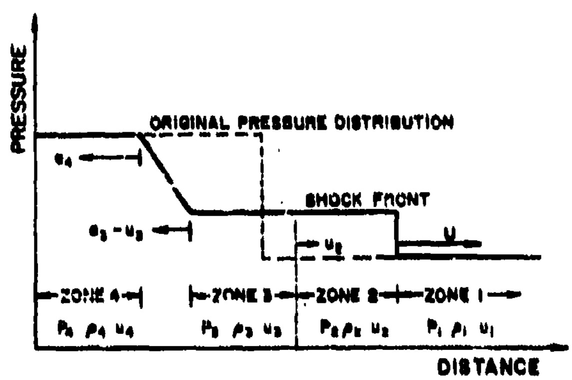


FIGURE 8 IDEAL PRESSURE DISTRIBUTION IN A SHOCK TUBE A SHORT TIME AFTER DISAPPEARANCE OF THE DIAPHRAGM.

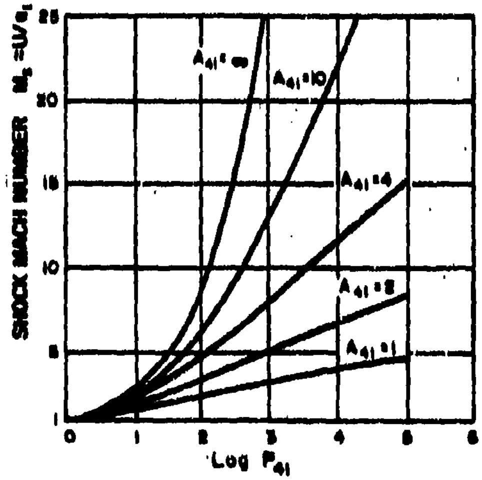


FIGURE 5. RELATION BETWEEN SHOCK MACH NUMBER AND P_{41} FOR SEVERAL VALUES OF A_{41}.

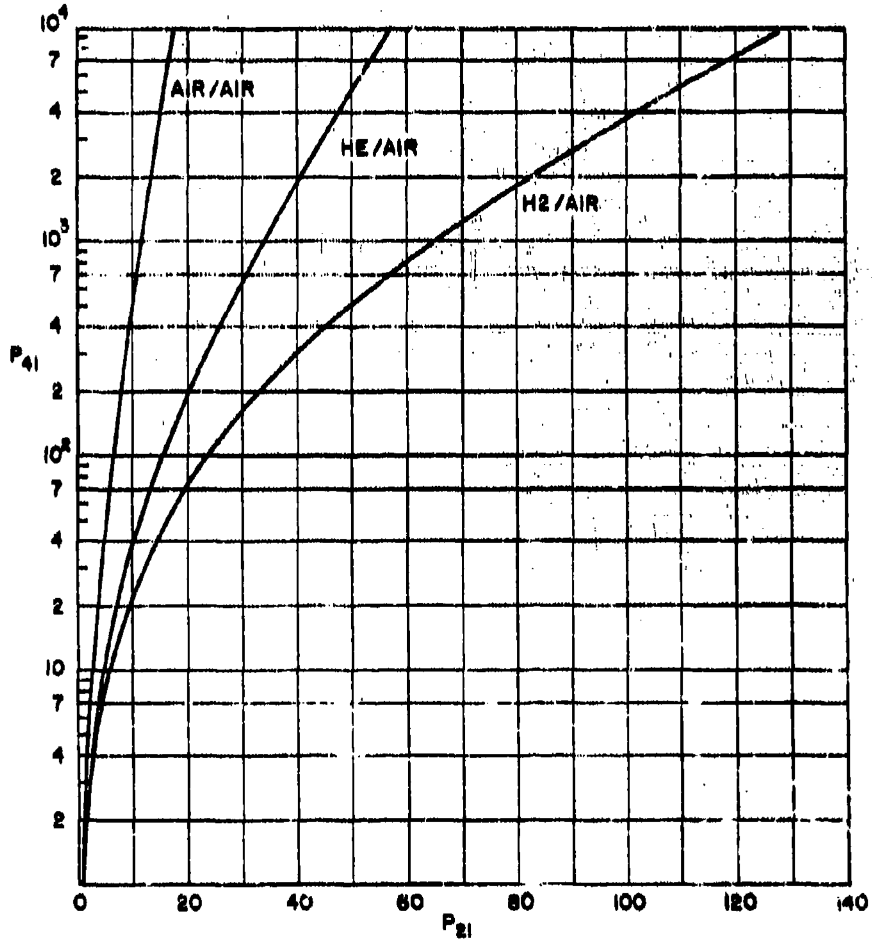


FIGURE 10 VARIATION OF SHOCK PRESSURE RATIO, P_{21} WITH DIAPHRAGM PRESSURE RATIO

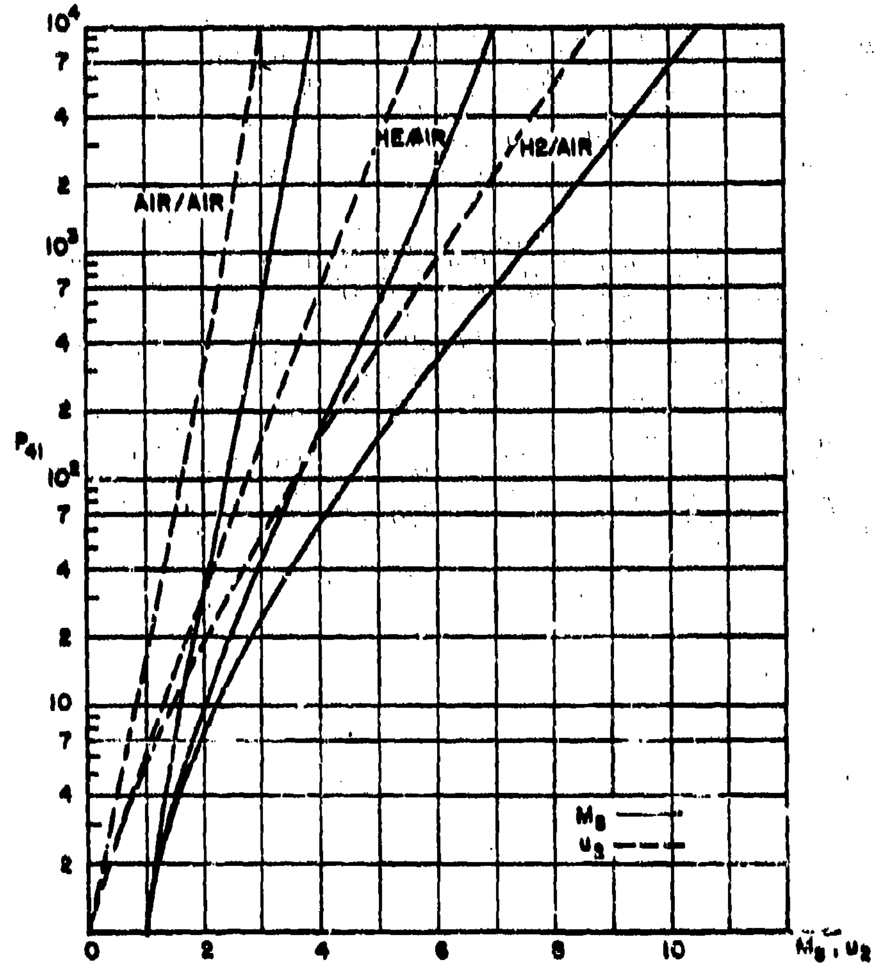


FIGURE 11 VARIATION OF SHOCK MACH NUMBER, M_s AND PARTICLE VELOCITY, u_2, WITH DIAPHRAGM PRESSURE RATIO, P_{41}

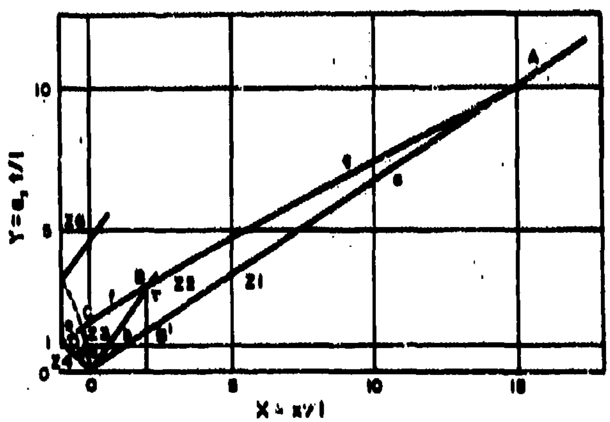


FIGURE 12 TIME-DISTANCE PLOT OF EARLY EVENTS OF WAVE PROPAGATION IN THE TUBE.

RELIABILITY CONCEPTS FOR MISSILE BATTERIES

Nicholas T. Wilburn
U. S. Army Electronics Research and Development Laboratory
Fort Monmouth, New Jersey

This session presents a long-awaited opportunity for a discussion of methods of determining the reliability of one-shot battery power supplies for missiles. Approaches and suggestions are urgently required as to appropriate test designs; the nature, extent and general advisability of making reliability predictions based on analysis of limited amounts of test data; and the consequences of assumptions made in regard to distributions and variance of the battery responses.

The batteries we are dealing with are one-shot types, installed in a missile to furnish power to guidance, control, or warhead systems. They are sto[illegible] in an inert condition in the missile and, upon command, must be activated almost instantaneously to provide power. Due to the inert state prior to activation, there is no way to test the operational readiness of an individual battery. The necessary reliability must be engineered into the battery during its development and maintained in production by adequate quality assurance techniques.

The problem areas in making reliability predictions may best be understood by first considering a typical missile battery and its electrical and environmental requirements. Electrical requirements are illustrated in Figure 1 (Figures are at the end of the article).

The shaded area is the time-voltage envelope which defines the electrical requirements. Following the activation impulse at time zero, the battery voltage must rise to the minimum under load by the end of the activation period, generally one second or less. The voltage under nominal resistance and under high or low resistance pulses must then remain between the minimum and maximum throughout the required service life throughout the specified temperature range and while being subjected to dynamic environments such as shock, vibration, acceleration, and spin. The battery design should of course be invulnerable to environments such as temperature, humidity, and transportation and handling stresses during long storage prior to activation.

Assuming that a battery has been developed for a specific application and that there is reasonable assurance, based on extensive testing during development, that the battery is capable of meeting all operating level electrical requirements while being subjected to the specified thermal and dynamic environments, the question is, how can the battery reliability be tested for and thus assured? By reliability we mean a high probability of the battery activating properly upon command and then fulfilling its mission which is, basically, meeting all specified requirements under any and all specified environmental stresses. The requirements are assumed to have been accurately stated.

Figure 2 lists some of the major problem areas. First are the reliability standards which must be established. These must, to be practical, represent some compromise between the desires of the missile people and the battery state of the art. A reliability of 99.999% at a confidence level of 99% may be desirable whereas a standard of 99.9% with 95% confidence may be an achievable goal.

Once the reliability standards have been determined, the design qualification procedures must be established. An adequate reliability testing concept is the major consideration. This will be discussed later at some length. The capabilities of test equipment must be considered, how well they will duplicate the missile environments. Test equipment which will simultaneously simulate more than one dynamic environment is normally not available. The single test concept is intriguing in this respect, an artificial environment which will contain the resultant forces of all of the major dynamic forces--shock, vibration, acceleration, etc. Unfortunately this is just a thought at present. And finally, what is the nature of the statistical distribution of the battery responses to the environments? The sample test results will be used to make reliability predictions based on the response distributions which are generally unknown, the nature of which must themselves be deduced from the sample results.

Finally when the design has been qualified and is ready for production, there are many important considerations to assure that no potentially defective units reach the field. These are, however, beyond the scope of this paper.

The major concern at present is the design qualification, testing a sample of batteries to determine if the reliability is adequate in order to go into production and field use, or if additional development is required to obtain the desired reliability. At this point I would like to describe briefly the various approaches that have been proposed and, in some instances, used by the military to determine "reliability". These are outlined in Figure 3.

The first method, testing at the "four times operating level", is a non-statistical approach which evolved as an early alternative to test-to-failure and which has been applied with considerable success. It involves testing a small sample (10 to 20) at four times the specified g level of a dynamic environment (shock, vibration, etc.). The method searches out mechanical weaknesses which can be corrected through redesign. The four times factor is arbitrary. It was felt that batteries capable of meeting this requirement would certainly have a high probability of meeting the specified g forces. Only one dynamic environment is applied at a time. The total sample size depends upon the number of different critical environmental conditions: high temperature shock, low temperature shock, high temperature vibration, etc. Temperature safety factors are investigated by testing beyond the specified temperature range in the unactivated condition. One serious disadvantage of this method is the inability to make a statistical prediction of the probable failure rate of the battery. Reliability must be expressed somewhat vaguely as "high engineering confidence in the design". Secondly it may be argued that if adequate safety factors have been considered in establishing the environmental requirements, then forcing the battery design to operate at higher stress levels may unnecessarily strengthen the design, possibly increasing cost, weight and size. Thirdly, testing only one dynamic environment at a time overlooks potentially destructive forces resulting from combinations of environments which are actually experienced.

The second method, presently being applied in the development of a missile battery, contains all of the elements of the first plus an analysis of the variance of performance parameters such as service life, activation time, maximum voltage, minimum voltage under pulses, etc. It is assumed that these responses are normally distributed. Therefore a standard may be established for the number of standard deviation units which a small test sample (15 to 25) must demonstrate between the sample mean and the operating requirement level for each performance parameter. This number, the k factor, is related to sample size, the desired maximum failure rate, and the confidence level used in making the reliability prediction. The total program thus gives a high engineering confidence in the battery design from a mechanical standpoint and also a high statistical confidence in the battery design meeting the electrical operating requirements while being subjected to each environment. Each dynamic environment is studied separately at either high or low temperature and at four times the specified g level. The possible disadvantages of the method are those as stated for the previous method plus the fact that an erroneous assumption of normality may weaken the value of the reliability predictions.

The third method is similar to the one just described except that instead of using an arbitrary multiple such as four, the specified g forces for the dynamic environments are applied. The method has the same potential disadvantages as the previous one in overlooking possible interactions between forces and in assuming normality of response distributions.

The fourth concept approaches the problem of reliability by determining the actual mean failure points of the battery design with respect to each environment, thermal or dynamic, determining the variability around the mean failure point and making a prediction of the battery capability at the requirement level for the individual environment. A method of this type was discussed by Mr. H. J. Langlie of the Ford Motor Company at the eighth conference at WRAIR last year and was the subject of much of Professor Chernoff's invited talk. In this approach the individual environmental test conditions are investigated in sequence. Thus interactions between environmental stresses are not considered. The assumption of normality is made in regard to the distribution of failure points in order for predictions to be made with small samples in the order of 15. The method can be applied only if the test equipment range is capable of inducing failures. This has proved to be a problem in many instances.

The last method which has been proposed is again based upon determining mean failure points and, based on analysis of variance, predicting the probability of failure at the requirement stress levels. The method visualizes the detrimental effects of increasingly higher environmental stress levels on the battery performance responses such that each performance parameter may be expressed as a function of the environment, thermal or dynamic, or as a function of as many environmental variables as may be studies simultaneously in accordance with the capability of test equipment. Tolerance intervals about the response surfaces thus generated can then be used for predicting the probability of failure at the requirement levels or at any settings of the variables throughout the test region. Actual failure need not be induced. The problems with this method are, again, making assumptions of normality and of uniform variance throughout the environmental test regions as well as the difficult problem of developing statistical techniques for obtaining multi-variate tolerance surfaces and being able to rely on predictions made from the testing of relatively small samples.

This general background of the attempts that have been made to determine the reliability of one-shot items should serve to illustrate some of the problems confronting us. The goal, basically, is to develop techniques with which we can come up with meaningful reliability prediction numbers from the testing of one-shot items where the sample sizes are necessarily small due to the relatively high cost of the individual items. An important question in determing our approach is, if we can't duplicate the specified missile environments in the testing laboratory in all respects with all of the impacts, vibrations and accelerations occuring in natural sequences or combinations, then are we deluding ourselves in the reliability statements we make and, in effect, playing a numbers game? Or can the statements we make following testing of bits and pieces of the total environmental picture have a real meaning which can be valuable in assessing the adequacy of a given design and assuring us that we have proceeded far enough in its development?

Next, if the reliability numbers we obtain can have some meaning, how do we get them? What is the best reliability testing concept? Do we, for example, study the variability of the battery performance levels at the specified environmental stress levels or do we find the failure points with respect to the environments and then determine that these failure points are sufficiently above the requirement levels? What statistical methods do we use or develop? How do we handle the problem of the types of distributions we are dealing with, generally unknown particularly when dealing with high g environmental stresses? What are the consequences of assuming certain distributions? How do we express the reliability? In facing these questions we must bear in mind that we don't want to force the development of a beast which is far beyond a reasonable and necessary size and cost. And, in our testing programs, we must remember that sample sizes must be limited by cost of the batteries, seldom less than one hundred dollars a piece.

These, then, are the general questions about reliability which we urgently need answers to. If the answer is that there is no answer, this too is important to know.

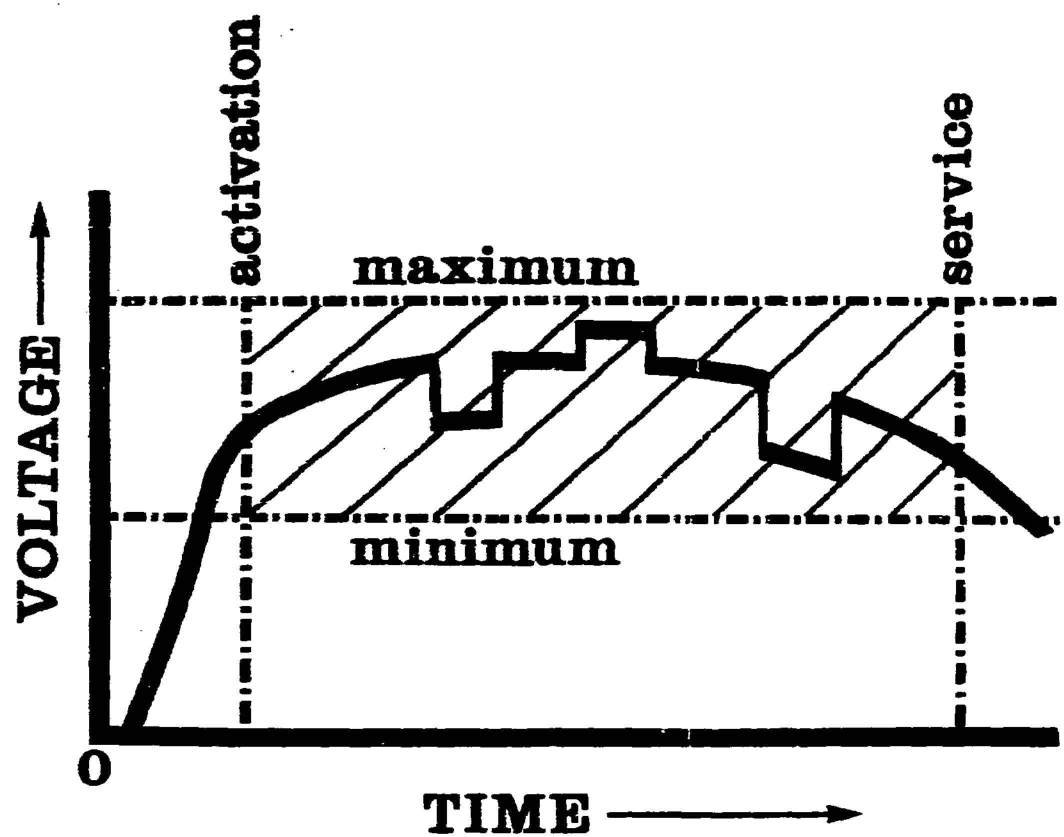


Figure 1

Reliability Problem Areas

Reliability Standards

Design Qualification Procedures

Testing Concept

Equipment Capability

Response Distributions

Production Considerations

Figure 2

Existing Approaches to Reliability

1. "Four Times Stress Level" Concept
2. 4x Concept with Performance Variability
3. Perf. Var. at Specified Stress Levels
4. Test-to-Failure [Langlie Method]
5. Test-to-Failure [Response Surfaces]

Figure 3

MONTE CARLO APPLICATION FOR DEVELOPING A DESIGN RELIABILITY GOAL COMPATIBLE WITH SMALL SAMPLE REQUIREMENTS

Ray Heathcock and Dale L. Burrows
George C. Marshall Space Flight Center
Propulsion & Vehicle Engineering Laboratory
Technical & Scientific Staff

__SUMMARY__. This report describes the application of Monte Carlo simulation to the construction of empirical sampling distribution of reliability estimates obtained by sampling from the classical stress and strength (load and failure) distributions, which are assumed Gaussian. Typical stress/strength distributions, representing specific values of reliability, were stored in a computer. From the distributions various sample sizes were taken and the resulting estimate of reliability computed. Iteration of this process resulted in the construction of empirical sampling distributions for specific values of reliability and specific sample sizes. Emphasis was placed on very high values of reliability (.99 to .99999999.) and on very small sample sizes (2 to 8) because a high reliability requirement coupled with a limited number of test articles is commonly imposed on space vehicle development programs.

Since the sampling distribution of reliability estimates was found to be very sensitive to the ratio of the standard deviations of the stress and strength distributions, sampling distributions were constructed for specific values of this ratio. Since the true ratio of standard deviations will seldom be known, the variation in the sampling distribution due to this ratio introduces a certain weakness into the application of the analysis. This weakness is discussed fully in the body of this report.

It is concluded from this investigation that the sampling distributions, constructed by Monte Carlo simulation, may be utilized to aid the designer in establishing a design reliability goal, place a confidence coefficient on reliability estimates, and to determine if sample stress and strength data demonstrates a specified reliability at a specified confidence level.

The primary purpose of this report is to present a method and examples of the use of the method. The ogives given in this report may be used in actual application, however, not indescriminately, since the ogives contain small inaccuracies due to the curve fitting procedure.

I. INTRODUCTION.

A. Background. In large space vehicle development, and in many other fields as well, a designer is often asked to design his equipment knowing that later a very small sample of items will be tested to determine if a specified minimum reliability is demonstrated at some fairly high confidence level. The designer is often quite perplexed as to the methods and reliability goals he should utilize in formulating his design. His own training and experience make him favor the safety factor approach, and yet statisticians and reliability engineers advocate safety margins, stress/strength relationships and other statistical approaches. The publications of Robert Lusser advocate the safety margin approach, (Ref. 1) while ARINC Research Corporation under contract NASA (Ref. 2) has advocated the use of stress/strength relationships as a reliability prediction technique. There are many other papers and publications which propose the use of statistical variations in stress and strength in design and subsequent analysis of the design. The application of a single distribution (strength) and a reliability boundary (upper limit of stress) is discussed in references 1 and 3.

B. Scope. This investigation attempts to shed some additional light on problems concerning the use of stress/strength statistics in design, reliability demonstration, and confidence limits. It is primarily concerned with the solution of statistical sampling problems which have no known theoretical solution and with the application of the data provided by these solutions. In this analysis, very small sample sizes in the range of 2 to 8 and high reliability values from .99 to .99999999 have been purposely used because space vehicle development programs are restricted to small samples and require high reliability values.

The authors wish to acknowledge the very capable assistance of Messrs. Robert Crafts, Joe Medlock and Matt Blue of the Computation Laboratory, MSFC, for programming the Monte Carlo simulation scheme, and of Mr. R. L. Bombara, Engine Projects Office, MSFC, whose technical advice and suggestions were very helpful.

II. DESCRIPTION.

A. Analysis. The reliability of many items in large space vehicles from piece parts to large structural elements, can be appropriately considered to be a function of a stress distribution and a strength distribution. A stress distribution is defined as a distribution of stresses to which the population of items will be subjected in actual use, and a strength distribution is defined as a distribution of stresses which will cause failure of the items. This analysis applies to the general problem of pitting a distribution of "what an item will do" against a distribution of "what it is required to do" in any performance parameter. The analysis is not restricted to the more common application of structural stress/strength analysis.

An example of the relationship between the stress and strength distributions of a typical situation is shown in Figure 1 (Figures are at the end of this article). The distributions are assumed to be normal, an assumption which is retained throughout this paper.

A randomly chosen value of stress (X_2) subtracted from a randomly chosen value of strength (X_1) gives a variate from the strength minus stress distribution. By repeating this process many times, a distribution of strength minus stress may be formed. It will appear as shown in Figure 2. The mean of this distribution is the mean of the strength distribution minus the mean of the stress distribution. The standard deviation of this distribution is the square root of the sum of the variances of the strength and stress distributions. At some point on this distribution, stress equals strength and a zero point appears on the $X_1 - X_2$ axis. Since any negative value of strength minus stress represents a failure, the area below the zero point represents unreliability, and the area above the zero point represents reliability. Mathematically this may be stated as follows (Ref. 4):

$$R = P_r(X_1 \geqq X_2)$$

$$R = P_r(X_1 - X_2 - M_{x_1} + M_{x_2} \geqq -M_{x_1} + M_{x_2})$$

$$R = P_r\left(\frac{X_1 - X_2 - M_{x_1} + M_{x_2}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}} \geqq \frac{M_{x_2} - M_{x_1}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}}\right)$$

$$R = P_r\left(Z \geqq \frac{M_{x_2} - M_{x_1}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}}\right) \text{ since } Z \text{ is a normal deviate}$$

In actual practice the parameters of the stress and strength distributions will rarely be known and must be estimated from sample data. Small sample estimates of the parameters may be used to estimate reliability as follows:

$$R_{est} = P_r \left(Z \geq \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{s_{x_1}^2 + s_{x_2}^2}} \right)$$ which can be obtained from the normal table of areas.

Since an estimate of the reliability of a particular situation can now be made, the next logical step is to describe the variations expected in this estimate due to sampling. In order to make use of the stress/strength statistical relationship, the sampling distribution of reliability estimates, based on this relationship, must be developed. Since there was no known theoretical solution to the description of this variation, a computer program was developed using Monte Carlo simulation to derive the empirical sampling distribution of the quantity

$$\frac{\bar{X}_2 - X_1}{\sqrt{s_{x_1}^2 + s_{x_2}^2}}$$

Very briefly, this simulation technique consists of the following steps:

(1) Store hypothetical stress and strength distributions in the computer
(2) Generate a pseudo random number
(3) Use this random number to get a random value of strength (X_1) repeating this process N times
(4) Compute a sample mean and standard deviation for strength
(5) Repeat the process for stress
(6) Compute

$$K = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{s_{x_1}^2 + s_{x_2}^2}}$$

(7) Form a histogram of values of K, which represents the sampling distribution. As many values of K as desired may be obtained from the program, dependent on the accuracy desired. 1000 values were used to obtain the information for this paper. The sampling distribution may be put in a cumulative form, termed an ogive, in order to be able to read K values corresponding to selected values of probability.

B. Results. The factors which influence the sampling distributions of K will now be discussed. First, sample size, of course, will influence it. An empirical sampling distribution must be generated for each sample size that will be used in actual practice. To give an idea of how the sampling distribution varies, as sample size varies, some ogives for various sample sizes have been developed.

The ogives in Figure 3, from left to right, represent values of reliability from .99 to .99999999. In other words, the first curve to the left represents the variation inherent in estimating a true .99 reliability using a specified sample size. This figure represents the case where $N_{x_1} = 8$ and $N_{x_2} = 8$ i.e., sample sizes of 8 from each of the strength and stress distributions. The standard deviations of stress and strength are equal in this case. The effect of varying these standard deviations will be discussed later. Other ogives for various sample sizes and equal sigmas appear in Figure 4 through 8.

As may be observed from Figure 3 through 8, the ogives vary quite radically with sample size, especially in the very small sample sizes used here. In order to use this type empirical data, ogives for the specific sample sizes used in a particular application must be developed.

The second important factor which causes the sampling distribution to vary is the ratio of the standard deviations of stress and strength. This ratio is defined as the smaller standard deviation divided by the larger standard deviation. Early in this program Monte Carlo results indicated that the sigma ratio had an important effect on the variation of the sampling distribution, and later, the theoretical application of reference 5 served to verify this conclusion. The results of the theoretical application showed that for equal sample sizes the degrees of freedom of the non-central t distribution is a function of sample size and the ratio of standard deviations of the stress and strength distributions. Figure 9 shows the variations in the sampling distributions for various ratios of the standard deviations with a fixed sample size of $N_{x_1} = N_{x_2} = 5$. Here only the upper portion of the distributions are shown since most applications will be concerned with high confidence and because the variation is most pronounced in this area. These curves were obtained by holding all variables in the computer program constant except the sigma ratios. The sample sizes were equal for stress and strength. As the figure indicates by convergence of the curves, there is little variation in the sampling distribution of ''K'' attributable to sigma ratios below the 80th percentile point of "K". Above this point there is considerable variation; the higher ratios result in sampling distributions which are skewed to a greater degree. Although ratios up to 1/100 were run on the Monte Carlo program, the practical range is probably between 1 and 1/10 for most applications. Even at the 1/10 ratio, however, there is a large variation; therefore, separate curves for the specific ratio must be utilized in actual application. A weakness, however, in this procedure is that in practice the actual ratio will seldom be known and therefore must be estimated from sample data. An F test for the ratio of two variances could be used to establish if there is any significant difference between the sigmas and possible, confidence limits for the ratio of two variances could be utilized for estimating the limits of the ratios. Further work is desirable in developing an approach to establishing a ratio from sample data which could be used for entering the appropriate set of curves.

Another important factor to consider is the case of unequal sigmas (stress and strength) and unequal sample sizes. The most skewed condition (long tail to the right) in the sampling distribution of "K" will result when the smaller sample is taken from the distribution which has the larger standard deviation. For instance, assume that the standard deviation of strength is twice that of stress and a sample of 8 is taken from the stress distribution and 5 from the strength distribution, i.e., the smaller sample is taken from the distribution with the larger sigma. What would happen if a reverse procedure were used and the large sample taken from the distribution having the larger sigma?

The ogive for this case is shown in Figure 10. The broken line represents the ogive for the large sample matched with the large sigma, and the solid line represents a rerun of the same case with the small sample with the large sigma. As can be seen, there is a significant variation in the two conditions. It is concluded, therefore, that this is another condition which must be included in the Monte Carlo output in order to utilize the results efficiently and accurately. This presents no problem, however, if the sigma ratio is known since Monte Carlo runs for the required conditions can be made. It is being presented merely to illustrate that it does have an influence.

The discussion, thus far, has dealt only with the mechanics of obtaining the sampling distribution of "K" and its variation as related to specific factors causing it. It is appropriate to discuss the application of such data.

C. Application. The criteria for determining whether sample data demonstrates a given reliability at a specified confidence will be developed in terms of "K" which is normal deviate corresponding to a specified area under the normal curve.

The basis for developing the demonstration criteria is shown in Figure 11 which depicts the distribution of $X_1 - X_2$ and $X_1 - X_2 - K_c\sqrt{s_{x_1}^2 + s_{x_2}^2}$. The Z shown in the figure is a normal deviate, the area above which (obtained from normal tables of areas) represents the reliability which it is required to demonstrate. K_c represents a normal deviate greater than Z which must be found and when applied will assure a demonstrated reliability with confidence C.

If a K_c that satisfies the following inequality

$$P_r(X_1 - X_2 - K_c\sqrt{s_{x_1}^2 + s_{x_2}^2} \leqq M_{x_1} - M_{x_2} - Z\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}) = C$$

(Ref. 3) can be found and applied as a criteria, a decision can be made as to whether or not the sample data demonstrates a specified reliability (Z). This inequality may be reduced to

$$P_r(X_1 - X_2 - K_c\sqrt{s_{x_1}^2 + s_{x_2}^2} \leqq 0) \quad \text{since} \quad M_{x_1} - M_{x_2} - Z\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} = 0$$

in this situation, because the Monte Carlo program was run on this basis. Manipulation of this inequality, as follows, mathematically gives the criteria:

$$P_r(X_1 - X_2 \leqq K_c\sqrt{s_{x_1}^2 + s_{x_2}^2}) = C$$

or

$$P_r\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_{x_1}^2 + s_{x_2}^2}} \leqq K_c\right) = C$$

Now K_c can be found by referring to the Monte Carlo developed ogives for the conditions of the problem (specified Z, N_{x_1}, N_{x_2}, σ_{x_1}, σ_{x_2}). Once K_c is found, the criteria for demonstration is as follows:

If the sample quantity $\dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_{x_1}^2 + s_{x_2}^2}} \geqq K_c$, reliability Z is demonstrated with the desired confidence. As an example, suppose samples of $N_{x_1} = 8$, $N_{x_2} = 8$ gives $\bar{X}_1 = 80{,}000$, $\bar{X}_2 = 60{,}000$, $S_{x_1} = 3000$, $S_{x_2} = 3000$ and the problem was to determine if .9999 reliability ($Z = 3.71$) was demonstrated at the 90% confidence level. The 90% point on the ogives for .9999, $N_{x_1} = N_{x_2} = 8$, $\sigma_{x_1} = \sigma_{x_2}$ is found to be 4.95 which is K_c.

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_{x_1}^2 + S_{x_2}^2}} = 4.7$$

therefore since $4.7 < 4.95$ reliability .9999 is not demonstrated at the 90% confidence level. The accuracy of the answer obtained is dependent on how close the true sigma ratio is to unity, since it was assumed that $\sigma_{x_1} = \sigma_{x_2}$ in this example.

Another application of the Monte Carlo results is the establishment of a lower confidence limit on a reliability estimate. The appropriate set of ogives is entered with the reliability estimate (reliability estimate expressed in terms of a K value) and a confidence is read for each reliability value represented by curve. If sample data ($N_{x_1} = N_{x_2} = 8$, $\sigma_{x_1} = \sigma_{x_2}$) gave a reliability estimate expressed in terms of $K = 3.5$ for a given sigma ratio and a given sample size, proceed as follows to arrive at a lower confidence limit:

(1) Refer to the series of ogives (figure 12) representing the sigma ratio and sample sizes applicable to the problem.

(2) On the horizontal scale locate a K value of 3.5 which was the sample estimate and draw a vertical line through the point.

(3) Where the vertical line intersects a curve, read a confidence for the reliability represented by that curve. For the example, Figure 12 shows 95% confidence in .99 reliability, 70% confidence in .999 reliability, etc. By generating a sufficient number of curves, a confidence coefficient for the reliability estimate in question, or at least close enough to it for practical application, can be obtained.

If in the initial design of an item, a designer knew exactly the stress distribution and knew exactly what strength distribution he could get, designing would be a simple problem and there would be no need for demonstration. However, this is not the case, and the designer has to make estimates of the distributions.

Assuming that the designer knew exactly the strength & stress distributions and was to design, knowing that a small sample of his items was to be tested later for demonstration purposes, it would be to his benefit to overdesign so that he would have a good chance say 90%, of having the small sample demonstrate the specified reliability. In order to arrive at how much he should overdesign, he can consult the Monte Carlo developed ogives and find a K_C such that if the sample K is greater than K_C, the sample data has demonstrated the required reliability at the desired confidence. Reference to Figure 13 will aid the reader in following this approach. If the designer wants a 90% chance of having a sample demonstrate the required reliability, he must design to a reliability represented by an ogive 90% of which is above the K_C point. This logic seems to be non sensical since if a designer knows the distributions he could just design to the reliability he desires and there would be no point in a demonstration program; however, this logic can be applied to the situation where the designer does not know the distributions but has some knowledge of them from experience or design calculation. Assuming a designer wants to design so that he has a 90% chance of demonstrating a specific reliability with a specified sample size, it can be concluded from the previous discussion that he must design to a reliability above that which gives him a 90% chance of demonstrating it with the specified sample size. How much above is a matter of engineering judgement, and depends upon how well the designer thinks he can estimate the distributions. This conclusion vividly points out the fact that a design goal (inherent design reliability) must be higher than the specific reliability which is to be demonstrated at a high degree of confidence. There are, of course, other factors which influence the establishment of reliability goals. Weight, cost and performance all should influence design decisions and must be properly considered as tradeoff factors against the statistically developed goal.

Since much of the foregoing discussion of reliability has been in terms of the normal deviate, Z, Figure 14 has been provided to enable the reader to determine Z directly from a numerical value of unreliability (1-R).

III. CONCLUSIONS & RECOMMENDATIONS.

A. Conclusions. This investigation has revealed that the statistical information afforded by very small samples from a stress/strength situation (even as low as 2) can be useful to the designer and the reliability engineer. It has also revealed that the sampling distribution of reliability estimates made by taking sample data from the stress/strength distributions is very sensitive to sample size and the ratio of the standard deviations of the stress and strength distributions.

It is concluded that the empirical sampling distribution of reliability estimates can be utilized by the designer or the reliability engineer as follows:

1. Reliability Demonstration-Given sample data from stress and strength distributions, a determination can be made as to whether the sample data demonstrates a specified reliability at a chosen confidence level.

2. Establishing Confidence Coefficients-Given sample data from a stress/strength situation, confidence coefficients for various values of reliability can be estimated, limited only by the number of curves that have been generated by Monte Carlo.

3. Recommend Design Goal-One can establish and recommend to the designer a design goal such that if he designs to it, a pre-chosen sample will demonstrate a specified reliability at a chosen confidence level, a specified percent of the time.

B. Recommendations for Future Investigation. In view of the number of promising applications discovered during this investigation and the weaknesses and limitations that are inherent in the nature and extent of this analysis, it is believed that the following areas are worthy of further investigation:

1. Since the ratio of standard deviations of the stress and strength distributions has a large affect on the variance of the reliability sampling distribution (Distribution of "K") and since in practice this ratio will seldom be known but must be estimated from sample data, a method should be derived for establishing a ratio which could be used for entering the appropriate set of empirical curves.

2. Extension of the analysis to various combinations of different types of distributions, other than normal, should be investigated by the Monte Carlo process since in these areas not even approximation to theoretical solutions are available.

REFERENCES

1. Lusser, Robert, Reliability through Safety Margins, United States Army Ordnance Missile Command, Redstone Arsenal, Alabama, 1958.

2. ARINC Research Corporation, Saturn Launch Vehicle Reliability Study, Marshall Space Flight Center, Huntsville, Alabama, 1960.

3. Bombara, Elwood L., Reliability of Compliance with One-Sided Specifications Limits when Data is Normally Distributed. Army Rocket and Guided Missile Agency Technical Report 2B1R, Redstone Arsenal, Alabama, 1961.

4. Normal, J. E., Estimating Reliability as a Function of Stress/Strength Data. Army Rocket & Guided Missile Agency Redstone Arsenal, Alabama, 1961.

5. Bombara, Elwood L., Probability that Stress is less than Strength at Prescribed Confidence Levels, for Normally Distributed Data, 1963.

LIST OF ILLUSTRATIONS

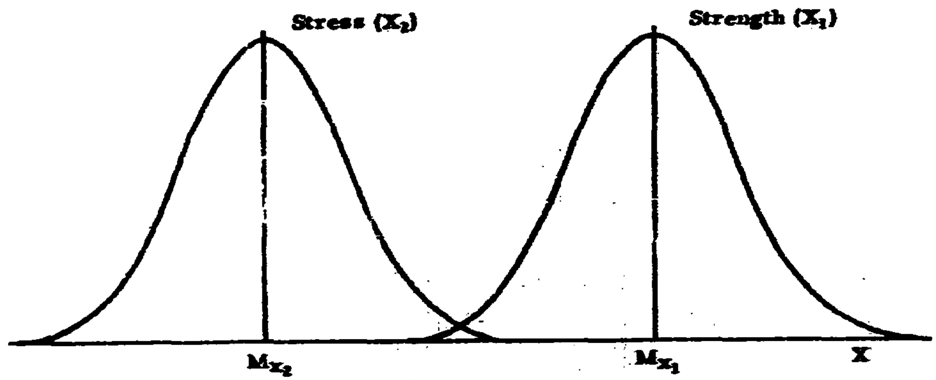
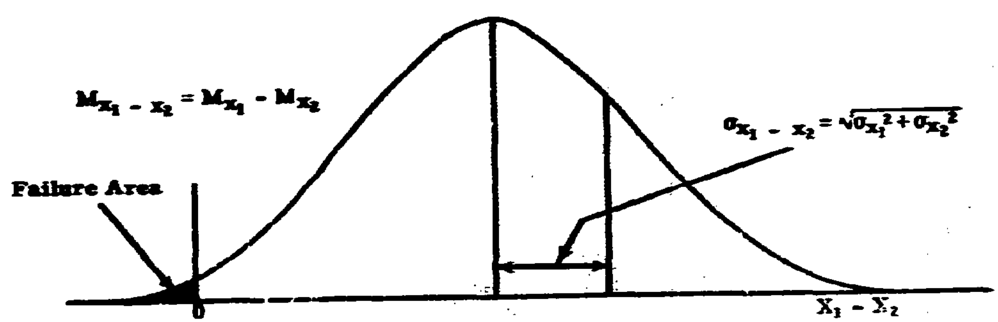


FIGURE 1. RELATIONSHIP OF STRENGTH AND STRESS DISTRIBUTIONS (NORMAL)



$$\text{Reliability} = P_r (X_1 > X_2)$$

$$R = P_r (X_1 - X_2 - M_{x_1} + M_{x_2} > -M_{x_1} + M_{x_2})$$

$$R = P_r \left(\frac{X_1 - X_2 - M_{x_1} + M_{x_2}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}} > \frac{M_{x_2} - M_{x_1}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}} \right)$$

$$R = P_r \left(Z > \frac{M_{x_2} - M_{x_1}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}} \right)$$

Where: Z = Normal deviate

$$R_{est} = P_r \left(Z > \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{S_{x_1}^2 + S_{x_2}^2}} \right)$$

Normal table of areas

FIGURE 2. RELIABILITY COMPUTATION - STRENGTH/STRESS DISTRIBUTIONS

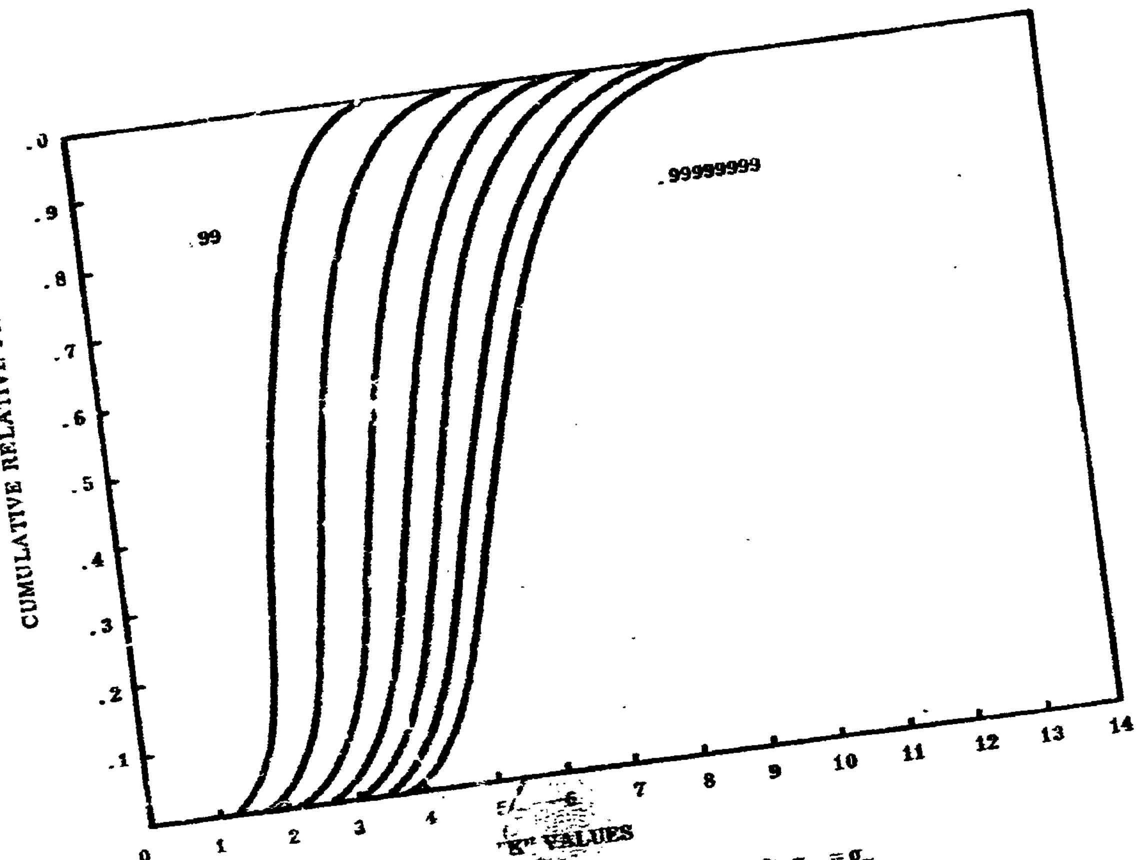


FIGURE 3. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 8$, $\sigma_{x_1} = \sigma_{x_2}$

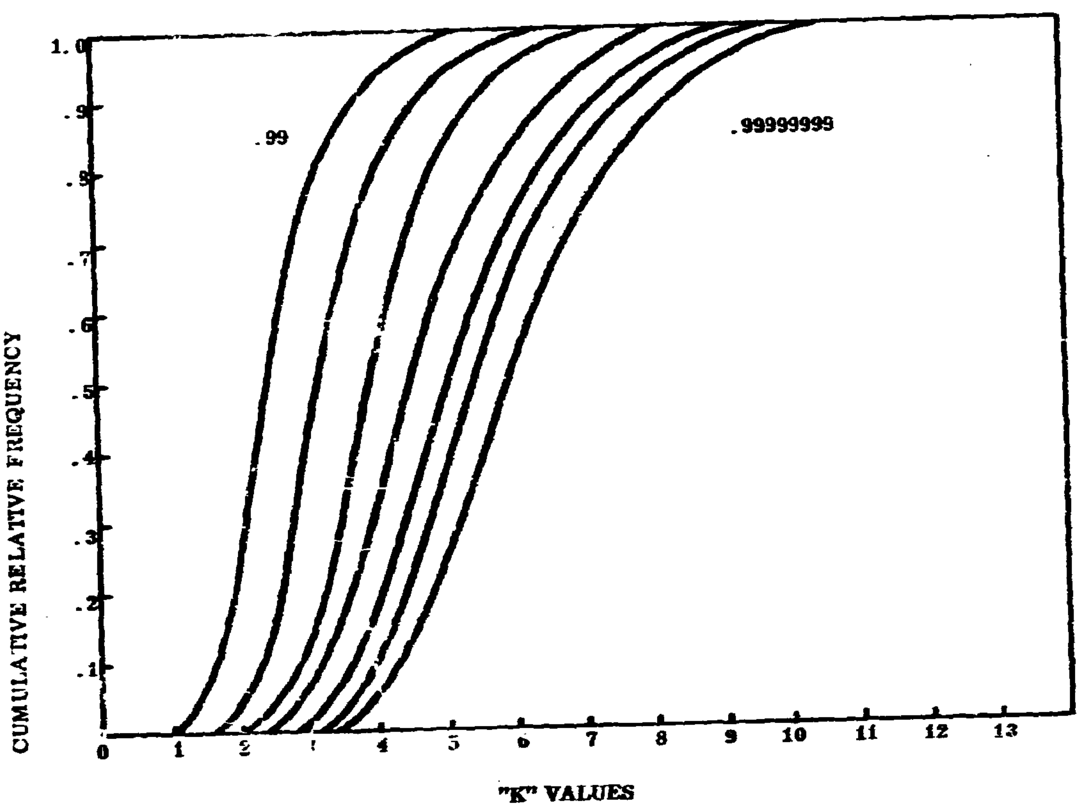


FIGURE 4. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 5$, $\sigma_{x_1} = \sigma_{x_2}$

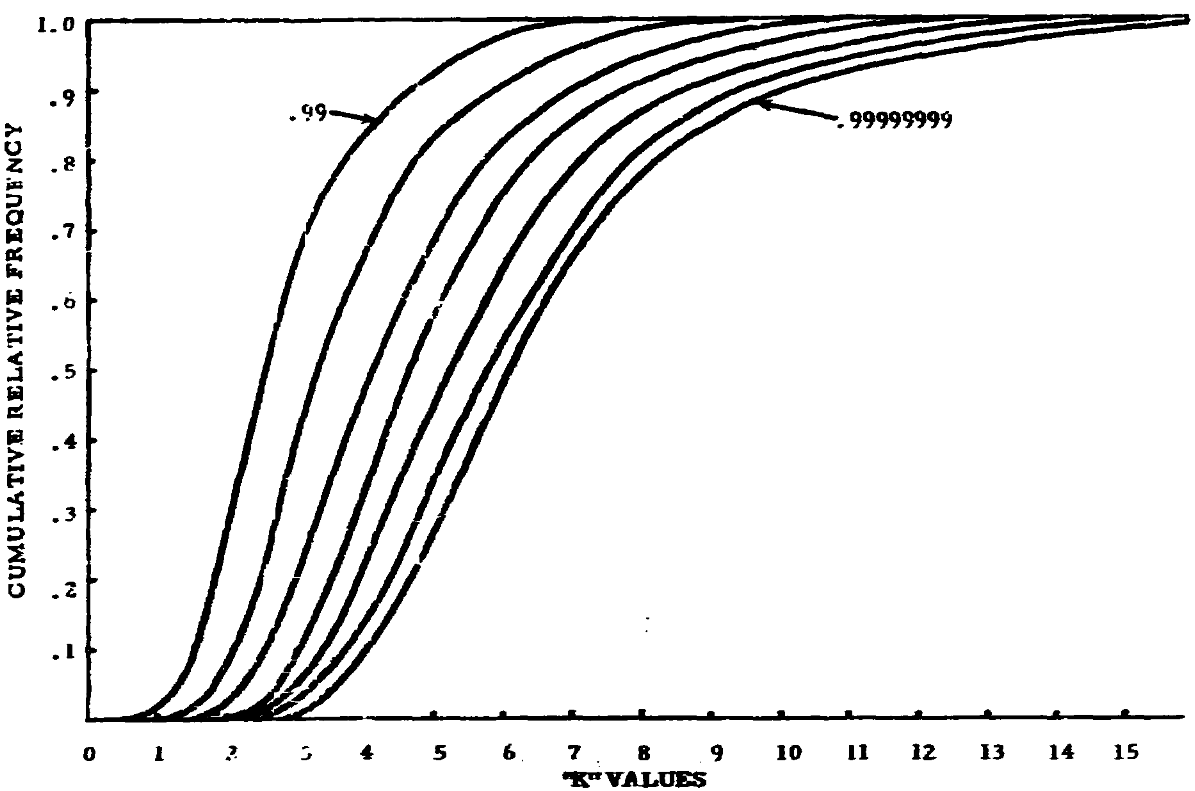


FIGURE 5. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 3$, $\sigma_{x_1} = \sigma_{x_2}$

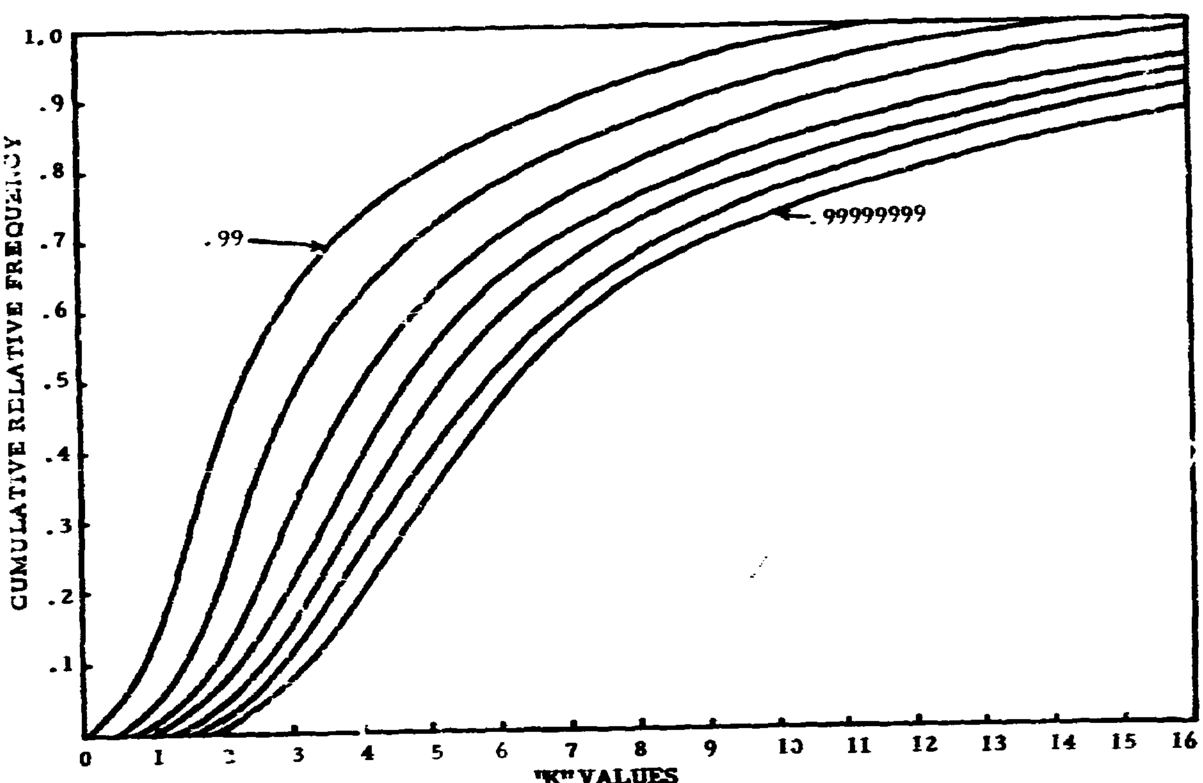


FIGURE 6. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 2$, $\sigma_{x_1} = \sigma_{x_2}$

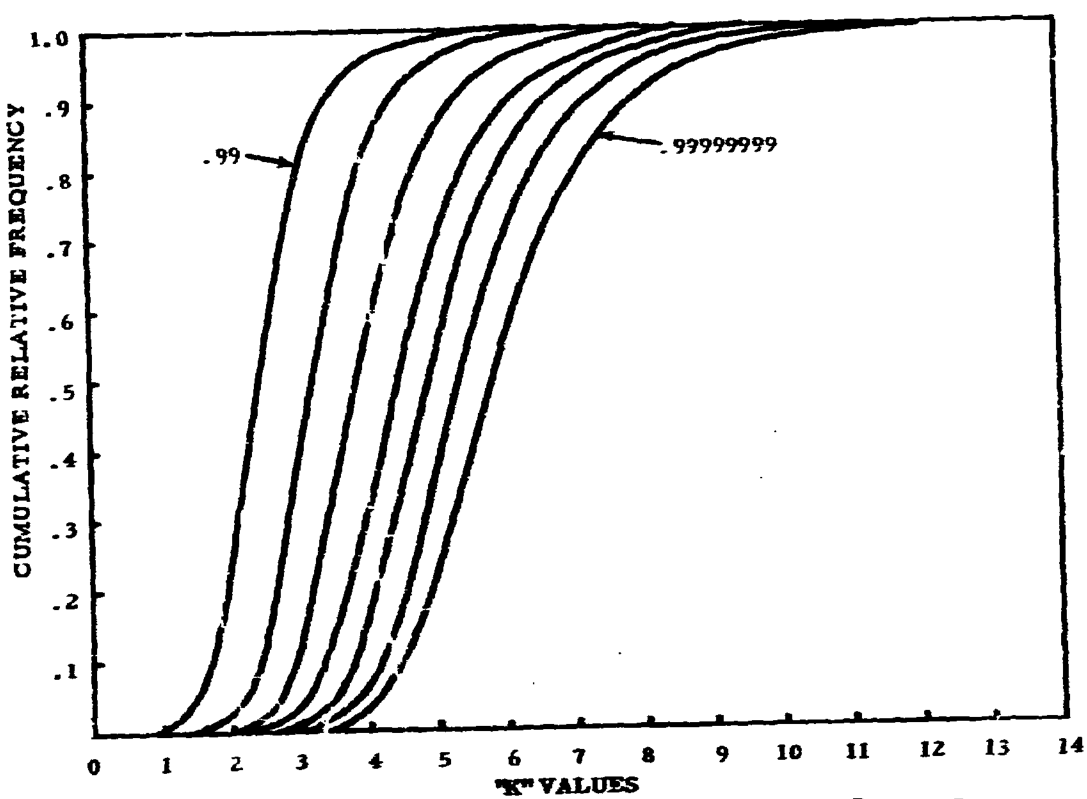


FIGURE 7. SAMPLING DIST. OF "K", $N_{x_1} = 8$, $N_{x_2} = 5$, $\sigma_{x_1} = \sigma_{x_2}$

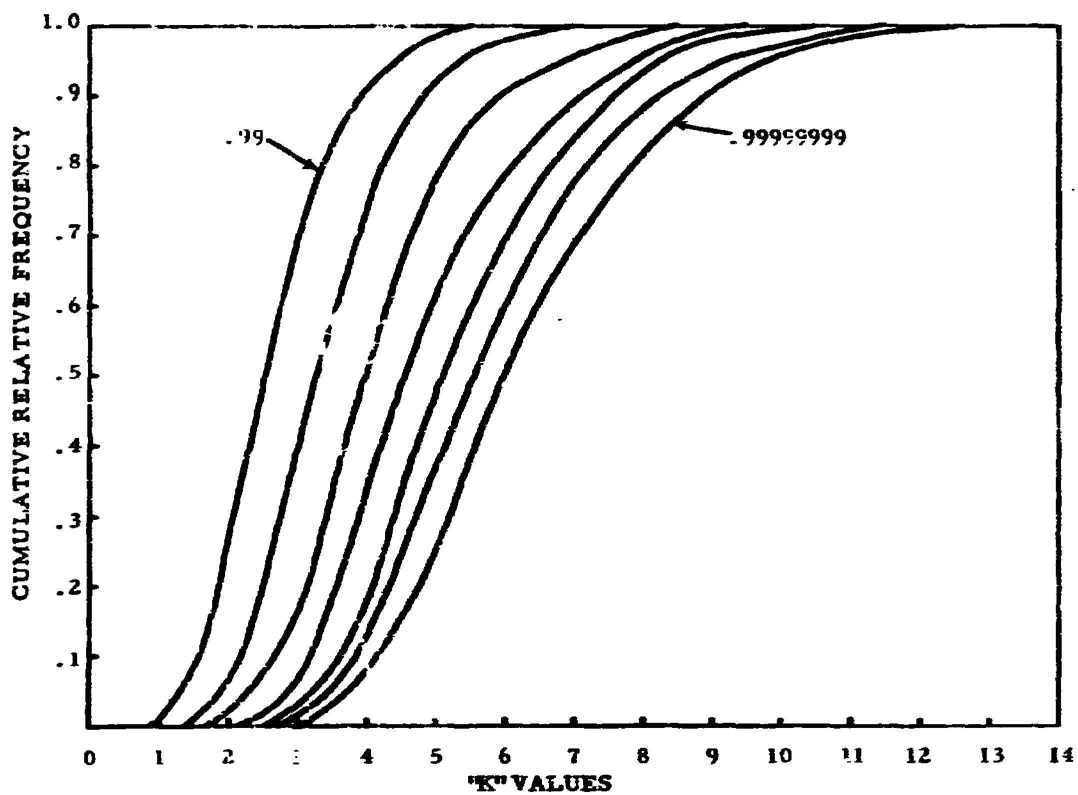


FIGURE 8. SAMPLING DIST. OF "K"; $N_{x_1} = 5$, $N_{x_2} = 3$, $\sigma_{x_1} = \sigma_{x_2}$

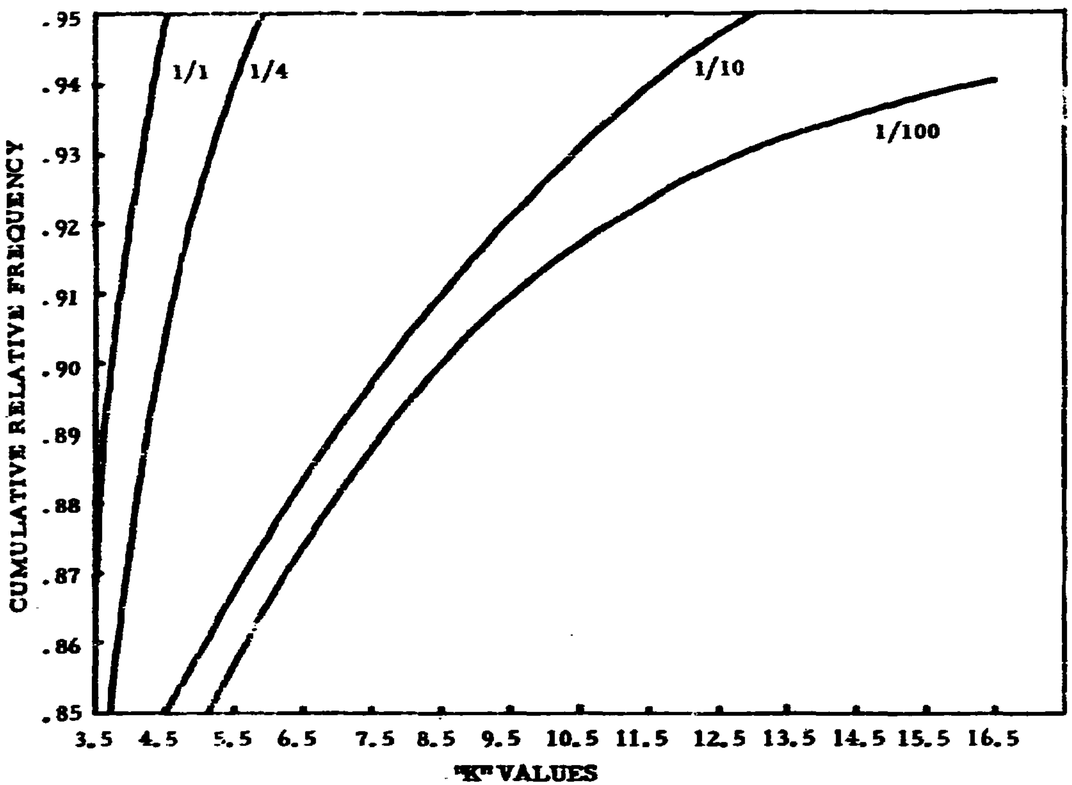


FIGURE 9. VARIATIONS IN DIST. OF "K" DUE TOSIGMA RATIO

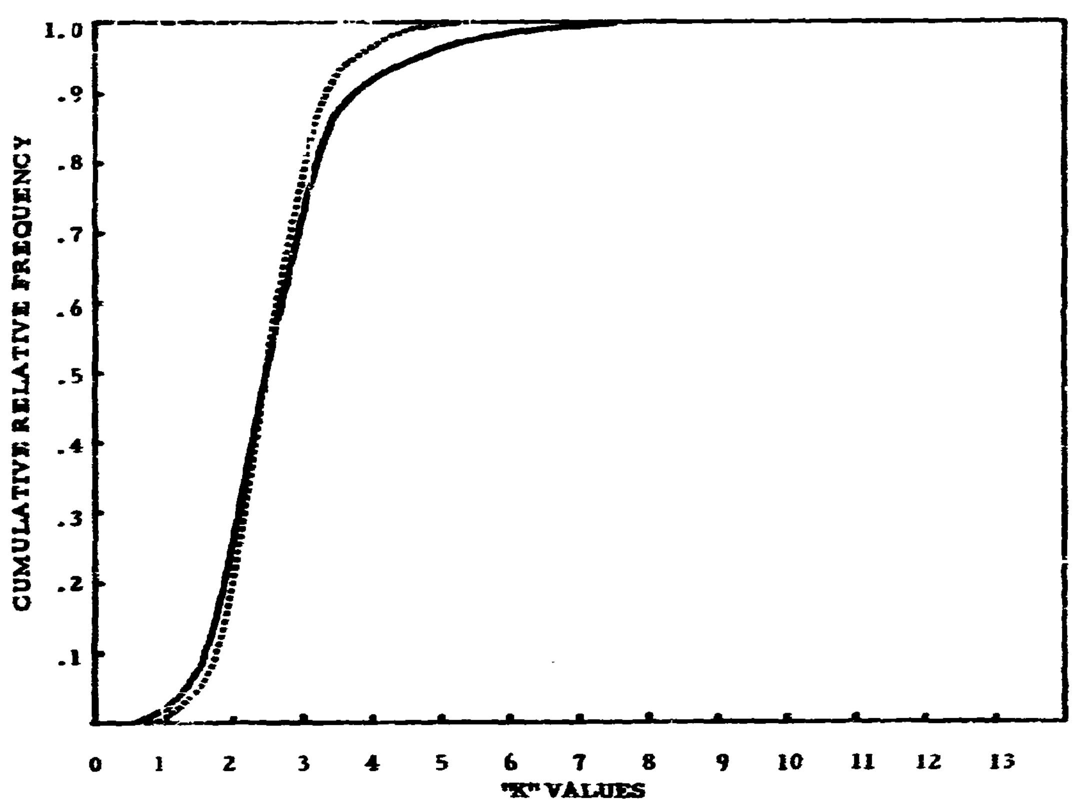
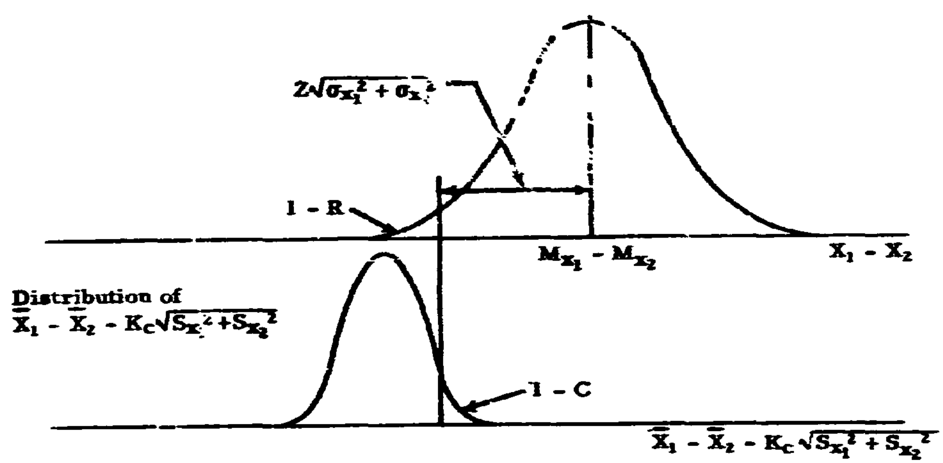


FIGURE 10. SAMPLING DIST. OF "K", $N_{x_1} \neq N_{x_2}$, $\sigma_{x_1} \neq \sigma_{x_2}$



$$P(\bar{X}_1 - \bar{X}_2 - K_C\sqrt{S_{x_1}^2 + S_{x_2}^2} \leq M_{x_1} - M_{x_2} - Z\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}) = C$$

$$P(\bar{X}_1 - \bar{X}_2 \leq K_C\sqrt{S_{x_1}^2 + S_{x_2}^2}) = C \quad \text{Since } M_{x_1} - M_{x_2} - Z\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} = 0$$

$$P\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_{x_1}^2 + S_{x_2}^2}} \leq K_C\right) = C \quad K_C \text{ is the 90\% point on the distribution of } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_{x_1}^2 + S_{x_2}^2}}$$

FIGURE 11. DISTRIBUTION OF $X_1 - X_2$ and $\bar{X}_1 - \bar{X}_2 - K_C\sqrt{S_{x_1}^2 + S_{x_2}^2}$

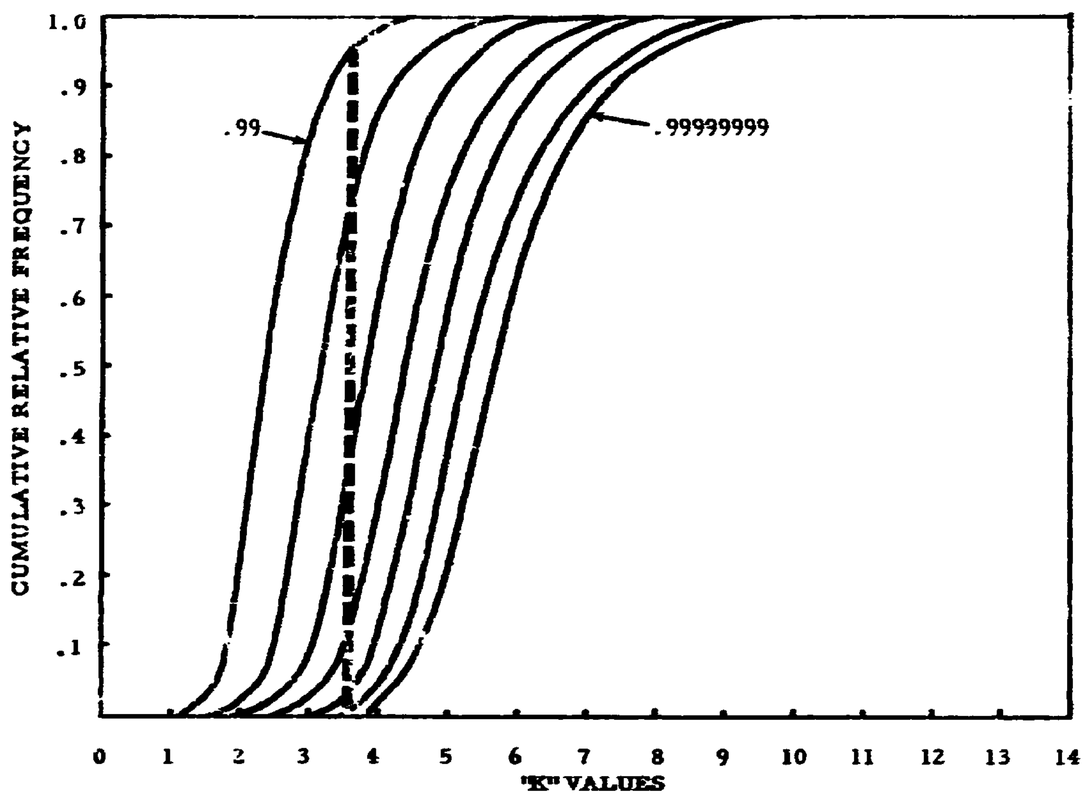


FIGURE 12. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 8$, $\sigma_{x_1} = \sigma_{x_2}$

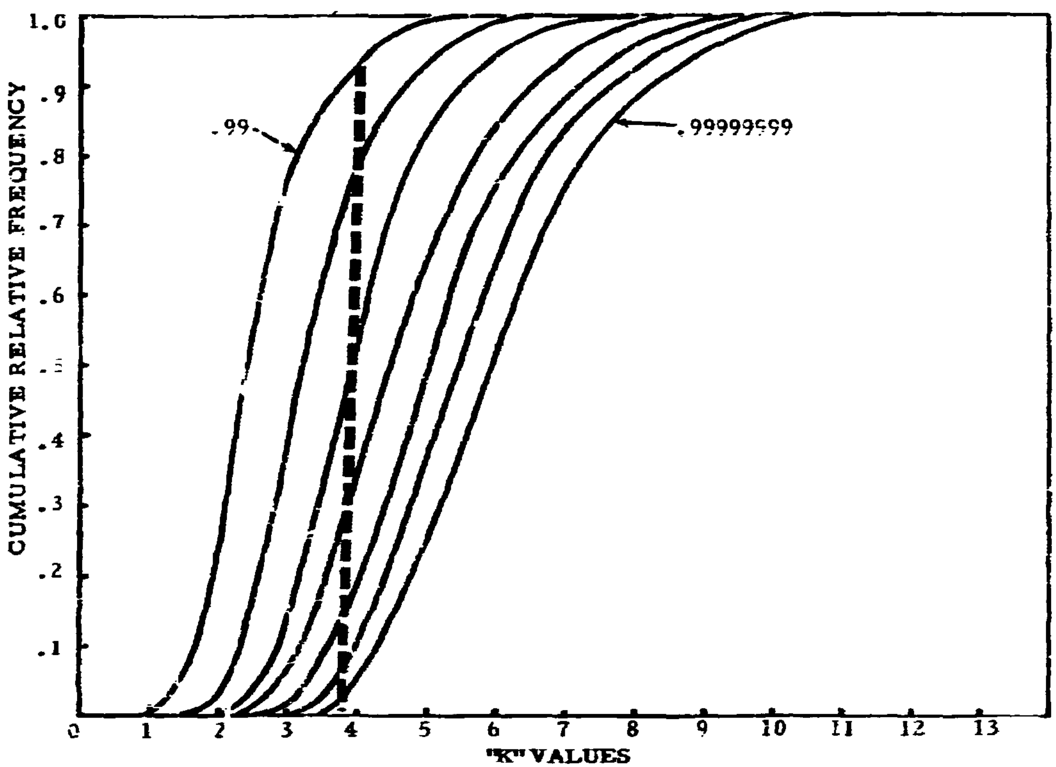


FIGURE 13. SAMPLING DIST. OF "K", $N_{x_1} = N_{x_2} = 5$, $\sigma_{x_1} = \sigma_{x_2}$

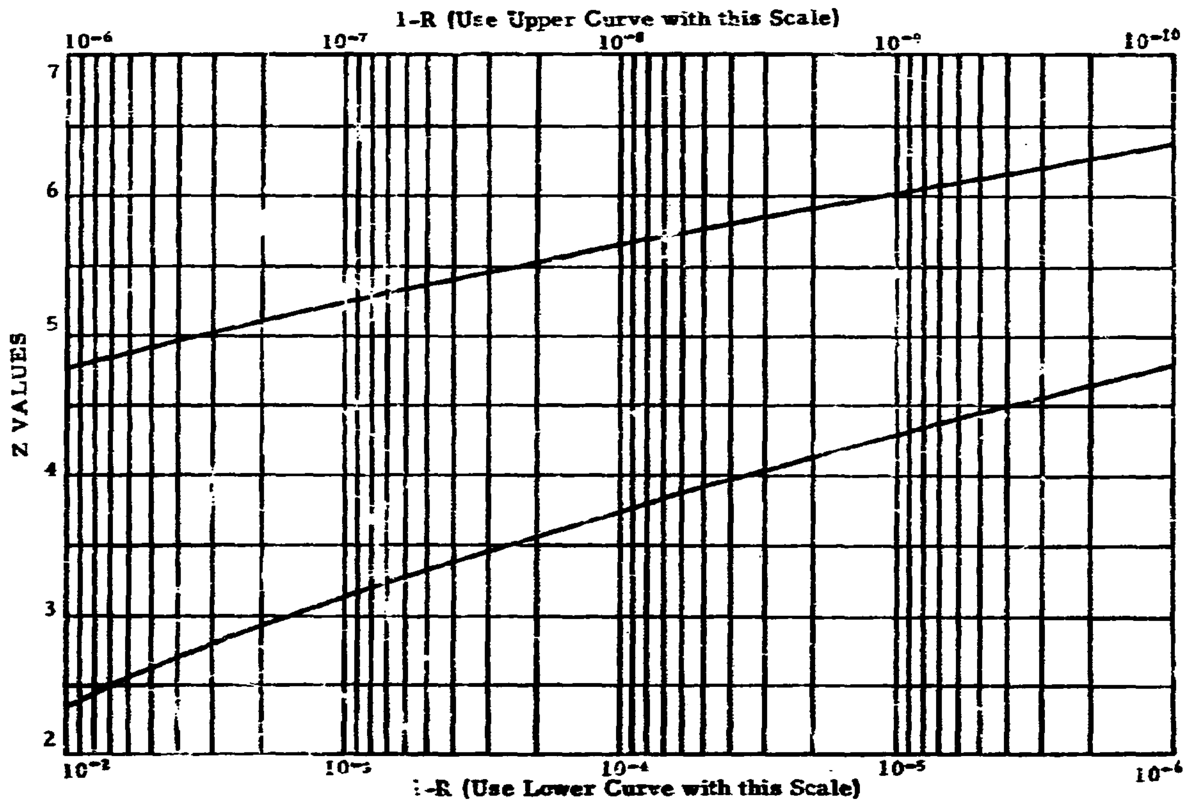


FIGURE 14. Graph of Z vs 1-R (Unreliability)

EXACT CONTINGENCY TABLE CALCULATIONS

Dorothy Berg, Morley Leyton*, and Clifford J. Maloney
Division of Biologics Standards
National Institutes of Health

1. INTRODUCTION. The great Dane is known throughout the civilized world for his famous rhetorical question: "To B or not to B ?" But he had a predecessor, equally famous, who 2000 years ago and more divided the universe not surprisingly into the even prior dichotomy A and not-A. These pioneering efforts were continued in many lands while Europe slept with the result that not 200 years ago science was able to divide the universe not only into A and not-A, but also simultaneously into B and not-B. We show this model in table 1, except that we have substituted the letters R and C (from row and column) for the Aristotelian labels A and B. Every element of the universe of discourse is presumed to be unambiguously assigned to one of the four cells in the body of the table.

We need not tell this audience that the probability of any given set of four cell frequencies among the set of all four-fold tables with specified row and column totals is given by the hypergeometric distribution

$$(1) \qquad P_i = \frac{c!\,C!\,r!\,R!}{f!\,g!\,h!\,k!\,N!}$$

Accordingly, reference (19) can be used to obtain significance levels for testing observed results. Tables for exact probabilities where

$$N = R + r \text{ and } r \leq R \leq 30$$

and less completely where

$$31 \leq R \leq 40, \quad r \leq R$$

---

* Now at the School of Public Hygiene, Johns Hopkins University

have recently appeared, (6). More exact and more complete tables are available in the UMT file of Mathematics of Computation at the David Taylor Model Basin, Washington 25, D. C., covering with (6) the range

$$r \leq R \leq 45$$

with table entries to four place accuracy. Tables for 2 X 3 contingency tables in the special case that all two cell marginal totals are equal, in the range 3 (1) 20 at levels of significance 0.05, 0.025, 0.01, 0.001 have recently appeared (3). Robertson (25) has published a computer program by the use of which exact probabilities can be calculated for any 2 X 2 contingency table. The computer program for handling the 2 X 3 case presumably is available from the authors (3).

Freeman and Halton (7) have published a computing method for calculating exact contingency probabilities for tables of any size. They treat also the calculation of interaction probabilities. Unfortunately the procedure is more theoretically than practically adequate.

It was early realised by Karl Pearson (23) that this probability could be approximated by the chi-square distribution provided the individual cell entries were not too small. Yates (32) supplied a correction which would make use of the chi-square approximations appropriate for smaller cell entries than would otherwise be possible, but pointed out that, where the cell entries were extremely unequal, even this correction would not be adequate until the minimum cell entry of the table should become quite large.

A long series of authors, among whom Haldane (9) was particularly prominent, have studied the question of developing adequate approximations to exact contingency probabilities when the chi-square is not useful. Others have considered various other aspects including studies of the power function of the test, choice of tests for interactions, and measures of the strength of the association between the factors when independence is not present. Particularly important from the standpoint of the exact calculation of contingency table probabilities is the paper of Kimball (16) where it is shown that the partition of contingency table chi-square can be reduced to the repeated calculation of 2 X 2 tables. A selected list of these papers is provided in the bibliography.

2. FIRST APPROACH. A table which arose in our work and which actually gave rise to this study is given in table 2. A set of data in which the reality of the three factor interaction was the point of interest is shown in table 3. The entries in table 2 are (some of them) too small to apply the simple chi-square test blindly; yet others are much too large for manual application of the Freeman-Halton procedure. The computer program for the 2 X 3 case (3) had not been announced (terms of availability were not stated) at that time, and a computer program for the 2 X 2 X 2 case has not yet appeared.

In an effort to devise a procedure for exact contingency table calculations which might be generalized to the general R X C case and which would be adapted to computer calculation, formula (1) for the probability of the particular set of cell entries which appear in table (1) was rearranged in the form

$$P_i = \binom{r}{f}\binom{R}{h} \div \binom{N}{C} \tag{2}$$

which brings out somewhat more clearly that the formula is a composite of binomial coefficients. The generating function can hence be shown in figure (1). [Figure 1 follows table 3 near the end of this article].
It is well known that the exponent on X plus the exponent on Y is a constant for all terms of a binomial expansion. In figure (1)

$$f + g = r \text{ and } h + k = R.$$

Further, if we multiply the two binomials term by term then for certain terms

$$f + h = c \text{ and } g + k = C$$

and the coefficient of this term is precisely the numerator of the probability of the table in question. For the particular case of a product of two binomials the product of one term from the first, by a unique term from the second yields the only pair whose exponents meet the marginal conditions, and hence is, by itself, (proportional to) the probability of that particular set of cell frequencies.

This suggests a very simple computing rule for calculating exact probabilities in a 2 X 2 contingency table. One writes out the numerators of the binomial expansion of the smallest marginal total of the table whose probability is desired. Let this smallest total be designated by r. Then the r + 1 terms of the expansion of this binomial are written in a row. In our notation R will then be greater than c. We next write out in a row beneath the preceding terms the r + 1 terms of $\binom{R}{h}$ in order, starting with c - r as the value for h under the right most element of the preceding expansion and proceedings to the left. These two rows are next multiplied element by corresponding element to give the r + 1 numerators of the complete probability distribution of the given 2 X 2 table. These terms are summed and each divided by that sum to yield the numerical value of the probabilities. The sum of the probabilities for those terms equal or more extreme than the observed table give the desired exact probability by which the independence of its main factors is tested.

The procedure may be illustrated by an example. Let r = 5, R = 10, c = 7, C = 8. We set out the complete expansion for the binomial of the fifth degree and under it the six terms from $\binom{10}{7}$ to $\binom{10}{2}$ inclusive.

$$\begin{matrix} 1 & 5 & 10 & 10 & 5 & 1 \\ \\ \frac{120}{120} & \frac{210}{1050} & \frac{252}{2520} & \frac{210}{2100} & \frac{120}{600} & \frac{45}{45} \end{matrix}$$

The respective probabilities for each table with a given value for f are:

f	0	1	2	3	4	5
p(f)	$\frac{120}{6435}$	$\frac{1050}{6435}$	$\frac{2520}{6435}$	$\frac{2100}{6435}$	$\frac{600}{6435}$	$\frac{45}{6435}$

For a two-tailed test, the probabilities for testing any observed table where the marginal totals are 5, 10, 7, and 8 is then a function of its entry in the cell defined by the 5 and 7 marginal totals only.

f	5	0	4	1	3	2
p(f)	$\frac{45}{6435}$	$\frac{165}{6435}$	$\frac{765}{6435}$	$\frac{1815}{6435}$	$\frac{3915}{6435}$	$\frac{6435}{6435}$

A one-tailed test for a 2 X 2 table would be exactly analogous. Exact probability calculations are, of course, not needed if the entries in every cell are large. If one marginal total is small and one large (on the same margin, say $r < < R$) the calculations can be greatly simplified by expressing each in terms of one of their number. For example, if

$$\binom{R}{c - r} = X_1$$

then

$$\binom{R}{c - r + 1} = \frac{C}{c - r + 1} X_1$$

$$\binom{R}{c - r + 2} = \frac{C - 1}{c - r + 2} \binom{R}{c - r + 1}$$

(3)

$$= \frac{C - 1}{c - r + 2} X_2$$

$$\vdots$$

$$\binom{R}{c} = \frac{C - r + 1}{c} X_r$$

As an example of these calculations, consider table (4). Here $r = 9$, $R = 681$, $0 \leq f \leq 9$. Since X_1 will eventually cancel out, set it equal to one. The calculations follow:

1	9	36	84	126	126	84	36	9	1

$$\frac{342}{340} X_8 \quad \frac{343}{339} X_7 \quad \frac{344}{338} X_6 \quad \frac{345}{337} X_5 \quad \frac{346}{336} X_4 \quad \frac{347}{335} X_3 \quad \frac{348}{334} X_2 \quad \frac{349}{333} X_1 \quad \frac{350}{332} \qquad 1$$

In systematically calculating the probabilities, a triple product is required; each element is multiplied by its right hand neighbor as well as by the number just above it in the preceding row. The term and cumulated frequencies as functions of f, the entry with smallest joint marginal totals can be expressed.

f	0	1	2	3	4	5	6	7	8	9
p(f)	0.002	0.019	0.075	0.171	0.251	0.244	0.157	0.065	0.015	0.002
P(f)	0.002	0.021	0.096	0.267	0.518	0.762	0.918	0.983	0.998	1

The procedure can be systematically extended to 2 X n tables, though the work involved quickly increases with n. The general formula for the probability of a specified 2 X n contingency table with fixed marginal totals, as given for example by Freeman and Halton, can be rearranged, as

$$(4) \qquad \frac{\binom{r}{f}\binom{s}{a}\binom{t}{b}\;.\;\;.\;\;.\;\binom{R}{h}}{\binom{N}{C}}$$

Where N is the table total; r, s, t, . . . R are the n marginal totals in the long direction; and C is one of the 2 marginal totals in the other direction. In this form, it is obvious that the calculations are of the same type as those required in the 2 X 2 case but that now the product of an appropriate term from each of n binomial expansions will be required. An extension of the generating function of (3) makes clear the terms from each binomial expansion admissable in each product.

For the 2 X 3 case, the calculations proceed as follows. Assume again that f is the smallest entry in the table and that r and c are the smallest marginal totals for rows and columns. (Of course, this is no restriction on the applicability of the method). Now f will range from 0 to r (assumed less than c). Each value of f will determine one cell of the 2 X 3 table exactly and thus reduce the remaining cells to the form of a 2 X 2 table with adjusted marginal totals (on the 3 cell margin).

Consider a 2 X 3 table* in which the 3 cell margins are 5 and 5 and the 2 cell margins are 2, 3, and 5. Then f can be 0, 1, and 2. We must calculate a 2 X 2 probability distribution for each of these values. Let now b play the role of f in the reduced table. It may take the values 0, 1, 2, and 3 (since the smallest remaining total is 3). Hence, in exact parallelism with our previous calculations, when f = 0, the qualifying marginal totals are 3 and 5, so that for f = 0, as b ranges from 0 to 3, h takes the values 5, 4, 3, and 2. In tabular form

b	0	1	2	3
f(b)	1	3	3	1
f(h)	1	5	10	10
	1	15	30	10

These are the numerators of the probabilities as they stand, since the factor from the binomial expansion of $\binom{r}{f}$ is unity. When, next, f takes the value 1, the marginal total 5 is reduced to 4 so that the calculation becomes

b	0	1	2	3
f(b)	1	3	3	1
f(h - 1)	5	10	10	5
	5	30	30	5

* Table 5.

Now, however, the numerator in the expansion of $\binom{r}{f}$ is 2, so that each of the above entries must be doubled to yield

10 60 60 10

Finally, when f = 2, the only remaining value, the contribution from $\binom{r}{f}$ is again unity and we have

h	0	1	2	3
f(h)	1	3	3	1
f(h - 2)	10	10	5	1
	10	30	15	1

These final products are proportional to the complete set of probabilities for admissable tables. They may be set out as in table (6). When divided by their total these entries are, of course, the probabilities themselves. Extension to 2 X 4 and higher tables, while increasingly laborious, is straightforward. Another option exists, however. From equation (4) it is obvious that the equation, being a running product of binomial coefficients can be broken at <u>any</u> point in the reduction process not merely after the first column, as was done above. Thus, a 2 X 4 could be "factored" into <u>two</u> 2 X 2 tables rather than successively into two 1 X 2 and one 2 X 2 table. A computer routine should be easy to devise to handle all cases up to 3 X 6. Further work along this line has been temporarily held up, however, since the approach seems less promising for tables, both dimensions of which exceed two.

3. <u>GENERAL R X C TABLE</u> The approach adopted in Section II, which treated several instances of the 2 X n contingency table exact probability calculations was to exploit the fact that the calculations could be reduced to the product of binomial coefficients. The general R X C table may be discussed in terms of a 3 X 3 table to simplify the notation. An assignment of symbols is shown in Table (7). The probability for any specific set of numerical values of the symbols is well known to be:

$$(5) \qquad P_i = \frac{\Pi C_j \cdot \Pi R_i}{a!\ b!\ c!\ d!\ e!\ f!\ g!\ h!\ i!\ N!}$$

In a manner exactly analogous to the reasoning used in the 2 X 2 case, this can be reduced to

$$(6) \qquad P_i = K \frac{R_1}{a!\,b!\,c!} \cdot \frac{R_2}{d!\,e!\,f!} \cdot \frac{R_3}{g!\,h!\,i!}$$

so that the generating function becomes

$$(7) \qquad G = \prod_i^3 (x + y + z)^{R_i}$$

This approach could serve as a straightforward method of calculating all probabilities for any 3 X 3 table, and could be extended in an obvious manner to the general R X C case.

The method suffers from two drawbacks. The number of possible terms grows rapidly, so that the amount of calculations grows even faster. Most of these tables will have extremely small probabilities, due to their very number. Hence, it seems wise to devise a method of approach in which only tables of appreciable probability will appear. This will be easy to insure provided (a) the various tables can be calculated in the order of decreasing probability and (b) a running sum of the probabilities of all tables calculated so far is maintained. By setting a criterion, say 0.99999, and terminating calculation when this total is reached, it is obvious that the last criterion is easy to ensure. Furthermore, even if a few tables of low probability are calculated, no harm is done other than the loss of machine time so that the method of ordering the calculations need not be perfect. A drawback, if perfect ordering is not achieved, is that additional machine time must be devoted to sorting to make the final output useful.

The procedure adopted in our current approach is composed of two portions. First, a single table, known to be of high probability is constructed. Further tables are constructed as perturbations in the entries of this table so selected that the resulting table is also likely to be of high probability. The starting information is the set of row and column totals. Hence the exact probabilities for an R X C contingency table is a table of R + C - 1 entry. Non-computing implications are discussed below. It is from these R + C given values that we construct three sets of tables needed in the calculations and in the output. First, as noted above, the initial complete table of known high probability is constructed as the only member of the first set. It must, however, always be held in memory whereas the members of the following sets may be sent to tape when calculations on them are completed. The calculation of this basic high probability table is done by first calculating the "expected" cell entries in the usual manner in applying the chi-square method:

$$X_{ij} = \frac{X_i X_j}{X} \tag{8}$$

These entries are in general not integers. The next step is to reduce all of them to integral form, but in such a manner that they sum to the given marginal totals. The second class of R X C tables consist of tables of adjustments by means of which other contingency tables with the specified row and column totals are formed by addition of the adjustment matrix to the original starting or basic matrix. For this purpose the rows and columns of the adjustment matrix itself must sum to zero. Individual entries may be small positive or negative integers or zero. The third class of matrices is composed of these sum matrices themselves, i.e. the set of admissable matrices other than the starting matrix itself. This is the set of matrices that must appear in the output if it is to be in most convenient form.

This approach has just now (July 1964) been programmed for the general 3 X 3 case so that, together with the approach of the preceding section, we should now be in a position to give exact contingency table probabilities for any 2 X n or 3 X 3 contingency table. An actual 3 X 3 table required 10 seconds of Honeywell 800 computer time so that, in this range at least, the approach looks hopeful. In the current program, even for the 3 X 3 case, we compute all admissable tables and have so far made no attempt to minimize memory requirements or computer time. As the program is extended to higher values of R and C this will become more necessary. On the other hand, as the dimensions of the table increase, the possibilities of adequate approximating formulae increase. The list of references at the end of this paper will indicate that no possibility of modifying the simple chi-square formula to extend the lower range of tabular cell frequencies has been overlooked. Unfortunately, no means exists for choosing between them in the absence of exact probabilities. One purpose, which we hope our studies will serve, is to permit such a definitive evaluation of the available approximating formulae.

The nub of the present computer technique is the formation of the set of adjustment matrices in the order of decreasing probability. This last is required if the calculations are to be truncated when all tables having appreciable probability have been calculated. Now the rows and columns of every adjustment matrix must sum to zero. This means that the sum of the positive adjustments must be balanced by the sum of the negative adjustments. This, in turn, means that the minimum number of adjustments in any row or column (if it contains an adjustment at all) must be two. On the other hand, any table formed by the minimal number of minimal adjustments (± 1) would seem to be itself a high probability table. To exploit these properties systematically it is intended to so arrange the rows and columns of the table that

$$(9) \qquad R_i \leq R_{i+1} \qquad\qquad C_j \leq C_{j+1}$$

for all i and j, which is obviously no restriction on the generality of the method. This can always be done as the row and column sums are the input to the calculation. Second, all possible corrections of ± 1 in at most two cells of any one row and column will be calculated first. Then all in at most four, then six and so on until every cell is adjusted up or down by one unit. If the probabilities of these tables do not sum to the cut off criterion, then all cases in which one cell is adjusted by two units will be similarly calculated systematically. During any one of these minor cycles it may prove worthwhile to introduce an individual table probability test equal say to 10^{-6} so that when a table with this low a probability is encountered the minor cycle (shifts in the row and or column of the adjustment) is terminated and the next adjustment formula applied. In the case of the 2 X 2 table, it is not difficult to distinguish between a one-tailed and a two-tailed test -- despite the fact that this is seldom done. In the general R X C table it becomes necessary to select among the possible definitions of a one-tailed test. However, whatever the definition, from a computer standpoint it reduces to deciding what terms to include in a summation of probabilities. For this reason we hope to print out probabilities for individual tables and study the sensitivity of changing the definition on acceptance levels. The results of this work will be reported subsequently.

Table 1

f	g	r
h	k	R
c	C	N

Symbol assignment for 2 x 2 contingency table.

Table 2

Population	I	II	III
Treated	23,973	101,637	56,613
Infected	3	49	25

Typical actual case of vaccine trial results in three treated cities.

Table 3

	Population			
	I		II	
Treatment	R	N	R	N
A	11	9	0	20
B	0	20	2	18

Actual 2 x 2 x 2 contingency table data

$$\frac{r!}{f!\,g!} X^f Y^g$$

$$\frac{R!}{h!\,k!} X^h Y^k$$

$$\frac{r!\,R!}{f!\,g!\,h!\,k!} X^c Y^C$$

Figure 1. Generating function for individual table frequencies of 2 x 2 contingency table.

Table 4

2	338	340
7	343	350
9	681	690

Hypothetical Table to illustrate exact probability calculation for as[redacted]etric contingency table. For explanation see text.

Table 5

t	b	h	5
i	j	k	5
2	3	5	10

Hypothetical 2 x 3 table. For explanation, see text,

Table 6

b \ f	0	1	2
0	1	10	10
1	15	60	30
2	30	60	15
3	10	10	1

Table 7

a	b	c	R_1
d	e	f	R_2
g	h	i	R_3
C_1	C_2	C_3	N

ACKNOWLEDGEMENTS

We are indebted to Drs. Marshall Hall, Jr. and Clyde Kramer for consultation, to Dr. Edward Batschelet for the procedure used for the computer generation of all admissable adjustment matrices in the 3 X 3 case, and to Mr. Andres Zellweger for programming the 3 X 3 method for the Honeywell 800. It is hoped that a discussion of the method will shortly appear in the computer literature.

REFERENCES

1. Bartlett, M. S. - "Contingency Table Interactions." Supp. J. of Roy. Stat. Soc., Vol. II, No. 2, 1935, pp. 248-252.

2. Bennett, B. M. and Hsu, P. "On the Power Function of the Exact Test for the 2 x 2 Contingency Table." Biometrika, Vol. 47, Parts 3 and 4, 1960, pp. 393-398.

3. Bennett, B. M. and Nakamura, E. "Tables for Testing Significance in a 2 x 3 Contingency Table." Technometrics, Vol. 5, No. 4, Nov. 63, pp. 501-511.

4. Bross, Irwin, D. J. and Kosten, Ethel L. - "Rapid Analysis of 2 x 2 Tables." J. Amer. Stat. Assoc., Vol. 52, No. 277, March 1957, pp. 18-28.

5. Darwin, J. - "On Corrections to the Chi-Square Distribution!" J. Roy. Stat. Soc., Vol. 20, 1958, pp. 387-392. Effect of discontinuity on chi-square significance levels.

6. Finney, D. J., et. al. "Tables for Testing Significance in a 2 x 2 Contingency Table." Cambridge University Press, 1963.

7. Freeman, G. H. and Halton, J. H. "Note on an Exact Treatment of Contingency Goodness of Fit and Other Problems of Significance." Biometrika 38, June 1951, pp. 141-149.

8. Good, I. J. - "Maximum Entropy for Hypothesis Formulation, Especially for Multi-Dimensional Contingency Tables." The Annals of Mathematical Statistics, Vol. 34, No. 3, Sept. 1963, pp. 911-934.

9. Haldane, J. B. S. - "The Precision of Observed Values of Small Frequencies." Biometrika, Vol. XXXV, Parts III and IV, Dec. 1948, pp. 297-300.

10 Harter, H. L. - "A New Table of the Percentage Points of the Chi-Square Distribution." 95th Central Regional Meeting, IMS, Madison, Wisconsin, June 14-15, 1963.

11. Hoel, Paul G. - "On the Chi-Square Distribution for Small Samples." Annals of Math. Stat., Vol. IX, No. 3, Sept. 1938, pp. 158-165. Second order approximation to the continuous distribution.

12. Irwin, J. O. - "Note on the Chi-Square Test for Goodnes of Fit." J. Roy. Stat. Soc., Part II, 1929, pp. 264-266.

13. Johnson, N. L. - "An Approximation to the Multinomial Distribution: Some Properties and Applications." Biometrika, Vol. 47, Parts 1 and 2, June 1960, pp. 93-102. Use of observed rather than expected frequencies in the denominator.

14. Kastenbaum, Marvin A. and Lamphiear, Donald E. - "Calculation of Chi-Square to Test the No Three Factor Interaction Hypothesis." Biometrics, Vol. 15, No. 1, March 1959, pp. 107-115.

15. Khamis, S. H. - "New Tables of the Chi-Square Cumulative Distribution." 34th Session, International Statistical Institute, Ottawa, Canada, August 21-29, 1963.

16. Kimball, A. W. - "Short-Cut Formulas for the Exact Partition of Chi-Square in Contingency Tables." Biometrics, Vol. 10, No. 4, Dec. 1954, pp. 452-458.

17. Kullback, S., Kupperman, M. and Ku, H. H. "Test for Contingency Tables and Markov Chains." Technometrics, Vol. 4, No. 4, Nov. 1962, pp. 573-608.

18. Lancaster, H. O. - "Complex Contingency Tables Treated by the Partition of Chi-Square." J. Roy. Stat. Soc., Series B (Methodological), Vol. XIII, No. 2, 1951, pp. 242-249.

19. Lieberman, Gerald, and Owen, Donald B. - "Tables of the Hypergeometric Probability Distribution." Stanford University Press, 1961. Can be used to obtain "exact" chi-square probabilities.

20. Mainland, Donald and Murray, L. M. - "Tables for Use in Fourfold Contingency Tests." Science, Vol. 116, Nov. 1952, pp. 591-594.

21. Maxwell, A. E. - "Analysing Qualitative Data." Metheun Monograph on Applied Probability and Statistics, 1961. Excellent intermediate level survey of subject.

22. Nass, C. A. G. - "The Chi-Square Test for Small Expectations in Contingency Tables; with Special Reference to Accidents and Absenteeism." Biometrika, Vol. 46, Parts 3 and 4, Dec. 1959, pp. 365-385. Second order approximation for R x C tables.

23. Pearson, Karl - "On the Theory of Contingency and its Relation to Association and Normal Correlation." 1904. Reprinted in Karl Pearson's early statistical papers.

24. Pearson, E. S. and Hartley, H. O. - "Biometrika Tables for Statisticians." Vol. 1, 2nd ed., pp. 188-193. Tables for 2 x 2 table significance tests.

25. Robertson, W. H. "Programming Fisher's Exact Method of Comparing Two Percentages." Technometrics, Vol. 2, No. 1, Feb. 1960, pp. 103-107.

26. Smith, H. Fairfield - "On Comparing Contingency Tables." The Philippine Statistician, Vol. VI, No. 2, June 1957, pp. 71-[illegible]

27. Taylor, William F. - "Some Peculiarities in Testing for the Difference between Proportions." Cleveland Meeting of Statistical Societies, Sept. 1963.

28. Watson, G. S. - "Some Recent Results in Chi-Square Goodness-of-fit Tests." Biometrics, Vol. 15, No. 3, Sept. 1959, pp. [illegible]-468.

29. Weintraub, Sol - "Cumulative Binomial Probabilities." J. Assoc. Computing Machinery, Vol. 9, No. 3, 1962, pp. 405-408.

30. Winsor, Charles P. - "Factorial Analysis of a Multiple Dichotomy." Human Biology, Vol. 20, No. 4, Dec. 1948, pp. 195-204. The Johns Hopkins Press. An approximation for extimation of sine of effects.

31. Wise, M. E. - "Multinomial Probabilities and the Chi-Square and X^2 Distributions." Biometrika, Vol. 50, Parts 1 and 2, pp. 145-154. Suggests arbitrary adjustment to the denominator of chi-square formula.

32. Yates, F. - "Contingency Tables Involving Small Numbers and the Chi-Square Test." Supp. J. Roy. Stat. Soc., Vol. 1, 1934, pp. 217-235.

MICROSPECTROSCOPY OF TISSUES

G. I. Lavin
Terminal Ballistic Laboratory, Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

ABSTRACT.

A consideration of non-destructive analytical procedures for the assessment of effects produced as the result of energy absorption by cellular (biological) systems.

NOTE.

This paper could not be released in time to appear in this publication. It is the hope of the editors of the Proceedings that this article will appear in the next issue of this series.

DESIGNS FOR THE SEQUENTIAL APPLICATION OF FACTORS[1]

Sidney Addelman
Research Triangle Institute

Some experimental plans that are appropriate when the levels of the factors of the 2^n factorial experiments are applied sequentially are presented. These plans permit the estimation of the main effects and two-factor interactions of the 2^n experiment and also the effects of the different possible orderings of the factors.

__INTRODUCTION__. When an experimental procedure involves the determination of the effects and interactions of several factors, a considerable advantage is gained if the experiment is designed so that the effect of changing any one factor can be assessed independently of the other factors. The effects of all factors on a characteristic of interest may be investigated simultaneously by varying each factor so that all, or a suitable subset of all, possible combinations of the factors are considered. An experiment in which this procedure is used is known as a factorial experiment. A factorial experiment consists of applying the chosen combinations of factors to the experimental units in a random manner and recording the yield of the response variable.

The standard statistical analyses of factorial experiments are appropriate when the factor levels of any treatment combination are simultaneously applied to an experimental unit. There are many situations when it is either impossible or impractical to apply the factors simultaneously to each experimental unit. In such situations the factor levels may be applied sequentially. If there is but one possible way to order the sequence of factors, or if the order in which the factors are applied does not affect the response, the experiment can be analyzed as if the factors were applied simultaneously. However, if the factors can be applied in several possible orders, and if the response due to the application of a combination of factors depends upon the particular order in which the factors are applied, the experiment should be planned so as to permit an evaluation of the order effects. The order in which factors are applied to experimental units may affect the responce due to differences in residual

[1]This research was supported by a grant from the U. S. Army Research Office-Durham.

or carry-over effects. A residual effect will occur when a factor is applied to an experimental unit before the effect of a previously applied factor has worn off. Thus, it is quite possible that the application of factor A before factor B will lead to a different response than the application of factor A after factor B.

This paper is concerned with the construction and analysis of experimental designs for the sequential application of the factors of a 2^n factorial arrangement. Throughout the paper it will be assumed that residual effects are permanent, i.e., if the effects of two factors are order dependent, the application of one factor will influence the effect of the second factor, no matter how many other factors are applied between them.

The experimental plans developed in this paper may be quite useful when the factors are environments. In laboratory simulation experiments it is often impossible or uneconomical to simultaneously apply all the environments to which an item may be subjected in actual use. If the factors are applied sequentially in a simulation experiment, important information on the residual effects can be gained by varying the order of the sequence. If there are alternative sequences of factors to which an item is subjected in actual use, one may determine the order of the sequence which is least detrimental to the item under test, by varying the order of the sequence in simulation experiments. These plans may also prove useful in experiments where items are treated with a sequence of chemicals to improve some characteristic (hardness, rust resistance, etc.) of the items.

PLANS WITH TWO FACTORS DEPENDENT ON ORDER. In this section we will consider experimental plans for the 2^n factorial arrangements where the order of only two of the n factors will affect the response. Some of the plans are constructed in such a manner that all factor effects of interest and the order effect are orthogonally estimable. For some plans the factor and order effects and the interactions are partially correlated. It is assumed that the low level of a factor (0) indicates that the factor was applied to the experimental unit at a level lower than the high level (1), i.e., the low level of a factor does not imply the absence of the factor. It is also assumed that if a factor has a residual effect on any subsequent factor, this effect will occur no matter which level of the factors are applied.

When only two factors are order dependent, the effect of ordering can be introduced as another factor. If the two factors that are order dependent are denoted by A and B, the ordering factor, X, may be introduced as follows. Let the 0 level of X indicate that factor A is applied before factor B, and let the 1 level of X indicate that factor A is applied after factor B. The experimental plan for a 2^n factorial arrangement where two of the n factors are order dependent may be deduced from a full or fractional replicate of a 2^{n+1} arrangement. Consider, for example, an experiment on three two-level factors A, B, and C, in which A and B are order dependent. Let X be introduced as the order factor. An experimental plan may be obtained by utilizing the 1/2 replicate of a 2^4 arrangement with the defining contrast $I = ABCX_0$. The treatment combinations for this plan are:

A	B	C	X
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Hence, the eight treatment combinations of the 2^3 arrangement in the order in which the factors are applied are: $a_0b_0c_0$, $b_0a_0c_1$, $b_1a_0c_0$, $a_0b_1c_1$, $b_0a_1c_0$, $a_1b_0c_1$, $a_1b_1c_0$, $b_1a_1c_1$. If the order factor, X, does not interact with A, B, or C, and all three-factor interactions are negligible, orthogonal estimates of μ, A, B, C, X, AB, AC, and BC may be obtained by a straightforward analysis.

If X interacts with A, B or C, a 3/4 replicate plan defined by the three defining contrasts

$$I = ABC_0 = CX_0 = ABX_0$$

$$I = ABC_0 = CX_1 = ABX_1$$

$$I = ABC_1 = CX_1 = ABX_0$$

will permit estimates of these interactions, where, for example, ABC_1 denotes the treatment combinations for which the sum of the levels of the factors A, B and C is equal to 1 modulo 2. The twelve treatment combinations in the order in which each factor is applied are:

$a_0b_0c_0$, $b_0a_0c_0$, $a_0b_1c_1$, $b_1a_0c_1$, $a_1b_0c_1$, $b_0a_1c_1$, $a_1b_1c_0$, $b_1a_1c_0$, $a_0b_0c_1$, $b_1a_0c_0$, $b_0a_1c_0$, $a_1b_1c_1$. The estimates of the effects of interest are correlated in four sets of three: $\hat{\mu}$, $\widehat{CX}$, $\widehat{ABC}$; $\hat{A}$, $\widehat{BC}$, $\widehat{BX}$; $\hat{B}$, $\widehat{AC}$, $\widehat{AX}$; $\hat{C}$, $\hat{X}$, $\widehat{AB}$. The method of analysis for 3/4 replicate plans is presented in Addelman [1].

<u>PLANS WITH SETS OF TWO FACTORS DEPENDENT ON ORDER</u>. When the response is affected by the order within each of k sets of two factors, the 2^k possible orderings may be accommodated in a 2^n factorial arrangement by introducing k ordering factors. The experimental plans will then consist of a fractional replicate of a 2^{n+k} factorial arrangement. Consider a 2^4 arrangement with factors denoted by A, B, C and D, where A and B are order dependent and C and D are order dependent. In this situation there are four possible orders, ABCD, BACD, ABDC, and BADC. Two ordering factors may be introduced as follows:

Let the 0 level of X denote the order AB,
the 1 level of X denote the order BA,
the 0 level of Y denote the order CD, and
the 1 level of Y denote the order DC.

If none of the factors A, B, C and D interacts with either X or Y, and if XY and all three-factor interactions are negligible, a 1/4 replicate of the 2^6 arrangement defined by $I = ABCX_0 = ABDY_0 = CDXY_0$ will yield orthogonal estimates of μ, A, B, C, D, X, Y, AB, AC, AD, BC, BD, and CD. The treatment combinations of the 2^4 experiment in their appropriate orders, as determined from the defining contrast, are:

$a_0b_0c_0d_0$	$b_1a_0d_0c_0$	$b_0a_1d_0c_0$	$a_1b_1c_0d_0$
$a_0b_0d_1c_0$	$b_1a_0c_0d_1$	$b_0a_1c_0d_1$	$a_1b_1d_1c_0$
$b_0a_0c_1d_0$	$a_0b_1d_0c_1$	$a_1b_0d_0c_1$	$b_1a_1c_1d_0$
$b_0a_0d_1c_1$	$a_0b_1c_1d_1$	$a_1b_0c_1d_1$	$b_1a_1d_1c_1$

If it is expected that the treatment factors interact with the order factors, a 3/8 replicate of the 2^6 arrangement will permit the estimation of these interactions, as well as the main effects and two-factor interactions that could be estimated with the 1/4 replicate of the 2^6 plan. For the irregular fraction plan to yield meaningful estimates, the interaction XY and all three-factor interactions must be negligible. The three 1/8 replicates of the 2^6 system are defined by

$$I = ABCDX_1 = ABY_0 = CDXY_1 = ACDY_0 = BXY_1 = BCD_0 = AX_1,$$

$$I = ABCDX_1 = ABY_0 = CDXY_1 = ACDY_1 = BXY_0 = BCD_1 = AX_0,$$

and $$I = ABCDX_1 = ABY_1 = CDXY_0 = ACDY_1 = BXY_0 = BCD_0 = AX_1,$$

respectively. The twenty-four treatment combinations of the 2^4 experiment in their appropriate orders of application are:

$b_0a_0c_0d_0$	$a_0b_0c_0d_1$	$b_0a_0d_0c_0$
$b_0a_0c_1d_1$	$a_0b_0c_1d_0$	$b_0a_0d_1c_1$
$b_1a_0d_1c_0$	$a_0b_1d_0c_0$	$b_1a_0c_0d_1$
$b_1a_0d_0c_1$	$a_0b_1d_1c_1$	$b_1a_0c_1d_0$
$a_1b_0c_0d_0$	$b_0a_1d_1c_0$	$a_1b_0c_0d_0$
$a_1b_0d_1c_1$	$b_0a_1d_0c_1$	$a_1b_0c_1d_1$
$a_1b_1c_0d_1$	$b_1a_1c_0d_0$	$a_1b_1d_1c_0$
$a_1b_1c_1d_0$	$b_1a_1c_1d_1$	$a_1b_1d_0c_1$

The least squares estimates of the estimable parameters are partially confounded in sets as follows:
$\hat{\mu}$ and $\widehat{AX}$; $\hat{A}$, $\hat{X}$, and $\widehat{BY}$; $\hat{B}$, $\widehat{AY}$, and $\widehat{CD}$; $\hat{C}$ and $\widehat{BD}$; $\hat{D}$ and $\widehat{BC}$; $\hat{Y}$, $\widehat{AB}$, and $\widehat{BX}$; $\widehat{AC}$, $\widehat{CX}$, and $\widehat{DY}$; $\widehat{CY}$, $\widehat{AD}$, and $\widehat{DX}$.

PLANS WHERE ONE FACTOR IS ORDER DEPENDENT WITH k OTHER FACTORS

When the response to a treatment combination of a 2^n factorial experiment is dependent on the position of one factor relative to k of the remaining $n-1$ factors, in the sequence of application, the 2^k possible orders may be accommodated by the introduction of k order factors. This type of plan can be illustrated by an experiment on four two-level factors, A, B, C, and D, where A is order dependent with B, C, and D. The three order factors X, Y, and Z may be defined as follows.

Let the 0 level of X denote the order AB,
the 1 level of X denote the order BA,

the 0 level of Y denote the order AC,
the 1 level of Y denote the order CA,
the 0 level of Z denote the order AD, and
the 1 level of Z denote the order DA.

The eight possible orderings are ABCD, BACD, CABD, BCAD, DABC, BDAC, CDAB, and BCDA.

The 2^4 arrangement with three order factors may be considered to be a 2^7 arrangement. A 1/8 replicate of the 2^7 arrangement defined by $I = ABXYZ_0 = BCXY_0 = ABCZ_0 = ABDX_0 = BDYZ_0 = ACDY_0 = CDXZ_0$ consists of the following 16 treatment combinations with each factor applied in the appropriate order.

$a_0b_0c_0d_0$	$b_0d_1a_0c_1$	$d_1a_1b_0c_0$	$b_0a_1c_1d_0$
$a_0b_1c_1d_1$	$b_1d_0a_0c_0$	$d_0a_1b_1c_1$	$b_1a_1c_0d_1$
$c_1d_0a_0b_0$	$b_0c_0a_0d_1$	$c_1a_1b_0d_1$	$b_0c_0d_0a_1$
$c_0d_1a_0b_1$	$b_1c_1a_0d_0$	$c_0a_1b_1d_0$	$b_1c_1d_1a_1$

If all interactions involving at least one order factor and all three-factor and higher order interactions are negligible, the above plan permits orthogonal estimation of μ, A, B, C, D, AB, AC, AD, BC, BD, CD, X, Y, and Z.

With the same assumptions as above, a plan may be constructed for the 2^5 arrangement on factors A, B, C, D, and E where factor A is order dependent with factors B, C, and D. This plan is derived from the treatment combinations of the 1/8 replicate of a 2^8 arrangement defined by $I = ABCDE_0 = ABCXY_0 = DEXY_0 = ADXZ_0 = BCEXZ_0 = BCDYZ_0 = AEYZ_0$, where X, Y, and Z are the order factors. If it is important to estimate all the two-factor interactions, a 3/16 replicate of the 2^8 would provide those estimates, some of which are partially confounded in sets of two or three. The generators of the defining contrast of the three 1/16 replicates which define such a plan are ABCDE, ABXYZ, ACX and BEY.

<u>PLANS WITH THREE FACTORS DEPENDENT ON ORDER</u>. In this section we consider plans for experiments in which all six permutations of the orderings of three factors may influence the response to a treatment combination. If the

three factors that are order dependent are denoted by A, B, and C, the six orders in which these three factors occur are ABC, ACB, BAC, BCA, CAB, and CBA. When there are more than three factors influencing the response, the factors other than those denoted by A, B, and C may occur any place in the sequence without affecting the response. The six orders may be accommodated in the experimental plan by introducing three order factors X, Y, and Z as follows:

Let the 0 level of X denote the order AB,
the 1 level of X denote the order BA,
the 0 level of Y denote the order AC,
the 1 level of Y denote the order CA,
the 0 level of Z denote the order BC, and
the 1 level of Z denote the order CB.

The order factors can be treated like three two-level factors. Of the $2^3 = 8$ treatment combinations of the three order factors, two combinations, 010 and 101, give impossible orders. Whenever these orders occur in the plan they must be changed. If the experiment consists of three factors A, B, and C which are all order dependent, an experimental plan may be obtained from the treatment combinations that comprise the 1/4 replicate of a 2^6 arrangement defined by $I = ABXY_0 = ACXZ_0 = BCYZ_0$, namely:

A	B	C	X	Y	Z
0	0	0	0	0	0
0	0	0	1	1	1
0	0	1	0	0	1
0	0	1	1	1	0
0	1	0	0	1	0*
0	1	0	1	0	1*
0	1	1	0	1	1
0	1	1	1	0	0
1	0	0	0	1	1
1	0	0	1	0	0
1	0	1	0	1	0*
1	0	1	1	0	1*
1	1	0	0	0	1
1	1	0	1	1	0
1	1	1	0	0	0
1	1	1	1	1	1

The four treatment combinations marked by an asterisk lead to impossible orderings and hence, the combination of the X, Y, Z levels must be changed in these treatment combinations. This can simply be accomplished by reversing the levels of Z in the four treatment combinations. The resulting treatment combinations of the 2^3 arrangement with appropriate orderings of the factor levels are:

$a_0b_0c_0$	$c_0a_0b_1$	$c_0a_1b_0$	$a_1c_0b_1$
$c_0b_0a_0$	$b_1a_0c_0$	$b_0a_1c_0$	$b_1c_0a_1$
$a_0c_1b_0$	$c_1a_0b_1$	$c_1a_1b_0$	$a_1b_1c_1$
$b_0c_1a_0$	$b_1a_0c_1$	$b_0a_1c_1$	$c_1b_1a_1$

This plan permits orthogonal estimates of all treatment effects, the order effects X and Y each being partially confounded with Z. The estimates of the treatment effects are simply:

$$\hat{A} = \frac{1}{8}[A], \quad \hat{B} = \frac{1}{8}[B], \quad \hat{C} = \frac{1}{8}[C], \quad \widehat{AB} = \frac{1}{8}[AB], \quad \widehat{AC} = \frac{1}{8}[AC],$$

$$\widehat{BC} = \frac{1}{8}[BC], \quad \widehat{ABC} = \frac{1}{8}[ABC],$$

where $[A]$ = (sum of treatment combinations with factor A at 1 level - sum of treatment combinations with factor A at 0 level),

$[AB]$ = (sum of treatment combinations whose expectations contain AB positively - sum of treatment combinations whose expectations contain AB negatively),

and so on.

The estimates of the order effects are given by

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{bmatrix} = \frac{1}{16} \cdot \begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} [X] \\ [Y] \\ [Z] \end{bmatrix}$$

The above estimates are valid if all interactions involving at least one order factor are negligible. If the two-factor interactions involving one order factor are not negligible, the sixteen treatment combinations of the previous plan are not adequate. When all two-factor interactions among the order factors and all three-factor and higher order interactions involving at least one order factor are negligible, a 3/8 replicate of the 2^6 arrangement will permit estimates of the remaining parameters. From the three 1/8 replicates defined by

$$I = XY_0 = YZ_0 = XZ_0 = ABC_1 = ABCXY_1 = ABCYZ_1 = ABCXZ_1 ,$$

$$I = XY_1 = YZ_0 = XZ_1 = ABC_0 = ABCXY_1 = ABCYZ_0 = ABCXZ_1 , \text{ and}$$

$$I = XY_0 = YZ_1 = XZ_1 = ABC_0 = ABCXY_0 = ABCYZ_1 = ABCXZ_1$$

the treatment combinations of the 2^3 arrangement in the appropriate sequences of application may be deduced. Some care was taken in selecting the three 1/8 replicates so that no treatment combination that resulted from the selection of the defining contrasts would involve impossible orderings. The 24 treatment combinations in their appropriate orders are:

$a_0b_0c_1$	$c_1b_1a_0$	$b_0a_0c_0$	$c_0a_0b_0$	$b_0c_0a_0$	$a_0c_0b_0$
$a_0b_1c_0$	$c_0b_1a_0$	$b_1a_0c_1$	$c_1a_0b_1$	$b_1c_1a_0$	$a_0c_1b_1$
$a_1b_0c_0$	$c_0b_0a_1$	$b_0a_1c_1$	$c_1a_1b_0$	$b_0c_1a_1$	$a_1c_1b_0$
$a_1b_1c_1$	$c_1b_1a_1$	$b_1a_1c_0$	$c_0a_1b_1$	$b_1c_0a_1$	$a_1c_0b_1$

The estimates are given by

$$\begin{bmatrix} 2\hat{\mu} \\ \widehat{ABC} \end{bmatrix} = \frac{1}{32} \cdot \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} T \\ [ABC] \end{bmatrix} ,$$ where T denotes the sum of all treatment combinations.

$$\begin{bmatrix} \hat{A} \\ \widehat{BC} \end{bmatrix} = \frac{1}{32} \cdot \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} [A] \\ [BC] \end{bmatrix} , \begin{bmatrix} \hat{B} \\ \widehat{AC} \end{bmatrix} = \frac{1}{32} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} [B] \\ [AC] \end{bmatrix} ,$$

$$\begin{bmatrix} \hat{C} \\ \widehat{AB} \end{bmatrix} = \frac{1}{32} \cdot \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} [C] \\ [AB] \end{bmatrix} ,$$

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} [X] \\ [Y] \\ [Z] \end{bmatrix}, \quad \begin{bmatrix} \widehat{AX} \\ \widehat{AY} \\ \widehat{AZ} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} [AX] \\ [AY] \\ [AZ] \end{bmatrix}$$

$$\begin{bmatrix} \widehat{BX} \\ \widehat{BY} \\ \widehat{BZ} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} [BX] \\ [BY] \\ [BZ] \end{bmatrix}, \quad \begin{bmatrix} \widehat{CX} \\ \widehat{CY} \\ \widehat{CZ} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} [CX] \\ [CY] \\ [CZ] \end{bmatrix}.$$

REFERENCE

Addelman, S. 1961. Irregular fractions of the 2^n factorial experiments, Technometrics 12:479-496.

2^p FACTORIAL EXPERIMENTS WITH THE FACTORS APPLIED SEQUENTIALLY

R. R. Prairie and W. J. Zimmer
Sandia Corporation

ABSTRACT. In the application of factorial experiments in industry, it is often necessary to apply the factors sequentially. For 2^p experiments in which (a) multiple testing on each experimental unit is possible and not degrading, (b) the effects of the factors are permanent, and (c) the low level of a factor is the absence of that factor, a design and an analysis are presented which allow the estimation and testing of order effects as well as the usual main and interaction effects of the factors.

INTRODUCTION. In the usual application of factorial experiments, the factor levels are applied simultaneously to the units and a response is recorded from each unit. However, when the factors are environments (e.g., vibration and mechanical shock), it is often impossible to apply the factor levels simultaneously. In this situation, the factor levels are applied in sequence.

When the factor levels must be applied sequentially, the order of application becomes a matter of importance. It is possible that the final performance of a unit that has been subjected to two factors may be very sensitive to the order in which the factors are applied.

In this paper 2^p experiments are considered where:

1. Each unit may be tested (p + 1) times (where p is the number of factors) without the testing itself having a degrading effect on the unit,
2. The high level of the factor is the application of the factor and the low level of the factor means the factor is not applied, and
3. The effects of the factors are permanent.

Thus, the magnitude of the response from a specific test will be affected by the factors applied to the unit up through that test plus the specific order of application of those factors. A design [called the Factor Sequencing Design (FSD)] and an analysis are presented, which allow the estimation of these order effects as well as the usual main and interaction effects of the factors.

DESIGN AND MODEL. Each factor V_ℓ [ℓ = 1, ..., p] has a high level [V_ℓ has been applied] and a low level [V_ℓ has not been applied. Thus, when a unit has the high level of V_ℓ applied to it, we shall say that it has received the V_ℓ factor or that the V_ℓ factor has been applied.] Since the effects of the factors are assumed to be permanent, a unit which has received the V_ℓ factor at some point must be considered from that point on as having the high level of V_ℓ. Further, each unit will eventually receive, in some order, each of the p factors.*

* The remainder of this paper has been reproduced photographically from the author's copy.

The type of design considered is one where r units are subjected to each of the p! possible orders of the p factors: $V_1, V_2, \ldots, V_p$. The p factors are applied in sequence and each unit is tested (p + 1) times, once prior to the application of any factor and once following the application of each of the p factors. Thus a notation for the response on the j^{th} test ($j = 1, \ldots, p + 1$) of the i^{th} unit ($i = 1, \ldots, r$) within or, in other words, eventually subjected to the k^{th} p-way order ($k = 1, \ldots, p!$) is:

$$X_{ijk}\left(\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}\right).$$

where $\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}$ indicate the j-1 factors which have been applied up through the j^{th} test and the order of application. If $j = 1$, i.e., no factors have been applied, the response is symbolized by $X_{i1k}(1)$. Although this notation is redundant, it was selected because of its convenience when describing the analysis. An example of the design is given for a 2^3 experiment in Table I.

TABLE I

Design of a 2^3 FSD in Six Different Orders

3-Way Order	Factor	Test 1	Factor	Test 2	Factor	Test 3	Factor	Test 4
Order 1:								
Unit 1	None	$X_{111}(1)$	V_1	$X_{121}(a_1)$	V_2	$X_{131}(a_1a_2)$	V_3	$X_{141}(a_1a_2a_3)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Unit r	None	$X_{r11}(1)$	V_1	$X_{r21}(a_1)$	V_2	$X_{r31}(a_1a_2)$	V_3	$X_{r41}(a_1a_2a_3)$
Order 2:								
Unit 1	None	$X_{112}(1)$	V_1	$X_{122}(a_1)$	V_3	$X_{132}(a_1a_3)$	V_2	$X_{142}(a_1a_3a_2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Unit r	None	$X_{r12}(1)$	V_1	$X_{r22}(a_1)$	V_3	$X_{r32}(a_1a_3)$	V_2	$X_{r42}(a_1a_3a_2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Order 6:								
Unit 1	None	$X_{116}(1)$	V_3	$X_{126}(a_3)$	V_2	$X_{136}(a_3a_2)$	V_1	$X_{146}(a_3a_2a_1)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Unit r	None	$X_{r16}(1)$	V_3	$X_{r26}(a_3)$	V_2	$X_{r36}(a_3a_2)$	V_1	$X_{r46}(a_3a_2a_1)$

The model that is used to represent a test for the 2^p FSD is

$$X_{ijk}\left(\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}\right) = m + u_{i(k)} + \xi_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}} + e_{ijk} \qquad (1)$$

$i = 1, 2, \ldots, r$

$j = 1, 2, \ldots, p + 1$

$k = 1, 2, \ldots, p!$

where

$X_{ijk}\left(\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}\right)$ = the response obtained on the j^{th} test from the i^{th} unit eventually subjected to the k^{th} p-way order of the factors where $\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}$ is a particular ordering of the $j-1$ factors applied up through the j^{th} test.

m = general mean.

$u_{i(k)}$ = effect of i^{th} unit eventually subjected to k^{th} p-way order of the p factors.

$\xi_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}}$ = effect of the application of $j-1$ factors in the specific order $\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}$; for $j = 1$, the symbol used is $\xi_{(1)}$.

e_{ijk} = random error associated with j^{th} test on the i^{th} unit eventually subjected to the k^{th} order of p factors.

For the 2^3 FSD some specific examples of the form of the model are:

$$
\begin{aligned}
X_{111}(1) &= m + u_{1(1)} + \xi_{(1)} + e_{111} \\
X_{121}(a_1) &= m + u_{1(1)} + \xi_{a_1} + e_{121} \\
X_{131}(a_1a_2) &= m + u_{1(1)} + \xi_{a_1a_2} + e_{131} \\
X_{141}(a_1a_2a_3) &= m + u_{1(1)} + \xi_{a_1a_2a_3} + e_{141} \\
&\;\vdots \\
X_{113}(1) &= m + u_{1(3)} + \xi_{(1)} + e_{113} \\
X_{123}(a_2) &= m + u_{1(3)} + \xi_{a_2} + e_{123} \\
X_{133}(a_2a_1) &= m + u_{1(3)} + \xi_{a_2a_1} + e_{133} \\
X_{143}(a_2a_1a_3) &= m + u_{1(3)} + \xi_{a_2a_1a_3} + e_{143} \\
&\;\vdots \\
X_{116}(1) &= m + u_{1(6)} + \xi_{(1)} + e_{116} \\
X_{126}(a_3) &= m + u_{1(6)} + \xi_{a_3} + e_{126} \\
X_{136}(a_3a_2) &= m + u_{1(6)} + \xi_{a_3a_2} + e_{136} \\
X_{146}(a_3a_2a_1) &= m + u_{1(6)} + \xi_{a_3a_2a_1} + e_{146}
\end{aligned}
$$

It will be assumed that

$$u_{i(k)} \sim NID(0, \sigma_u^2), \; e_{ijk} \sim NID(0, \sigma_e^2)$$

and that each $\xi_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}}$ is the sum of two parameters, one associated with the factors involved and one with the order of the factors, i.e.,

$$\xi_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}} = \eta_{\alpha_{s_1'}, \ldots, \alpha_{s_{j-1}'}} + \delta_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}}$$

where $\eta_{\alpha_{s_1'}, \ldots, \alpha_{s_{j-1}'}}$ is the parameter associated with the j-1 factors ($s_1', \ldots, s_{j-1}'$ are the $s_1, \ldots, s_{j-1}$ ordered from the smallest through the largest) and $\delta_{\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}}$ is the parameter associated with the specific order of application of the j-1 factors.

It is also assumed that the ξ's sum to zero.

From Equation 1 it can be seen that

$$\operatorname{cov}\left[X_{ijk}\left(\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}\right), X_{i'j'k'}\left(\alpha_{s_1}, \ldots, \alpha_{s_{j-1}}\right)\right]$$

$$= \begin{cases} \sigma_e^2 + \sigma_u^2, & i = i',\ j = j',\ k = k' \\ \sigma_u^2, & i = i',\ j \neq j',\ k = k' \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

Analysis

Since the model assumed is linear, the proposed analysis is based on the method of least squares. From Equation 1 it is apparent that $E(X) = A^*\xi^*$ where X is the $r(p + 1)p! \times 1$ column vector of observations, A^* is the $r(p + 1)p! \times \left[1 + \sum_{i=0}^{p} \frac{p!}{(p - i)!}\right]$ design matrix and ξ^* is the $\left[1 + \sum_{i=0}^{p} \frac{p!}{(p - i)!}\right] \times 1$ column vector of parameters. Now the rank of A^* is $\sum_{i=0}^{p} \frac{p!}{(p-i)!}$ which is one less than the number of parameters. Since the number of estimable parameters is equal to the rank of A^*, a reparameterization must be performed. To accomplish this a new vector ξ^+ is defined. Then $E(X) = A^* M \xi^+$, where

$$\xi^{+'} = \left[m + \xi_{(1)}, \xi_{a_1} - \xi_{(1)}, \ldots, \xi_{a_1 a_2}, \ldots, a_p - \xi_{(1)}, \ldots, \xi_{a_p a_{p-1}}, \ldots, a_1 - \xi_{(1)}\right],$$

is of order $\sum_{i=0}^{p} \frac{p!}{(p-i)!}$, M is the matrix of reparameterization, and $A^* M$ has rank $\sum_{i=0}^{p} \frac{p!}{(p-i)!}$.

It is not yet possible to obtain the best estimate of ξ^+ from

$$\hat{\xi}^+ = \left[(A^* M)'(A^* M)\right]^{-1}(A^* M)'X$$

because the Gauss-Markov (G-M) theorem is not applicable, as the X's are not independently distributed [see Equation 2].

In matrix notation Equation 2 is written:

$$\text{Var}(X) = \sigma_e^2 I + \sigma_u^2 \operatorname{diag}\{U\} \tag{3}$$

where I is the identity matrix of order $(p + 1)rp!$ and diag $\{U\}$ is a square diagonal matrix of order $(p + 1)rp!$ for which each diagonal element is the submatrix U, a square matrix of order $(p + 1)$ all of whose elements are 1.

To satisfy the conditions of the G-M theorem, a linear transformation,

$$Z = BX$$

can be found such that $\text{var}(Z) = \sigma_e^2 I$, where

$$Z' = \left[Z_{111}, Z_{121}, \ldots, Z_{1p1}, Z_{211}, \ldots, Z_{2p1}, \ldots, Z_{r1p!}, \ldots, Z_{rpp!}\right]$$

Since

$$\text{var}(Z) = B[\text{var}(X)]B' = B\left[\sigma_e^2 I + \sigma_U^2 \text{ diag } \{U\}\right]B' , \tag{4}$$

a matrix, B, which possesses the properties $BB' = I$ and $B\left[\text{diag } \{U\}\right]B' = 0$, will yield $\text{var}(Z) = \sigma_e^2 I$. A particular matrix that yields the desired result is a $rpp! \times r(p + 1)p!$ matrix B where

$$B = \overbrace{\begin{bmatrix} p \text{ rows} \{ \overbrace{[\beta]}^{(p+1) \text{ columns}} & [0] & \cdots & & [0] \\ p \text{ rows} \{ \overbrace{[0]}^{(p+1) \text{ columns}} & [\beta] & \cdots & & [0] \\ [0] & [0] & [\beta] & \cdots & [0] \\ \vdots & \vdots & & \ddots & \\ [0] & [0] & & & [\beta] \end{bmatrix}}^{r(p+1)p! \text{ columns}} \Bigg\} \; rpp! \text{ rows}$$

and where

$$\beta = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{p(p+1)}} & \frac{1}{\sqrt{p(p+1)}} & \frac{1}{\sqrt{p(p+1)}} & \cdots & -\frac{p}{\sqrt{p(p+1)}} \end{bmatrix}$$

Then

$$E\ (Z) = B[E\ (X)] = BA^{*}M\xi^{+} ,$$

and since the first column of $BA^{*}M$ is all zeros, set

$$E\ (Z) = A\xi$$

where A is a $\left[rpp! \times \sum_{i=1}^{p} \frac{p!}{(p-i)!} \right]$ matrix and is $BA^{*}M$ with the first column, which is a column of zeros, deleted and ξ is the parameter vector ξ^{+} reduced by deleting the first element.

Since $Z = PX$ is a linear transformation of normally distributed variables, the Z's are normally distributed. As $\text{var}(Z) = \sigma_e^2 I$, the Z's satisfy the conditions of the G-M theorem. Thus, the best linear unbiased estimate of ξ is

$$\hat{\xi} = (A'A)^{-1}A'Z \ . \tag{5}$$

An analysis of variance associated with Equation 5 can be given. This analysis of variance appears in Table II.

TABLE II

Preliminary Analysis of Variance Including All Parameters

Source of Variation (SV)	Degrees of Freedom (DF)	Sum of Squares (SS)
Treatments (factors and orders)	$\sum_{i=1}^{p} \frac{p!}{(p-i)!}$	$SS_{V+S} = Z'\left(A(A'A)^{-1}A'\right)Z$
Error	$\sum_{i=1}^{p-2} \frac{p!}{(p-i)!}\left[r(p-i)! - 1\right] + 2(r-1)p!$	$SS_E = Z'\left(I - A(A'A)^{-1}A'\right)Z$
Total	$rpp!$	$SS_T = Z'Z$

By neglecting order, a new parameter vector is generated. Denote this new vector as η. For example, with a 2^3 FSD

$$\eta' = \Big[\eta_{\alpha_1} - \eta_{(1)}, \eta_{\alpha_2} - \eta_{(1)}, \eta_{\alpha_3} - \eta_{(1)}, \eta_{\alpha_1\alpha_2} - \eta_{(1)}$$

$$\eta_{\alpha_1\alpha_3} - \eta_{(1)}, \eta_{\alpha_2\alpha_3} - \eta_{(1)}, \eta_{\alpha_1\alpha_2\alpha_3} - \eta_{(1)}\Big] \ .$$

The best estimate of η is

$$\hat{\eta} = (A_1'A_1)^{-1}A_1'Z , \tag{6}$$

where A_1 is $\left[rpp! \times 2^p - 1\right]$ matrix and is $BA_1^*M_1$ with the first column deleted, where A_1^* is the design matrix associated with η and M_1 is the matrix of reparameterization associated with η .

The corresponding analysis of variance is given in Table III.

TABLE III

Preliminary Analysis of Variance
Disregarding Order Parameters

SV	DF	SS
Factorial effects	$2^p - 1 = \sum_{i=1}^{p} \binom{p}{i}$	$Z'\left[A_1(A_1'A_1)^{-1}A_1'\right]Z$
Residual	$rpp! - \sum_{i=1}^{p} \binom{p}{i}$	$Z'\left[I - A_1(A_1'A_1)^{-1}A_1'\right]Z$
Total	$rpp!$	$Z'Z$

Then, by means of the principle of conditional error, the sums of squares, due to the factorial effects, the order effects, and the error, can be obtained as shown in Table IV.

TABLE XV

Analysis of Variance for Factorial and Order Effects

SV	DF	SS
Factors	$2^p - 1$	$Z'\left[A_1(A_1'A_1)^{-1}A_1'\right]Z$
Orders	$\sum_{i=2}^{p} \binom{p}{i}(i! - 1)$	$Z'\left[A(A'A)^{-1}A' - A_1(A_1'A_1)^{-1}A_1'\right]Z$
Error	$\sum_{i=1}^{p-2} \frac{p!}{(p-i)!}\left[r(p-i)! - 1\right] + 2(r-1)p!$	$Z'\left[I - A(A'A)^{-1}A'\right]Z$
Total	$ryp!$	$Z'Z$

It can be shown that the sums of squares are independently distributed as $\chi^2\sigma^2$, and the sum of squares due to order is clean of factorial effects. Thus, the significance of the order effects can be tested by the usual F-test and if order effects are negligible the factorial effects can also be tested by the F-test.

In the case where order effect is found negligible, any given factorial effect, say E_v, would be estimated by

$$\hat{E}_v = d'\hat{\eta} = C'Z ,$$

where $\hat{\eta}$ is given by Equation 6 and d' is the vector which gives the appropriate linear combination among the parameters. The sum of squares corresponding to that effect is $(C'Z)^2/C'C$.

In the case where the order of application of the factors has a significant effect, one would presumably investigate different contrasts among parameters of ξ, as estimated by Equation 5, to determine which are contributing to the order effect. This can be done by calculating expressions of the form $g\hat{\xi}$, where g is a vector defining a specific order contrast of interest. If order is significant, one should be careful about inferences concerning factorial effects.

It is interesting to note that when $p \geq 4$, the number of units needed for the FSD ($p!$, or 24 when $p = 4$) is greater than the number (2^p, or 16 when $p = 4$) needed for the usual 2^p factorial. In this case, however, the advantage of the FSD over the usual factorial is the information obtained on the order effects, and the increased efficiency of estimation due to the removal of unit variation from the error. Furthermore, it is believed that a fractional replicate of the FSD, using fewer units than the usual factorial, would still provide estimates of both the factorial effects and the order effects. This aspect of the FSD is presently being investigated.

Also, for the case where there is no order effect, it is possible to construct a design which requires only p units and for which it is possible to estimate all $2^p - 1$ factorial effects. This design is similar to a Latin square, where each unit is tested $p + 1$ times.

The general FSD technique described above will now be illustrated by two examples of a 2^3 experiment.

Example 1 of a 2^3 FSD

Consider a 2^3 FSD whose design and results are shown in Table V.

TABLE V

Results From One Replication of a 2^3 FSD

3-Way Order	Factor	Test 1	Factor	Test 2	Factor	Test 3	Factor	Test 4
Order 1								
Unit 1	None	56.258	V_1	56.579	V_2	52.661	V_3	51.315
Order 2								
Unit 1	None	55.500	V_1	57.461	V_3	57.475	V_2	50.396
Order 3								
Unit 1	None	58.515	V_2	56.323	V_1	62.023	V_3	61.673
Order 4								
Unit 1	None	56.583	V_2	56.924	V_3	56.111	V_1	62.085
Order 5								
Unit 1	None	54.217	V_3	55.914	V_1	53.974	V_2	49.734
Order 6								
Unit 1	None	56.034	V_3	57.895	V_2	55.440	V_1	62.863

The model used to represent any response is given by Equation 1:

$$X_{ijk}\left(\alpha_{\ell_1}, \ldots, \alpha_{\ell_{j-1}}\right) = m + u_{i(k)} + \xi_{\alpha_{\ell_1}, \ldots, \alpha_{\ell_{j-1}}} + e_{ijk}$$

where $i = 1$; $j = 1, 2, 3, 4$; $k = 1, 2, 3, 4, 5, 6$.

Now,

$$Z = BX = \begin{bmatrix} \frac{1}{\sqrt{2}}(56.258 - 56.579) \\ \frac{1}{\sqrt{6}}(56.258 + 56.579 - 2[52.661]) \\ \frac{1}{\sqrt{12}}(56.258 + 56.579 + 52.661 - 3[51.315]) \\ \vdots \\ \frac{1}{\sqrt{12}}(56.034 + 57.895 + 55.440 - 3[62.863]) \end{bmatrix}$$

and

$$Z' = [-0.227, 3.068, 3.335, -1.387, -0.812, 5.556, 1.550, -3.739, -2.355, -0.241, 0.524, -4.803, -1.200, 0.891, 4.285, -1.316, 1.245, -5.548] .$$

Then the best estimate of:

$$\xi' = \left[\xi_{\alpha_1} - \xi_{(1)}, \xi_{\alpha_2} - \xi_{(1)}, \ldots, \xi_{\alpha_3\alpha_1\alpha_2} - \xi_{(1)}, \xi_{\alpha_3\alpha_2\alpha_1} - \xi_{(1)}\right]$$

is [see Equation 5]

$$\hat{\xi} = (A'A)^{-1}A'Z \tag{7}$$

and

$$\hat{\gamma}' = [1.141, -0.985, 1.779, -3.187, 4.141, 1.565, -0.202,$$

$$-1.105, -0.635, -4.533, -5.514, 3.791, 4.869, -4.422, 6.788] .$$

An analysis of variance associated with Equation 7 and corresponding to that of Table II is given in Table VI.

TABLE VI

Preliminary Analysis of Variance for a 2^3 FSD Including All Parameters

SV	DF	SS
Treatments (factors and orders)	15	$Z'[A(A'A)^{-1}A']Z = 151.50$
Error	3	$Z'[I - A(A'A)^{-1}A']Z = 2.28$
Total	18	$Z'Z = 154.18$

By neglecting order, an analysis of variance corresponding to that of Table III can be carried out, as shown in Table VII.

TABLE VII

Preliminary Analysis of Variance for a 2^3 FSD Disregarding Order Parameters

SV	DF	SS
Factorial Effects	7	$Z'\left[A_1(A_1'A_1)^{-1}A_1'\right]Z = 56.98$
Residual	11	$Z'\left[I - A_1(A_1'A_1)^{-1}A_1'\right]Z = 97.20$
Total	18	$Z'Z = 154.18$

In Table VIII the analysis of variance corresponding to Table IV for factorial and order effects is given.

TABLE VIII

Analysis of Variance for Factorial and Order Effects from 2^3 FSD

SV	DF	SS
Factors	7	$Z'\left[A_1(A_1'A_1)^{-1}A_1'\right]Z = 56.99$
Orders	8	$Z'\left[A(A'A)^{-1}A' - A_1(A_1'A_1)^{-1}A_1'\right]Z = 94.91$
Error	3	$Z'\left[I - A(A'A)^{-1}A'\right] = 2.28$
Total	18	$Z'Z = 154.18$

The test of significance for order effects is

$$F_{8,3} = \frac{3(94.91)}{8(2.28)} = 15.60,$$

which is significant at the 0.025 level.

If order effects are judged significant, one could make comparisons among the order parameters as a further step in the analysis. One possible set of comparisons would yield information on V_1V_2 versus V_2V_1, V_1V_3 versus V_3V_1, and V_2V_3 versus V_3V_2. These effects could be estimated by $c_1'\hat{\ell}$, $c_2'\hat{\ell}$, and $c_3'\hat{\ell}$, respectively, where

$$C = \begin{bmatrix} c_1' \\ c_2' \\ c_3' \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & 0 & 1 & 0 & -1 \end{bmatrix}$$

The computed value of $C\hat{\ell}$ is

$$C\hat{\ell} = \begin{bmatrix} -8.954 \\ -0.134 \\ 0.469 \end{bmatrix}$$

with estimated variance-covariance matrix

$$\widehat{\text{Var}}\,(C\hat{\ell}) = C(A'A)^{-1}C'\hat{\sigma}_e^2 = \begin{bmatrix} 1.097 & -0.253 & -0.253 \\ & 1.097 & -0.253 \\ & & 1.097 \end{bmatrix}$$

Tests of significance for these particular order effects could be performed by the "F" test or, equivalently, by the usual "t" test where.

$$t_{(n_e)} = \frac{\text{estimated effect}}{\text{standard error of estimate}}$$

and n_e is the number of degrees of freedom associated with error. The tests are

$$V_1V_2 \text{ versus } V_2V_1 : t_{(3)} = \frac{-8.954}{\sqrt{1.097}} = -8.552$$

$$V_1V_3 \text{ versus } V_3V_1 : t_{(3)} = \frac{-0.134}{\sqrt{1.097}} = -0.127$$

$$V_2V_3 \text{ versus } V_3V_2 : t_{(3)} = \frac{0.469}{\sqrt{1.097}} = 0.448 .$$

It is apparent that there is a large effect resulting from V_1V_2 versus V_2V_1. With the large order effect, the interpretation of the factorial effects is difficult.

Example 2 of a 2^3 FSD

Consider a 2^3 FSD, the design and results of which are shown in Table IX.

TABLE IX

Results From One Replication of a 2^3 FSD

3-Way Order	Factor	Test 1	Factor	Test 2	Factor	Test 3	Factor	Test 4
Order 1								
Unit 1	None	9.219	V_1	15.061	V_2	14.950	V_3	17.550
Order 2								
Unit 1	None	11.095	V_1	15.964	V_3	15.732	V_2	14.033
Order 3								
Unit 1	None	9.541	V_2	8.517	V_1	15.213	V_3	13.922
Order 4								
Unit 1	None	9.104	V_2	11.484	V_3	9.109	V_1	14.153
Order 5								
Unit 1	None	9.045	V_3	9.173	V_1	14.836	V_2	14.701
Order 6								
Unit 1	None	12.596	V_3	9.226	V_2	10.591	V_1	15.806

The transformation

$$Z = BX$$

yields

$$Z' = [-4.131, -2.294, -3.874, -3.443, -1.798, 0.200, 0.724, -5.049, -2.452,$$
$$-1.683, 0.968, -3.684, -0.091, -4.676, -3.190, 2.303, 0.261, -4.332] .$$

Then the best estimate of

$$\xi' = \left[\xi_{\alpha_1} - \xi_{(1)}, \xi_{\alpha_2} - \xi_{(1)}, \xi_{\alpha_3} - \xi_{(1)}, \ldots \xi_{\alpha_3\alpha_2\alpha_1} - \xi_{(1)}\right]$$

is [see Equation 5]

$$\hat{\xi} = (A'A)^{-1}A'Z \tag{8}$$

and

$$\hat{\xi}' = [5.356, 0.678, -1.621, 5.488, 6.523, 4.880, 4.916, -0.846,$$

$$-1.130, 8.088, 3.181, 5.232, 4.198, 4.781, 4.085] .$$

The analysis of variance corresponding to that of Table IV for factorial and order effects is given in Table X.

TABLE X

Analysis of Variance for Factorial and Order Effects from 2^3 FSD

SV	DF	SS
Factors	7	$Z\left[A_1(A_1'A_1)^{-1}A_1'\right]Z' = 141.200$
Orders	8	$Z\left[A(A'A)^{-1}A' - A_1(A_1'A_1)^{-1}A_1'\right]Z' = 11.001$
Error	3	$Z\left[I - A(A'A)^{-1}A'\right]Z' = 6.193$
Total	18	$Z'Z = 158.394$

The test of significance for order effects,

$$F_{8,3} = \frac{3(11.001)}{8(6.193)} = 0.666 \ ,$$

indicates that the effect of order is negligible.

Since the order effect is concluded to be negligible, the best estimate of

$$\eta' = \big[\eta_{a_1} - \eta_{(1)}, \eta_{a_2} - \eta_{(1)}, \eta_{a_3} - \eta_{(1)}, \eta_{a_1a_2} - \eta_{(1)},$$

$$\eta_{a_1a_3} - \eta_{(1)}, \eta_{a_2a_3} - \eta_{(1)}, \eta_{a_1a_2a_3} - \eta_{(1)}\big]$$

is

$$\hat{\eta} = (A_1'A_1)^{-1}A_1'Z \ ,$$

and

$$\hat{\eta}' = \big[5.160, 0.584, 1.331, 5.380, 5.229, -0.694, 4.927\big] \ .$$

The estimates of the factor effects are then provided by

$$\hat{E} = I\hat{\eta}$$

or

$$\begin{bmatrix} \hat{\ell}_{V_1} \\ \hat{\ell}_{V_2} \\ \hat{\ell}_{V_3} \\ \hat{\ell}_{V_1V_2} \\ \hat{\ell}_{V_1V_3} \\ \hat{\ell}_{V_2V_3} \\ \hat{\ell}_{V_1V_2V_3} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{bmatrix} \hat{\eta} = \begin{bmatrix} 5.534 \\ 0.285 \\ -0.748 \\ -0.326 \\ 0.536 \\ -0.117 \\ -0.144 \end{bmatrix}$$

The variance of $\hat{\ell}$ is estimated by

$$\widehat{Var}(\hat{\ell}) = L(A_1'A_1)^{-1}L'\hat{\sigma}_e^2$$

$$= \begin{bmatrix} 0.636 & -0.189 & -0.189 & 0 & 0 & 0 & -0.086 \\ & 0.636 & -0.189 & 0 & 0 & 0 & -0.086 \\ & & 0.636 & 0 & 0 & 0 & -0.086 \\ & & & 0.479 & -0.111 & -0.111 & 0 \\ & & & & 0.479 & -0.111 & 0 \\ & & & & & 0.479 & 0 \\ & & & & & & 0.430 \end{bmatrix}$$

Tests of significance for the factors and interactions are

$$V_1 : t_{(3)} = \frac{5.534}{\sqrt{0.636}} = 6.939$$

$$V_2 : t_{(3)} = \frac{0.285}{\sqrt{0.636}} = 0.357$$

$$V_3 : t_{(3)} = \frac{-0.748}{\sqrt{0.636}} = -0.938$$

$$V_1V_2 : t_{(3)} = \frac{-0.326}{\sqrt{0.479}} = -0.471$$

$$V_1V_3 : t_{(3)} = \frac{0.556}{\sqrt{0.479}} = 0.803$$

$$V_2V_3 : t_{(3)} = \frac{-0.117}{\sqrt{0.479}} = -0.169$$

$$V_1V_2V_3 : t_{(3)} = \frac{-0.144}{\sqrt{0.430}} = -0.220$$

If the 0.05 level of significance were used, the only significant effect would be the main effect V_1.

It is enligntening to compare the usual 2^3 factorial design (FD), 8r units, with an FSD, 6r units, where, for each design, a unit is tested four times. The variance of the main and interaction effects for an FD as described is

$$\text{var (FD)} = \frac{0.125}{r}\left[\sigma_e^2 + 4\sigma_u^2\right].$$

The variance of main effects for an FSD is $\frac{0.308330}{r}\sigma_e^2$. Hence, the relative efficiency of the FSD to the FD for main effects is

$$E_r = \frac{\text{var (FD)}}{\text{var (FSD)}} = 0.405410\left[1 + 4\frac{\sigma_u^2}{\sigma_e^2}\right].$$

Hence, if $\sigma_u^2/\sigma_e^2 > 0.37$, the FSD is more efficient. The variances of the interaction effects for an FSD are less than the variance of the main effects. Thus, the advantage of the FSD is even more pronounced for the interaction effects.

ESTIMATION OF ERROR SPECTRA FROM THE CROSS-AUTOCOVARIANCE FUNCTIONS OF DIFFERENCES

D. B. Duncan
Johns Hopkins University

and

W. T. Wells
Pan American World Airways
Guided Missiles Range Division
Patrick Air Force Base, Florida

I. INTRODUCTION. The problem here is that of estimating the error (noise) spectra of several tracking systems all of which are simultaneously observing the same trajectory (signal). It extends the work of a previous memorandum (Duncan and Carroll, [2]) from processes of one dimension to processes of several dimensions. The basic principle is similar in that the estimates are obtained under the requirement that they be independent of the trajectory. A slight change in approach is made; the estimates are obtained through differences between the processes rather than through their average processes. Either way leads to a mathematically identical estimate; the change has been made as a matter of convenience.

The basic ideas can be formulated in a relatively simple context as follows. Let

$$y_1 = m + e_1, \quad y_2 = m + e_2, \quad y_3 = m + e_3$$

be three uncorrelated observations, all with the same expected mean m and with errors having the variances

$$v_1 = v(e_1), \quad v_2 = v(e_2), \quad v_3 = v(e_3).$$

In this simple context the problem is that of estimating v_1, v_2, v_3 without knowledge of m.

Form the differences

$$d_{12} = y_1 - y_2 = e_1 - e_3 , \quad d_{13} = y_1 - y_3 = e_1 - e_3$$

and then the product

$$(1.1) \qquad p_1 = d_{12}d_{13} = e_1^2 - e_1e_2 - e_1e_3 + e_2e_3 .$$

Since the errors are uncorrelated it is immediately apparent that p_1 is an unbiased estimator of v_1 and is not dependent in any way on the common expected mean m. Similarly

$$p_2 = d_{21}d_{23} \quad \text{and} \quad p_3 = d_{31}d_{32}$$

where $d_{ij} = d_i - d_j$ provide corresponding unbiased estimates of v_2 and v_3.

This approach, the essence of which is due to Grubbs [3], is the basic feature of the method here considered for estimating the noise spectra of three or more systems independently of the underlying common signal.

The variances or standard errors of the spectral estimates obtained indirectly through the differences are naturally not as low as those of the estimates which would be obtained directly from the error processes themselves if the latter were known free of the trajectory. The variances of the indirect estimates can be obtained by a technique which in the simple context work as follows.

Since the errors themselves are uncorrelated, the terms in the right hand side of (1.1) are all uncorrelated and hence

$$v(p_1) = v(e_1^2) + v(e_1e_2) + v(e_1e_3) + v(e_2e_3).$$

Working further with each term

$$v(e_i^2) = m_{i4} - v_i^2 ,$$

$$v(e_i e_j) = v_i v_j \text{ , each } i, j,$$

and thus

$$v(p_1) = m_{14} - v_1^2 + v_1 v_2 + v_1 v_3 + v_2 v_3 .$$

If the errors are Gaussian, the fourth moment m_{14} is $3v_1^2$ and $v(p_1)$ and the other corresponding variances become

$$v(p_1) = 2v_1^2 + v_1 v_2 + v_1 v_3 + v_2 v_3$$

(1.2) $$v(p_2) = 2v_2^2 + v_1 v_2 + v_1 v_3 + v_2 v_3$$

$$v(p_4) = 2v_3^2 + v_1 v_2 + v_1 v_3 + v_2 v_3$$

II. THE PROBLEM. The main problem will be now stated in more detail. Let $t = 1, \ldots, n$

$$\{y_{1t} = m_t + e_{1t}\} , \{y_{2t} = m_t + e_{2t}\} \text{ and } \{y_{3t} = m_t + e_{3t}\} ,$$

be three processes of observations. Each has the same underlying process of expected means $\{m_t\}$. The error processes $\{e_{1t}\}$, $\{e_{2t}\}$ and $\{e_{3t}\}$ have means zero (as their name implies), are stationary with spectral densities $S_1(f)$, $S_2(f)$ and $S_3(f)$, and are uncorrelated one with another. Each process is a vector process with k components. That is

$$y_{1t} = \begin{bmatrix} y_{11t} \\ y_{12t} \\ \cdots \\ y_{1kt} \end{bmatrix}, \quad m_t = \begin{bmatrix} m_{1t} \\ m_{2t} \\ \cdots \\ m_{kt} \end{bmatrix}$$

etc., which without loss of generality, we will often discuss with $k = 2$ or $k = 3$. For $k > 1$ the spectral densities at each frequency form $k \times k$ complex matrices which we shall write as

$$(3.1) \quad S_i(f) = R_i(f) + iQ_i(f), \qquad i = 1, 2, 3$$

or, in more detail for say a two-dimensional process with

$$\begin{bmatrix} S_{i11}(f) & S_{i12}(f) \\ S_{i21}(f) & S_{i22}(f) \end{bmatrix} = \begin{bmatrix} R_{i11}(f) & R_{i12}(f) \\ R_{i21}(f) & R_{i22}(f) \end{bmatrix} + i \begin{bmatrix} Q_{i11}(f) & Q_{i12}(f) \\ Q_{i21}(f) & Q_{i21}(f) \end{bmatrix} .$$

The i coefficient for the imaginary matrix $Q(f)$ is not to be confused with the subscript i for the i^{th} system.

The problem is to get unbiased estimates

$$(3.2) \quad \tilde{S}_i(f) = \tilde{R}_i(f) + i\tilde{Q}_i(f), \qquad i = 1, 2, 3$$

for the densities (3.1) which depend in no way on the mean process $\{m_t\}$. A secondary problem is that of estimating the variances or standard errors of the elements of the estimates $\tilde{R}_i(f)$ and $\tilde{Q}_i(f)$, $i = 1, 2, 3$.

III. THE METHOD. The method for the main problem is first to get the difference processes

$$d_{12t} = y_{1t} - y_{2t}, \quad d_{13t} = y_{1t} - y_{2t}, \quad t = 1, \ldots, n$$

and then to compute the cross-autocovariances,

$$C_{11h} = \frac{1}{n-h} \sum_{t=1}^{n-h} d_{12t} d'_{13(t+h)} ,$$

$$C_{12h} = \frac{1}{n-h} \sum_{t=1}^{n-h} d_{13t} d'_{12(t+h)} , \quad h = 0, \ldots, m$$

each of which is a k x k matrix.

Then following a k-variate generalization of the Blackman-Tukey [1] spectral density estimation method (with Hanning smoothing) the autocovariances are next transformed to raw spectral density estimates

$$\tilde{\tilde{S}}_1(j) = \tilde{\tilde{R}}_1(j) + i\tilde{\tilde{Q}}_1(j) ,$$

where

$$\tilde{\tilde{R}}(j) = 2\Delta t \left[C_{110} + \sum_{h=1}^{m-1} (C_{11h} + C_{12h}) \cos \frac{\pi hj}{m} + \cos(\pi j) \right] ,$$

$$\tilde{\tilde{Q}}(j) = 2\Delta t \left[\sum_{h=1}^{m-1} (C_{11h} - C_{12h}) \sin \frac{\pi hj}{m} \right] , \quad j = 0, \ldots, m$$

The real parts are then smoothed to the final form

$$\tilde{R}_1(o) = \frac{1}{2}\tilde{\tilde{R}}_1(o) + \frac{1}{2}\tilde{\tilde{R}}_1(1)$$

$$\tilde{R}_1(j) = \frac{1}{4}\tilde{\tilde{R}}_1(j-1) + \frac{1}{2}\tilde{\tilde{R}}_1(j) + \frac{1}{4}\tilde{\tilde{R}}_1(j+1), \quad j = 1, \ldots, m-1$$

$$\tilde{R}_1(m) = \frac{1}{2}\tilde{\tilde{R}}_1(m-1) + \frac{1}{2}\tilde{\tilde{R}}_1(m) .$$

The imaginary parts are smoothed in the same way. The combined forms

$$\tilde{S}_1(j) = \tilde{R}_1(j) + i\tilde{Q}_1(j)$$

at $j = 0, \ldots, m$ cycles per record length of n points estimate the spectral densities

$$S_1(f) = R_1(f) + iQ_1(f)$$

at $f = \Delta tj/2m$, cycles per second, where Δt is the time interval between data points.

Similar estimates are next obtained for $S_2(f)$ and $S_3(f)$. The starting difference processes using the same notation are d_{21t}, d_{23t}, and d_{31t}, d_{32t}; the cross autocovariances are C_{21h}, C_{22h}, and C_{31h}, C_{32h}, $h = 0, \ldots, m$ and the final estimates are

$$\tilde{S}_2(f) = \tilde{R}_2(f) + i\tilde{Q}_2(f) \quad \text{and} \quad \tilde{S}_3(f) = \tilde{R}_3(f) + i\tilde{Q}_3(f)$$

respectively.

IV. AN EXAMPLE. Three independent stationary Gaussian error 3x1 vector processes $\{e_{1t}\}$, $\{e_{2t}\}$, $\{e_{3t}\}$, of $n = 2,000$ time points each were generated using the recursive formula

$$e_t = \underset{3\times3}{A} w_t - \underset{3\times3}{B_1} e_{t-1} - \underset{3\times3}{B_2} e_{t-2} - \underset{3\times3}{B_3} e_{t-3} \, .$$

The processes w_t from which these were derived consisted of independent Gaussian standard white noise 3×1 vectors. The values of the coefficient matrices for each process were as shown in Table 1.

TABLE 1

Coefficients for Generating

Error Processes

	Process (1)			Process (2)			Process (3)		
A	1	0	0	1	0	0	1	0	0
	0	1	0	0	1.1	.2	0	1.1	.2
	0	0	1	0	0	1.2	0	0	1.2
B_1		0		-.6	0	0	-.4	0	0
				0	-.7	-.1	0	-.45	-.05
				0	0	-.5	0	0	-.35
B_2		0			0		-.4	0	0
							0	-.45	-.05
							0	0	-.35
B_3		0			0		.16	0	0
							0	.17	.02
							0	0	.15

Error process 1 is itself simply standard white noise. The other two are autoregressive processes of order 1 and 3 respectively. To get things started in the latter two cases, values of e_t for $t \leq 0$ were filled with white noise. To eliminate transient effects as much as seemed desirable, 40 vector observations were discarded from the beginning of the processes. The computing work for generating the white noise, the error processes, and subsequent calculations will be written up in reports by D. B. Duncan, E. E. McGehee and S. B. Burkett, and by L. B. Collins and A. Rinaldi who have been cooperating in the programming work. (As currently planned, these reports will appear as Pan American Technical Staff Memos and RCA Data Reduction Programming Memos respectively.)

Three sets of spectral density estimates $S_1(f)$, $S_2(f)$ and $S_3(f)$ were obtained by the cross-spectrum-of-differences method described in Section III. (In the example no mean process $\{m_t\}$ was added to each of the error processes since the differences could be and were obtained just as well directly from the errors, e. g. $\{d_{12t} = y_{1t} - y_{2t} = e_{1t} - e_{2t}\}$.) The maximum lag taken in the example was $m = 20$; the time interval assumed was $\Delta t = .05$; estimates of density were thus obtained at $f = 0, .5, \ldots, 10$ cycles per second. The results are shown in nine tables for each of the processes $\{e_{1t}\}$, $\{e_{2t}\}$ and $\{e_{3t}\}$. Tables 1.1, ..., 1.9 show the results for the spectral densities of process one, Table 1.1 is for the 1, 1 element, Table 1.2 for the 1,2 element, and so on up to Table 1.9 for the 3,3 element. The columns marked (5), Indirect Estimates, S 2, show the real part of the estimate $\hat{R}_2(f)$ under the Fortran notation SR2 and the imaginary part $\hat{Q}_2(f)$ under SI2, for each of the frequency f, denoted in Fortran under F, Column (1). Tables 2.1, ..., 2.9 and 3.1, ..., 3.9 in Column (5) show the corresponding results for processes two and three.

For comparison, the columns headed (2), Expected Values, S, show the real, SR, and imaginary part, SI, of the spectral density, $S(f)$, being estimated; the columns headed (3), Direct Estimates, S1, show the real, SR1, and imaginary part, SI1, of the estimates obtained directly from the individual processes, which are available in a simulation study like this, but which, of course, would not be available in a real problem. The direct estimates represent an optimum which one could not hope to improve upon. The standard errors for the real and imaginary parts of both the direct estimates and the indirect estimates have been computed by the formulae developed in the subsequent sections and are shown in columns (4) and (6) under the Fortran notations ER1, EI1 and ER2, EI2 respectively.

V. <u>EXPECTATIONS AND VARIANCES OF ESTIMATES</u>. In this section we derive the expectations (showing that they are unbiased) and the approximate variances of the spectral density estimates obtained through the difference processes. In so doing, to help introduce the methods, we also first derive the well-known expectations and approximate variances of the direct estimates. In each part the work is considered in detail for the case $r = 1$, where r is the ratio $r = n/m$, in which case a spectral density estimate can be regarded approximately as the squared modulus of a single complex random variable.

The results then carry over to the case $r > 1$ by virtue of the fact (Jenkins [4]) that an estimate in this case can be expressed approximately as the mean of r uncorrelated estimates of the case $r = 1$ form. Thus the expectation is the same and the variance is obtained simply by dividing by r.

5.1 DIRECT ESTIMATES, UNIVARIATE PROCESS. Aim: To show that

$$E\left[\hat{S}(f)\right] = S(f), \quad v\ \hat{S}(f) = \frac{1}{r} S^2(f)$$

where $\hat{S}(f)$ represents a direct, say, Blackman and Tukey [1] estimate of any one of the error spectral density functions.

Case $r = 1$: Here we can write

$$\hat{S}(f) \underset{\text{approx.}}{=} z\bar{z}$$

where

$$z = \frac{1}{\sqrt{2}}(x + i\,y),\ E(x) = E(y) = 0,\ v(x) = v(y) = S(f),$$

and $c(x, y) = 0$, (using $c(\cdot\,,\cdot)$ for $c(\cdot\,,\cdot)$).

[Note in passing: This implies

$$E(z) = \frac{1}{\sqrt{2}}(E(x) + i\,E(y)) = 0,$$

$$V(z) = E(z - E(x))\,\overline{(z - E(z))},\ \text{by definition}$$

$$= E(z\bar{z}) = \frac{1}{2}E(x^2 + y^2)$$

$$= \frac{1}{2}(S(f) + S(f)) = S(f).\ ]$$

Thus,

$$E\left[\hat{S}(f)\right] = E(z\bar{z}) = v(z) = S(f)$$

which established the first result. Next,

$$v\left[\hat{S}(f)\right] = v\left[\frac{1}{2}(x^2 + y^2)\right]$$

If x and y (and thus z) are Gaussian, which we shall also assume, we have

$$v(x^2) = v(y^2) = 2S^2(f),$$

Also x, y are independent from which $c(x^2, y^2) = 0$. Hence

$$v\left[\hat{S}(f)\right] = \frac{1}{4}(2S^2(f) + 2S^2(f)) = S^2(f)$$

which establishes the second and final result.

Case $r > 1$. Here the expectation is the same as for case $r = 1$, the variance is obtained approximately by dividing by $r = n/n_1$, thus

$$E\left[\hat{S}(f)\right] = S(f), \quad v\left[\hat{S}(f)\right] = \frac{1}{r}S^2(f).$$

5.2 DIRECT ESTIMATES, K-VARIATE PROCESS. Aim: Put $\hat{S}_{ij}(f) = \hat{R}_{ij}(f) + i\hat{Q}_{ij}(f)$ for the direct estimate of the ij^{th} cross-spectral density $S_{ij}(f) = R_{ij}(f) + iQ_{ij}(f)$ for any one of the error processes. Then it is desired to show that

$$E\left[\hat{R}_{ij}(f)\right] = \hat{R}_{ij}(f), \qquad E\left[\hat{Q}_{ij}(f)\right] = Q_{ij}(f)$$

and

$$v\left[\hat{R}_{ij}(f)\right] = \frac{1}{2r}(A + B)$$
$$v\left[\hat{Q}_{ij}(f)\right] = \frac{1}{2r}(A - B).$$

where

$$A = R_{ii}(f)\,R_{jj}(f) \quad \text{and} \quad B = R_{ij}^2(f) - Q_{ij}^2(f).$$

[Note: Strictly speaking, the development here for $r = 1$, is not correct for the extreme frequencies $f = 0$ and $f = m/2$. At each of these frequencies the spectral variable z is real and may be written as $z = x$.

From this

$$E\left[\hat{S}(f)\right] = v(z) = S(f)$$

and $\hat{S}(f)$ is unbiased as before, However, for the variance

$$V\left[\hat{S}(f)\right] = v(x^2) = 2S^2(f)$$

which is double the previous value $S^2(f)$. The standard error is $\sqrt{2}S(f)$.

Because of the smoothing used in the more general type of spectral estimate for the cases $r > 1$, the standard errors $(1/\sqrt{r})S(f)$ for $f = 0$ and $m/2$ are not low by nearly as much as the factor $1/\sqrt{2}$ and this formula has been used for all frequencies.

Remarks of this type apply to the derivations of all standard errors in this report at the extreme frequencies $f = 0$ and $m/2$. In looking at Tables 1.1, . . . , 3.9 the standard errors shown for both the direct and indirect estimate at $f = 0$ and $m/2$ are on the low side for this reason.]

<u>Case $r = 1$:</u> Here we can write

$$\hat{S}_{ij}(f) = z_i \bar{z}_j$$

where

$$z_k = \frac{1}{\sqrt{2}}(x_k + i\,y_k), \quad E(x_k) = 0, \quad v(x_k) = v(y_k) = R_{kk}(f),$$

$$c(x_k, y_k) = 0, \quad k = i, j; \quad c(x_i, x_j) = c(y_i, y_j) = R_{ij}(f),$$

$$-c(x_i, y_j) = c(x_j, y_i) = Q_{ij}(f).$$

Thus

$$\hat{R}_{ij}(f) = \mathrm{Re}(\hat{S}_{ij}(f)) = \frac{1}{2}(x_i x_j + y_i y_j),$$

$$\hat{Q}_{ij}(f) = \mathrm{Im}(\hat{S}_{ij}(f)) = \frac{1}{2}(-x_i y_j + x_j y_i).$$

Now

$$E\left[\hat{R}_{ij}(f)\right] = \frac{1}{2}(E(x_i x_j) + E(y_i y_j)) = \frac{1}{2}(2\,R_{ij}(f)) = R_{ij}(f)$$

$$E\left[\hat{Q}_{ij}(f)\right] = \frac{1}{2}(-E(x_i y_j) + E(x_j y_i)) = \frac{1}{2}(2\,Q_{ij}(f)) = Q_{ij}(f)$$

which establishes the result on unbiasedness.

Next,

$$v\left[\hat{R}_{ij}(f)\right] = \frac{1}{4}(v(x_i x_j) + 2\,c(x_i x_j,\ y_i y_j) + v(y_i y_j)).$$

To expand this we will use a result due to Isserlis quoted, for example, by Blackman and Tukey [1, p. 100]: If a, b, c, d are four Gaussian variables with zero means, then

$$c(ab,\ cd) = E(ac)\,E(bd) + E(ad)\,E(bc).$$

First,

$$v(x_i x_j) = c(x_i x_j,\ x_i x_j)$$

$$= R_{ii}(f)\, R_{jj}(f) + R_{ij}^2(f),$$

$$c(x_i x_j,\ y_i y_j) = 0 - Q_{ij}^2(f),$$

$$v(y_i y_j) = R_{ii}(f)\, R_{jj}(f) + R_{ij}^2(f).$$

Combining, we get

$$v(\hat{R}_{ij}(f)) = \frac{1}{4}(2\, R_{ii}(f)\, R_{jj}(f) + 2\, R_{ij}^2(f) - 2\, Q_{ij}^2(f)) = \frac{1}{2}(A + B),$$

which establishes the first result on variances.

Next,

$$v\left[\hat{Q}_{ij}(f)\right] = \frac{1}{4}(v(x_i y_j) - 2c(x_i y_j,\ x_j y_i) + v(x_j y_i))$$

in which

$$v(x_i y_j) = R_{ii}(f)\, R_{jj}(f) + Q_{ij}^2(f) = v(x_j y_i),$$

$$c(x_i y_j,\ x_j y_i) = R_{ij}^2(f) + 0,$$

giving

$$v\left[\hat{Q}_{ij}(f)\right] = \frac{1}{2}(A - B).$$

which concludes case $r = 1$.

<u>Case $r > 1$</u>: Here the results are the same as for $r = 1$ except for

the additional $1/r$ factor in the variances, establishing the desired results.

Diagonal terms: For these it is noted that

$$v\left[\hat{R}_{ii}(f)\right] = \frac{1}{2r}(2R_{ii}^2(f) = \frac{1}{r}R_{ii}^2(f)$$

$$v\left[\hat{Q}_{ii}(f)\right] = 0$$

which agree, as they should, with the univariate case results.

5.3 INDIRECT ESTIMATES, K-VARIATE PROCESSES. Aim: Putting

$$\tilde{S}_{ijk}(f) = \tilde{R}_{ijk}(f) + i\tilde{Q}_{ijk}(f)$$

for the indirect estimates defined in section 3, for the $j\ k^{th}$ spectral estimate at frequency (f) for system i; $i = 1, 2, 3; j, k = 1, \ldots, K$, it is desired to show that

$$E\left[\tilde{S}_{ijk}(f)\right] = S_{ijk}(f),$$

$$\text{(i. e.}\quad E\left[\tilde{R}_{ijk}(f)\right] = R_{ijk}(f), \qquad E\left[\tilde{Q}_{ijk}(f)\right] = Q_{ijk}(f)),$$

and that

$$v\left[\tilde{R}_{ijk}(f)\right] = \frac{1}{2r}(A + B + C)$$

$$v\left[\tilde{Q}_{ijk}(f)\right] = \frac{1}{2r}(A - B + C)$$

where

$$A = R_{ijj}R_{ikk}, \quad B = R_{ijk}^2 - Q_{ijk}^2$$

and, for $i = 1$,

$$C = R_{1jj}R_{3kk} + R_{2jj}R_{1kk} + R_{2jj}R_{3kk},$$

for $i = 2$,

$$C = R_{2jj}R_{3kk} + R_{1jj}R_{2kk} + R_{1jj}R_{3kk},$$

and for $i = 3$,

$$C = R_{3jj}R_{2kk} + R_{1jj}R_{3kk} + R_{1jj}R_{2kk}.$$

Case $r = 1$: Here we can write for system 1 [i. e. $i = 1$],

$$\tilde{S}_{1jk}(f) = z_{12j}\bar{z}_{13k} = (z_{1j} - z_{2j})\,\overline{(z_{1k} - z_{3k})}$$

$$= z_{1j}\bar{z}_{1k} - z_{1j}\bar{z}_{3k} - z_{2j}\bar{z}_{1k} + z_{2j}\bar{z}_{3k}$$

where

$$z_{hj} = \frac{1}{\sqrt{2}}(x_{hj} + i\,y_{hk}), \quad z_{ij} = \frac{1}{\sqrt{2}}(x_{ij} + iy_{ik})$$

and the real x's and y's have the same means, variances and covariances within each system (i. e., for $h = i$) as given in the previous subsection. In addition all covariances across systems (i. e., for $h \neq i$) are zero.

First, dealing with real and imaginary parts together,

$$E(\tilde{S}_{1jk}(f)) = E(z_{1j}\bar{z}_{1k}) - 0 - 0 + 0 = S_{1jk}$$

which establishes the result on unbiasedness. Next,

$$v(\tilde{S}_{1jk}(f)) = v(z_{1j}\bar{z}_{1k}) + v(z_{1j}\bar{z}_{3k}) + v(z_{2j}\bar{z}_{1k}) + v(z_{2j}\bar{z}_{3k}),$$

all cross product terms being zero because of the zero crossproducts between systems. Dealing with just the real parts of the right hand side, term by term, we have

$$v(\mathrm{Re}(z_{1j}\bar{z}_{1k})) = v(\tilde{R}_{1jk}) \qquad \text{(cf. previous subsection)}$$

$$= \frac{1}{2}A + \frac{1}{2}B, \qquad (i = 1),$$

$$v(\mathrm{Re}(z_{1j}\bar{z}_{3k})) = v(\tfrac{1}{2}(x_{1j}x_{3k} + y_{1j}y_{3k}))$$

$$= \frac{1}{4}(R_{1jj}R_{3kk} + 2(0 + 0) + R_{1jj}R_{3kk})$$

$$= \frac{1}{4}R_{1jj}R_{3kk}.$$

Similarly

$$v(\mathrm{Re}(z_{2j}\bar{z}_{1k})) = \frac{1}{2}R_{2jj}R_{1kk}$$

$$v(\mathrm{Re}(z_{2j}\bar{z}_{3k})) = \frac{1}{2}R_{2jj}R_{3kk}.$$

Collecting terms

$$v(\tilde{R}_{1jk}) = \frac{1}{2}A + \frac{1}{2}B + \frac{1}{2}C, \qquad (i = 1).$$

This establishes the result for $v(\tilde{R}_{1jk})$. Similar steps lead to the corresponding results for $i = 2, 3$.

Finally, for the imaginary terms

$$v(\mathrm{Im}(z_{1j}\overline{z_{1k}})) = v(\hat{Q}_{1jk}) \qquad \text{(cf. previous subsection)}$$

$$= \frac{1}{2}A - \frac{1}{2}B, \qquad (i = 1),$$

$$v(\mathrm{Im}(z_{1j}\overline{z}_{3k})) = v(\tfrac{1}{2}(-x_{1j}y_{3k} + y_{1j}x_{3k}))$$

$$= \frac{1}{4}(R_{1jj}R_{3kk} + 2(0 + 0) + R_{1jj}R_{3kk})$$

$$= \frac{1}{2}R_{1jj}R_{3kk}, \text{ etc.}$$

Collecting terms

$$v(\tilde{Q}_{1jk}) = \frac{1}{2}A - \frac{1}{2}B + \frac{1}{2}C, \qquad (i = 1).$$

This, and similar steps for $i = 2, 3$ establish the $(r = 1)$ - case result on variances.

Case $r > 1$: The $(r = 1)$ - case result on expectations together with the variance results divided by r, establish the full $(r > 1)$ - case results.

5.4 SUMMARY OF RESULTS ON VARIANCES. The results of this section which have been used in computing the standard errors (square roots of the variances) in Tables 1.1, ..., 1.9 may be summarized as follows:

For direct estimates

$$v(\hat{R}_{ijk}(f)) = \frac{1}{2r}(A + B)$$

$$v(\hat{Q}_{ijk}(f)) = \frac{1}{2r}(A - B)$$

where

$$A = R_{ijj}(f)R_{ikk}(f), \qquad B = R^2_{ijk}(f) - Q^2_{ijk}(f)$$

$$i = 2, 2, \text{and} \ j, k = 1, \ldots, K$$

For a diagonal term $(j = k)$ or a univariate $(K = 1)$ process these appear as

$$v(\hat{R}_{ijj}(f)) = \frac{1}{r} R^2_{ijj}(f)$$

$$v(\hat{Q}_{ijj}(f)) = 0$$

the latter being reasonable since $\hat{Q}_i = 0, \ i = 1, 2, 3$

<u>For the indirect estimates</u>

$$v(\tilde{R}_{ijk}(f)) = \frac{1}{2r}(A + B + C)$$

$$v(\tilde{Q}_{ijk}(f)) = \frac{1}{2r}(A - B + C)$$

where A and B are as already defined and for $i = 1$,

$$C = R_{1jj} R_{3kk} + R_{2jj} R_{1kk} + R_{2jj} R_{3kk},$$

for $i = 2$

$$C = R_{2jj} R_{3kk} + R_{1jj} R_{2kk} + R_{1jj} R_{3kk}.$$

and for $i = 3$

$$C = R_{3jj} R_{2kk} + R_{1jj} R_{3kk} + R_{1jj} R_{2kk},$$

$$i = 1, 2, 3; j, \ k = 1, \ldots, K.$$

For a diagonal term $(j = k)$ or a univariate $K = 1$ process these appear as

$$v(\tilde{R}_{ijj}(f)) = \frac{1}{r}(R^2_{ijj}(f) + \frac{1}{2}C)$$

$$v(\tilde{Q}_{ijj}(f)) = \frac{1}{2r}C$$

where

$$C = R_{1jj} R_{2jj} + R_{1jj} R_{3jj} + R_{2jj} R_{3jj}.$$

VI. <u>DISCUSSION AND CONCLUDING REMARKS</u>. The numerical results in Tables 1.1, ..., 3.9 serve to illustrate and emphasize the properties of unbiasedness and the theoretical variances for the indirect estimator obtained in section V. These bear out the conclusions of the previous study [2] in the univariate case $K = 1$, showing that they also apply in the more general cases $K > 1$.

The indirect estimates $\tilde{S}$ (or S2 as they appear in the tables) are good (have a relatively low standard error) at most points and especially at frequencies at which powers in the other systems are lower than the one being estimated. They are not so good however at frequencies at which the reverse is true. At frequencies at which the powers in the other systems are much larger than the one being estimated. The noise added to the spectral estimate from the other systems by having to work through the differences involving the latter can be seriously large at such points.

Work is proceeding on an iterative approach to minimize this difficulty. This can be briefly indicated in principle as follows. After one step consisting of the above procedure and on the basis of the knowledge so gained about the error spectrum in each case, each system is filtered to give individually optimum independent estimates of the trajectories

$$\left\{\tilde{Y}_{it}:\ t = 1, \ldots, n\right\}, \quad i = 1, 2, 3.$$

The indirect spectrum estimation procedure can then be repeated in a second step based on the differences

$$\left\{d_{12t} = Y_{1t} - \tilde{Y}_{2t},\ d_{13t} = Y_{1t} - \tilde{Y}_{3t}:\ t = 1, \ldots, n\right\} \text{ for system 1,}$$

$$\left\{d_{21t} = Y_{2t} - \tilde{Y}_{1t},\ d_{23t} = Y_{2t} - \tilde{Y}_{3t}:\ t = 1, \ldots, n\right\} \text{ for system 2,}$$

$$\left\{d_{31t} = Y_{3t} - \tilde{Y}_{1t},\ d_{32t} = Y_{3t} - \tilde{Y}_{2t}:\ t = 1, \ldots, n\right\} \text{ for system 3.}$$

If the new spectral estimates show much change, the same step can be repeated again and several times thereafter if necessary.

The same method expressed in terms of residuals (as in [2]) extends readily to cases with more than three systems. The greater the number of systems the less the cross transmission of noise to the indirect spectral estimate.

To sum up, the given method will, in itself, provide a good spectral estimator in many situations. In most of the remainder it promises to provide a valuable first step in iterations leading to a good estimator.

BIBLIOGRAPHY

1) R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra. Dover, New York, 1958.

2) David B. Duncan and Charles L. Carroll, Jr. "Studies in the Simultaneous Estimation of a Trajectory from Several Tracking Systems". No. 1. Estimates of Error Spectra Derived from Estimates of the Spectra of Residuals. MTV-TM-62-13.

3) F. E. Grubbs. "On Estimating Precision of Measuring Instruments and Product Variability". Journal of the American Statistical Association, 43, pp 234-264, 1948.

4) G. M. Jenkins, "General Considerations in the Analysis of Spectra", Technometrics, 3, 11, 133-166, 1961.

TABLE 1.1 ERROR SPECTRAL DENSITY, PROCESS 1
ELEMENT 1,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.1000	0.	0.0853	-0.	0.0100	0.	0.1132	-0.0493	0.0567	0.0558
0.5	0.1000	0.	0.0859	-0.	0.0100	0.	0.1068	-0.0344	0.0516	0.0506
1.0	0.1000	0.	0.0915	-0.	0.0100	0.	0.0890	0.0127	0.0411	0.0399
1.5	0.1000	0.	0.0972	0.	0.0100	0.	0.0847	0.0196	0.0315	0.0299
2.0	0.1000	0.	0.0991	-0.	0.0100	0.	0.1032	-0.0100	0.0247	0.0226
2.5	0.1000	0.	0.0955	0.	0.0100	0.	0.1132	-0.0004	0.0203	0.0177
3.0	0.1000	0.	0.0918	-0.	0.0100	0.	0.1057	0.0091	0.0175	0.0144
3.5	0.1000	0.	0.0969	0.	0.0100	0.	0.1027	-0.0048	0.0157	0.0121
4.0	0.1000	0.	0.1001	-0.	0.0100	0.	0.1072	-0.0079	0.0145	0.0106
4.5	0.1000	0.	0.0920	0.	0.0100	0.	0.1000	0.0063	0.0138	0.0094
5.0	0.1000	0.	0.0854	-0.	0.0100	0.	0.0924	0.0089	0.0132	0.0097
5.5	0.1000	0.	0.0951	0.	0.0100	0.	0.0981	0.0034	0.0129	0.0081
6.0	0.1000	0.	0.1055	-0.	0.0100	0.	0.1037	0.0112	0.0126	0.0077
6.5	0.1000	0.	0.1107	0.	0.0100	0.	0.1119	0.0156	0.0125	0.0075
7.0	0.1000	0.	0.1101	-0.	0.0100	0.	0.1154	0.0077	0.0125	0.0075
7.5	0.1000	0.	0.0986	0.	0.0100	0.	0.1102	0.0038	0.0126	0.0077
8.0	0.1000	0.	0.0877	-0.	0.0100	0.	0.1019	0.0039	0.0128	0.0080
8.5	0.1000	0.	0.0870	0.	0.0100	0.	0.0910	0.0019	0.0132	0.0087
9.0	0.1000	0.	0.0944	-0.	0.0100	0.	0.0932	0.0019	0.0138	0.0095
9.5	0.1000	0.	0.0941	0.	0.0100	0.	0.0915	0.0051	0.0145	0.0104
10.0	0.1000	0.	0.0897	-0.	0.0100	0.	0.0831	0.0068	0.0148	0.0109

TABLE 1.2 ERROR SPECTRAL DENSITY. PROCESS 1
ELEMENT 1,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	-0.0011	-0.0019	0.0071	0.0071	-0.0034	0.0880	0.0850	0.0850
0.5	0.	0.	-0.0042	-0.0039	0.0071	0.0071	-0.0376	0.0879	0.0719	0.0719
1.0	0.	0.	-0.0014	-0.0061	0.0071	0.0071	-0.0450	0.0526	0.0504	0.0504
1.5	0.	0.	-0.0020	-0.0018	0.0071	0.0071	-0.0148	0.0206	0.0350	0.0350
2.0	0.	0.	0.0074	0.0020	0.0071	0.0071	0.0132	0.0161	0.0256	0.0256
2.5	0.	0.	0.0019	0.0006	0.0071	0.0071	0.0147	0.0117	0.0200	0.0200
3.0	0.	0.	-0.0053	-0.0015	0.0071	0.0071	-0.0058	0.0051	0.0166	0.0166
3.5	0.	0.	-0.0051	-0.0054	0.0071	0.0071	0.0019	-0.0134	0.0144	0.0144
4.0	0.	0.	-0.0014	-0.0032	0.0071	0.0071	0.0066	-0.0122	0.0130	0.0130
4.5	0.	0.	-0.0036	0.0025	0.0071	0.0071	0.0032	0.0036	0.0123	0.0123
5.0	0.	0.	-0.0044	0.0073	0.0071	0.0071	-0.0075	0.0151	0.0114	0.0114
5.5	0.	0.	-0.0064	0.0158	0.0071	0.0071	-0.0182	0.0240	0.0110	0.0110
6.0	0.	0.	-0.0034	0.0167	0.0071	0.0071	-0.0157	0.0203	0.0107	0.0107
6.5	0.	0.	-0.0073	0.0086	0.0071	0.0071	0.0015	0.0102	0.0105	0.0105
7.0	0.	0.	-0.0036	-0.0023	0.0071	0.0071	0.0040	-0.0006	0.0106	0.0106
7.5	0.	0.	-0.0040	-0.0083	0.0071	0.0071	0.0096	-0.0094	0.0107	0.0107
8.0	0.	0.	0.0018	-0.0025	0.0071	0.0071	0.0053	-0.0026	0.0111	0.0111
8.5	0.	0.	0.0021	-0.0025	0.0071	0.0071	0.0065	-0.0014	0.0117	0.0117
9.0	0.	0.	0.0025	-0.0070	0.0071	0.0071	0.0027	0.0036	0.0126	0.0126
9.5	0.	0.	0.0029	-0.0037	0.0071	0.0071	0.0034	0.0132	0.0135	0.0135
10.0	0.	0.	0.0033	[illegible]	0.0071	0.0071	0.0055	0.0169	0.0140	0.0140

TABLE 1.3 ERROR SPECTRAL DENSITY. PROCESS 1
ELEMENT 1.3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0035	0.0045	0.0071	0.0071	0.0865	0.0685	0.0542	0.0542
0.5	0.	0.	0.0025	0.0026	0.0071	0.0071	0.0474	0.0608	0.0508	0.0508
1.0	0.	0.	0.0035	-0.0012	0.0071	0.0071	-0.0015	0.0110	0.0427	0.0427
1.5	0.	0.	-0.0000	0.0034	0.0071	0.0071	-0.0129	-0.0045	0.0340	0.0340
2.0	0.	0.	0.0010	0.0045	0.0071	0.0071	-0.0056	0.0156	0.0269	0.0269
2.5	0.	0.	0.0056	-0.0082	0.0071	0.0071	0.0095	-0.0126	0.0217	0.0217
3.0	0.	0.	0.0084	-0.0132	0.0071	0.0071	0.0078	-0.0256	0.0181	0.0181
3.5	0.	0.	0.0019	-0.0052	0.0071	0.0071	-0.0077	-0.0139	0.0157	0.0157
4.0	0.	0.	-0.0041	-0.0078	0.0071	0.0071	-0.0141	-0.0103	0.0140	0.0140
4.5	0.	0.	-0.0012	-0.0024	0.0071	0.0071	-0.0073	0.0047	0.0129	0.0129
5.0	0.	0.	0.0018	0.0061	0.0071	0.0071	-0.0029	0.0006	0.0121	0.0121
5.5	0.	0.	0.0037	0.0012	0.0071	0.0071	0.0037	-0.0194	0.0116	0.0116
6.0	0.	0.	0.0045	-0.0069	0.0071	0.0071	0.0102	-0.0295	0.0113	0.0113
6.5	0.	0.	-0.0031	0.0006	0.0071	0.0071	-0.0042	-0.0146	0.0111	0.0111
7.0	0.	0.	-0.0068	0.0065	0.0071	0.0071	-0.0114	0.0045	0.0112	0.0112
7.5	0.	0.	0.0023	0.0003	0.0071	0.0071	0.0015	0.0016	0.0113	0.0113
8.0	0.	0.	0.0070	-0.0073	0.0071	0.0071	-0.0030	-0.0136	0.0117	0.0117
8.5	0.	0.	0.0017	-0.0064	0.0071	0.0071	-0.0097	-0.0137	0.0123	0.0123
9.0	0.	0.	-0.0056	-0.0014	0.0071	0.0071	-0.0097	0.0012	0.0132	0.0132
9.5	0.	0.	0.0030	-0.0005	0.0071	0.0071	-0.0149	0.0058	0.0140	0.0140
10.0	0.	0.	0.0052	-0.0011	0.0071	0.0071	-0.0216	0.0023	0.0144	0.0144

TABLE 1.4 ERROR SPECTRAL DENSITY. PROCESS I
ELEMENT 2,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	-0.0011	0.0019	0.0071	0.0071	-0.0483	0.0722	0.0850	0.0850
0.5	0.	0.	-0.0042	0.0039	0.0071	0.0071	-0.0247	0.0757	0.0738	0.0738
1.0	0.	0.	-0.0074	0.0061	0.0071	0.0071	-0.0132	0.0537	0.0533	0.0533
1.5	0.	0.	-0.0020	0.0010	0.0071	0.0071	-0.0012	0.0333	0.0372	0.0372
2.0	0.	0.	0.0074	-0.0023	0.0071	0.0071	0.0306	0.0189	0.0271	0.0271
2.5	0.	0.	0.0049	-0.0005	0.0071	0.0071	0.0244	-0.0052	0.0210	0.0210
3.0	0.	0.	-0.0054	0.0013	0.0071	0.0071	0.0059	-0.0148	0.0172	0.0172
3.5	0.	0.	-0.0051	0.0059	0.0071	0.0071	0.0071	-0.0065	0.0148	0.0148
4.0	0.	0.	-0.0014	0.0032	0.0071	0.0071	0.0152	-0.0032	0.0133	0.0133
4.5	0.	0.	-0.0033	-0.0025	0.0071	0.0071	0.0165	-0.0089	0.0122	0.0122
5.0	0.	0.	-0.0041	-0.0073	0.0071	0.0071	-0.0026	-0.0146	0.0115	0.0115
5.5	0.	0.	-0.0064	-0.0158	0.0071	0.0071	-0.0168	-0.0245	0.0110	0.0110
6.0	0.	0.	-0.0034	-0.0167	0.0071	0.0071	-0.0212	-0.0215	0.0107	0.0107
6.5	0.	0.	-0.0073	-0.0086	0.0071	0.0071	-0.0136	-0.0041	0.0105	0.0105
7.0	0.	0.	-0.0036	0.0023	0.0071	0.0071	-0.0064	0.0078	0.0105	0.0105
7.5	0.	0.	-0.0046	0.0083	0.0071	0.0071	0.0033	0.0079	0.0106	0.0106
8.0	0.	0.	0.0018	0.0026	0.0071	0.0071	0.0079	0.0027	0.0109	0.0109
8.5	0.	0.	0.0021	0.0025	0.0071	0.0071	0.0081	0.0037	0.0113	0.0113
9.0	0.	0.	0.0025	0.0070	0.0071	0.0071	0.0038	0.0042	0.0120	0.0120
9.5	0.	0.	0.0027	0.0037	0.0071	0.0071	-0.0064	0.0034	0.0126	0.0126
10.0	0.	0.	0.0037	0.0008	0.0071	0.0071	-0.0139	-0.0011	0.0131	0.0131

TABLE 1.5 ERROR SPECTRAL DENSITY. PROCESS 1
ELEMENT 2,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.1000	0.	0.1037	0.	0.0100	0.	0.0177	-0.0356	0.1288	0.1284
0.5	0.1000	0.	0.09[illegible]3	0.	0.0100	0.	0.0674	-0.0291	0.1040	0.1035
1.0	0.1000	0.	0.0942	0.	0.0100	0.	0.1199	-0.0054	0.0667	0.0659
1.5	0.1000	0.	0.1112	-0.	0.0100	0.	0.1392	0.0400	0.0430	0.0419
2.0	0.1000	0.	0.1136	0.	0.0100	0.	0.1459	0.0672	0.0302	0.0285
2.5	0.1000	0.	0.0939	-0.	0.0100	0.	0.1021	0.0375	0.0232	0.0209
3.0	0.1000	0.	0.0867	0.	0.0100	0.	0.0585	0.0084	0.0192	0.0164
3.5	0.1000	0.	0.0999	-0.	0.0100	0.	0.0683	0.0048	0.0168	0.0135
4.0	0.1000	0.	0.1146	0.	0.0100	0.	0.0977	-0.0013	0.0153	0.0116
4.5	0.1000	0.	0.1166	-0.	0.0100	0.	0.1050	-0.0080	0.0143	0.0123
5.0	0.1000	0.	0.1071	0.	0.0100	0.	0.0938	-0.0042	0.0137	0.0093
5.5	0.1000	0.	0.0958	-0.	0.0100	0.	0.0993	0.0063	0.0133	0.0087
6.0	0.1000	0.	0.0949	0.	0.0100	0.	0.1048	0.0029	0.0130	0.0083
6.5	0.1000	0.	0.0917	-0.	0.0100	0.	0.0921	-0.0011	0.0129	0.0081
7.0	0.1000	0.	0.0845	0.	0.0100	0.	0.0866	-0.0007	0.0128	0.0081
7.5	0.1000	0.	0.0823	-0.	0.0100	0.	0.0790	-0.0061	0.0130	0.0082
8.0	0.1000	0.	0.0903	0.	0.0100	0.	0.0884	-0.0025	0.0133	0.0087
8.5	0.1000	0.	0.1082	-0.	0.0100	0.	0.1165	0.0049	0.0138	0.0095
9.0	0.1000	0.	0.1158	0.	0.0100	0.	0.1276	0.0035	0.0145	0.0106
9.5	0.1000	0.	0.1107	-0.	0.0100	0.	0.1183	0.0005	0.0154	0.0117
10.0	0.1000	0.	0.1066	0.	0.0100	0.	0.1122	-0.00[illegible]9	0.0159	0.0123

TABLE 1.6 ERROR SPECTRAL DENSITY. PROCESS :
ELEMENT 2,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0012	0.0022	0.0071	0.0071	-0.0883	0.0176	0.0320	0.0820
0.5	0.	0.	0.0016	0.0020	0.0071	0.0071	-0.0715	0.0096	0.0732	0.0732
1.0	0.	0.	0.0018	0.0029	0.0071	0.0071	-0.0219	-0.0105	0.0562	0.0562
1.5	0.	0.	0.0012	0.0051	0.0071	0.0071	0.0324	-0.0038	0.0412	0.0412
2.0	0.	0.	0.0012	0.0044	0.0071	0.0071	0.0475	0.0225	0.0308	0.0308
2.5	0.	0.	-0.0002	0.0029	0.0071	0.0071	0.0187	0.0220	0.0239	0.0239
3.0	0.	0.	-0.0013	0.0071	0.0071	0.0071	-0.0033	0.0131	0.0195	0.0195
3.5	0.	0.	-0.0022	0.0058	0.0071	0.0071	-0.0058	0.0064	0.0166	0.0166
4.0	0.	0.	-0.0057	-0.0055	0.0071	0.0071	-0.0003	-0.0021	0.0146	0.0146
4.5	0.	0.	-0.0053	-0.0083	0.0071	0.0071	0.0070	-0.0013	0.0133	0.0133
5.0	0.	0.	-0.0014	-0.0045	0.0071	0.0071	0.0040	-0.0071	0.0125	0.0125
5.5	0.	0.	0.0005	-0.0029	0.0071	0.0071	-0.0037	-0.0195	0.0119	0.0119
6.0	0.	0.	-0.0006	0.0017	0.0071	0.0071	-0.0194	-0.0133	0.0115	0.0115
6.5	0.	0.	-0.0020	0.0054	0.0071	0.0071	-0.0093	-0.0017	0.0113	0.0113
7.0	0.	0.	-0.0031	0.0032	0.0071	0.0071	-0.0001	0.0048	0.0113	0.0113
7.5	0.	0.	-0.0063	-0.0003	0.0071	0.0071	0.0021	0.0030	0.0115	0.0115
8.0	0.	0.	-0.0046	-0.0072	0.0071	0.0071	0.0110	-0.0107	0.0119	0.0119
8.5	0.	0.	0.0079	-0.0140	0.0071	0.0071	0.0204	-0.0111	0.0125	0.0125
9.0	0.	0.	0.0120	-0.0075	0.0071	0.0071	0.0213	-0.0027	0.0133	0.0133
9.5	0.	0.	0.0066	0.0005	0.0071	0.0071	0.0145	-0.0029	0.0142	0.0142
10.0	0.	0.	0.0057	0.0014	0.0071	0.0071	0.0104	-0.0042	0.0146	0.0146

TABLE 1.7 ERROR SPECTRAL DENSITY, PROCESS I
ELEMENT 3,1

(1) FREC. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0035	-0.0045	0.0071	0.0071	0.0257	-0.0244	0.0543	0.0543
0.5	0.	0.	0.0023	-0.0026	0.0071	0.0071	-0.0127	-0.0270	0.0502	0.0502
1.0	0.	0.	0.0025	0.0010	0.0071	0.0071	-0.0466	-0.0272	0.0414	0.0414
1.5	0.	0.	-0.0003	-0.0031	0.0071	0.0071	-0.0175	-0.0163	0.0326	0.0326
2.0	0.	0.	0.0010	-0.0045	0.0071	0.0071	-0.0090	-0.0035	0.0259	0.0259
2.5	0.	0.	0.0056	0.0082	0.0071	0.0071	-0.0236	-0.0039	0.0212	0.0212
3.0	0.	0.	0.0084	0.0132	0.0071	0.0071	-0.0071	-0.0019	0.0180	0.0180
3.5	0.	0.	0.0019	0.0092	0.0071	0.0071	0.0115	0.0063	0.0158	0.0158
4.0	0.	0.	-0.0041	0.0078	0.0071	0.0071	0.0200	0.0090	0.0143	0.0143
4.5	0.	0.	-0.0012	0.0024	0.0071	0.0071	0.0228	-0.0023	0.0132	[illegible]
5.0	0.	0.	0.0018	-0.0061	0.0071	0.0071	0.0159	-0.0193	0.0124	0.0124
5.5	0.	0.	0.0037	-0.0012	0.0071	0.0071	0.0120	-0.0161	0.0119	0.0119
6.0	0.	0.	0.0045	0.0068	0.0071	0.0071	0.0098	-0.0013	0.0116	0.0116
6.5	0.	0.	-0.0031	-0.0006	0.0071	0.0071	-0.0060	0.0046	0.0114	0.0114
7.0	0.	0.	-0.0065	-0.0065	0.0071	0.0071	-0.0143	-0.0036	0.0113	0.0113
7.5	0.	0.	0.0023	-0.0003	0.0071	0.0071	0.0045	-0.0004	0.0114	0.0114
8.0	0.	0.	0.0070	0.0073	0.0071	0.0071	0.0140	0.0200	0.0117	0.0117
8.5	0.	0.	0.0017	0.0054	0.0071	0.0071	0.0096	0.0171	0.0122	0.0122
9.0	0.	0.	-0.0006	[illegible]	0.0071	0.0071	0.0116	0.0073	0.0129	0.0129
9.5	0.	0.	0.0039	[illegible]	0.0071	0.0071	0.0053	0.0121	0.0137	0.0137
10.0	0.	0.	0.0058	0.0011	0.0071	0.0071	-0.0027	0.0145	0.0141	0.0141

TABLE 1.8 ERROR SPECTRAL DENSITY. PROCESS 1
ELEMENT 3,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0012	-0.0022	0.0071	0.0071	0.0236	0.0466	0.0821	0.0821
0.5	0.	0.	0.0016	-0.0020	0.0071	0.0071	-0.0110	0.0122	0.0706	0.0706
1.0	0.	0.	0.0018	-0.0029	0.0071	0.0071	-0.0612	-0.0153	0.0515	0.0515
1.5	0.	0.	0.0012	-0.0051	0.0071	0.0071	-0.0592	0.0282	0.0372	0.0372
2.0	0.	0.	0.0012	-0.0044	0.0071	0.0071	-0.0276	0.0474	0.0280	0.0280
2.5	0.	0.	-0.0002	-0.0029	0.0071	0.0071	-0.0104	0.0059	0.0223	0.0223
3.0	0.	0.	-0.0013	-0.0071	0.0071	0.0071	-0.0024	-0.0263	0.0136	0.0186
3.5	0.	0.	-0.0022	-0.0058	0.0071	0.0071	-0.0004	-0.0223	0.0162	0.0162
4.0	0.	0.	-0.0037	0.0055	0.0071	0.0071	0.0037	-0.0033	0.0146	0.0146
4.5	0.	0.	-0.0053	0.0063	0.0071	0.0071	0.0158	-0.0055	0.0135	0.0135
5.0	0.	0.	-0.0014	0.0045	0.0071	0.0071	0.0235	-0.0029	0.0127	0.0127
5.5	0.	0.	0.0005	0.0029	0.0071	0.0071	0.0212	-0.0044	0.0121	0.0121
6.0	0.	0.	-0.0006	-0.0017	0.0071	0.0071	0.0125	-0.0133	0.0118	0.0118
6.5	0.	0.	-0.0020	-0.0054	0.0071	0.0071	0.0066	-0.0201	0.0116	0.0116
7.0	0.	0.	-0.0031	-0.0032	0.0071	0.0071	-0.0053	-0.0199	0.0116	0.0116
7.5	0.	0.	-0.0063	0.0003	0.0071	0.0071	-0.0120	-0.0107	0.0117	0.0117
8.0	0.	0.	-0.0046	0.0072	0.0071	0.0071	0.0010	-0.0073	0.0121	0.0121
8.5	0.	0.	0.0079	0.0149	0.0071	0.0071	0.0194	0.0022	0.0127	0.0127
9.0	0.	0.	0.0120	0.0075	0.0071	0.0071	0.0244	0.0083	0.0136	0.0136
9.5	0.	0.	0.0066	-0.0005	0.0071	0.0071	0.0138	0.0140	0.0147	0.0147
10.0	0.	0.	0.0057	-0.0014	0.0071	0.0071	0.0068	0.0142	0.0152	0.0152

TABLE 1.9 ERROR SPECTRAL DENSITY. PROCESS 1
ELEMENT 3,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.1000	0.	0.0973	0.	0.0100	0.	0.0910	0.0473	0.0528	0.0519
0.5	0.1000	0.	0.0942	0.	0.0100	0.	0.0651	0.0383	0.0504	0.0494
1.0	0.1000	0.	0.0895	0.	0.0100	0.	0.0509	0.0086	0.0442	0.0430
1.5	0.1000	0.	0.0940	-0.	0.0100	0.	0.0875	0.0154	0.0368	0.0354
2.0	0.1000	0.	0.1016	0.	0.0100	0.	0.1014	0.0441	0.0302	0.0285
2.5	0.1000	0.	0.0996	-0.	0.0100	0.	0.0870	0.0354	0.0251	0.0231
3.0	0.1000	0.	0.0959	0.	0.0100	0.	0.0941	0.0136	0.0215	0.0190
3.5	0.1000	0.	0.0962	-0.	0.0100	0.	0.0966	-0.0170	0.0190	0.0161
4.0	0.1000	0.	0.0950	0.	0.0100	0.	0.0891	-0.0260	0.0173	0.0141
4.5	0.1000	0.	0.0887	-0.	0.0100	0.	0.0932	-0.0080	0.0161	0.0126
5.0	0.1000	0.	0.0900	0.	0.0100	0.	0.0978	0.0021	0.0152	0.0115
5.5	0.1000	0.	0.0996	-0.	0.0100	0.	0.0944	-0.0007	0.0147	0.0108
6.0	0.1000	0.	0.0958	0.	0.0100	0.	0.0914	-0.0113	0.0143	0.0103
6.5	0.1000	0.	0.0930	-0.	0.0100	0.	0.1048	-0.0156	0.0142	0.0100
7.0	0.1000	0.	0.1006	0.	0.0100	0.	0.1111	-0.0085	0.0141	0.0100
7.5	0.1000	0.	0.1004	-0.	0.0100	0.	0.0948	0.0015	0.0143	0.0102
8.0	0.1000	0.	0.1019	0.	0.0100	0.	0.0883	0.0051	0.0146	0.0106
8.5	0.1000	0.	0.1035	-0.	0.0100	0.	0.0988	0.0014	0.0152	0.0114
9.0	0.1000	0.	0.0995	0.	0.0100	0.	0.1065	0.0052	0.0160	0.0125
9.5	0.1000	0.	0.1040	-0.	0.0100	0.	0.1071	0.0187	0.0168	0.0135
10.0	0.1000	0.	0.1098	0.	0.0100	0.	0.1067	0.0199	0.0172	0.0140

TABLE 2.1 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 1,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.6250	0.	0.7298	0.	0.0625	0.	0.6723	0.0493	0.0838	0.0553
0.5	0.5722	0.	0.5837	0.	0.0572	0.	0.5632	0.0344	0.0764	0.0506
1.0	0.4572	0.	0.4155	0.	0.0457	0.	0.4141	-0.0127	0.0607	0.0399
1.5	0.3439	0.	0.3664	-0.	0.0344	0.	0.3454	-0.0196	0.0455	0.0299
2.0	0.2570	0.	0.2943	0.	0.0257	0.	0.2500	0.0100	0.0342	0.0226
2.5	0.1955	0.	0.1872	-0.	0.0196	0.	0.1548	0.0004	0.0264	0.0177
3.0	0.1528	-0.	0.1415	0.	0.0153	0.	0.1243	-0.0091	0.0210	0.0144
3.5	0.1227	-0.	0.1224	-0.	0.0123	0.	0.1081	0.0048	0.0172	0.0121
4.0	0.1011	-0.	0.0944	0.	0.0101	0.	0.0869	0.0079	0.0146	0.0106
4.5	0.0853	-0.	0.0757	-0.	0.0085	0.	0.0690	-0.0063	0.0127	0.0094
5.0	0.0735	-0.	0.0683	0.	0.0074	0.	0.0607	-0.0089	0.0114	0.0087
5.5	0.0646	-0.	0.0640	-0.	0.0065	0.	0.0569	-0.0034	0.0104	0.0081
6.0	0.0578	-0.	0.0550	0.	0.0056	0.	0.0446	-0.0112	0.0097	0.0077
6.5	0.0525	-0.	0.0484	-0.	0.0052	0.	0.0433	-0.0156	0.0092	0.0075
7.0	0.0484	-0.	0.0450	0.	0.0048	0.	0.0471	-0.0077	0.0089	0.0075
7.5	0.0453	-0.	0.0446	-0.	0.0045	0.	0.0442	-0.0038	0.0089	0.0077
8.0	0.0429	-0.	0.0472	0.	0.0043	0.	0.0432	-0.0039	0.0091	0.0080
8.5	0.0412	-0.	0.0448	-0.	0.0041	0.	0.0453	-0.0019	0.0096	0.0087
9.0	0.0400	-0.	0.0420	0.	0.0040	0.	0.0436	-0.0019	0.0103	0.0095
9.5	0.0393	-0.	0.0461	-0.	0.0039	0.	0.0465	-0.0051	0.0112	0.0104
10.0	0.0391	0.	0.0495	0.	0.0039	0.	0.0521	-0.0068	0.0115	0.0109

TABLE 2.2 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 1,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	-0.	0.0328	0.0160	0.0598	0.0698	0.0101	-0.0503	0.1119	0.1119
0.5	0.	0.	0.0184	0.0210	0.0610	0.0610	0.0238	-0.0439	0.0959	0.0959
1.0	0.	0.	0.0031	0.0167	0.0445	0.0445	0.0313	-0.0207	0.0683	0.0683
1.5	-0.	0.	-0.0145	-0.0049	0.0308	0.0308	0.0035	-0.0195	0.0472	0.0472
2.0	-0.	0.	-0.0348	-0.0187	0.0218	0.0218	-0.0362	-0.0398	0.0337	0.0337
2.5	-0.	0.	-0.0203	-0.0118	0.0159	0.0159	-0.0287	-0.0412	0.0252	0.0252
3.0	-0.	0.	-0.0036	0.0004	0.0121	0.0121	-0.0006	-0.0209	0.0198	0.0198
3.5	-0.	0.	0.0019	0.0005	0.0095	0.0095	0.0147	0.0050	0.0162	0.0162
4.0	-0.	0.	0.0092	-0.0010	0.0077	0.0077	0.0262	0.0137	0.0137	0.0137
4.5	-0.	0.	0.0059	0.0029	0.0065	0.0065	0.0115	0.0021	0.0120	[illegible]
5.0	-0.	0.	0.0035	0.0022	0.0055	0.0055	0.0036	0.0045	0.0107	[illegible]
5.5	-0.	0.	0.0012	0.0007	0.0046	0.0048	0.0039	0.0074	0.0099	0.0099
6.0	-0.	0.	-0.0021	0.0002	0.0043	0.0043	0.0016	0.0056	0.0093	[illegible]
6.5	-0.	0.	0.0003	-0.0016	0.0039	0.0039	-0.0040	-0.0063	0.0089	[illegible]
7.0	-0.	0.	0.0028	0.0011	0.0036	0.0036	-0.0003	-0.0095	0.0087	[illegible]
7.5	-0.	0.	0.0021	0.0043	0.0033	0.0033	0.0086	0.0025	0.0098	[illegible]
8.0	-0.	0.	0.0018	0.0009	0.0032	0.0032	0.0050	0.0054	0.0092	0.0092
8.5	-0.	0.	0.0017	-0.0010	0.0030	0.0039	-0.0016	-0.0093	0.0099	0.0099
9.0	-0.	0.	0.0009	0.0031	0.0029	0.0029	-0.0014	-0.0057	0.0109	[illegible]
9.5	-0.	0.	-0.0023	0.0041	0.0029	0.0029	-0.0034	-0.0132	0.0120	0.0120
10.0	-0.	-0.	-0.0045	0.0025	0.0029	0.0029	-0.0059	-0.0155	0.0125	0.0125

TABLE 2.3 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 1,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S'		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	-0.	0.0358	-0.0141	0.0424	0.0424	0.0208	-0.0731	0.0683	0.0683
0.5	0.	0.	0.0270	-0.0406	0.0396	0.0396	0.0162	-0.0905	0.0639	0.0639
1.0	0.	0.	-0.0379	-0.0542	0.0332	0.0332	0.0051	-0.0514	0.0537	0.0537
1.5	0.	0.	-0.0164	-0.0044	0.0263	0.0263	0.0096	0.0156	0.0427	0.0427
2.0	0.	0.	-0.0221	0.0219	0.0205	0.0205	0.0014	0.0006	0.0336	0.0336
2.5	0.	0.	-0.0236	0.0073	0.0161	0.0161	-0.0123	-0.0115	0.0267	0.0267
3.0	0.	0.	-0.0233	0.0065	0.0129	0.0129	-0.0074	0.0112	0.0219	0.0218
3.5	-0.	0.	-0.0219	0.0017	0.0105	0.0105	-0.0049	0.0076	0.0183	0.0183
4.0	-0.	0.	-0.0095	-0.0063	0.0086	0.0088	0.0100	-0.0136	0.0158	0.0158
4.5	-0.	0.	-0.0015	-0.0246	0.0075	0.0075	0.0227	-0.0127	0.0140	0.0140
5.0	-0.	0.	0.0016	0.0004	0.0065	0.0065	0.0178	0.0126	0.0127	0.0127
5.5	-0.	0.	0.0018	0.0039	0.0059	0.0058	0.0051	0.0340	0.0117	0.0117
6.0	-0.	0.	0.0022	-0.0012	0.0052	0.0052	0.0009	0.0246	0.0110	0.0110
6.5	-0.	0.	0.0081	-0.0054	0.0047	0.0047	0.0052	-0.0021	0.0106	0.0106
7.0	-0.	0.	0.0074	-0.0045	0.0044	0.0044	0.0042	-0.0114	0.0104	0.0104
7.5	-0.	0.	0.0011	-0.0023	0.0041	0.0041	0.0052	-0.0061	0.0105	0.0105
8.0	-0.	0.	-0.0022	-0.0014	0.0039	0.0039	0.0128	0.0027	0.0108	0.0108
8.5	-0.	0.	-0.0035	-0.0012	0.0037	0.0037	0.0152	0.0021	0.0114	0.0114
9.0	-0.	0.	-0.0046	0.0005	0.0036	0.0036	0.0070	-0.0100	0.0122	0.0122
9.5	-0.	0.	-0.0079	0.0016	0.0036	0.0036	0.0057	-0.0085	0.0131	0.0131
10.0	-0.	-0.	-0.0107	0.0014	0.0035	0.0035	0.0104	-0.0016	0.0135	0.0135

TABLE 2.4 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 2.1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0328	-0.0160	0.0698	0.0698	0.0551	-0.1100	0.1077	0.1077
0.5	0.	0.	0.0134	-0.0210	0.0610	0.0610	0.0179	-0.1197	0.0935	0.0935
1.0	0.	0.	0.0034	-0.0167	0.0445	0.0445	0.0055	-0.0856	0.0675	0.0675
1.5	-0.	-0.	-0.0145	0.0049	0.0308	0.0308	-0.0100	-0.0314	0.0467	0.0467
2.0	-0.	-0.	-0.0348	0.0187	0.0218	0.0218	-0.0537	0.0048	0.0332	0.0332
2.5	-0.	-0.	-0.0200	0.0118	0.0159	0.0159	-0.0384	0.0348	0.0247	0.0247
3.0	-0.	-0.	-0.0036	-0.0004	0.0121	0.0121	-0.0074	0.0307	0.0193	0.0193
3.5	-0.	-0.	0.0019	-0.0005	0.0095	0.0095	0.0095	0.0149	0.0157	0.0157
4.0	-0.	-0.	0.0092	0.0010	0.0077	0.0077	0.0176	0.0017	0.0133	0.0133
4.5	-0.	-0.	0.0059	-0.0029	0.0065	0.0065	0.0011	-0.0037	0.0116	0.0116
5.0	-0.	-0.	0.0035	-0.0022	0.0055	0.0055	-0.0044	-0.0050	0.0104	0.0104
5.5	-0.	-0.	0.0012	-0.0007	0.0048	0.0048	0.0024	-0.0069	0.0095	0.0095
6.0	-0.	-0.	-0.0021	-0.0002	0.0043	0.0043	0.0071	-0.0044	0.0089	0.0089
6.5	-0.	-0.	0.0008	0.0016	0.0039	0.0039	0.0112	0.0092	0.0086	0.0086
7.0	-0.	-0.	0.0028	-0.0011	0.0036	0.0036	0.0100	0.0023	0.0084	0.0084
7.5	-0.	-0.	0.0021	-0.0043	0.0033	0.0033	0.0060	-0.0010	0.0084	0.0084
8.0	-0.	-0.	0.0018	-0.0009	0.0032	0.0032	0.0024	-0.0055	0.0087	0.0087
8.5	-0.	-0.	0.0017	0.0010	0.0030	0.0030	-0.0031	-0.0020	0.0093	0.0093
9.0	-0.	-0.	0.0009	-0.0031	0.0029	0.0029	-0.0025	0.0009	0.0101	0.0101
9.5	-0.	-0.	-0.0023	-0.0041	0.0029	0.0029	0.0064	-0.0003	0.0109	0.0109
10.0	-0.	-0.	-0.0045	-0.0025	0.0029	0.0029	0.0115	-0.0002	0.0113	0.0113

TABLE 2.5 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 2,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	1.5596	-0.	1.1278	-0.	0.1560	0.	1.1672	0.0356	0.2020	0.1284
0.5	1.3001	0.	1.0[illegible]	-0.	0.1300	0.	1.0001	0.0291	0.1662	0.1035
1.0	0.8646	0.	0.7433	-0.	0.0865	0.	0.7072	0.0054	0.1087	0.0659
1.5	0.5534	0.	0.5283	0.	0.0553	0.	0.5242	-0.0400	0.0694	0.0419
2.0	0.3683	0.	0.3780	-0.	0.0368	0.	0.3762	-0.0672	0.0466	0.0285
2.5	0.2589	-0.	0.2579	0.	0.0259	0.	0.2498	-0.0375	0.0333	0.0229
3.0	0.1916	-0.	0.1864	-0.	0.0192	0.	0.1929	-0.0084	0.0252	0.0164
3.5	0.1481	-0.	0.1393	0.	0.0148	0.	0.1524	-0.0048	0.0200	0.0135
4.0	0.1187	-0.	0.1088	-0.	0.0119	0.	0.1191	0.0013	0.0165	0.0116
4.5	0.0982	-0.	0.0917	0.	0.0098	0.	0.0937	0.0080	0.0142	0.0103
5.0	0.0834	-0.	0.0793	-0.	0.0083	0.	0.0748	0.0042	0.0125	0.0093
5.5	0.0724	-0.	0.0749	0.	0.0072	0.	0.0694	-0.0063	0.0113	0.0087
6.0	0.0642	-0.	0.0738	-0.	0.0064	0.	0.0733	-0.0029	0.0105	0.0083
6.5	0.0579	-0.	0.0627	0.	0.0058	0.	0.0645	0.0011	0.0099	0.0081
7.0	0.0532	-0.	0.0484	-0.	0.0053	0.	0.0491	0.0007	0.0096	0.0081
7.5	0.0495	-0.	0.0486	0.	0.0050	0.	0.0475	0.0061	0.0096	0.0082
8.0	0.0468	-0.	0.0531	-0.	0.0047	0.	0.0423	0.0025	0.0099	0.0087
8.5	0.0448	-0.	0.0464	0.	0.0045	0.	0.0296	-0.0049	0.0105	0.0095
9.0	0.0434	-0.	0.0362	-0.	0.0043	0.	0.0290	-0.0035	0.0114	0.0106
9.5	0.0426	-0.	0.0306	0.	0.0043	0.	0.0341	-0.0005	0.0125	0.0117
10.0	0.0424	0.	0.0285	-0.	0.0042	0.	0.0353	0.0009	0.0130	0.0123

TABLE 2.6 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 2,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.3520	-0.	0.1361	-0.0079	0.0715	0.0622	0.2074	-0.0062	0.1063	0.1003
0.5	0.2879	-0.1068	0.1532	-0.0301	0.0627	0.0567	0.2098	-0.0249	0.0942	0.0903
1.0	0.1823	-0.1252	0.1502	-0.0649	0.0466	0.0447	0.1477	-0.0532	0.0713	0.0791
1.5	0.1090	-0.1001	0.0924	-0.0713	0.0335	0.0332	0.0490	-0.0751	0.0519	0.0517
2.0	0.0677	-0.0716	0.0529	-0.0699	0.0245	0.0246	0.0174	-0.0950	0.0384	0.0385
2.5	0.0447	-0.0500	0.0453	-0.0642	0.0185	0.0186	0.0336	-0.0728	0.0294	0.0295
3.0	0.0314	-0.0352	0.0381	-0.0276	0.0144	0.0145	0.0363	-0.0167	0.0234	0.0235
3.5	0.0232	-0.0252	0.0210	0.0030	0.0116	0.0116	0.0200	0.0108	0.0193	0.0194
4.0	0.0180	-0.0184	0.0105	-0.0032	0.0095	0.0095	0.0143	-0.0058	0.0165	0.0165
4.5	0.0144	-0.0136	0.0055	-0.0123	0.0080	0.0080	0.0139	-0.0273	0.0145	0.0145
5.0	0.0120	-0.0103	0.0051	-0.0059	0.0069	0.0069	0.0110	-0.0161	0.0132	0.0130
5.5	0.0102	-0.0079	0.0143	-0.0026	0.0061	0.0061	0.0226	0.0053	0.0120	0.0120
6.0	0.0089	-0.0061	0.0246	-0.0041	0.0055	0.0054	0.0404	0.0119	0.0113	0.0112
6.5	0.0080	-0.0047	0.0230	-0.0041	0.0050	0.0049	0.0265	0.0078	0.0108	0.0108
7.0	0.0072	-0.0037	0.0127	-0.0022	0.0046	0.0045	-0.0002	0.0009	0.0106	0.0106
7.5	0.0067	-0.0028	0.0087	-0.0010	0.0043	0.0042	-0.0059	0.0031	0.0106	0.0106
8.0	0.0063	-0.0021	0.0117	-0.0043	0.0041	0.0040	-0.0031	0.0074	0.0109	0.0109
8.5	0.0060	-0.0015	0.0091	-0.0054	0.0039	0.0039	-0.0011	0.0025	0.0115	0.0115
9.0	0.0058	-0.0010	0.0060	-0.0026	0.0038	0.0037	0.0048	-0.0001	0.0124	0.0124
9.5	0.0057	-0.0025	0.0037	-0.0023	0.0037	0.0037	0.0040	-0.0003	0.0133	0.0133
10.0	0.0056	0.	0.0009	-0.0031	0.0037	0.0037	0.0024	-0.0016	0.0137	0.0137

TABLE 2.7 ERROR SPECTRAL DENSITY, PROCESS 2
ELEMENT 3,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0358	0.0141	0.0424	0.0424	0.0815	0.0291	0.0687	0.0687
0.5	0.	0.	0.0210	0.0406	0.0396	0.0396	0.0762	0.0567	0.0637	0.0637
1.0	0.	0.	-0.0079	0.0542	0.0332	0.0332	0.0502	0.0676	0.0524	0.0524
1.5	0.	0.	-0.0154	0.0044	0.0263	0.0263	0.0142	0.0053	0.0409	0.0409
2.0	0.	0.	-0.0221	-0.0219	0.0205	0.0205	0.0048	-0.0126	0.0317	0.0317
2.5	0.	0.	-0.0265	-0.0073	0.0161	0.0161	0.0207	0.0280	0.0249	0.0249
3.0	0.	0.	-0.0233	-0.0066	0.0129	0.0129	0.0076	0.0173	0.0201	0.0201
3.5	-0.	-0.	-0.0210	-0.0017	0.0105	0.0105	-0.0240	0.0059	0.0167	0.0167
4.0	-0.	-0.	-0.0095	0.0063	0.0088	0.0088	-0.0241	0.0150	0.0143	0.0143
4.5	-0.	-0.	-0.0016	0.0046	0.0075	0.0075	-0.0074	0.0110	0.0125	0.0125
5.0	-0.	-0.	0.0016	-0.0004	0.0065	0.0065	-0.0039	0.0065	0.0112	0.0112
5.5	-0.	-0.	0.0018	-0.0030	0.0058	0.0058	-0.0032	0.0016	0.0103	0.0103
6.0	-0.	-0.	0.0022	0.0012	0.0052	0.0052	0.0014	0.0062	0.0097	0.0097
6.5	-0.	-0.	0.0081	0.0054	0.0047	0.0047	0.0071	0.0121	0.0093	0.0093
7.0	-0.	-0.	0.0074	0.0045	0.0044	0.0044	0.0071	0.0105	0.0091	0.0091
7.5	-0.	-0.	0.0011	0.0023	0.0041	0.0041	0.0022	0.0050	0.0091	0.0091
8.0	-0.	-0.	-0.0022	0.0011	0.0039	0.0039	-0.0613	-0.0091	0.0094	0.0094
8.5	-0.	-0.	-0.0035	0.0012	0.0037	0.0037	-0.0042	-0.0055	0.0100	0.0100
9.0	-0.	-0.	-0.0046	-0.0005	0.0036	0.0036	-0.0142	0.0015	0.0108	0.0108
9.5	-0.	-0.	-0.0079	-0.0016	0.0036	0.0036	-0.0146	-0.0095	0.0117	0.0117
10.0	-0.	-0.	-0.0107	-0.0014	0.0035	0.0035	-0.0085	-0.0153	0.0122	0.0122

TABLE 2.8 ERROR SPECTRAL DENSITY, PROCESS 2
ELEMENT 3,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.3520	-0.	0.1361	0.0079	0.0715	0.0622	0.0954	-0.0579	0.1109	0.1051
0.5	0.2879	0.1068	0.1582	0.0301	0.0627	0.0567	0.1493	0.0031	0.0961	0.0923
1.0	0.1820	0.1252	0.1502	0.0649	0.0466	0.0447	0.1870	0.0790	0.0704	0.0692
1.5	0.1090	0.1003	0.0924	0.0713	0.0335	0.0332	0.1407	0.0507	0.0523	0.0521
2.0	0.0677	0.0716	0.0539	0.0699	0.0245	0.0246	0.0926	0.0251	0.0362	0.0359
2.5	0.0447	0.0500	0.0452	0.0642	0.0185	0.0186	0.0628	0.0448	0.0280	0.0281
3.0	0.0314	0.0352	0.0361	0.0276	0.0144	0.0145	0.0355	0.0304	0.0222	0.0222
3.5	0.0232	0.0252	0.0210	-0.0030	0.0116	0.0116	0.0145	0.0050	0.0132	0.0132
4.0	0.0130	0.0124	0.0105	0.0032	0.0095	0.0095	0.0143	0.0162	0.0154	0.0154
4.5	0.0144	0.0136	0.0055	0.0123	0.0080	0.0080	0.0030	0.0341	0.0134	0.0134
5.0	0.0120	0.0103	0.0051	0.0059	0.0069	0.0069	-0.0066	0.0262	0.0120	0.0120
5.5	0.0102	0.0079	0.0143	0.0026	0.0061	0.0061	-0.0073	0.0186	0.0109	0.0109
6.0	0.0069	0.0051	0.0246	0.0041	0.0055	0.0054	0.0034	0.0143	0.0102	0.0102
6.5	0.0080	0.0247	0.0237	0.0041	0.0050	0.0049	0.0136	0.0139	0.0098	0.0098
7.0	0.0072	0.0037	0.0127	0.0022	0.0046	0.0045	0.0050	0.0143	0.0096	0.0096
7.5	0.0067	0.0028	0.0087	0.0010	0.0043	0.0042	0.0083	0.0045	0.0097	0.0096
8.0	0.0063	0.0021	0.0117	0.0043	0.0041	0.0040	0.0070	0.0036	0.0100	0.0100
8.5	0.0060	0.0015	0.0091	0.0054	0.0039	0.0039	-0.0021	0.0264	0.0107	0.0107
9.0	0.0058	0.0010	0.0060	0.0026	0.0036	0.0037	0.0017	-0.0055	0.0118	0.0113
9.5	0.0057	0.0005	0.0037	0.0023	0.0037	0.0037	0.0047	-0.0109	0.0130	0.0139
10.0	0.0056	0.	0.0069	0.0031	0.0037	0.0037	0.0041	-0.0084	0.0136	0.0136

TABLE 2.9 ERROR SPECTRAL DENSITY. PROCESS 2
ELEMENT 3,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.5760	-0.	0.5005	-0.	0.0576	0.	0.4996	-0.0473	0.0775	0.0519
0.5	0.5499	0.	0.5029	-0.	0.0549	0.	0.5270	-0.0383	0.0738	0.0494
1.0	0.4817	0.	0.4618	-0.	0.0482	0.	0.5028	-0.0036	0.0646	0.0430
1.5	0.4011	0.	0.3528	0.	0.0401	0.	0.3613	-0.0154	0.0535	0.0354
2.0	0.3265	0.	0.2738	-0.	0.0327	0.	0.2650	-0.0441	0.0433	0.0285
2.5	0.2652	0.	0.2627	0.	0.0265	0.	0.2660	-0.0354	0.0351	0.0231
3.0	0.2175	0.	0.2393	-0.	0.0217	0.	0.2458	-0.0136	0.0289	0.0190
3.5	0.1809	-0.	0.1959	0.	0.0181	0.	0.2077	0.0170	0.0242	0.0161
4.0	0.1530	-0.	0.1671	-0.	0.0153	0.	0.1840	0.0260	0.0208	0.0141
4.5	0.1317	-0.	0.1438	0.	0.0132	0.	0.1554	0.0080	0.0182	0.0126
5.0	0.1152	-0.	0.1162	-0.	0.0115	0.	0.1149	-0.0021	0.0163	0.0115
5.5	0.1024	-0.	0.1034	0.	0.0102	0.	0.0941	0.0007	0.0149	0.0108
6.0	0.0924	-0.	0.1034	-0.	0.0092	0.	0.0989	0.0113	0.0138	0.0103
6.5	0.0845	-0.	0.0942	0.	0.0085	0.	0.0944	0.0156	0.0131	0.0100
7.0	0.0784	-0.	0.0805	-0.	0.0078	0.	0.0792	0.0085	0.0127	0.0100
7.5	0.0736	-0.	0.0741	0.	0.0074	0.	0.0700	-0.0015	0.0126	0.0102
8.0	0.0699	-0.	0.0783	-0.	0.0070	0.	0.0882	-0.0051	0.0127	0.0106
8.5	0.0673	-0.	0.0748	0.	0.0067	0.	0.0972	-0.0014	0.0132	0.0114
9.0	0.0654	-0.	0.0670	-0.	0.0065	0.	0.0800	-0.0082	0.0141	0.0125
9.5	0.0644	-0.	0.0692	0.	0.0064	0.	0.0690	-0.0187	0.0150	0.0135
10.0	0.0640	0.	0.0724	-0.	0.0064	0.	0.0670	-0.0198	0.0154	0.0140

TABLE 3.1 ERROR SPECTRAL DENSITY, PROCESS 3
ELEMENT 1,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.7716	0.	0.6283	0.	0.0772	0.	0.6122	-0.0493	0.0952	0.0558
0.5	0.6774	0.	0.6011	0.	0.0677	0.	0.5800	-0.0344	0.0846	0.0506
1.0	0.4886	0.	0.4999	0.	0.0489	0.	0.4866	0.0127	0.0631	0.0399
1.5	0.3242	-0.	0.3376	-0.	0.0324	0.	0.3474	0.0196	0.0441	0.0299
2.0	0.2136	-0.	0.2165	0.	0.0214	0.	0.2405	-0.0100	0.0311	0.0226
2.5	0.1451	-0.	0.1563	-0.	0.0145	0.	0.1719	-0.0014	0.0229	0.0177
3.0	0.1030	0.	0.1080	0.	0.0103	0.	0.1121	0.0091	0.0177	0.0144
3.5	0.0770	0.	0.0775	-0.	0.0077	0.	0.0798	-0.0048	0.0144	0.0121
4.0	0.0606	0.	0.0633	0.	0.0061	0.	0.0584	-0.0079	0.0122	0.0106
4.5	0.0503	-0.	0.0483	0.	0.0050	0.	0.0451	0.0063	0.0107	0.0094
5.0	0.0440	-0.	0.0390	0.	0.0044	0.	0.0375	0.0089	0.0097	0.0087
5.5	0.0405	-0.	0.0378	-0.	0.0041	0.	0.0379	0.0034	0.0091	0.0081
6.0	0.0393	-0.	0.0409	0.	0.0039	0.	0.0456	0.0112	0.0087	0.0077
6.5	0.0403	-0.	0.0436	-0.	0.0040	0.	0.0461	0.0156	0.0086	0.0075
7.0	0.0436	0.	0.0433	0.	0.0044	0.	0.0427	0.0077	0.0087	0.0075
7.5	0.0500	0.	0.0463	-0.	0.0050	0.	0.0380	0.0038	0.0092	0.0077
8.0	0.0606	0.	0.0597	0.	0.0061	0.	0.0543	0.0039	0.0101	0.0080
8.5	0.0774	0.	0.0825	-0.	0.0077	0.	0.0862	0.0019	0.0116	0.0087
9.0	0.1015	-0.	0.1083	0.	0.0102	0.	0.1188	0.0019	0.0139	0.0095
9.5	0.1285	-0.	0.1299	-0.	0.0128	0.	0.1253	0.0051	0.0166	0.0104
10.0	0.1417	0.	0.1384	0.	0.0142	0.	0.1194	0.0068	0.0179	0.0109

TABLE 3.2 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 1,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	-0.	-0.0572	0.0135	0.0855	0.0855	0.0048	0.0572	0.1227	0.1227
0.5	0.	0.	-0.0478	0.0249	0.0698	0.0698	-0.0283	0.0703	0.1031	0.1031
1.0	-0.	0.	-0.0265	0.0233	0.0445	0.0445	-0.0375	0.0653	0.0702	0.0702
1.5	-0.	0.	-0.0370	0.0138	0.0271	0.0271	-0.0128	0.0589	0.0462	0.0462
2.0	-0.	-0.	-0.0027	0.0097	0.0169	0.0169	-0.0122	0.0614	0.0316	0.0316
2.5	-0.	-0.	-0.0037	0.0093	0.0112	0.0112	-0.0067	0.0301	0.0230	0.0230
3.0	0.	0.	0.0030	0.0054	0.0078	0.0078	0.0010	0.0009	0.0177	0.0177
3.5	0.	0.	0.0011	0.0035	0.0058	0.0058	-0.0073	-0.0085	0.0144	0.0144
4.0	-0.	0.	-0.0036	0.0016	0.0046	0.0046	-0.0181	-0.0083	0.0123	0.0123
4.5	-0.	0.	-0.0004	-0.0020	0.0038	0.0038	-0.0062	-0.0012	0.0108	0.0108
5.0	-0.	0.	0.0069	-0.0039	0.0033	0.0033	0.0084	-0.0011	0.0098	0.0098
5.5	-0.	0.	0.0029	-0.0024	0.0030	0.0030	0.0103	-0.0108	0.0091	0.0091
6.0	-0.	0.	0.0041	-0.0025	0.0030	0.0030	0.0072	-0.0105	0.0087	0.0087
6.5	-0.	0.	0.0063	-0.0042	0.0030	0.0030	0.0011	-0.0093	0.0085	0.0085
7.0	-0.	-0.	-0.0019	-0.0024	0.0033	0.0033	-0.0070	0.0067	0.0086	0.0086
7.5	-0.	-0.	-0.0017	0.0013	0.0038	0.0038	-0.0137	0.0028	0.0090	0.0090
8.0	0.	0.	0.0024	0.0000	0.0046	0.0046	-0.0075	-0.0034	0.0097	0.0097
8.5	0.	0.	0.0064	-0.0052	0.0060	0.0060	0.0084	-0.0040	0.0110	0.0110
9.0	-0.	0.	0.0020	-0.0120	0.0080	0.0080	0.0101	-0.0038	0.0130	0.0130
9.5	-0.	0.	-0.0075	-0.0134	0.0102	0.0102	-0.0032	-0.0004	0.0153	0.0153
10.0	-0.	-0.	-0.0117	-0.0108	0.0114	0.0114	-0.0104	0.0021	0.0165	0.0165

TABLE 3.3 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 1,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	-0.	0.0592	0.0133	0.0524	0.0524	0.0456	-0.0044	0.0749	0.0749
0.5	0.	0.	0.0493	0.0181	0.0475	0.0475	0.0736	-0.0124	0.0687	0.0687
1.0	0.	0.	0.0364	0.0254	0.0368	0.0368	0.0920	-0.0191	0.0552	0.0552
1.5	-0.	0.	0.0241	0.0288	0.0261	0.0261	0.0634	0.0114	0.0417	0.0417
2.0	-0.	0.	0.0027	0.0184	0.0180	0.0180	0.0211	0.0308	0.0315	0.0315
2.5	-0.	0.	-0.0075	-0.0015	0.0126	0.0126	0.0052	-0.0002	0.0243	0.0243
3.0	0.	0.	-0.0362	-0.0075	0.0091	0.0091	-0.0100	-0.0127	0.0196	0.0196
3.5	0.	0.	-0.0386	-0.0057	0.0069	0.0068	-0.0202	-0.0067	0.0163	0.0163
4.0	0.	0.	-0.0060	-0.0056	0.0054	0.0054	-0.0223	0.0007	0.0141	0.0141
4.5	0.	0.	-0.0030	-0.0097	0.0045	0.0045	-0.0265	0.0074	0.0126	0.0126
5.0	-0.	0.	-0.0040	0.0318	0.0039	0.0039	-0.0209	-0.0082	0.0115	0.0115
5.5	-0.	0.	-0.0010	0.0005	0.0036	0.0036	-0.0067	-0.0243	0.0108	0.0108
6.0	-0.	0.	0.0029	0.0009	0.0035	0.0035	0.0071	-0.0184	0.0104	0.0104
6.5	-0.	0.	0.0023	-0.0004	0.0036	0.0036	0.0043	-0.0012	0.0102	0.0102
7.0	-0.	-0.	-0.0005	-0.0042	0.0038	0.0038	0.0035	-0.0009	0.0102	0.0102
7.5	-0.	-0.	-0.0030	-0.0047	0.0044	0.0044	-0.0079	-0.0043	0.0106	0.0106
8.0	0.	0.	-0.0043	-0.0036	0.0053	0.0053	-0.0162	-0.0055	0.0114	0.0114
8.5	0.	0.	-0.0053	-0.0047	0.0067	0.0067	-0.0174	-0.0029	0.0127	0.0127
9.0	-0.	0.	-0.0067	0.0002	0.0087	0.0087	-0.0145	0.0143	0.0147	0.0147
9.5	-0.	0.	-0.0241	0.0085	0.0108	0.0108	-0.0112	0.0202	0.0168	0.0168
10.0	-0.	-0.	-0.0033	0.0169	0.0119	0.0119	-0.0099	0.0137	0.0173	0.0173

TABLE 3.4 ERROR SPECTRAL DENSITY, PROCESS 3
ELEMENT 2,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	-0.0572	-0.0135	0.0855	0.0855	-0.0451	0.1039	0.1180	0.1180
0.5	0.	0.	-0.0478	-0.0249	0.0698	0.0698	-0.0224	0.0933	0.0930	0.0930
1.0	-0.	-0.	-0.0265	-0.0283	0.0445	0.0445	-0.0067	0.0409	0.0657	0.0657
1.5	-0.	-0.	-0.0071	-0.0138	0.0271	0.0271	0.0068	-0.0083	0.0429	0.0429
2.0	-0.	-0.	-0.0027	-0.0097	0.0169	0.0169	0.0053	-0.0264	0.0294	0.0294
2.5	-0.	-0.	-0.0037	-0.0093	0.0112	0.0112	0.0030	-0.0237	0.0215	0.0215
3.0	0.	-0.	0.0070	-0.0054	0.0078	0.0078	0.0078	-0.0196	0.0167	0.0167
3.5	0.	-0.	0.0013	-0.0035	0.0058	0.0058	-0.0021	-0.0114	0.0137	0.0137
4.0	-0.	0.	-0.0036	-0.0016	0.0046	0.0046	-0.0095	-0.0071	0.0117	0.0117
4.5	-0.	0.	-0.0004	0.0020	0.0038	0.0038	0.0041	-0.0042	0.0103	0.0103
5.0	-0.	0.	0.0009	0.0039	0.0033	0.0033	0.0134	0.0016	0.0094	0.0094
5.5	-0.	-0.	0.0029	0.0024	0.0030	0.0030	0.0118	0.0103	0.0088	0.0088
6.0	-0.	-0.	0.0041	0.0025	0.0030	0.0030	0.0017	0.0094	0.0084	0.0084
6.5	-0.	-0.	0.0005	0.0042	0.0030	0.0030	-0.0140	0.0064	0.0082	0.0082
7.0	-0.	-0.	-0.0019	0.0024	0.0033	0.0033	-0.0173	0.0005	0.0083	0.0083
7.5	-0.	-0.	-0.0017	-0.0013	0.0038	0.0038	-0.0111	-0.0043	0.0087	0.0087
8.0	0.	-0.	0.0024	-0.0005	0.0046	0.0046	-0.0049	0.0035	0.0094	0.0094
8.5	0.	-0.	0.0364	0.0052	0.0060	0.0060	0.0100	0.0064	0.0107	0.0107
9.0	-0.	0.	0.0020	0.0120	0.0080	0.0080	0.0112	0.0086	0.0126	0.0125
9.5	-0.	0.	-0.0075	0.0134	0.0102	0.0102	-0.0130	0.0142	0.0149	0.0149
10.0	-0.	-0.	-0.0117	0.0109	0.0114	0.0114	-0.0278	0.0137	0.0160	0.0160

TABLE 3.5 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 2.2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	1.8942	-0.	1.7677	0.	0.1894	0.	1.6537	-0.0356	0.2289	0.1284
0.5	1.4372	0.	1.4053	0.	0.1457	0.	1.4150	-0.0291	0.1771	0.1935
1.0	0.8118	-0.	0.8110	0.	0.0912	0.	0.7526	-0.0054	0.1046	0.0659
1.5	0.4514	-0.	0.4626	-0.	0.0451	0.	0.4221	0.0409	0.0616	0.0419
2.0	0.2682	-0.	0.2350	0.	0.0263	0.	0.2694	0.0672	0.0391	0.0285
2.5	0.1721	0.	0.1862	-0.	0.0172	0.	0.1984	0.0375	0.0271	0.0229
3.0	0.1136	0.	0.1253	0.	0.0119	0.	0.1559	0.0024	0.0202	0.0164
3.5	0.0873	0.	0.0969	-0.	0.0087	0.	0.1280	0.0048	0.0161	0.0135
4.0	0.0683	0.	0.0801	0.	0.0068	0.	0.0971	-0.0013	0.0134	0.0116
4.5	0.0566	-0.	0.0567	-0.	0.0057	0.	0.0677	-0.0080	0.0117	[illegible]
5.0	0.0495	-0.	0.0453	0.	0.0050	0.	[illegible]	-0.0042	0.0106	0.0093
5.5	0.0457	-0.	0.0449	-0.	0.0046	0.	0.0425	0.0063	0.0098	0.0097
6.0	0.0445	-0.	0.0439	0.	0.0045	0.	0.0390	0.0029	0.0094	0.0083
6.5	0.0459	-0.	0.0472	-0.	0.0046	0.	0.0430	-0.0011	0.0093	0.0081
7.0	0.0500	0.	0.0532	0.	[illegible]	0.	0.0517	-0.0007	0.0095	0.0081
7.5	0.0579	0.	0.0595	-0.	0.0058	0.	0.0611	-0.0061	0.0101	0.0082
8.0	0.0712	0.	0.0676	0.	0.0071	0.	0.0742	-0.0025	0.0112	0.0087
8.5	0.0928	0.	0.0846	-0.	0.0093	0.	0.0912	0.0049	0.0133	0.0095
9.0	0.1252	-0.	0.1129	0.	0.0125	0.	0.1142	0.0035	0.0164	0.0106
9.5	0.1635	-0.	0.1344	-0.	0.0164	0.	0.1357	0.0005	0.0201	0.0117
10.0	0.1832	0.	0.1405	0.	0.0183	0.	0.1436	-0.0009	0.0221	0.0123

TABLE 3.6 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 2,3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.4082	-0.	0.4779	-0.1227	0.0876	0.0762	0.4464	-0.0850	0.1169	0.1095
0.5	0.3032	-0.1422	0.3783	-0.1390	0.0717	0.0666	0.3625	-0.1262	0.0985	0.0949
1.0	0.1519	-0.1366	0.21[illegible]6	-0.0973	0.0475	0.0470	0.2476	-0.1113	0.0690	0.0634
1.5	0.0238	-0.0886	0.1293	-0.0333	0.0307	0.0309	0.1983	-0.0172	0.0476	0.0476
2.0	0.0465	-0.0512	0.0868	-0.0394	0.0201	0.0203	0.1431	0.0009	0.0340	0.0341
2.5	0.0222	-0.0290	0.0496	-0.0367	0.0137	0.0137	0.0729	-0.0324	0.0256	0.0257
3.0	0.0186	-0.0167	0.0210	-0.0134	0.0097	0.0097	0.0227	-0.0397	0.0203	0.0203
3.5	0.0133	-0.0109	0.009[illegible]	-0.0095	0.0073	0.0072	0.0094	-0.0289	0.0168	0.0168
4.0	0.0103	-0.0062	0.0100	-0.0083	0.0058	0.0057	0.0087	-0.0185	0.0145	0.0145
4.5	0.0084	-0.0039	0.0090	-0.0039	0.0048	0.0047	-0.0047	-0.0025	0.0129	0.0129
5.0	0.0073	-0.0025	0.0045	-0.0074	0.0042	0.0041	-0.0160	0.0023	0.0118	0.0117
5.5	0.0068	-0.0016	0.0026	-0.0059	0.003[illegible]	0.0038	-0.0197	-0.0049	0.0110	0.0113
6.0	0.0065	-0.0008	0.0044	-0.0020	0.0037	0.0037	-0.0224	-0.0101	0.0106	0.0105
6.5	0.0069	-0.000[illegible]	0.0056	-0.0009	0.0032	0.0038	-0.0133	-0.01[illegible]6	0.0104	0.0104
7.0	0.0077	0.0005	0.0089	0.0036	0.0041	0.0041	0.0087	-0.0001	0.0105	0.0104
7.5	0.0092	0.0013	0.0148	0.0075	0.0048	0.0047	0.0262	0.0026	0.0109	0.0109
8.0	0.0119	0.0025	0.0163	0.0094	0.0058	0.0057	0.0312	-0.0064	0.0118	0.0117
8.5	0.0163	0.004[illegible]	0.0142	0.0043	0.0074	0.0072	0.0206	-0.0137	0.0133	0.0132
9.0	0.0239	0.0056	0.0177	-0.0073	0.0398	0.0395	0.0114	-0.0132	0.0156	0.0154
9.5	0.0335	0.0049	0.0255	-0.0065	0.0124	0.0120	0.0199	-0.0033	0.0182	0.0179
10.0	0.0389	0.	0.0292	-0.0014	0.0138	0.0132	0.0277	0.0036	0.0195	0.0191

TABLE 3.7 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 3,1

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.	0.	0.0592	-0.0133	0.0524	0.0524	-0.0152	0.0485	0.0753	0.0753
0.5	0.	0.	0.0493	-0.0181	0.0475	0.0475	0.0136	0.0462	0.0692	0.0692
1.0	0.	0.	0.0365	-0.0254	0.0368	0.0368	0.0539	0.0029	0.0556	0.0556
1.5	-0.	-0.	0.0247	-0.0288	0.0261	0.0261	0.0589	-0.0323	0.0417	0.0417
2.0	-0.	-0.	0.0027	-0.0184	0.0180	0.0180	0.0167	-0.0187	0.0308	0.0308
2.5	-0.	-0.	-0.0075	0.0015	0.0126	0.0126	-0.0279	-0.0163	0.0233	0.0233
3.0	0.	-0.	-0.0062	0.0075	0.0091	0.0091	-0.0250	-0.0158	0.0182	0.0182
3.5	0.	-0.	-0.0086	0.0057	0.0068	0.0068	-0.0011	-0.0059	0.0149	0.0149
4.0	0.	-0.	-0.0060	0.0056	0.0054	0.0054	0.0118	-0.0021	0.0126	0.0126
4.5	0.	-0.	-0.0030	0.0007	0.0045	0.0045	0.0036	-0.0057	0.0110	0.0110
5.0	-0.	0.	-0.0040	-0.0018	0.0039	0.0039	-0.0022	-0.0103	0.0100	0.0100
5.5	-0.	-0.	-0.0010	-0.0004	0.0036	0.0036	0.0017	-0.0113	0.0093	0.0093
6.0	-0.	-0.	0.0029	-0.0009	0.0035	0.0035	0.0016	-0.0124	0.0089	0.0089
6.5	-0.	-0.	0.0025	0.0004	0.0036	0.0036	0.0024	-0.0089	0.0087	0.0087
7.0	-0.	-0.	-0.0005	0.0042	0.0038	0.0038	0.0007	0.0018	0.0088	0.0088
7.5	-0.	-0.	-0.0030	0.0047	0.0044	0.0044	-0.0050	0.0074	0.0092	0.0092
8.0	0.	-0.	-0.0043	0.0036	0.0053	0.0053	-0.0041	0.0120	0.0100	0.0100
8.5	0.	-0.	-0.0053	0.0047	0.0067	0.0067	0.0019	0.0063	0.0113	0.0113
9.0	-0.	0.	-0.0067	-0.0002	0.0087	0.0087	0.0067	-0.0058	0.0132	0.0132
9.5	-0.	0.	-0.0041	-0.0085	0.0108	0.0108	0.0090	-0.0023	0.0154	0.0154
10.0	-0.	-0.	-0.0008	-0.0109	0.0119	0.0119	0.0090	0.0031	0.0164	0.0164

TABLE 3.8 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 3,2

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.4062	-0.	0.4779	0.1227	0.0870	0.0768	0.5584	0.1491	0.1216	0.1148
0.5	0.3032	0.1422	0.3785	0.1390	0.0717	0.0666	0.4230	0.1480	0.1041	0.1006
1.0	0.1619	0.1365	0.2155	0.0973	0.0478	0.0470	0.2383	0.0854	0.0743	0.0738
1.5	0.0833	0.0886	0.1293	0.0333	0.0307	0.0309	0.1267	0.0416	0.0511	0.0512
2.0	0.0465	0.0512	0.0368	0.0394	0.0201	0.0203	0.0679	0.0690	0.0358	0.0359
2.5	0.0282	0.0290	[illegible]	0.0367	0.0137	0.0137	0.0438	0.0603	0.0262	0.0262
3.0	0.0186	0.0167	0.0210	0.0134	0.0097	0.0097	0.0236	0.0250	0.0201	0.0201
3.5	0.0133	0.0100	0.0093	0.0096	0.0073	0.0072	0.0149	0.0130	0.0161	0.0161
4.0	0.0103	0.0062	0.0100	0.0083	0.0058	0.0057	0.0127	0.0081	[illegible]	0.0135
4.5	0.0084	0.0039	0.0095	0.0039	0.0048	0.0047	0.0041	-0.0043	0.0118	0.0118
5.0	0.0073	0.0025	0.0045	0.0074	0.0042	0.0041	0.0036	-0.0120	0.0106	0.0106
5.5	0.0068	0.0016	0.0026	0.0059	0.0038	0.0038	0.0102	-0.0190	0.0096	0.0098
6.0	0.0066	0.0008	0.0044	0.0020	0.0037	0.0037	0.0095	-0.0161	0.0094	0.0094
6.5	0.0069	0.0002	0.0056	0.0009	0.0038	0.0038	0.0026	-0.0112	0.0092	0.0092
7.0	0.0077	-0.0035	0.0089	-0.0036	0.0041	0.0041	0.0035	-0.0150	0.0093	0.0093
7.5	0.0092	-0.0013	0.0148	-0.0075	0.0042	0.0047	0.0121	-0.0103	0.0098	0.0097
8.0	0.0118	-0.0025	0.0163	-0.0074	0.0058	0.0057	0.0211	-0.0046	0.0107	0.0106
8.5	0.0163	-0.0040	0.0142	-0.0043	0.0074	0.0072	0.0196	0.0048	0.0122	0.0121
9.0	0.0238	-0.0056	0.0177	0.0073	0.0098	0.0095	0.0145	0.0189	0.0146	0.0144
9.5	0.0335	-0.0047	0.0255	0.0068	0.0124	0.0120	0.0192	0.0144	0.0172	0.0169
10.0	0.0388	0.	0.0292	0.0014	0.0138	0.0132	0.0240	0.0064	0.0186	0.0132

TABLE 3.9 ERROR SPECTRAL DENSITY. PROCESS 3
ELEMENT 3.3

(1) FREQ. F	(2) EXPECTED VALUES S		(3) DIRECT ESTIMATES S1		(4) STANDARD ERRORS E1		(5) INDIRECT ESTIMATES S2		(6) STANDARD ERRORS E2	
	SR	SI	SR1	SI1	ER1	EI1	SR2	SI2	ER2	EI2
0.	0.7111	-0.	0.7862	-0.	0.0711	0.	0.8344	0.0473	0.0880	0.0519
0.5	0.6654	0.	0.7145	-0.	0.0666	0.	0.7565	0.0383	0.0829	0.0494
1.0	0.5539	0.	0.5408	-0.	0.0554	0.	0.5949	0.0086	0.0701	0.0430
1.5	0.4237	-0.	0.3878	0.	0.0421	0.	0.4320	0.0154	0.0550	0.0354
2.0	0.3044	-0.	0.3042	-0.	0.0304	0.	0.3114	0.0441	0.0417	0.0265
2.5	0.2184	-0.	0.2278	0.	0.0218	0.	0.2281	0.0354	0.0318	0.0231
3.0	0.1597	0.	0.1594	-0.	0.0160	0.	0.1597	0.0136	0.0248	0.0190
3.5	0.1210	0.	0.1284	0.	0.0121	0.	0.1394	-0.0170	0.0202	0.0161
4.0	0.0958	0.	0.1138	-0.	0.0095	0.	0.1305	-0.0260	0.0170	0.0141
4.5	0.0796	0.	0.0933	0.	0.0080	0.	0.0927	-0.0083	0.0149	0.0126
5.0	0.0695	-0.	0.0714	-0.	0.0069	0.	0.0696	0.0021	0.0134	0.0115
5.5	0.0638	-0.	0.0647	0.	0.0064	0.	0.0725	-0.0307	0.0125	0.0108
6.0	0.0617	-0.	0.0663	-0.	0.0062	0.	0.0675	-0.0113	0.0120	0.0103
6.5	0.0628	-0.	0.0648	0.	0.0063	0.	0.0571	-0.0156	0.0118	0.0100
7.0	0.0676	0.	0.0665	-0.	0.0068	0.	0.0631	-0.0065	0.0120	0.0100
7.5	0.0768	0.	0.0742	0.	0.0077	0.	0.0766	0.0015	0.0127	0.0102
8.0	0.0920	0.	0.0943	-0.	0.0092	0.	0.0854	0.0051	0.0141	0.0106
8.5	0.1155	0.	0.1165	0.	0.0116	0.	0.0916	0.0014	0.0162	0.0114
9.0	0.1481	-0.	0.1360	-0.	0.0148	0.	0.1054	0.0082	0.0194	0.0125
9.5	0.1828	-0.	0.1740	0.	0.0183	0.	0.1648	0.0187	0.0227	0.0135
10.0	0.1993	0.	0.2062	-0.	0.0199	0.	0.2072	0.0195	0.0243	0.0140

ICNT AND KCN 2863 0

REALISTIC EVALUATION OF THE PRECISION AND ACCURACY OF INSTRUMENT CALIBRATION SYSTEMS*

Churchill Eisenhart
National Bureau of Standards, Washington, D. C.

ABSTRACT. Calibration of instruments and standards is a refined form of measurement. Measurement of some property of a thing is an operation that yields as an end result a number that indicates how much of the property the thing has. Measurement is ordinarily a repeatable operation, so that it is appropriate to regard measurement as a production process, the "product" being the numbers, i.e., the measurements, that it yields; and to apply to measurement processes in the laboratory the concepts and techniques of statistical process control that have proved so useful in the quality control of industrial production.

Viewed thus it becomes evident that a particular measurement operation cannot be regarded as constituting a measurement process unless statistical stability of the type known as a state of statistical control has been attained. In order to determine whether a particular measurement operation is, or is not, in a state of statistical control it is necessary to be definite on what variations of procedure, apparatus, environmental conditions, observers, operators, etc., are allowable in "repeated applications" of what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. To be realistic, the "allowable variations" must be of sufficient scope to bracket the circumstances likely to be met in practice. Furthermore, any experimental program that aims to determine the standard deviation of a measurement process as an indication of its precision, must be based on appropriate random sampling of this likely range of circumstances.

Ordinarily the accuracy of a measurement process may be characterized by giving (a) the standard deviation of the process and (b) credible bounds to its likely overall systematic error. Determination of credible bounds to the combined effect of recognized potential sources of systematic error always involves some arbitrariness, not only in the placing of reasonable bounds on the systematic error likely to be contributed by each particular assignable

---

*Presented at the 1962 Standards Laboratory Conference, National Bureau of Standards, Boulder, Col., August 8-10, 1962. Published in the Journal of Research of the National Bureau of Standards--C, Engineering and Instrumentation, vol. 67C, No. 2, April-June 1963.

cause, but also in the manner in which these individual contributions are combined. Consequently, the "inaccuracy" of end results of measurement cannot be expressed by "confidence limits" corresponding to a definite numerical "confidence level," except in those rare instances in which the possible overall systematic error of a final result is negligible in comparison with its imprecision.

1. INTRODUCTION. Calibration of instruments and standards is basically a refined form of measurement. Measurement is the assignment of numbers to material things to represent the relations existing among them with respect to particular porperties. One always measures properties of things, not the things themselves. In practice, measurement of some property of a thing ordinarily takes the form of a sequence of steps or operations that yields as an end result a number that indicates how much of this property the thing has, for someone to use for a specific purpose. The end result may be the outcome of a single reading of an instrument. More often it is some kind of average, e.g., the arithmetic mean of a number of independent determinations of the same magnitude, or the final result of a least squares "reduction" of measurements of a number of different quantities that bear known relations to each other in accordance with a definite experimental plan. In general, the purpose for which the answer is needed determines the accuracy required and ordinarily also the method of measurement employed.

Specification of the apparatus and auxiliary equipment to be used, the operations to be performed, the sequence in which they are to be executed, and the conditions under which they are respectively to be carried out--these instructions collectively serve to define a method of measurement. A measurement process is the realization of a method of measurement in terms of particular apparatus and equipment of the prescribed kinds, particular conditions that at best only approximate the conditions prescribed, and particular persons as operators and observers.

It has long been recognized that, in undertaking to apply a particular method of measurement, a degree of consistency among repeated measurements of a single quantity needs to be attained before the method of measurement concerned can be regarded as meaningfully realized, i.e., before a measurement process can be said to have been established that is a realization of the method of mesurement concerned. Indeed, consistency or statistical stability of a very special kind is required: to qualify as a measurement process a measurement operation must have attained what is known in industrial quality control language as a state of statistical control. Until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all. And when it has attained a state of statistical control there may still remain the question of whether it is faithful to the method of measurement of which it is intended to be a realization.

The systematic error, or bias, of a measurement process refers to its tendency to measure something other than what was intended; and is determined by the magnitude of the difference $\mu - \tau$ between the process average or limiting mean μ associated with measurement of a particular quantity by the measurement process concerned and the true value τ of the magnitude of this quantity. On first thought, the "true value" of the magnitude of a particular quantity appears to be a simple straightforward concept. On careful analysis, however, it becomes evident that the "true value" of the magnitude of a quantity is intimately linked to the purposes for which knowledge of the magnitude of this quantity is needed, and cannot, in the final analysis, be meaningfully and usefully defined in isolation from these needs.

The precision of a measurement process refers to, and is determined by the degree of mutual agreement characteristic of independent measurements of a single quantity yielded by repeated applications of the process under specified conditions; and its accuracy refers to, and is determined by, the degree of agreement of such measurements with the true value of the magnitude of the quantity concerned. In brief "accuracy" has to do with closeness to the truth; "precision," only with closeness together.

Systematic error, precision, and accuracy are inherent characteristics of a measurement process and not of a particular measurement yielded by the process. We may also speak of the systematic error, precision, and accuracy of a particular method of measurement that has the capability of statistical control. But these terms are not defined for a measurement operation that is not in a state of statistical control.

The precision, or more correctly, the <u>im</u>precision of a measurement process is ordinarily summarized by the standard deviation of the process, which expresses the characteristic disagreement of repeated measurements of a single quantity by the process concerned, and thus serves to indicate by how much a particular measurement is likely to differ from other values that the same measurement process might have provided in this instance, or might yield on remeasurement of the same quantity on another occasion. Unfortunately, there does not exist any single comprehensive measure of the accuracy (or <u>in</u>accuracy) of a measurement process analogous to the standard deviation as a measure of its imprecision.

To characterize the accuracy of a measurement process it is necessary, therefore, to indicate (a) its systematic error or bias, (b) its precision (or imprecision)--and, strictly speaking, also, (c) the form of the distribution of the individual measurements about the process average. Such is the unavoidable situation if one is to concern one s self with individual measurements yielded by any particular measurement process. Fortunately, however, "final results" are ordinarily some kind of average or adjusted value derived from a set of independent measurements, and when four or more independent measurements are involved, such adjusted values tend to be normally distributed to a very good approximation, so that the accuracy of such final results can ordinarily be characterized satisfactorily by indicating (a) their imprecision as expressed by their standard error, and (b) the systematic error of the process by which they were obtained.

The error of any single measurement or adjusted value of a particular quantity is, by definition, the difference between the measurement or adjusted value concerned and the true value of the magnitude of this quantity. The error of any particular measurement or adjusted value is, therefore, a fixed number; and this number will ordinarily be unknown and unknowable, because the true value of the magnitude of the quantity concerned is ordinarily unknown and unknowable. Limits to the error of a single measurement or adjusted value may, however, be inferred from (a) the precision, and (b) bounds on the systematic error of the measurement process by which it was produced--but not without risk of being incorrect, because, quite apart from the inexactness with which bounds are commonly placed on a systematic error of a measurement process, such limits are applicable to the error of the single measurement or adjusted value, not as a unique individual outcome, but only as a typical case of the errors characteristic of such measurements of the same quantity that might have been, or might be, yielded by the same measurement process under the same conditions.

Since the precision of a measurement process is determined by the characteristic "closeness together" of successive independent measurements of a single magnitude generated by repeated application of the process under specified conditions, and its bias or systematic error is determined by the direction and amount by which such measurements tend to differ from the true value of the magnitude of the quantity concerned, it is necessary to be clear on what variations of procedure, apparatus, environmental conditions, observers, etc., are allowable in "repeated applications" or what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. If whatever measures of the precision and bias of a measurement process we may adopt are to provide a realistic indation of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Furthermore, any experimental program that aims to determine the precision, and thence the accuracy of a measurement process, must be based on an appropriate random sampling of this "range of circumstances," if the usual tools of statistical analysis are to be strictly applicable.

When adequate random sampling of the appropriate "range of circumstances" is not feasible, or even possible, then it is necessary (a) to compute, by extrapolation from available data, a more or less subjective estimate of the precision of the measurement process concerned, to serve as a substitute for a direct experimental measure of this characteristic, and (b) to assign more or less subjective bounds to the systematic error of the measurement process. To the extent that such at least partially subjective computations are involved, the resulting evaluation of the overall accuracy of a measurement process "is based on subject-matter knowledge and skill, general information, and intuition--but not on statistical methodology" [Cochran et al. 1953, p. 693]. Consequently, in such cases the statistically precise concept of a family of "confidence intervals" associated with a definite "confidence level" or "confidence coefficient" is not applicable.

The foregoing points and certain other related matters are discussed in greater detail in the succeeding sections, together with an indication of procedures for the realistic evaluation of precision and accuracy of established procedures for the calibration of instruments and standards that minimize as much as possible the subjective elements of such an evaluation. To the extent that complete elimination of the subjective element is not always possible, the responsibility for an important and sometimes the most difficult part of the evaluation is shifted from the shoulders of the statistician to the shoulders of the subject matter "expert."

2. MEASUREMENT.

2.1. Nature and Object. Measurement is the assignment of numbers to material things to represent the relations existing among them with respect to particular properties. The number assigned to some particular property serves to represent the relative amount of this property associated with the object concerned.

Measurement always pertains to properties of things, not to the things themselves. Thus we cannot measure a meter bar, but can and usually do, measure its length; and we could also measure its mass, its denity, and perhaps, also its hardness.

The object of measurement is twofold: first, symbolic representation of properties of things as a basis for conceptual analysis; and second, to effect the representation in a form amenable to the powerful tools of mathematical analysis. The decisive feature is symbolic representation of properties, for which and numerals are not the only usable symbols.

In practice the assignment of a numerical magnitude to a particular property of a thing is ordinarily accomplished by comparison with a set of standards, or by comparison either of the quantity itself, or of some transform of it, with a previously calibrated scale. Thus, length measurements are usually made by directly comparing the length concerned with a calibrated bar or tape; and mass measurements, by directly comparing the weight of a given mass with the weight of a set of standard masses, by means of a balance; but force measurements are usually carried out in terms of some transform, such as by reading on a calibrated scale the extension that the force produces in a spring, or the deflection that it produces in a proving ring; and temperature measurements are usually performed in terms of some transform, such as by reading on a calibrated scale the expansion of a column of mercury, or the electrical resistance of a platinum wire.

2.2. Qualitative and Quantitative Aspects. As Walter A. Shewhart, father of statistical control charts, has remarked:

"It is important to realize...that there are two aspects of an operation of measurement; one is quantitative and the other qualitative. One consists of numbers or pointer readings such as the observed lengths in *n* measurements of the length of a line, and the other consists of the physical manipulations of physical things by someone, in accord with instructions that we shall assume to be describable in words constituting a text." [Shewhart 1939, p. 130.]

More specifically, the qualitative factors involved in the measurement of a quantity are: the apparatus and auxiliary equipment (e.g., reagents, batteries or other source of electrical energy, etc.) employed; the operators and observers, if any, involved; the operations performed, together with the sequence in which, and the conditions under which, they are respectively carried out.

2.3. Correction and Adjustment of Observations. The numbers obtained as "readings" on a calibrated scale are ordinarily the end product of everyday measurement in the trades and in the home. In scientific work there are usually two important additional quantitative aspects of measurement: (1) correction of the readings, or their transforms, to compensate for known deviations from ideal execution of the prescribed operations, and for non-negligible effects of variations in uncontrolled variables; and (2) adjustment of "raw" or corrected measurements of particular quantities to obtain values of these quantities that conform to restrictions upon, or interrelations among, the magnitudes of these quantities imposed by the nature of the problem.

Thus, it may not be practicable or economically feasible to take readings at exactly the prescribed temperatures; but quite practicable and feasible to bring and hold the temperature within narrow neighborhoods of the prescribed values and to record the actual temperatures to which the respective readings correspond. In such cases, if the deviations from the prescribed temperatures are not negligible, "temperature corrections" based on appropriate theory are usually applied to the respective readings to bring them to the values that presumable would have been observed if the temperature in each instance had been exactly as prescribed.

In practice, however, the objective just stated is rarely, if ever, actually achieved. Any "temperature corrections" applied could be expected to bring the respective readings "to the values that presumably would have been observed if the temperature in each instance had been exactly as prescribed" if and only if these "temperature corrections" made appropriate allowances for all of the effects of the deviations of the actual temperatures from those prescribed. "Temperature corrections" ordinarily correct only for particular effects of the deviations of the actual temperatures from their prescribed values; not for all of the effects on the readings traceable to deviations of the actual temperatures from those prescribed. Thus Michelson utilized "temperature corrections" in his 1879 investigation of the speed of light; but his results exhibit a dependence on temperature after "temperature correction." The "temperature corrections" applied corrected only for the effects of thermal expansion due to variations in temperature and not also for changes in the index of refraction of the air due to changes in the humidity of the air, which in June and July at Annapolis is highly correlated with temperature. Corrections applied in practice are usually of more limited scope than the names that they are given appear to indicate.

Adjustment of observations is fundamentally different from their "correction." When two or more related quantities are measured individually, the resulting measured values usually fail to satisfy the constraints on their magnitudes implied by the given interrelations among the quantities concerned. In such cases these "raw" measured values are mutually contradictory, and require adjustment in order to be usable for the purpose intended. Thus, measured values of the three cyclic differences (A - B), (B - C), and (C - A) between the lengths of three nominally equivalent gage blocks are mutually contradictory, and strictly speaking are not usable as values of these differences, unless they sum to zero.

The primary goal of adjustment is to derive from such inconsistent measurements, if possible, adjusted values for the quantities concerned that do satisfy the constraints on their magnitudes imposed by the nature of the quantities themselves and by the existing interrelations among them. A second objective is to select from all possible sets of adjusted values the set that is the "best"--or, at least, a set that is "good enough" for the intended purpose--in some well-defined sense. Thus, in the above case of the measured differences between the lengths of three gage blocks, an adjustment could be effected by ignoring the measured value of one of the differences entirely, say, the difference (C - A), and taking the negative of the sum of the other two as its adjusted value,

$$\text{Adj}(C - A) = -\left[(A - B) + (B - C)\right].$$

This will certainly assure that the sum of all three values, (A - B) + (B - C) + Adj(C - A), is zero, as required, and is clearly equivalent to ascribing all of the excess or deficit to the replaced measurement, (C - A). Alternatively, one might prefer to distribute the necessary total adjustment $-\left[(A - B) + (B - C) + (C - A)\right]$ equally over the individual measured differences, to obtain the following set of adjusted values:

$$\begin{aligned}\text{Adj}(A - B) &= (A - B) - \tfrac{1}{3}\left[(A - B) + (B - C) + (C - A)\right]\\ &= \tfrac{1}{3}\left[2(A - B) - (B - C) - (C - A)\right]\\ \text{Adj}(B - C) &= \tfrac{1}{3}\left[2(B - C) - (A - B) - (C - A)\right]\\ \text{Adj}(C - A) &= \tfrac{1}{3}\left[2(C - A) - (A - B) - (B - C)\right].\end{aligned}$$

Clearly, the sum of these three adjusted values must always be zero, as required, regardless of the values of the original individual measured differences. Furthermore, most persons, I believe, would consider this latter adjustment the better; and under certain conditions with respect to the "law of error" governing the original measured differences, it is indeed the "best."

Note that no adjustment problem existed at the stage when only two of these differences had been measured whichever they were, for then the third could be obtained by subtraction. As a general principle, when no more observations are taken than are sufficient to provide one value of each of the unknown quantities involved, then the results so obtained are usable at least-- they may not be "best." On the other hand, when additional observations are taken, leading to "over determination" and consequent contradiction of the fundamental properties of, or the basic relationships among the quantities concerned, then the respective observations must be regarded as contradicting one another. When this happens the observations themselves, or values derived from them, must be replaced by adjusted values such that all contradiction is removed. "This is a logical necessity, since we cannot accept for truth that which is contradictory or leads to contradictory results," [Chauvenet 1868, p. 472.]

2.4. Scheduling the Taking of Measurements. Having done what one can to remove extraneous sources of error, and to make the basic measurements as precise and as free from systematic error as possible, it is frequently possible not only to increase the precision of the end results of major interest but also to simultaneously decrease their sensitivity to sources of possible systematic error, by careful scheduling of the measurements required. An instance is provided by the traditional procedure for calibrating liquid-in-glass thermometers [Waidner and Dickinson 1907, p. 702, NPL 1957, pp. 29-30; Swindells 1959, pp. 11-12]. Instead of attempting to hold the temperature of the comparison bath constant, a very difficult objective to achieve, the heat input to the bath is so adjusted that its temperature is slowly increasing at a steady rate, and then readings of, say, four test thermometers and two standards are taken in accordance with the schedule

$$S_1T_1T_2T_3T_4S_2S_2T_4T_3T_2T_1S_1$$

the readings being spaced uniformly in time so that the arithmetic mean of the two readings of any one thermometer will correspond to the temperature of the comparison bath at the midpoint of the period. Such scheduling of measurement taking operations so that the effects of the specific types of departures from perfect control of conditions and procedure will have an opportunity to balance out is one of the principal aims of the art and science of statistical design of experiments. For additional physical science examples, see, for instance, Youden [1951a; and 1954-1959].

2.5. Measurement as a Production Process. We may summarize our discussion of measurement up to this point, as follows: Measurement of some property of a thing in practice always takes the form of a sequence of steps or operations that yield as an end result a number that serves to represent the amount or quantity of some particular property of a thing--a number that indicates how much of this property the thing has, for someone to use for a specific purpose. The end result may be the outcome of a single reading of an instrument, with or without corrections for departures from prescribed conditions. More often it is some kind of average or adjusted value, e.g., the arithmetic mean of a number of independent determinations of the same magnitude, or the final result of, say, a least squares "reduction" of measurements of a number of different quantities that have known relations to the quantity of interest.

Measurement of some property of a thing is ordinarily a repeatable operation. This is certainly the case for the types of measurement ordinarily met in the calibration of standards and instruments. It is instructive, therefore, to regard measurement as a production process, the "product" being the numbers, that is, the measurements that it yields; and to compare and contrast measurement processes in the laboratory with mass production processes in industry. For the moment it will suffice to note (a) that when successive amounts of units of "raw material" are processed by a particular mass production process, the output is a series of nominally identical items of product--of the particular type produced by the mass production operation, i.e., by the method of production concerned; and (b) that when successive objects are measured by a particular measurement process, the individual items of "product" produced consist of the numbers assigned to the respective objects to represent the relative amounts that they possess of the property determined by the method of measurement involved.

2.6. Methods of Measurement and Measurement Processes. Specification of the apparatus and auxiliary equipment to be used, the operations to be performed, the sequence in which they are to be carried out, and the conditions under which they are respectively to be carried out--these instructions collectively serve to define a method of measurement. To the extent that corrections may be required they are an integral part of measurement. The types of corrections that will ordinarily need to be made, and specific procedures for making them, should be included among "the operations to be performed." Likewise, the essential adjustments required should be noted, and specific procedures for making them incorporated in the specification of a method of measurement.

A measurement process is the realization of a method of measurement in terms of particular apparatus and equipment of the prescribed kinds, particular conditions that at best only approximate the conditions prescribed, and particular persons as operators and observers [ASTM 1961, p. 1758; Murphy 1961, p. 264]. Of course, there will often be a question whether a particular measurement process is loyal to the method of measurement of which it is intended to be a realization; or whether two different measurement processes can be considered to be realizations of the same method of measurement.

To begin with, written specifications of methods of measurement often contain absolutely precise instructions which, however, cannot be carried out (repeatedly) with complete exactitude in practice; for example, "move the two parallel cross hairs of the micrometer of the microscope until the graduation line of the standard is centered between them." The accuracy with which such instructions can be carried out in practice will always depend upon "the circumstances"; in the case cited, on the skill of the operator, the quality of the graduation line of the standard, the quality of the screw of the micrometer, the parallelism of the cross hairs, etc. To the extent that the written specification of a method of measurement involves absolutely precise instructions that cannot be carried out with complete exactitude in practice there are certain to be discrepancies between a method of measurement and its realization by a particular measurement process.

In addition, the specification of a method of measurement often includes a number of imprecise instructions, such as "raise the temperature slowly," "stir well before taking a reading," "make sure that the tubing is clean," etc. Not only are such instructions inherently vague, but also in any given instance they must be understood in terms of the general level of refinement characteristic of the context in which they occur. Thus, "make sure that the tubing is clean" is not an absolutely definite instruction; to some people this would mean simply that the tubing should be clean enough to drink liquids through; in some laboratory work it might be interpreted to mean mechanically washed and scoured so as to be free from dirt and other ordinary solid matter (but not cleansed also with chemical solvents to remove more stubborn contaminants); to an advanced experimental physicist it may mean not merely mechanically washed and chemically cleansed, but also "out gassed" by being heated to and held at a high temperature, near the softening point, for an hour or so. All will agree, I believe, that it would be exceedingly difficult to make such instructions absolutely difinite with a convenient number of words. To the extent that the specification of a method of measurement includes instructions that are not absolutely definite, there will be room for differences between measurement processes that are intended to be realization of the very same method of measurement.

Recognition of the difficulty of achieving absolute definiteness in the specification of a method of measurement does not imply that "any old set" of instructions will serve to define a method of measurement. Quite the contrary. To qualify as a specification of a method of measurement, a set of instructions must be sufficiently definite to insure statistical stability of repeated measurements of a single quantity, that is, derived measurement processes must be capable of meeting the criteria of statistical control [Shewhart 1939, p. 131; Murphy 1961, p. 265; ASTM 1961, p. 1758]. To elucidation of the meaning of, and need for this requirement we now turn.

3. PROPERTIES OF MEASUREMENT PROCESSES

3.1. Requirement of Statistical Control. The need for attaining a degree of consistency among repeated measurements of a single quantity before the method of measurement concerned can be regarded as meaningful has certainly been recognized for a long, long time. Thus Galileo, describing his famous experiment on the acceleration of gravity in which he allowed a ball to roll different distances down an inclined plane wrote:

"... si lasciava (come dico) scendere per il detto canale la palla, notando, nel modo che appresso dirò, il temp che consumava nello scorrerlo tutto, replicando il medesimo atto molte volte per assicurarsi bene della quantità del temp, nel quale non si trovava mai differenza ne anco della decima parte d'una battuta di polso. Fatta e stabilita precisamente tale operazione, facemmo scender la medisima palla solamente per la quarta parte della lunghezza di esso canale ..."[1] [Galileo 1638, Third Day; Nat'l. ed., p. 213.]

Something more than mere "consistency" is required, however, as Shewhart points out eloquently in his very important chapter on "The Specification of Accuracy and Precision" [Shewhart 1939, ch IV]. He begins by noting that the description given by R. A. Millikan [1903, pp. 195-196] of a method for determining the surface tension T of a liquid from measurements of the

[1] I am grateful to my collegue Ugo Fano for the following literal translation:
"... we let, as I was saying, the ball descend through said channel, recording, in a manner presently to be described, the time it took in traversing it all, repeating the same action many times to make really sure of the magnitude of time, in which one never found a difference of even a tenth of a pulsebeat. Having done and established precisely such operation, we let the same ball descend only for the fourth part of the length of the same channel;..."

force of tension *F* of a film of the liquid contains the following instruction with regard to the basic readings from which measurements of *T* are derived: "Continue this operation until a number of consistent readings can be obtained." Shewhart then comments on this as follows:

"... the text describing the operation does not say to carry out such and such physical operations and call the result a measurement of *T*. Instead, it says in effect not to call the result a measurement of *T* until one has attained a certain degree of consistency among the observed values of *F* and hence among those of *T*. Although this requirement is not always explicitly stated in specifications of the operation of measurements as it was here, I think it is always implied. Likewise, I think it is always assumed that there can be too much consistency or uniformity among the observed values as, for example, if a large number of measurements of the surface tension of a liquid were found to be identical. What is wanted but not explicitly described is a specific kind and degree of consistency.

"... it should be noted that the advice to repeat the operation of measuring surface tension until a number of consistent readings have been obtained is indefinite in that it does not indicate how many readings shall be taken before applying a test for consistency, nor what kind of test of consistency is to be applied to the numbers or pointer readings.... One of the objects of this chapter is to see how far one can go toward improving this situation by providing an operationally definite criterion that preliminary observations must meet before they are to be considered consistent in the sense implied in the instruction cited above.

"... Before doing this, however, we must give attention not so much to the consistency of the *n* observed values already obtained by *n* repetitions of the operation of measurement as we do to the reproducibility of the operation as determined by the numbers in the potentially infinite sequence corresponding to an infinite number of repetitions of this operation. No one would care very much how consistent the first *n* preliminary observations were if nothing could be validly inferred from this as to what future observations would show. Hence, it seems to me that the characteristics of the numerical aspects of an operation that is of greatest practical interest is its reproducibility within tolerance limits throughout the infinite sequence. The limit to which we may go in this direction is to attain a state of statistical control. The attempt to attain a certain kind of consistency within the first *n* observed values is merely a means of attaining reproducibility within limits throughout the whole of the sequence." [Shewhart 1939, pp. 131-132.]

The point that Shewhart makes forcefully, and stresses repeatedly later in the same chapter, is that the first n measurements of a given quantity generated by a particular measurement process provide a logical basis for predicting the behavior of further measurements of the same quantity by the same measurement process if and only if these n measurements may be regarded as a random sample from a "population" or "universe" of all conceivable measurements of the given quantity by the measurement process concerned; that is, in the language of mathematical statistics, if and only if the n measurements in hand may be regarded as "observed values" of a sequence of random variables characterized by a probability distribution identified with the measurement process concerned, and related through the values of one or more of its parameters to the magnitude of the quantity measured.

It should be noted especially that nothing is said about the mathematical form of the probability distribution of these random variables. The important thing is that there be one. W. Edwards Deming has put this clearly and forcefully in these words:

"In applying statistical theory, the main consideration is not what the shape of the universe is, but whether there is any universe at all. No universe can be assumed, nor ... statistical theory ... applied unless the observations show statistical control. In this state the samples when cumulated over a suitable interval of time give a distribution of a particular shape, and this shape is reproduced hour after hour, day after day, so long as the process remains in statistical control--i.e., exhibits the properties of randomness. In a state of control, n observations may be regarded as a sample from the universe of whatever shape it is. A big enough sample, or enough small samples, enables the statistician to make meaningful and useful predictions about future samples. This is as much as statistical theory can do.

"... Very often the experimenter, instead of rushing in to apply [statistical methods] should be more concerned about attaining statistical control and asking himself whether any predictions at all (the only purpose of his experiment), by statistical theory or otherwise, can be made." [Deming 1950, pp. 502-503.]

Shewhart was well aware of the fact that from a set of n measurements in hand it is not possible to decide with absolute certainty whether they do or do not constitute a random sample from some definite statistical "population" characterized by a probability distribution. He, therefore, proposed [Shewhart 1939, pp. 146-147] that in any particular instance one should "decide to act for the present as if"[2] the measurements in hand (and their immediate successors) were a simple random sample from a definite statistical population--i.e., in the language of mathematical statistics, were "observed values" of independent identically distributed random variables--only if the measurements in hand met the requirements of the small-samples version of Criterion I of his previous book [Shewhart 1931, pp. 309-318] and of certain additional test of randomness that he described explicitly for the first time in his contribution to the University of Pennsylvania Bicentennial Conference in September 1940 [Shewhart, 1941]. In other words, Shewhart proposed that one should consider a measurement process to be--i.e., should "decide to act for the present as if" the process were--in a state of (simple) statistical control, only if the measurements in hand show no evidence of lack of statistical control when analyzed for randomness in the order in which they were taken by the control chart techniques for averages and standard deviations that he found so valuable in industrial process control and by certain additional tests for randomness based on "runs above and below average" and "runs up and down."

Simpson[3] did not prove that taking of the Arithmetic Mean was the best thing to do but merely that it is good. However, in accomplishing this goal he did something much more important: he took the bold step of regarding errors of measurement, not as unique unrelated magnitudes unamenable to

---

[2]This very explicit phraseology is due to John W. Tukey [1960, p. 424].

[3]Thomas Simpson, in his now famous letter [Simpson 1755] to the President of the Royal Society of London "on the Advantage of taking the Mean of a Number of Observations, in practical Astronomy," was the first to consider repeated measurements of a single quantity by a given measurement process as observed values of independent random variables having the same probability distribution. His conclusion is of interest in itself:

"Upon the whole of which it appears, that the taking of the Mean of a number of observations, greatly diminishes the chances for all the smaller errors, and cuts off almost all possibility of any great ones: which last consideration, alone, seem sufficient to recommend the use of the method, not only to astronomers, but to all others concerned in making of experiments of any kind (to which the above reasoning is equally applicable). And the more observations or experiments there are made, the less will the conclusion be liable to err, provided they admit of being repeated under the same circumstances."

mathematical analysis, but as distributed in accordance with a probability distribution that was an intrinsic property of the measurement process itself. He thus opened the way to a mathematical theory of measurement based on the mathematical theory of probability; and, in particular, to the formulation and development of the Method of Least Squares in essentially its present day form by Gauss (1809, 1821) and Laplace (1812).

"Student" (William Sealy Gosset 1876-1937) pioneer statistical consultant and "father" of the "theory of small samples," was certainly among the first to stress the importance of randomness in measurement and experimentation. Thus, he began his revolutionary 1908 paper on "The probable error of a mean" with these remarks:

"Any experiment may be regarded as forming an individual of a 'population' of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

"Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong." [Student 1908, p. 1.]

None of these writers, nor any of their contemporaries, however, provided "an operationally definite criterion that preliminary observations must meet" before we take it upon ourselves "to act for the present as if" they and their immediate successors were random samples from a "population" or "universe" of all conceivable measurements of the given quantity by the measurements of the given quantity by the measurement process concerned. Provision of such a criterion is Shewhart's major contribution.

Experience shows that in the case of measurement processes the ideal of strict statistical control that Shewhart prescribes is usually very difficult to attain, just as in the case of industrial production processes. Indeed, many measurement processes simply do not and, it would seem, cannot be made to conform to this ideal of producing successive measurements of a single quantity that can be considered to be "observed values" of independent identically distributed random variables.[4] The nature of the "trouble" was stated succinctly by Student in 1917 when, speaking of physical and chemical determinations, he wrote:

> "After considerable experience I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if the repetitions had been made on different days." [Student 1917, p. 415.]

In other words, production of measurements seems to be like the production of paint; and just as in the case of paint, if one must cover a large surface all of which is visible simultaneously, one will do well to use paint all from the same batch, so in the case of measurements, if a scientist or metrologist "wishes to impress his clients" he will "arrange to do repetition analyses as nearly as possible at the same time." [Student 1927, p. 155.]

Fortunately, just as one may blend paint from several batches to obtain a more uniform color, and one which is, presumable, closer to the "process average," so also may a scientist or metrologist "if he wishes to diminish his real error, ... separate [his measurements] by as wide an interval of time as possible" [Student, loc. cit.] and then take an appropriate average of them as his determination. Consequently, if we are to permit such averaging as an allowable step in a fully specified measurement process (see sec. 2.6 above), then we are obliged to recognize both within-day and between-day components of variation, and accept such a complex measurement process as being in a state of statistical control overall, or as we shall say, in a state of COMPLEX statistical control, when the components of within-day and between-day variation are both in a state of statistical control in Shewhart's strict sense, which we shall term SIMPLE statistical control. In more complex situations, one may be obliged to recognize more than two "layers" of variation

---

[4] Looking at the matter from a fundamental viewpoint, perhaps we should say, not that Shewhart's ideal of strict statistical control is unattainable in the case of such measurement processes, but rather that the degree of approximation to this ideal can be made as close as one chooses, if one is willing to pay the price. In other words, how close one chooses to bring a measurement process to the ideal of strict statistical control is, in any given instance, basically an economic matter, taking into account, of course, not only the immediate purpose(s) for which the measurements are intended but also the other uses to which they may be put. (Compare Simon [1946, p. 566] and Eisenhart [1952, p. 554]).

and, sometimes, more than a single component of variation within a given "layer."

Adopting this more general concept of statistical control, R. B. Murphy of the Bell Telephone Laboratories in his essay "On the Meaning of Precision and Accuracy" [Murphy 1961], published in advance of the issuance by the American Society for Testing and Materials of its Tentative Recommended Practice with respect to the "Use of the Terms Precision and Accuracy as Applied to Measurement of a Property of a Material" [ASTM 1961], remarks:

"Following through with this line of thought borrowed from quality control, we shall add a requirement that an effort to follow a test method ought not to be known as a measurement process unless it is capable of statistical control. Capability of control means that either the measurements are the product of an identifiable statistical universe or an orderly array of such universes or, if not, the physical causes preventing such identification may themselves be identified and, if desired, isolated and suppressed. Incapability of control implies that the results of measurement are not to be trusted as indications of the physical property at hand--in short, we are not in any verifiable sense measuring anything.... Without this limitation of the notion of measurement process, one is unable to go on to give meaning to those statistical measures which are basic to any discussion of precision and accuracy." [Murphy 1961, pp. 264-265.]

3.2. Postulate of Measurement and the Concept of a Limiting Mean. A conspicuous characteristic of measurement is disagreement of repeated measurements of the same quantity. Experience shows that, when high accuracy is sought, repeated measurements of the same quantity by a particular measurement process does not yield uniformly the same number.[5] We explain these discordances by saying that the individual measurements are

[5]The qualification "when high accuracy is sought" is essential; for if using an ordinary two-pan chemical balance we measure and record the mass of a small metallic object only to the nearest gram, then we would expect all of our measurements to be the same--except in the equivocal case of a mass equal, or very nearly equal, to an odd multiple of 1/2 g, and such equivocal cases can be resolved easily by adding a 1/2 g mass to one pan. Full accordance of measurements clearly cannot be taken as incontestable evidence of high accuracy; but rather should be regarded as evidence of limited accuracy.

affected by **errors**, which we interpret to be the manifestations of variations in the execution of the process of measurement resulting from "the imperfections of instruments, and of organs of sense," and from the difficulty of achieving (or even specifying with a convenient number of words) the ideal of perfect control of conditions and procedure.

This "cussedness of measurements" brings us face to face with a fundamental question: In what sense can we say that the measurements yielded by a particular measurement process serve to determine a unique magnitude, when experience shows that repeated measurement of a single quantity by this process yields a sequence of nonidentical numbers. What is the value thus determined?

The answer takes the form of a postulate about measurement processes that has been expressed by N. Ernest Dorsey, as follows:

"The mean of a family of measurements--of a number of measurements for a given quantity carried out by the same apparatus, procedure and observer--approaches a definite value as the number of measurements is indefinitely increased. Otherwise, they could not properly be called measurements of a given quantity. In the theory of errors, this limiting mean is frequently called the 'true' value, although it bears no necessary relation to the true quaesitum, to the actual value of the quantity that the observer desires to measure. This has often confused the unwary. Let us call it the limiting mean." [Dorsey 1944, p. 4; Dorsey and Eisenhart 1953, p. 103.]

In my lectures at the National Bureau of Standards, and elsewhere, I have termed this--or rather a slightly rephrased version of it-- the Postulate of Measurement. A mathematical basis for it is provided by the Strong Law of Large Numbers, a theorem in the mathematical theory of probability discovered during the present century. See, for example, Feller [1957, pp. 243-245, 374], Gnedenko [1962, pp. 241-249], or Parzen [1960, p. 420].

Needless to say, by a "family of measurements" Dorsey means, not a succession of "raw" readings, but rather a succession of adjusted or corrected values which, by virtue of adjustment or correction, can rightfully be considered to be determinations of a single magnitude.

a. Mathematical Formulation

The foregoing can be expressed mathematically as follows: on some particular occasion, say the ith, we may take a number of successive measurements of a single quantity by a given measurement process under certain specified circumstances. Let

$$(1) \qquad x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots$$

denote the sequence of measurements so generated. Conceptually at least, this sequence could be continued indefinitely. Likewise, on different occasions we might start a new sequence, using the same measurement procedure and applying it to measurement of the same quantity under the same fixed set of circumstances. Each such fresh "start" would correspond to a different value of i. If, for example, the measurement process concerned is statistically stable in the sense of being in a state of statistical control as defined by Shewhart [1939], then the Strong Law of Large Numbers will be applicable and we may expect the sequence of cumulative arithmetic means on the ith occasion, namely,

$$(2) \qquad \bar{x}_{in} = (x_{i1} + x_{i2} + \ldots + x_{in})/n, \quad (n = 1, 2, \ldots),$$

to converge to μ, a number that constitutes the limiting mean associated with the quantity measured by this measurement process under the circumstances concerned, but independent of the "occasion," that is, independent of the value of "i." The Strong Law of Large Numbers does not guarantee that the sequence (2) for a particular value of "i" will converge to μ as the number of observations n on this occasion tends to infinity, but simply states that among the family of such sequences corresponding to a large number of different starts, (i = 1, 2, ...), the instances of nonconvergence to μ will be rare exceptions. In other words, if the measurement process with which one is concerned satisfies the conditions for validity of the Strong Law of Large Numbers, then in practice one is almost certain to be working with a "good" sequence--one for which (2) would converge to μ if the number of observations were continued indefinitely--but "bad" occasions can occur, though rarely. Thus, the Postulate of Measurement expresses something better than an "on-the-average" property--it expresses an "in-almost-all-cases" property. Furthermore, this limiting mean μ, the value of which each individual measurement x is trying to express, can be regarded not only as the mean or "center of gravity" of the infinite conceptual population of all measurements x that might conceivably be generated by the measurement process concerned under the specified circumstances, but also as the value of the quantity concerned as determined by this measurement process.

b. Aim of the Postulate

The sole aim of the Postulate of Measurement is axiomatic acceptance of the existence of a limit approached by the arithmetic mean of a finite number

$\underline{n}$ of measurements generated by any measurement process as $\underline{n} \to \infty$. It says nothing about how the "best" estimate of this limiting mean is to be obtained from a finite number of such observations. The Postulate is an answer to the need of the practical man for a justification of his desire to consider the sequence of nonidentical numbers that he obtains when he attempts to measure a quantity "by the same method under like circumstances" as pertaining to a single magnitude, in spite of the evident discordance of its elements. The Postulate aims to satisfy this need by telling him that if he were to continue taking more and still more measurements on this quantity "by the same method under like circumstances" ad infinitum, and were to calculate their cumulative arithmetic means at successive stages of this undertaking, then he would find that the successive terms of this sequence of cumulative arithmetic means would settle down to a narrower and ever narrower neighborhood of some definite number which he could then accept as the value of the magnitude that his first few measurements were striving to express.

c. Importance of Limiting Mean

The concept of a limiting mean associated with the measurement of a given quantity by a particular measurement process that is in a state of statistical control is important because by means of statistical methods based on the mathematical theory of probability we can make quantitative inferential statements, with known chances of error, about the magnitude of this limiting mean from a set of measurements of the given quantity by the measurement process concerned. The magnitude of the limiting mean associated with the measurement of a given quantity by a particular measurement process must be carefully distinguished from the true magnitude of the quantity measured, about which we may be tempted to make similar inferential statements. Insofar as we make statistical inferences from a set of measurements, we make them with respect to a property of the measurement process involved under the circumstances concerned. The step from quantitative inferential statements about the limiting mean associated with the measurement of a given quantity by a particular measurement process, to quantitative statements about the true magnitude of the quantity concerned, may be based on subject matter knowledge and skill, general information and intuition--but not on statistical methodology. (Compare Cochran, Mosteller, and Tukey [1953, pp. 692-693].)

3.3. Definition of the Error of a Measurement, and of the Systematic Error, Precision, and Accuracy of a Measurement Process.

a. Error of a Single Measurement or Adjusted Value

The error of any measurement of a particular quantity is, by definition, the difference between the measurement concerned and the true value of the magnitude of this quantity, taken positive or negative accordingly as the measurement is greater or less than the true value. In other words, if x denotes a single measurement of a quantity, or an adjusted value derived from a specific set of individual measurements, and τ is the true value of the magnitude of the quantity concerned, then, by definition,

$$\text{the error of } x \text{ as a measurement of } \tau = x - \tau.$$

The error of any particular measurement or adjusted value, x, is, therefore, a fixed number. The numerical magnitude and sign of this number will ordinarily be unknown and unknowable, because the true value of the magnitude of the quantity concerned is ordinarily unknown and unknowable. Limits to the error of a single measurement or adjusted value may, however, be inferred from (a) the precision, and (b) bounds on the systematic error, of the measurement process by which it was produced--but not without risk of being incorrect, because, quite apart from the inexactness with which bounds are commonly placed on the systematic error of a measurement process, such limits are applicable to the error of a single measurement or adjusted value, not as a unique individual outcome, but only as a typical case of the errors characteristic of measurements of the same quantity that might have been, or might be, yielded by the same measurement process under the same conditions.

b. Systematic Error of a Measurement Process

When the limiting mean μ associated with measurement of the magnitude of a quantity by a particular measurement process does not agree with the true value τ of the magnitude concerned, the measurement process is said to have a systematic error, or bias, of magnitude $\mu - \tau$.

The systematic error of a measurement process will ordinarily have both constant and variable components. Consider, for example, measurement of the distance between two points by means of a graduated metal tape [Holman 1892, p. 9]. Possible causes of systematic error that immediately come to mind are:

(1) Mistakes in numbering the scale divisions of the tape;
(2) irregular spacing of the divisions of the tape;
(3) sag of tape;
(4) stretch of tape;
(5) temperature not that for which the tape was calibrated.

For any single distance, the effects of (1) and (2) will be constant; and the effects of (3) and (4) will undoubtedly each contain a constant component characteristic of the distance concerned. Some of these effects will be of one sign, some of the other, and their algebraic sum will determine the constant error of this measurement process with respect to the particular distance concerned. Furthermore, the "constant error" of this measurement process will be different (at least, conceptually) for different distances measured.

In the case of repeated measurement of a single distance, the effect of (5), and at least portions of the effects of (3) and (4), may be expected to vary from one "occasion" to the next (e.g., from day to day), thus contributing variable components to the systematic error of the process.

A large fraction of the variable contributions of (3) and (4) could, and in practice no doubt would, be removed by stretching the tape by a spring balance or other means so that it is always under the same tension. The stretch corresponding to a particular distance would then be nearly the same at all times, and a fixed correction could be made for most of the sag corresponding to this distance. Furthermore, the effect of (5) could, and in practice probably would, be reduced by determining the temperature of the tape at various points along its length and applying a temperature correction. By comparison of the tape with a standard, the error arising from (1) could be eliminated entirely, and corrections determined as a basis for eliminating, or at least, reducing the effect of (2).

As in the foregoing example there are usually certain obvious sources of systematic error. Unfortunately, there are generally additional sources of systematic error, the detection, diagnosis, and eradication of which call for much patience and acumen on the part of the observer. The work involved in their detection, diagnosis, and eradication often far exceeds that of taking the final measurements, and is sometimes discouraging to the experienced observer as well as to the beginner. Fortunately, there are various statistical tools that are helpful in this connection, and Olmstead [1952] has found that of these the two most effective and universally useful are the average ($\bar{x}$) and range (R) charts of industrial quality control. (For details on the construction and use of $\bar{x}$- and R-charts, see, for example, the ASTM Manual on Quality Control of Materials [ASTM 1951, pp. 61-63 and p. 83]; or American Standards Z1.2-1958 and Z1.3-1958 [ASA 1958b, ASA 1958c].)

c. Concept of True Value

In the foregoing we have defined the error of a measurement x to be the difference $x - \tau$ between the measurement and the true value τ of the magnitude of the quantity concerned; and the systematic error, or bias, of a measurement process as the difference $\mu - \tau$ between the limiting mean associated with the measurement of a particular quantity by the measurement process concerned, and the true value τ of the magnitude of this quantity. This immediately raises the question: Just how is the "true value" of the magnitude of a particular property of some thing defined? In the final analysis, the "true value" of the magnitude of a quantity is defined by agreement among experts on an exemplar method for the measurement of its magnitude--it is the limiting mean of a conceptual exemplar process that is an ideal realization of the agreed-upon exemplar method. And the refinement to which one should go in specifying the exemplar process will depend on the purposes for which a determination of the magnitude of the quantity concerned is needed--not just the immediate purpose for which measurements are to be taken but also the other uses to which these measurements, or a final adjusted value derived therefrom, may possibly be put.

Consider, for example, the "true value" of the length of a particular gage block. In our minds we envisage the gage block as a rectangular parallelepiped, and its length is, of course, the distance between its two "end" faces. But it is practically certain that the particular gage block in question is not an exact rectangular parallelepiped; and that its two end faces are not planes, nor even absolutely smooth surfaces. Shall we define the "true length" of this gage block to be the distance between the "tops" of the highest "mountains" at each end, i.e., the distance between the two "outermost points" at each end? If so, is this distance to be measured diagonally, if necessary, or parallel to the "lengthwise axis" of the gage block? If the latter, then we have the problem of how this "length-wise axis" is to be defined, especially in the case of a thin gage block whose length corresponds to what would ordinarily be considered to be its thickness. Or shall we be, perhaps, more sophisticated, and envisage a "mean plane" at each end, which in general will not be parallel to each other, and define the length of this gage block to be the distance between two particular points on these planes. If we choose the "outermost points" we again have the problem of the direction in which the distance is to be measured. Alternatively, we might define the length of this gage block to be the distance between two strictly parallel and conceptually perfect optical flats "just touching" the gage block at each end. If so, then is the "true distance" between these flats defined in terms of wavelengths of light via the techniques of optical interferometry the "true length" of the gage block appropriate to the purposes for which the gage block is to be used, namely, to calibrate gages and to determine the lengths of other objects by mechanical comparisons? Furthermore, it is clear, that the intrinsic difficulty of defining the "true value" of the length of a particular gage block is not eliminated if, instead, we undertake to define the "true value" of the difference in length of two particular gage blocks, one of which is a standard, the accepted value of those length is, say, m microinches exactly, by industry, national or international agreement.

Similar difficulties arise, of course, in the definition of the "true value" of the mass of a mass standard, one of which has been resolved by international agreement. In defining the "true value" of the mass of a particular metallic mass standard, shall the mass of this particular standard be envisaged as the mass of its metallic substance alone, relative to the International Prototype Kilogram, or as the mass of its metallic substance plus the mass of the air and water vapor absorbed upon its surface under standard conditions? The difference amounts to about 45 μg in the case of a platinum-iridium standard kilogram, and becomes critical in the case of 500 mg standards. The mass of a mass standard is, therefore, specified in measurement science to be the mass of the metallic substance of the standard plus the mass of the average volume of air absorbed upon its surface under standard conditions. Definition of the "true value" of the mass of a mass standard, and a fortiori, of the difference in mass of two mass standards is, therefore, a very complex matter.

W. Edwards Deming uses the expression "preferred procedure" for what we have termed an "exemplar method," and very sagely remarks that "a preferred procedure is distinguished by the fact that it supposedly gives or would give results nearest to what are needed for a particular end; and also by the fact that it is more expensive or more time consuming, or even impossible to carry out," adding that "as a preferred procedure is always subject to modification or obsolescence, we are forced to conclude that neither the accuracy nor the bias of any procedure can ever be known in a logical sense." [Deming 1950, pp. 15-17.]

It should be evident from the foregoing that the "true value" of the magnitude of some property of a thing or system cannot be defined with complete absolute exactitude.

As Cassius J. Keyser has remarked, "Absolute certainty is a privilege of uneducated minds--and fanatics. It is, for scientific folk, an unattainable ideal." [Keyser 1922, p. 120.] The degree of refinement to which one will or ought to go in a particular instance will depend on the uses for which knowledge of the magnitude of the property concerned is needed. The "true value" of the length of a piece of cloth in everyday commerce is certainly a fuzzy concept. "Certainly we are not going to specify that the cloth shall be measured while suspended horizontally under a tension of x pounds, at an ambient temperature of y degrees and a relative humidity of z percent" [Simon 1946, p. 654]. On the other hand, a moderate degree of refinement is necessary in defining the "true length" and "true width" of the recessed area in a window sash to which a pane of glass is to be fitted. Considerably greater refinement is needed in the definition of the "true value" of the length of a gage block, of the mass of a mass standard or of the frequency of a frequency standard--and in the last mentioned case there is not today, I understand, complete agreement among experts on the matter.

Indeed, as is evident from the foregoing, the "true value" of the magnitude of a particular quantity is intimately linked to the purposes for which a value of the magnitude of this quantity is needed, and its "true value" cannot, in the final analysis, be defined meaningfully and usefully in isolation from these needs. Therefore, as this fact becomes more widely recognized in science and engineering, I hope that the traditional term "true value" will be discarded in measurement theory and practice, and replaced by some more appropriate term such as "target value"[6] that conveys the idea of being the value that one would like to obtain for the purpose in hand, without any implication that it is some sort of permanent constant preexisting and transcending any use that we may have for it. I have retained the traditional expression "true value" in the sequel because of its greater familiarity but shall always mean by it the relevant "target value."

d. Concepts of the Precision and Accuracy of a Measurement Process

By the precision of a measurement process we mean the degree of mutual agreement characteristic of independent measurements of a single quantity yielded by repeated applications of the process under specified conditions; and by its accuracy the degree of agreement of such measurements with the true value of the magnitude of the quantity concerned. In other words, the accuracy of a measurement process refers to, and is determined by the degree of conformity to the truth that is characteristic of independent measurements of a single quantity produced (or producible) by the repeated applications of the process under specified conditions; whereas its precision refers solely to, and is determined solely by the degree of conformity to each other characteristic of such measurements, irrespective of whether they tend to be close or far from the truth. Thus, accuracy has to do with closeness to the

[6] "We admit the existence of systematic error--of a difference between the quantity measured (the measured quantity) and the quantity of interest (the target quantity). We ask the observations about the measured quantity. We ask our subject matter knowledge, intuition, and general information about the relation between the measured quantity and the target quantity." [Cochran, et al. 1954, p. 33.]

".... Some people prefer the term 'true value', although others excoriate it as philosophically unsound.

"We could also call the reference level a 'target value'. In a way this is a bad term because it implies that it is something we want to find through the measurement process rather than something we ought to find because, like Mt. Everest, it is there. Unfortunately our desires can influence our notion of what is true, and we can even unconsciously bring the latter into agreement with the former; my use of the term 'target value' is not meant to imply that I think it legitimate to equate what we would like to see with what is there." [Murphy 1961, p. 265.]

truth; precision, only with closeness together.

This distinction between the meanings of the terms "accuracy" and "precision" as applied to measurement processes and measuring instruments is consistent with the etymological roots of these words. "Etymologically the term 'accurate' has a Latin origin meaning 'to take pains with' and refers to the care bestowed upon a human effort to make such effort what it ought to be, and 'accuracy' in common dictionary parlance implies freedom from mistakes or exact conformity to truth. 'Precise,' on the other hand, has its origin in a term meaning 'cutoff, brief, concise'; and 'precision' is supposed to imply the property of determinate limitations or being exactly and sharply defined." [Shewhart 1939, p. 124.] Thus one can properly speak of a national, state, or local law as being "precise," but not as being "accurate"--to what truth can it conform? On the other hand, if one spoke of a particular translation as being "accurate" this would imply a high degree of fidelity to the original "attained by the exercise of care." Whereas, to speak of it as being "precise," would imply merely that it is unambiguous, without indicating whether it is or is not correct.[7]

In spite of the distinct difference between the etymological meanings of the terms "accuracy" and "precision," they are treated as synonyms in many standard dictionaries; and Merriam-Webster [1942], after drawing the helpful distinctions quoted in the foregoing footnote, promptly topples the structure so carefully built by adding "scrupulous exactness" as an alternative meaning of "precise." Consequently it is not surprising that "There are probably few words as loosely used by scientists as precision and accuracy.--It is not unusual to find them used interchangeably in scientific writings." [Schrock 1950, p. 10.]

---

[7]It is sometimes helpful to distinguish between "correct," "accurate," and "exact": "CORRECT, the most colorless term, implies scarcely more than freedom from the fault or error, as judged by some (usually) conventional or acknowledged standard;... ACCURATE implies, more positively, fidelity to fact or truth attained by the exercise of care;... EXACT emphasizes the strictness or rigor of the agreement, which neither exceeds nor falls short of the fact, standard or truth;... PRECISE stresses rather sharpness of definition or delimitation..." [Merriam-Webster 1942 p. 203].

On the other hand, as Shewhart has remarked:

"Careful writers in the theory of errors, of course, have always insisted that accuracy involves in some way or other the difference between what is observed and what is true, whereas precision involves the concept of reproducibility of what is observed. Thus Laws, writing on electrical measurements, says:[8] 'Every experimenter must form his own estimate of the accuracy, or approach to the absolute truth obtained by the use of his instruments and processes of measurement. He must remember that a high precision, or agreement of the results among themselves, is no indication that the quantity under measurement has been accurately determined.' As another example we may take the following comment from a recent and authoritative treatise on chemical analysis:[9] 'The analyst should form the habit of estimating the probable accuracy of his work. It is a common mistake to confuse accuracy and precision. Accuracy is a measure of the degree of correctness. Precision is a measure of reproducibility in the hands of a given operator.'" [Shewhart 1939, pp. 124-125.]

More recently, Lundell, Hoffman, and their associates at the National Bureau of Standards have re-emphasized the importance of the distinction between "precision" and "accuracy":

"In discussions of chemical analysis, the terms precision and accuracy are often used interchangeably and therefore incorrectly, for precision is a measure of reproducibility, whereas accuracy is a measure of correctness. The analyst is vitally interested in both, for his results must be sufficiently accurate for the purpose in mind, and he cannot achieve accuracy without precision, especially since his reported result is often based on one determination and rarely on more than three determinations. The recipient of the analysis is interested in accuracy alone, and only in accuracy sufficient for his purposes." [Hillebrand et al., 1953, p. 3.]

---

[8] Frank A. Laws, Electrical Measurements, p. 593 (McGraw-Hill, New York, N. Y., 1917).

[9] G. E. F. Lundell and J. I. Hoffman, Outlines of Methods of Chemical Analysis, p. 220 (John Wiley and Sons, New York, N. Y., 1938).

It is most unfortunate that in everyday parlance we often speak of "accuracy and precision," because accuracy requires precision, but precision does not necessarily imply accuracy.

"It is, in fact, interesting to compare the measurement situation with that of a marksman aiming at a target. We would call him a precise marksman if, in firing a sequence of rounds, he were able to place all his shots in a rather small circle on the target. Any other rifleman unable to group his shots in such a small circle would naturally be regarded as less precise. Most people would accept this characterization whether either rifleman hits the bull's-eye or not.

"Surely all would agree that if our man hits or nearly hits the bull's-eye on all occasions, he should be called an accurate marksman. Unhappily, he may be a very precise marksman, but if his rifle is out of adjustment, perhaps the small circle of shots is centered at a point some distance from the bull's-eye. In that case we might regard him as an inaccurate marksman. Perhaps we should say that he is a potentially accurate marksman firing with a faulty rifle, but speaking categorically, we should have to say that the results were inaccurate." [Murphy 1961, p. 265.]

It follows from what has been said thus far that "if the precisions of two processes are the same but the biases are different, the process of smaller bias may be said to have higher accuracy while if the biases are both negligible, the process of higher precision may be said to have higher accuracy." Unfortunately, "in other cases such a simple comparison may be impossible." [ASTM 1961, p. 1760.]

To fully appreciate the preceding statement--and especially the difficulty of comparing accuracies in some cases--let us consider figures 1 and 2, in which the origins of the scales correspond to the true value of the quantity measured, so that the curves shown may be regarded as depicting the distributions of errors of the measurements yielded by a selection of different measurement processes. Consider first the three symmetrical distributions in the top half of figure 1. All three of these distributions are centered on zero, so that these measurement processes have no bias. It is evident that the process of highest precision, c, is also the process of highest accuracy; and that the process of least precision, a, is also the process of least accuracy. Since curve b in the upper half of figure 1 and curve d in the lower half have identical size and shape, the corresponding processes have the same precision; but process b is without bias, whereas

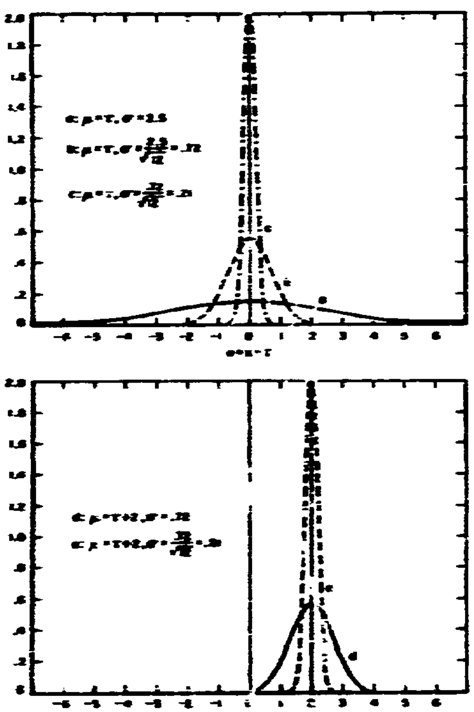


FIGURE 1. Distributions of errors of some biased and unbiased measurement processes of various precisions

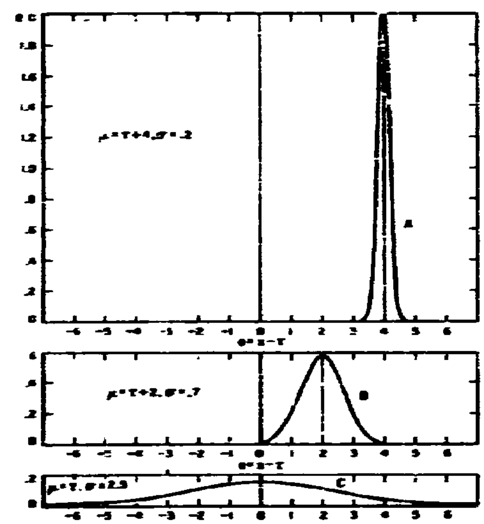


FIGURE 2. Three measurement processes differing from each other with respect to both precision and accuracy.

process d has a positive bias of two units, so that process b is clearly the more accurate. (In particular we may note that whereas it is practically certain that process b will not yield a measurement deviating from the truth by more than two units, exactly one-half of the measurements yielded by process d will deviate from the truth by this much or more.) Similar remarks clearly apply to processes c and e corresponding to curve c in the upper half and curve e in the lower half of figure 1, but in this instance the superiority of process c relative to process e with respect to accuracy is even more marked. (In particular, we may note that whereas it is practically certain that no measurement yielded by process c will deviate from the truth by as much as one unit, it is practically certain that every measurement yielded by process e will deviate from the truth by more than one unit.)

Figure 2, which is essentially the same as one given by General Simon [1946, fig. 1], portrays three measurement processes A, B, and C, differing from each other with respect to both precision and bias. Comparison of these three processes with respect to accuracy is not quite so simple. First, it is evident that, although process A has greater precision than process B, process B is the more accurate of the two. (In particular, it is practically certain that none of the measurements yielded by process B will deviate from the truth by more than 4 units, whereas 50 percent of the measurements from process A will deviate from the truth by four units or more.) Next, is process B more (or less) accurate than process C which is unbiased, but has a very low precision? Process B has a positive bias of two units, but has sufficiently greater precision than process C to also have greater accuracy than process C. (While approximately 50 percent of the measurements yielded by process C will deviate from the truth by more than two units (in either direction), and exactly 50 percent of the measurements yielded by process B will deviate from the truth by two units or more (in the positive direction only), it cannot be ignored that about 10 percent of the measurements yielded by process C will deviate from the truth by four units or more whereas it is practically certain that no measurement yielded by process B will deviate from the truth by as much as four units.) Similarly, it may be argued that process A, in spite of its bias, has greater accuracy than process C "since the range in measurements of C more than covers the corresponding ranges of A or B." [Simon 1946, p. 654.] While this conclusion that of the three measurement processes depicted in figures 2, process C has the least accuracy, may not be entirely acceptable to some persons, it is consistent with Gauss' dictum, in a letter to F. W. Bessel, to the effect that maximizing the probability of a zero error is less important than minimizing the "average" injurious effects of errors in general. [C. F. Gauss, 1839, pp. 146-147.]

Before leaving figure 2, we must not fail to join General Simon in remarking that "the average of a large number of measurements from [process] C will be more accurate than a similar average from either A or B" [Simon 1946, p. 654]. This point is actually illustrated in our figure 1: the three curves in the top half of figure 1 portray the distributions of errors of single measurements (curve a) of averages of 12 measurements (curve b) and averages of 144 measurements (curve c) from process C ; and curves d and e in the lower half show the distributions of errors of individual measurements (curve d), and of averages of 12 measurements (curve e) from process B, respectively. It is evident that averages of 12 measurements from process C (curve b in upper portion of fig. 1) have not only greater accuracy than individual measurements from process B (curve d in lower portion of the figure), but also greater accuracy than averages of 12 measurements from process B (curve e in lower portion).

On the other hand, it is obvious that, if our choice is between individual measurements from process C (curve a) and averages of 12 measurements from process B (curve e), the latter will clearly provide greater accuracy. In brief, a procedure with a small bias and a high precision can be more accurate than an unbiased procedure of low precision. It is important to realize this, for in practical life it is often far better to always be quite close to the true value than to deviate all over the place in individual cases, but strictly correct "on the average," like the duck hunter who put one swarm of shot ahead of the duck, and one swarm behind, lost his quarry, but had the dubious satisfaction of knowing that in theory he had hit it "on the average." This we must remember: in practical life we rarely make a very large number of measurements of a given type--we can't wait to be right on the average--our measurements must stand up in individual cases as often as possible.

Despite the foregoing, freedom from bias, that is, freedom from "large" bias, is a desirable characteristic of a measurement process. After all we want our measurements to yield us a determination that we can use as a substitute for the unknown value of a particular magnitude whose value we need for some purpose--we don't want a determination of the value of some other magnitude whose relation to the one we need is indefinitely known.

In view of the difficulty of comparing with respect to accuracy measurement processes that differ both in bias and precision, some writers have elected to take the easy way out by defining "accuracy" to be equivalent to absence of bias, saying that two measurement processes having different biases, the process of smaller bias is the more "accurate" regardless of the relation of their respective precisions. (See, for example, Beers [1953, p. 4], Ostle [1954, p. 4], and Schenck [1961, p. 4, p. 14].) While the adoption of the concept of "accuracy" certainly makes the discussion of "accuracy" and "precision" simpler for the authors concerned, this practice is contrary to the principle of "conservation of linguistic resources," as R. B. Murphy puts it, adding: "It seems to me that the terms 'bias' and 'systematic error' are adequate to cover the situation with which they are concerned. If, nevertheless, we add the term 'accuracy' to apply again in this restricted sense, we are left wordless--at the moment at least--when it comes to the idea of over-all error. From the point of view of the need for a term it is hard to defend the view that accuracy should concern itself solely with bias... [and] there is overwhelming evidence that we need a term at least for the concept of over-all error." [Murphy 1961, pp. 265-266.]

3.4. Mathematical Specification of the Precision of a Measurement Process.

a. Simple Statistical Control

Let us now consider the mathematical definition of the precision of a measurement process under a fixed set of circumstances. By definition, the precision of a measurement process has to do with the "closeness together" that is typical of successive measurements of a single quantity generated by applications of the process under these fixed conditions. Otherwise expressed, it has to do with the typical "closeness together" of the two individual measurements constituting an arbitrary pair. If the expression "typical 'closeness together'" is to be meaningful, the measurements generated by repeated application of the process to the measurement of a single quantity must be homogeneous in some sense. Therefore, for the moment, let us assume that the measurement process is in a state of simple statistical control, so that the successive measurements in each of the sequences (1), (i = 1, 2, 3, ...), generated by the process may all be regarded as "observed" values of independent identically distributed random variables.

Just as we may regard each individual measurement x_{ij} in a particular sequence (1) as striving to express the value of the limiting mean μ, so also we may regard each individual difference $x_{ij} - x_{ik}$, $j \neq k$, as striving to express the characteristic spread between an arbitrary pair of measurements, x' and x", say. For this purpose the signs of these differences are clearly irrelevant. Therefore, by analogy with our use of a sequence of cumulative arithmetic means, (2), to achieve a mathematical formulation of the concept of a limiting mean associated with measurement of a given quantity by a particular measurement process, let us adopt the sequence of cumulative arithmetic means of the squares of the $n(n-1)/2$ distinct differences among the first n measurements of a particular sequence (1), for example, the sequence

$$(3)\qquad \overline{(d^2)}_{in} = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} (x_{ij} - x_{ik})^2 , \qquad (n = 2, 3, \ldots),$$

as the basis of a mathematical formulation of the concept of the precision of a measurement process.

The necessary and sufficient condition for almost sure convergence of the sequence (3) to a finite limit, say Δ^2, is that the Strong Law of Large Numbers be applicable to the sequence.

$$(4)\qquad x_{i1}^2, x_{i2}^2, \ldots, x_{ij}^2, \ldots,$$

consisting of the squares of the corresponding terms of the original sequence (1). (Boundedness of the x's in addition to statistical control is, for example, sufficient to ensure that the sequence (4) will also obey the Strong Law of Large Numbers.) If the Strong Law of Large Numbers is applicable to the sequence of squares (4), and if the measurement process is in a state of simple statistical control, then the cumulative arithmetic means of the squares of the measurements, that is, the sequence

$$(5)\qquad \overline{(x^2)}_{in} = \sum_{j=1}^{n} x_{ij}^2/n, \qquad (n = 1, 2, \ldots),$$

will almost surely tend to a limit, say S, the magnitude of which will depend on the quantity measured, the measurement process involved, but not on the "occasion" (identified by the subscript "i"). By virtue of an algebraic identity that is well known to students of mathematical inequalities, namely,

$$(6)\qquad n\sum_{j=1}^{n} a_j^2 - \left(\sum_{j=1}^{n} a_j\right)^2 = \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}(a_j - a_k)^2, \qquad (n \geq 2)$$

and of the fact that the right-hand side of (6) is always positive except when the a's are all equal, it is easily seen, on dividing both sides of (6) by n^2, that S will always exceed μ^2, the square of the (almost sure) limit of the sequence (2), so that we may write $S = \mu^2 + \sigma^2$, with $\sigma^2 > 0$. Furthermore, applying the algebraic identity (6) in reverse to the right-hand side of (3) yields the following relationship between the corresponding terms of sequences (3), (5), and (1):

$$(7)\qquad \overline{(d^2)}_{1n} = 2\left(\frac{n}{n-1}\right)\left\{\overline{(x^2)}_{1n} - \overline{(x_{1n})}^2\right\} > 0, \qquad (n \geq 2).$$

Hence, if a measurement process is in a state of simple statistical control and the Strong Law of Large Numbers is applicable to a __sequence__ of squared measurements (4), then the sequence $\overline{(d^2)}_{1n}$, defined by (3), will, in view of (7), tend almost surely to a finite limit $\Delta^2 = 2\sigma^2$. Thus we see that σ^2, termed the __variance__ of the measurement process, is the mean value of one-half of the squared difference between two arbitrary measurements x' and x'', that is,

$$(8)\qquad \sigma^2 = \frac{1}{2}\overline{(x' - x'')^2}$$

and provides an indication of the __imprecision__ of the process. The square root of the variance, σ, is termed the __standard deviation__ of the process.

It is natural, therefore, on the basis of a single sequence of __n__ measurements of a single quantity, to take

$$(9)\qquad s^2 = \frac{1}{2}\overline{(d^2)} = \frac{1}{n(n-1)}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}(x_j - x_k)^2 = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1}$$

as the sample estimate of the underlying variance σ^2; and the square root, s, as the sample estimate of σ.[10]

From (9), since $\bar{x} = \bar{x}_n$ tends (almost surely) to μ, it is evident that σ^2 is also the mean value of the squared deviations of individual measurements from the limiting mean μ of the process, that is $\sigma^2 = \overline{(x - \mu)^2}$, so that the standard deviation σ may be regarded, in the language of mechanics, as the radius of gyration of the distribution of all possible measurements x about μ, the limiting mean of the process.

Remark: Mathematically the foregoing discussion can be carried out equally well in terms of the absolute (unsigned) values of the differences instead of in terms of their squares. Such an approach is, mathematically speaking, somewhat more general in that it requires for its validity merely that the Strong Law of Large Numbers be applicable to the sequence $|x_{i1}|, |x_{i2}|, \ldots, |x_{ij}|, \ldots$ of absolute values of the x_{ij} rather than to the sequence (4) of their squares. From the practical viewpoint, however, this greater generality is entirely illusory, and the mathematics of absolute values of variables is always more cumbersome than the mathematics of their squares. For example, the arithmetic mean of the absolute values of the n(n - 1)/2 distinct differences among n measurements, i.e.,

$$\overline{|d|}_n = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} |x_j - x_k| \tag{10}$$

is not expressible as a multiple of the sum of the absolute deviations of the measurements from their mean $\sum |x_i - \bar{x}|$, and for large values of n the

---

[10] From the algebraic identity (6), it is evident that the practice in some circles of dividing $\sum_{j=1}^{n} (x - \bar{x})^2$ by n, instead of n - 1, amounts to including each of the distinct squared differences $(x_j - x_k)^2$, $j \neq k$, twice in the summation, together with n identically zero terms $(x_j - x_k)^2$, $j = k$, each included once, and then dividing by n^2, the total number of terms (real and phantom) involved. Viewed in this light it would seem that division by n - 1 is more reasonable, in that the inclusion of identically zero terms in the formulation of a measure of variation is a bit unreasonable.

evaluation of (10) presents computational difficulties. The approach in terms of the absolute values of the differences also has the disadvantage from the practical viewpoint that, as we shall see in a moment, components of imprecision are additive in terms of squared quantities such as σ^2, so that in this sense the variance σ^2 is a more appropriate measure of the dispersion of the x's about their limiting mean μ than is σ itself.

Ordinarily, the magnitude of σ^2 (and, hence, of σ), unlike that of μ, depends only on the measurement process concerned and the circumstances under which it is applied, and not also on the magnitude of the quantity measured--otherwise we could not speak of a measurement process having a variance, or a standard deviation.

Since the precision of the process obviously decreases as the value of σ (or, of σ^2) increases, and vice versa, it is necessary to take some inverse function of σ as a measure of the precision of process. To conform with traditional usage it is necessary to regard the precision of a measurement process as inversely proportional to its standard deviation σ which is, therefore, a measure of the imprecision of the process. Thus, Gauss, writing in 1809, remarked that his constant $h = 1/\sigma\sqrt{2}$ could properly be considered to be a measure of the precision of the observations because if, for example $h' = 2h$, that is, if $\sigma' = 1/2\,\sigma$, then "a double error can be committed in the former system with the same facility as a single error in the latter, in which case, according to the common way of speaking, a double degree of precision is attributed to the latter observations."[11]

The fact of the matter is, however, that:

"... different fields have particularly favorite ways of expressing precision. Most of these measures are multiples of the standard deviation; it is not always clear which multiple is meant....

---

[11] "Ceterum constans h tamquam mensure praecisionis observationum considerari poterit. ... Quodsi igitur e.g., h' = 2h, aeque facile in systemate priori error duplex committi poterit; ac simplex in posteriori, in quo casu observationibus posterioribus secundum vulgarem loquendi morem praecisio duplex tribuitur." [Gauss 1809, Art. 178; 1871, p. 233; English translation, 1857, pp. 259-260.]

"Some consider it unfortunate that precision should be stated as a multiple of standard deviation, since precision should increase as standard deviation decreases. Indeed, it would be more exact to say that standard deviation is a measure of imprecision. However, sensitivity, as we have previously indicated, suffers from this logical inversion without hurt. Perhaps we can best avoid this by saying that standard deviation is an index of precision. The habit of saying 'The precision is ...' is deeply rooted, and there would be understandable impatience with the notion that standard deviation should be numerically inverted before being quoted in a statement of precision." [Murphy 1961, pp. 266-267.]

In consequence the ASTM has, at least tentatively, taken the following position:

"The numerical value of any commonly used index of precision will be smaller the more closely bunched are the individual measurements of a process. As more causes are added to the system, the greater the numerical value of the index of precision will ordinarily become. If the same index of precision is used on two different processes based on the same method or intended to measure the same physical property, the process that has the smaller value of the index of precision is said to have higher precision. Thus, although the more usual indexes of precision are really direct measures of imprecision, this inversion of reference has been firmly established by custom. The value of the selected index of precision of a process is referred to simply as its precision or its stated precision." [ASTM 1961, p. 1759.]

As we have remarked previously, in practical work the end result of measuring some quantity or calibrating an instrument for a standard rarely consists of a single measurement of the quantity of interest. More often it is some kind of average or adjusted value, for example, the arithmetic mean of a number of independent measurements of the quantity of interest. Let us, therefore, consider the statistical properties of a sequence of arithmetic means of successive nonoverlapping groups of n measurements each from a sequence (1) of individual measurements yielded by a measurement process on a particular occasion. In other words, let us consider the sequence

(11) $$\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{im}, \ldots$$

of distinct arithmetic means of n measurements each

$$\bar{x}_{im} = \frac{1}{n} \sum_{j=(m-1)n+1}^{mn} x_{ij}, \qquad (m = 1, 2, \ldots), \tag{12}$$

derived from a sequence (1) of individual measurements of a single quantity produced, or at least conceptually producible, by the measurement process concerned on, say, the ith occasion. If the "underlying measurement process" giving rise to the individual measurements x_{ij} is in a state of simple statistical control, then the "extended measurement process" giving rise to the averages $\bar{x}_{im}$ will also be in a state of simple statistical control. Consequently, the mathematical analysis of section 3.2, but with the averages $\bar{x}_{im}$ in place of the individual measurements x_{ij}, will carry through without other change. Let $\mu_{\bar{x}}$ denote the limiting mean thus associated with the "extended measurement process" giving rise to the averages $\bar{x}_{im}$ as its "individual" measurements. Since the cumulative arithmetic mean of the first m terms of the sequence (11) is the same as the cumulative arithmetic mean of the first mn terms of the sequence (1) of individual measurements, it is clear that the limiting mean $\mu_{\bar{x}}$ associated with the sequence of averages (11) is the same as the limiting mean associated with the original sequence (1) of individual measurements, that is,

$$\mu_{\bar{x}} = \mu_{x} = \mu \tag{13}$$

Similarly, the mathematical analysis at the beginning of the present section, but with the individual measurements x_{ij} in (3) thru (9), replaced by the averages $\bar{x}_{im}$, carries through essentially as before. Let $\sigma^2_{\bar{x}}$ denote the variance thus associated with the "extended measurement process" giving rise to the sequence of averages (11). As in the case of the variance σ^2 of individual measurements, so also may $\sigma^2_{\bar{x}}$ be interpreted as the overall mean value of the squared deviation of "individual" averages $\bar{x}$ from the limiting mean $\mu_{\bar{x}}$ of the "extended process," that is,

$$\sigma^2_{\bar{x}} = \overline{(\bar{x} - \mu_{\bar{x}})^2} = \overline{(\bar{x} - \mu)^2} \tag{14}$$

By virtue of the algebraic identity

(15)
$$(\bar{x} - \mu)^2 = \left[\frac{1}{n} \sum_{j=1}^{n} x_j - \mu \right]^2 = \left[\frac{1}{n} \sum_{j=1}^{n} (x_j - \mu) \right]^2$$
$$= \frac{1}{n^2} \left[\sum_{j=1}^{n} (x_j - \mu)^2 + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} (x_j - \mu)(x_k - \mu) \right]$$

it is readily seen that

(16)
$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{\sigma^2}{n} .$$

(The mean value of a sum is always the sum of the mean values of its individual terms, so that the overall mean value of the first summation inside the brackets in the last line of (15) is simply $n \sigma_x^2$. Furthermore, in the case of independent identically distributed measurements, the overall mean value of the term involving the double summation is 0.)

Since, from (16), $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, it is seen that the precision of the arithmetic mean of n independent measurements is proportional to $\sqrt{n}$. Hence the arithmetic mean of 4 independent measurements has double the precision of a single measurement; the mean of 9 independent measurements, thrice the precision of a single measurement; and 144 independent measurements will be required if their arithmetic mean is to have a 12-fold increase in precision over a single measurement. (But to ask for a 12-fold increase in precision is to ask for a very considerable improvement indeed, as can be seen from a comparison of curves a and c in the top half of fig. 1.)

To serve as a reminder of the distinction between the standard deviation of an individual measurement and the standard deviation of a mean $\bar{x}$, it is customary to refer to σ as the "standard deviation" of a single measurement x, and to $\sigma_{\bar{x}}$ as the "standard error" of the (arithmetic) mean $\bar{x}$.

b. Within-Occasions Control

In the foregoing it has been assumed that the individual measurements comprising the sequences (1) corresponding to the respective "occasions," (i = 1, 2, ...), could all be regarded as "observed values" of independent identically distributed random variables, that is, that the measurement process concerned was in a state of simple statistical control. When such is the case then any subset of n measurements is strictly comparable to any other subset of n measurements, and any two such subsets can be combined and regarded validly as a single set of 2n measurements. Unfortunately, as Student's comment quoted on page 484 above clearly implies, such complete homogeneity of measurement is rarely if ever met in practice. More often the situation is as described by Sir George Biddell Airly, British Astronomer Royal 1835-1881, in (to my knowledge) the first elementary book on the theory of errors and combination of observations in the English language [Airy 1861, p. 92]:

"When successive series of observations are made, day after day, of the same measurable quantity, which is either invariable... or admits of being reduced by calculation to an invariable quantity...; and when every known instrumental correction has been applied...; still it will sometimes be found that the result obtained on one day differs from the result obtained on another day by a larger quantity than could have been anticipated. The idea then presents itself, that possibly there has been on some one day, or on every day, some cause, special to the day, which has produced a Constant Error in the measures of that day."

Sir George, however, cautions against jumping to conclusions on the basis of only a few observations:

"The existence of a daily constant error...ought not to be lightly assumed. When observations are made on only two or three days, and the number of observations on each day is not extremely great, the mere fact, of accordance on each day and discordance from day to day, is not sufficient to prove a constant error. [And we should interject here that under such circumstances apparent over-all accordance is not sufficient to prove the absence of daily constant errors either.] The existence of an accordance analogous to a 'round of luck' in ordinary chances is sufficiently probable.... More extensive experience, however, may give greater confidence to the assumption of constant errors...first, it ought, in general to be established that there is possibility of error, constant on one day but varying from day to day...." [Airy 1861, p. 93.]

The most useful statistical tools for this purpose are the control-chart techniques of the industrial quality control engineer. If in such a situation, a series of measurements obtained by measurement of a single quantity a number of times on each of several different days or "occasions" by a particular measurement process is plotted in the form of a control chart for individuals [ASTM 1951, pp. 76-78, and pp. 101, 105], the individual measurements so plotted will be seen to consist of "sections" identifiable with the subsequences (i) corresponding to the respective "occasions," (i = 1, 2, 3, ...), with the measurements within sections pair-wise closer together on the average than two measurements one of which comes from one section and the other from another. Such a series of measurements is clearly "out of control." If now parallel $\bar{X}$- and R-charts are constructed from these data, based on a series of samples of equal size from within the respective "occasions" or "sections" only, i.e., excluding means $\bar{X}$ and ranges R of any samples that "straddle" two occasions, and the points on the resulting $\bar{X}$-chart are clearly "out of control," then we may infer the existence of day-by-day components of error, constant, perhaps, on one day, but varying from day to day.

If points on the R-chart constructed as described are "out of control" also, then the measurement operation concerned is in a completely unstable condition and cannot be described validly as a "measurement process" at all. On the other hand, if the $\bar{X}$-chart is "out of control," but the R-chart is "in control," then we may regard the measurement process as being in a state of within-occasions control. ("It is usually not safe to conclude that a state of control exists unless the plotted points for at least 25 successive subgroups fall within the 3-sigma control limits. In addition, if not more than 1 out of 35 successive points, or not more than 2 out of 100, fall outside the 3-sigma control limits, a state of control may ordinarily be assumed to exist." [ASA 1958c, p. 16.]) In such a situation we postulate the existence of (at least, conceptually) different limiting means μ_i for the respective "occasions" (i = 1, 2, ...), and a common within-occasions variance σ_w^2.

An unbiased estimate of the with-occasions standard deviation σ_w can be obtained, if desired, from the average range $\bar{R}$ used in constructing the R-chart, by means of the formula

$$\text{(17)} \qquad \text{unbiased estimate of } \sigma_w = \bar{R}/d_2$$

where d_2 is the factor given in the d_2 column of table B_2 of [ASTM 1951, p. 115] corresponding to the sample or subgroup size n used in constructing the R-chart.

Alternatively, if desired, an unbiased estimate of σ_w^2 can be obtained directly from the measurements involved by means of the formula

$$(18) \qquad \text{unbiased estimate of } \sigma_w^2 = s_w^2 = \frac{\sum_{h=1}^{k} \sum_{j=1}^{n} (x_{hj} - \bar{x}_h)^2}{k(n-1)}$$

where x_{hj} denotes the jth measurement and $\bar{x}_h$ the arithmetic mean of the n measurements of the hth subgroup, respectively, and k is the number of subgroups involved in constructing the R-chart.

c. Complex or Multistage Control

When a measurement process is not in a state of simple statistical control that satisfies the criteria of within-occasions control, that is, when the $\bar{x}$-chart (and control chart for individuals) are clearly "out of control," but the 25 or more subgroup ranges plotted on the R-chart exhibit control, then it is usually of importance to ascertain whether the measurement process concerned is possibly in a state of complex or multistage statistical control. For this purpose four or more measurements from each of at least 25 different occasions will be needed. Taking one sample of n successive measurements, ($4 \leq n \leq 10$), from the available measurements corresponding to each of, say, k(≥ 25) different "occasions," evaluate the arithmetic means $\bar{x}_i$ of these samples, ($i = 1, 2, \ldots, k$), and treating these averages as INDIVIDUAL measurements construct a control chart for these "individuals" and parallel $\bar{x}$- and R-charts as described in [ASTM 1951, Example 22, p. 101]. If the points plotted on these three control charts exhibit control, then we "act for the present as if" the measurement process concerned is in a state of complex or multistage statistical control and regard the limiting means μ_i for the respective "occasions," ($i = 1, 2, \ldots$) as being in a state of simple statistical control with a limiting mean μ and variance σ_b^2, termed the between-occasions component of variance.

If in such a situation we were to form cumulative arithmetic means such as (3) of the squares of all distinct differences between arbitrary pairs of measurements from <u>within</u> each of the respective "occasions," then such cumulative arithmetic means of squares of differences would almost surely tend to $2\sigma_w^2$ in the limit as the number of pairs included tends to infinity, where σ_w^2 is the "within-occasions variance" mentioned above in connection with "within-occasions control." If, on the other hand we were to form similar cumulative arithmetic means of the squares of differences between arbitrary pairs consisting in each instance of one measurement from each of two different sections, then such a cumulative arithmetic mean of squared differences would tend almost certainly to $2(\sigma_w^2 + \sigma_b^2)$ as the number of "occasions" sampled tends to infinity, where σ_b^2 is the above mentioned "between-occasions variance," i.e., the variance of the limiting means μ_i for the respective "occasions" about their limiting mean μ.

If in utilizing measurements from a measurement process that is in such a state of complex statistical control, one forms an average $\bar{x}_N$ that is the arithmetic mean of a total of N = kn measurements, composed of n measurements from each of k different "occasions," then the variance of $\bar{x}_N$ will be

$$\text{(19)} \qquad \sigma_{\bar{x}_N}^2 = \overline{(\bar{x}_N - \mu)^2} = \frac{1}{k}\left(\sigma_b^2 + \frac{\sigma_w^2}{n}\right).$$

From (19) it is clear that, if σ_b^2 is at all sizable compared to σ_w^2, then, for fixed N = kn, $\bar{x}_N$ will have greater precision <u>as a determination of μ</u> when based on a large number k of different occasions, with only a small number n of measurements from each occasion. Finally, setting k = 1, we see that the mean $\bar{x}_i$ of n measurements all taken on the same occasion <u>considered as a determination of the overall limiting mean μ</u> has an <u>overall variance</u> $\sigma_{\bar{x}}^2 = \sigma_b^2 + (\sigma_w^2/n)$; but <u>considered as a determination of μ_i</u>, the limiting mean for the ith occasion, its variance is only σ_w^2/n. In other words, the "standard error" of a mean such as $\bar{x}_i$ is not unique, but depends on the purpose for which it is to be used.

An unbiased estimate of the overall standard deviation $\sigma_{\bar{x}_i}$ of the arithmetic mean of n measurements taken on a single "occasion" may be obtained by the procedure of formula (17) above, if desired, using the average range $\bar{R}$ employed in constructing the R-chart corresponding to the groups of averages $\bar{x}_{in}$.

Alternatively, an unbiased estimate of the overall variance $\sigma^2_{\bar{x}}$ can be obtained directly from the means $\bar{x}_i$ used in constructing the $\bar{x}$ -chart, by using the formula

$$(20) \qquad s^2_{\bar{x}} = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{\bar{x}})^2}{k - 1}$$

where $\bar{x}_i$ is the arithmetic mean of the n successive observations from the ith "occasion," (i = 1, 2, . . ., k) and $\bar{\bar{x}}$ is the arithmetic mean of those k means.

The foregoing concept of a state of <u>complex or multistage statistical control</u> can be extended readily to more complex truly "multistage" situations involving three or more "levels" or random variation.

Finally, it is evident from the foregoing that when a measurement process is in a state of complex or multistage statistical control, then the difference between two individual measurements (or the arithmetic means of n measurements) corresponding to two different "occasions" will include the difference $\mu_i - \mu_{i'}$, between the limiting means corresponding to the two particular occasions involved. In so far as such a comparison is regarded as a unique individual case, the difference $\mu_i - \mu_{i'}$, is a fixed constant and hence a systematic error affecting this comparison. On the other hand, if the difference between these two individual measurements (or those two arithmetic means) is regarded only as a typical instance of the outcomes that might be yielded by the same measurement process on <u>other</u> pairs of occasions, then the difference $\mu_i - \mu_{i'}$ may be regarded as a random component having a zero mean and variance $2\sigma^2_b$.

If goes without saying, of course, that if a control-chart analysis of the type described above is undertaken for the purpose of ascertaining whether the process is in a state of complex control, but the points plotted on the $\bar{x}$-chart are clearly "out of control," then the measurement process concerned cannot be regarded as statistically stable from occasion to occasion, and should be used only for <u>comparative measurement</u> within-occasions. Even when such a measurement process is used solely for comparative measurement within "occasions," it needs to be shown that comparative measurements of <u>fixed differences</u> are in a state of (simple or complex) statistical control, if this measurement process is to be generally valid in any absolute sense. Thus in the case of the thermometer calibration procedure mentioned in section 2.4 above, one needs to examine the results of repeated measurement, occasion after occasion, of the difference between two standard thermometers $\underline{S}_1$ and $\underline{S}_2$ of proven stability in order to determine whether the process is or is not in a state of simple or complex statistical control.

3.5. <u>Difficulty of Characterizing the Accuracy of a Measurement Process.</u> Unfortunately, there does not exist any single comprehensive measure of the accuracy (or <u>in</u>accuracy) of a measurement process (analogous to the standard deviation as a measure of its imprecision) that is really satisfactory. This difficulty stems from the fact that "accuracy," like "true value," seems to be a reasonably definite concept on first thought, but as soon as one attempts to specify exactly what one means by "accuracy" in a particular situation, the concept becomes illusive; and in attempting to resolve the matter one comes face to face, sooner or later, with the question: "Accurate" for what <u>purpose</u>?

Gauss, in his second development (1821-1823) of the Method of Least Squares clearly recognized the difficulty of characterizing sharply the "accuracy" of any particular procedure:

"Quippe quaestio haec per rei naturam aliquid vagi implicat, quod limitibus circumscribi nisi per principium aliquatenus arbitrarium nequit... neque demonstrationibus mathematicis decidenda, sed libero tantum arbitrio remittenda."[12] [Gauss 1823, Part I, Art. 6.]

---

[12] I am grateful to may colleague Frank Alt for the following literal translation of these phrases:

"For this question implies, by the very nature of the matter, something vague which cannot be clearly delimited except by somewhat arbitrary principle ... nor can it be decided by mathematical demostrations, but must be left to mere arbitrary judgment."

Gauss himself proposed loc. cit. that the <u>mean square error</u> of a procedure -- that is, $\sigma^2 + (\mu - \tau)^2$, where σ is its <u>standard deviation</u>; and $\mu - \tau$, its <u>bias</u> be used to characterize its accuracy. While <u>mean square error</u> is a useful criterion for comparing the relative accuracies of measurement processes differing widely in both precision and bias, it clearly does not "tell the whole story." For example, if one were to accept the principle that measurement processes having the same mean square error were equally "accurate," then one would be obliged to consider the measurement processes corresponding to the three curves shown in figure 3 as being of equal accuracy, whereas for many purposes one would regard process C (portrayed to the right) as the "most accurate," in spite of the fact that the chances of scoring a "bull's eye" or "near miss" are greater in the case of process A shown in the upper left.

Alternatively, if one were to say that two measurement processes were equally accurate when exactly the same proportion P of the measurements of each lay within $\pm\sigma$ units from the true value, then for $P = 0.5$ one would be obliged to say that the measurement processes corresponding to curves e and d in the lower half of figure 1 were equally accurate, and that the measurement process corresponding to curve a in the upper half of the same figure was slightly more accurate than either e or d. Or, taking $P = 0.95$, one would be obliged to say that the measurement processes corresponding to the three curves shown in figure 4 were equally accurate. From these, and other cases easily constructed, it is readily seen that it is unsatisfactory to regard two measurement processes as being equally accurate if the same specified fraction P of the measurements produced by each lie within the same distance from the true value.

Thus one is led by the force of necessity to the inescapable conclusion that ordinarily (at least) two numbers are needed to adequately characterize the accuracy of a measurement process. And this has been recognized by the American Society for Testing and Materials in their recent recommendations [ASTM 1961, pp. 1759-1760]:

"Generally the index of accuracy will consist of two or more different numbers. Since the concept of accuracy embraces not only the concept of precision but also the idea of more or less consistent deviation from the reference level (systematic error or bias), it is preferable to describe accuracy by separate values indicating precision and bias."

The fact of the matter is that two numbers ordinarily suffice only because the "end results" of measurement and calibration programs are usually averages or adjusted values based on a number of independent "primary measurements," and such averages and adjusted values tend to be normally distributed to a very good approximation when four or more "primary measurements" are involved. This is illustrated by figure 5, which shows the distributions of individual measurements of two unbiased measurement processes with identical standard deviations but having uniform and normal "laws of error," respectively, together with the corresponding distributions of arithmetic means of 4 independent measurements from these respective processes--these latter two distributions are depicted by a single curve because the differences between the two distributions concerned are far less than can be resolved on a chart drawn to this scale. Since both of the processes concerned are unbiased, "accuracy" thus becomes only a matter of "precision"--or does it?--both curves for n = 1 have the same standard deviation, do they reflect equal "accuracy"? Would not the answer depend on the advantages to be gained from the small errors balanced against the seriousness of large errors, in relation to the purpose for which a single measurement from one or the other is needed? But "the problem" disappears nicely if averages of 4 measurements are to be used.

4. EVALUATION OF THE PRECISION, AND OF CREDIBLE BOUNDS TO THE SYTEMATIC ERROR OF A MEASUREMENT PROCESS.

As we have just seen, two numbers are ordinarily needed to characterize the accuracy of a measurement process, the one indicating its precision, and the other its bias. In practice, however, the bias of a measurement process is unknown and unknowable because the "true values" of quantities measured are almost always unknown and unknowable. The principle exception is when one is measuring a difference that is by hypothesis identically zero. If the bias of a measurement process could be, and were known exactly, then one would of course subtract it off as a "correction" and thus dispose of it entirely. Since ordinarily we cannot expect to know the exact magnitude of the bias of a measurement process, we are forced in practice to settle for credible bounds to its likely magnitude--much as did Steyning and the thief in chapter VI of Kipling's story, Captains Courageous: "Steyning tuk him for the reason that the thief tuk the hot stove--bekaze for there was nothing else that season". Consequently, neither the bias nor the accuracy of any measurement process, or method of measurement, can ever be known in a logical sense. The precision of a measurement process, however, can be measured and known. (Compare Deming [1950, p. 17].)

4.1. Evaluation of the Precision of a Measurement Process. In the foregoing we have stressed that a measurement operation to qualify as a measurement process must have attained a state of statistical control; and that until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all. It is also clear, from our discussion of the control-chart techniques for determining whether in any given instance one is entitled to "act for the present as if" a state of statistical control has been attained, that a fairly large amount of experience with a particular measurement process is needed before one can resolve the question in the affirmative. Once a measurement process has attained a state of statistical control and so long as it remains in this state, then an estimate of the standard deviation of the process can be obtained from the data employed in establishing control, as we have indicated above.

Since the precision of a measurement process refers to, and is determined by the characteristic "closeness together" of successive independent measurements of a single magnitude generated by repeated application of the process under specified conditions, it is clearly necessary in determining whether a measurement operation is or is not in a state of statistical control, and in evaluating its precision to be reasonably definite on what variations of procedure, apparatus, environmental conditions, observers, operators, etc., are allowable in "repeated applications" of what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. If whatever measure of the precision and bounds to the bias of the measurement process we may adopt are to provide a realistic indication of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Scientists and engineers commonly append "probable errors" or "standard errors" to the results of their experiments and tests. These measures of imprecision are supposed to indicate the extent of the reproducibility of these experiments or tests under "essentially the same conditions," but there are great doubts whether the "probable errors" and "standard errors" generally presented actually have this meaning. The fault in most cases is not with the statistical formulas and procedures used to compute such probable errors or standard errors from the measurements in hand, but rather with the limited scope of the "conditions" sampled in taking the measurements.

a. Concept of a "Repetition" of a Measurement

As a very minimum, a "repetition" of a measurement by the same measurement process should "leave the door open" to, and in no way inhibit changes of the sort that would occur if, on termination of a given series of measurements, the data sheets were stolen and the experimenter were to repeat the series as closely as possible with the same apparatus and auxiliary equipment following the same instructions. In contrast, a "repetition" by the same method of measurement should permit and in no way inhibit the natural occurrence of such changes as will occur if the experimenter were to mail to a friend complete details of the apparatus, auxiliary equipment, and experimental procedure employed--i.e., the written text specification that defines the "method of measurement" concerned--and the friend, using apparatus and auxiliary equipment of the same kind, and following the procedural instuctions received to the best of his ability, were then, after a little practice, to attempt a repetition of the measurement of the same quantity. Such are the extremes, but there is a "gray region" between in which there is not to be found a sharp line of demarcation between the "areas" corresponding to "repetition" by the same measurement process, and to "repetition" by the same method of measurement.

Let us consider "repetitions" by the same measurement process more fully. Such repetitions will undoubtedly be carried out in the same place, i.e., in the same laboratory, because if it is to be the same measurement process, the very same apparatus must be used. But a "repetition" cannot be carried out at the same time. How great a lapse of time should be allowed, nay required, between "repetitions"? This is a crucial question. Student gives an answer in a passage from which we quoted above [Student 1917, p. 415]:

"Perhaps I may be permitted to restate my opinion as to the best way of judging the accuracy of physical or chemical determinations.

"After considerable experience I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if the repetitions had been made on different days.

"It also follows that if the probable error is calculated from a number of observations made close together in point of time, much of the secular error will be left out and for general use the probable error will be too small.

"Where then the materials are sufficiently stable it is well to run a number of determinations on the same material through any series of routine determinations which have to be made, spreading them over the whole period."

Another important question is : Are "repetitions" by the same measurement process, to be limited to repetitions by the same observers and operators, using the same auxiliary equipment (bottles of reagents, etc.); or enlarged to include repetitions with nominally equivalent auxiliary equipment, by various but equivalently trained observers and operators? I believe that everyone will agree that substitution, and certainly replacement, of bottles of reagents, of batteries as sources of electrical energy, etc., by "nominally equivalent materials" must be allowed. And any calibration laboratory having a large amount of "business" will certainly, in the long run at any rate, have to face up to allowing changes, even replacement of observers and operators--and, ultimately, even of apparatus.

A very crucial question, not always faced squarely, is: in complete "repetitions" by the same measurement process, are such "repetitions" to be limited to those intervals of time over which the apparatus is used "as is" and "undisturbed," or extended to include the additional variations that almost always manifest themselves when the apparatus is disassembled, cleaned, reassembled, and readjusted? Unless such disassembly, cleaning, reassembly, and readjustment of apparatus is permitted among the allowable variations affecting a "repetition" by the same measurement process, then there is very little hope of achieving satisfactory agreement between two or more measurement processes in the same laboratory that differ only in their identification with different pieces of apparatus of the same kind. In practice it is found that statistical control can be attained and maintained under such a broad concept of "repetition" only through the use of reference standards of proven stability. Furthermore, by thus more squarely facing the issue of the scope of variations allowable with respect to "repetitions" by the same measurement process, we shall go a long way toward narrowing the gap between a "repetition" by the same measurement process and by the same method of measurement.

As we have said before, if whatever measures of the precision and bias of a measurement process we may adopt are to provide a realistic indication of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Furthermore, any experimental program that aims to determine the precision and systematic error, and thence the accuracy of a measurement process, must be based on an appropriate random sampling of this "range of circumstances," if the usual tools of statistical analysis are to be strictly applicable. Or as Student put it, "the experiments must be capable of being considered to be a random sample of the population to which the conclusions are to be applied. Neglect of this rule has led to the estimate of the value of statistics which is expressed in the crescendo 'lies, damned lies, statistics'." [Student 1926, p. 711.]

When adequate random sampling of the appropriate "range of circumstances" is not feasible, or even possible, then it is necessary to compute, by extrapolation from available data, a more or less subjective estimate of the "precision" of the end results of a measurement operation, to serve as a substitute for a direct experimental measure of their "reproducibility." Youden [1962d] calls this "approach the 'paper way' of obtaining an estimate of the [precision]." Its validity, if any, "is based on subject-matter knowledge and skill, general information, and intuition--but not on statistical methodology" [Cochran et al. 1953, p. 693].

b. Some Examples of Realistic "Repetitions"

As Student remarked [1917, p. 415], "The best way of judging the accuracy of physical or chemical determination ... [when] the materials are sufficiently stable... is ... to run a number of determinations on the same material thru any series of routine determinations which have to be made, spreading them over the whole period." To this end, as well as to provide an overall check on procedure, on the stability of reference standards, and to guard against mistakes, it is common practice in many calibration procedures, to utilize two or more reference standards as part of the regular calibration procedure.

The calibration procedure for liquid-in-glass thermometers, referred to in section 2.4 above, is a case in point. A measurement of the difference between the two standards S_1 and S_2 is obtained as by-product of the calibration of the four test thermometers T_1, T_2, T_3, and T_4 in terms of the (corrected) readings of the two standards. It is such remeasurements of the difference between a pair of standard thermometers from "occasion" to "occasion" that constitutes realistic "repetitions" of the calibration procedure. The data yielded by these "repetitions" are of exactly the type needed (a) to ascertain whether or not the process is in a state of statistical control; and if so, (b) to determine its overall standard deviation.

Similarly, in the calibration of laboratory standards of mass at the National Bureau of Standards, "known standard weights are calibrated side-by-side with [the] unknown weights" [Almer et al., 1962, p. 33]. Indeed, weights whose values are otherwise determined "are not said to have been 'calibrated'. That term is reserved for measurements based on at least two mass standards." [loc. cit., p. 43.] In the specimen work sheets exhibited by Almer et al., the auxiliary standards involved are those from the Bureau's "NH series" of reference standards known by the designations NH50, NH20, $NH10_1$ respectively. It is the measurements obtained in routine calibrations of the differences between the values of these standards and their accepted values that not only provide valuable checks on day-to-day procedure, but also serve as the basis for determination of the overall standard deviation of this calibration process.

A third example is provided by the method followed at the National Bureau of Standards for testing alternating-current watthour meters, which has been described in some detail by Spinks and Zapf [1954]. Four reference watthour metters are involved. One of these, termed "the Standard Watthour Meter," is located in the device portrayed in figure 1 of the paper by Spinks and Zapf. The other three are located in a terperature-controlled cabinet. A "test" of a watthour meter sent to the Bureau involves not only a comparison of this watthour meter with the Standard Watthour Meter, but also comparisons of each of the Comparison Standard Watthour Meters with the Standard Watthour Meter. It is from the data yielded by these intercomparisons of the Standard Watthour Meter and the Comparison Standard Watthour Meters that the standard deviation of this test procedure is evaluated. Spinks and Zapf's section of "Precision and Accuracy Attainable" is notable for its exceptional lucidity as well as for its completeness with respect to relevant details.

Some additional examples of realistic "repetitions" are discussed by Youden [1962c].

1.2. Treatment of Inaccuracy Due to Systematic Errors of Assignable Origins but of Unknown Magnitudes. As we remarked in section 3.3b above, the systematic error of a measurement process will ordinarily have both constant and variable components. For convenience of exposition, it is customary to regard the individual components of the overall systematic error of a measurement or calibration process as elemental or constituent "systematic errors" and to refer to them simply as "systematic errors," for short. Included among such "systematic errors" affecting a particular measurement or calibration process are: "... all those errors which cannot be regarded as fortuitous, as partaking of the nature of chance. They are characteristic of the system involved in the work; they may arise from errors in theory or in standards, from imperfections in the apparatus or in the observer, from false assumptions, etc. To them, the statistical theory of error does not apply." [Dorsey 1944, p. 6; Dorsey and Eisenhart 1953, p. 104.]

The overall systematic error of a measurement process ordinarily consists of elemental "systematic errors" due to both assignable and unassignable causes. Those of unknown (not thought of, not yet identified, or as yet undiscovered) origin are always to be feared; allowances can be made only for those of recognized origin.

Since the "known" systematic errors affecting a measurement process ascribable to specific origins are ordinarily determinate in origin only, their individual values ordinarily being unknown both with respect to sign and magnitude, it is not possible to evaluate their algebraic sum and thereby arrive at a value for the overall systematic error of the measurement process concerned. In consequence, it is necessary to arrive at bounds for each of the individual components of systematic error that may be expected to yield nonnegligible contributions, and then from these bounds arrive at credible bounds to their combined effect on the measurement process concerned. Both of these steps are fraught with difficulties.

Determination of reasonable bounds to the systematic error likely to be contributed by a particular origin or assignable cause necessarily involves an element of judgment, and the limits cannot be set in exactitude. By assigning ridiculously wide limits, one could be practically certain that the actual error due to a particular cause would never lie outside of these limits. But such limits are not likely to be very helpful. The narrower the range between the assigned limits, the greater the uneasiness one feels that the assigned limits will not include whatever systematic error is contributed by the cause in question. But a decision has to be made; and on the basis of theory, other related measurements, a careful study of the situation in hand, especially its sensitivity to small changes in the factor concerned, and so forth, "the experimenter presently will feel justified in saying that he feels, or believes, or is of the opinion," that the systematic error due to the particular source in question does not exceed such and such limits, "meaning thereby, since he makes no claim to omniscience, that he has found no reason for believing" that it exceeds these limits. In other words, "nothing has come to light in the course of the work to indicate" that the systematic error concerned lies outside the stated range. [Dorsey 1944, pp. 9-10; Dorsey and Eisenhart, 1953, pp. 105-107.]

This being done to each of the recognized potential sources of systematic error, the problem remains how to determine credible bounds to their combined effect. Before considering this problem in detail, it will be helpful to digress for a moment, to consider an instructive example relating to the combined effect of constant errors in an everyday situation.

a. An Instructive Example

Consider the hypothetical situation of an individual who is comparing his checkbook balance with his bank statement. To this end he needs to know the total value of his checks outstanding. Loathing addition, or perhaps, simply to save time, he adds up only the dollars, neglecting the cents, and thus arrives at a total of, say, \$312, for 20 checks outstanding. Adding a correction of 50 cents per check, or \$10 in all, he takes \$322 as his estimate. Within what limits should he consider the error of this estimate to lie?

The round-off error cannot exceed ±50 cents per check, so that barring mistakes in addition, he can be absolutely certain that the total error of his estimate does not exceed ±\$10. But these are extremely pessimistic limits: they correspond to every check being in error by the maximum possible amount and all in the same direction. (Actually the maximum possible positive error is 49 cents per check or +\$9.80 in all.)

To be conservative, but not so pessimistic, one might "allow" a maximum error of ±50 cents per check, but consider it reasonable to regard their signs as being equally likely to be plus or minus. In this way one would be led to conclude "with probability 0.95" that the total error lies between ±\$7.00; or "with probability 0.99," between ±\$8.00, as shown in the column headed "Binomial in table 1, for n = 20. The "saving" by this procedure is clearly not great.

Alternatively, one might consider it to be more "realistic" to regard the individual errors as independently and uniformly distributed between -50 cents and +50 cents, concluding "with probability 0.95" that the total error does not exceed ±$2.53; or "with probability 0.99," is not greater then ±$3.33--as shown in the columns under the heading "uniform" in table 1. It is clear that a considerable reduction in the estimate of the total error is achieved by this approach.

Strictly speaking, the foregoing analyses via the theory of probability are both inapplicable to the problem at hand: each round-off error is a fixed number between ±50 cents, and their sum is a fixed number between ±$10. If it were true that round-off errors in such cases were uniformly distributed between ±50 cents, then, if one made a habit of evaluating limits of error according to this procedure, one could expect the limits of error so calculated to include the true total error in 95 percent, or 99 percent of the instances in which this procedure was used in the long run. Round-off errors in such cases are almost certainly not uniformly distributed between ±50 cents. (Many items are priced these days at $2.98 etc., and this will distort the distribution of the cents-portion of one's bills but added sales taxes no doubt have a "smoothing" effect.)

Nevertheless, I believe that you will agree that if, in the hypothetical case under discussion, the checkbook balance, with an allowance of $322 for checks outstanding, failed to agree with the bank statement to within $2.53 (or $3.33), our "friend" would do well to check into the matter more thoroughly. And, alternatively, if his checkbook balance so adjusted, and the bank statement, agreed to with $2.53 (or $3.33), it would be reasonably "safe" for him to "act for the present as if" his balance and the bank statement were in agreement. (See Eisenhart [1947a, p. 218] for discussion of a similar example relating to computation with logarithms.)

b. Combination of Allowances for Systematic Errors

The foregoing example suggests that a similar procedure be used for arriving at credible limits to the likely overall effect of systematic errors due to a number of different origins. A number of additional difficulties confront us, however, in this case. To begin with, in view of the inexactness with which bounds can ordinarily be placed on each of the individual

Table 1. *Limits of error of a sum of n items indicated by various methods of evaluation*

n	Absolute ±	Binomial		Uniform		Triangular		Normal, $2\sigma=0.5$		Normal, $3\sigma=0.5$	
		0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±
1	0.50	0.50	0.50	0.48	0.50	0.39	0.45	0.49	0.64	0.33	0.43
2	1.00	1.00	1.00	0.78	0.90	0.56	0.71	0.69	0.91	0.46	0.61
3	1.50	1.50	1.50	0.97	1.19	0.69	0.88	0.85	1.12	0.57	0.74
4	2.00	2.00	2.00	1.12	1.41	0.80	1.03	0.98	1.29	0.65	0.86
5	2.50	2.50	2.50	1.25	1.60	0.89	1.15	1.10	1.44	0.73	0.96
6	3.00	2.50	3.00	1.38	1.76	0.98	1.29	1.20	1.58	0.80	1.05
7	3.50	3.00	3.50	1.49	1.91	1.06	1.39	1.30	1.70	0.86	1.14
8	4.00	3.50	3.50	1.59	2.05	1.13	1.49	1.39	1.82	0.92	1.21
9	4.50	3.50	4.00	1.69	2.18	1.20	1.58	1.47	1.93	0.98	1.29
10	5.00	4.00	4.50	1.78	2.31	1.28	1.66	1.55	2.04	1.03	1.36
15	7.50	5.50	6.00	2.19	2.88	1.55	2.04	1.90	2.49	1.27	1.69
20	10.00	7.00	8.00	2.53	3.33	1.79	2.35	2.19	2.88	1.46	1.92
25	12.50	8.50	9.50	2.83	3.72	2.00	2.63	2.45	3.22	1.63	2.15
30	15.00	10.00	11.00	3.07	4.03	2.19	2.88	2.68	3.53	1.79	2.35
40	20.00	13.00	14.00	3.58	4.70	2.53	3.33	3.10	4.07	2.07	2.72
50	25.00	16.00	17.00	4.00	5.26	2.83	3.72	3.46	4.55	2.31	3.04
60	30.00	19.00	20.00	4.38	5.76	3.10	4.07	3.80	4.99	2.53	3.33

components of systematic error, it is not possible to say with absolute certainty that their combined effect lies between the sum of the positive bounds and the sum of the negative bounds.

Second, even if it were possible to scale the situation so that the bounds for each of the components of systematic error was the same, say, $\pm\Delta$, there would still remain the problem of translation into an appropriate probability calculus. Most persons would, I believe, regard the "binomial" approach (corresponding to equal probability of maximum error in either direction), as too pessimistic; and the approach via a uniform distribution of error, as a bit conservative, on the grounds that one intuitively feels that the individual errors are somewhat more likely to lie near the centers than near the ends of their respective ranges. Therefore, one might attempt to simulate this "feeling" by assuming the "law of error" to be an isosceles triangle centered at zero and ends at $\pm\Delta$; or, more daringly, by assuming the "law of error" to be approximately normal with Δ corresponding to 2 "σ" or even 3 "σ."

Unfortunately whatever "probability limits" may be placed upon the combined effects of several independent systematic errors by these procedures are quite sensitive to the assumption made at this stage, as is evident from table 1. Therefore, anyone who uses one of these methods for the "combination of errors" should indicate explicitly which of these (or an alternative method) he has used. When (a) the number of systematic errors to be combined is large, (b) the respective ranges are approximately equal in size, and (c) one feels "fairly sure" that the individual errors do not fall outside of their respective ranges, then my personal feeling is that the "uniform" method is probably a wee bit conservative but "safe"; the triangular method is a bit "too daring"; the normal method with "σ" = $\Delta/3$ ordinarily "much too daring"; but the normal method with "σ" = $\Delta/2$, probably "not too daring." When (b) and (c) hold but _n_ is small, then it will probably be safe to use the "uniform" method with "Δ" taken equal to the average of the individual ranges. Other cases, e.g., when n is large but, say, one or two of the ranges is (are) much larger than the others and tend(s) to dominate the situation, requires special consideration which is beyond the scope of the present paper.

4.3. <u>Expression of the Inaccuracy of a Measurement Process</u>. By whatever means credible bounds to the likely overall systematic error of the measurement process are obtained they should not be combined (by simple addition, by "quadrature," or otherwise) with an experimentally determined measure of its standard deviation to obtain an overall index of its accuracy (or, more correctly, of its inaccuracy). Rather (a) the standard deviation of the process and (b) credible bounds to its systematic error should be stated separately, because, as we showed in figure 3, a measurement process having standard deviation $\sigma = 0.25$ and a bias $\Delta = \sqrt{15/16} = 0.97$ is for most purposes "more accurate" than a measurement process having zero bias and standard deviation $\sigma = 1$, so that a process with $\sigma = 0.25$ and a bias less than ± 0.97 will *a fortiori* be "more accurate."

Finally, if the uncertainties in the assigned value of a national standard or of some fundamental constant of nature (e.g., in the volt as maintained at the National Bureau of Standards, or in the speed of light c, or in the acceleration of gravity g on the Potsdam basis) is an important potential source of systematic error affecting the measurement process, no allowance for possible systematic error from this source should be included ordinarily in evaluating overall bounds to the systematic error of the measurement process. Since the error concerned, what ever it is, affects all results obtained by the method of measurement involved, to include an allowance for this error would be to make everybody's results appear unduly inaccurate relative to each other. Instead, in such instances one should state (a) that results obtained by the measurement process concerned are in terms of the volt (or the watthour, or the kilogram, etc.) "as maintained at the National Bureau of Standards" [McNish and Cameron 1960, p. 102], or "correspond to the speed of light $c = 2.997925 \times 10^{10}$ cm/sec. exactly," say; and (b) that the indicated bounds to the systematic error of the process are exclusive of whatever errors may be present from this (or these) source(s). Given such information, experts can make such additional allowances, as may be needed, in fundamental scientific work; and comparative measurements within science and industry within the United States will not appear to be less accurate than they very likely are for the purposes for which they are to be used.

It is a pleasure to acknowledge the technical assistance of Janace A. Speckman in several phases of the preparation of this paper.

5. BIBLIOGRAPHY

Airy, George Biddel (1861), On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations (Macmillan and Co., Cambridge and London).

Almer, H. E., L. B. Macurdy, H. S. Peiser, and E. A. Weon (1962), Weight calibration schemes for two knife-edge direct-reading balances, J. Research NBS 66C (Eng. and Instr.) no. 1, pp. 33-44.

American Standards Association (1958a), Guide for quality control, American Standard Z 1.1-1958. (American Standards Association, 70 East Forty-fifth St., New York 17, N. Y.).

American Standards Association (1958b), Control chart method of analyzing data, American Standard Z 1.2-1958. (American Standards Association, 7- East Forty-fifth St., New York 17, N. Y.).

American Standards Association (1958c), Control chart method of controlling quality during production. American Standard Z 1.3-1958. (American Standards Association, 70 East Forty-fifth St., New York 17, N. Y.).

American Society for Testing Materials (1951), ASTM Manual on Quality Control of Materials Special Technical Publication 15-C (American Society for Testing Materials 1916 Race St., Philadelphia 3).

American Society for Testing and Materials (1961), Use of the terms precision and accuracy as applied to measurement of a property of a material, ASTM Designation: E 177-61T. Reprinted from ASTM Standards, Pt 11, pp. 1758-1766.

Baird, D. C. (1962), Experimentation: An Introduction to Measurement Theory and Experiment Design, (Prentice-Hall, Inc., Englewood Cliffs, N. J.).

Beers, Yardley (1953), Introduction to the Theory of Error, (Addison-Wesley Publishing Co., Cambridge 42, Mass.).

Bicking, Charles A. (1952), The reliability of measured values--an illustrative example. Photogrammetric Engineering XVIII, pp. 554-558.

Cameron, J. M. (1951), The use of components of variance in preparing schedules for the sampling of baled wool, Biometrics 7, pp. 83-96.

Chauvenet, William (1868), A Manual of Spherical and Practical Astronomy Vol. II, 4th edition, (J. B. Lippincott and Co., Philadelphia).

Cochran, William G., Frederick Mosteller, and John W. Tukey (1953), Statistical problems of the Kinsey report, J. Am. Stat. Assoc. 48, pp. 673-716.

Cochran, William, G., Frederick Mosteller, and John W. Tukey (1954), Principles of sampling, J. Am. Stat. Assoc. 49, pp. 13-35.

Crow, Edwin L. (1960), An analysis of the accumulated error in a hierarchy of calibrations. IRE Trans. Instr. I-9, pp. 105-114.

Deming, W. Edwards and Raymond T. Birge (1937), On the Statistical Theory of Errors, reprinted from Reviews of Modern Physics 6, pp. 119-161 (1934) with additional notes dated 1937 (The Graduate School, U. S. Department of Agriculture, Washington 25, D. C.).

Deming, W. Edwards (1943), Statistical Adjustment of Data (John Wiley & Sons, Inc., New York, N. Y.).

Deming, W. Edwards (1950), Some Theory of Sampling (John Wiley & Sons, New York, N. Y.).

Dorsey, N. Ernest (1944), The velocity of light, Transactions American Philosophical Society XXXIV, pp. 1-110.

Dorsey, N. Ernest and Churchill Eisenhart (1953), On absolute measurement, The Scientific Monthly LXXVII, pp. 103-109.

Eisenhart, Churchill (1947a), Effects of rounding or grouping data, Chapter 4 of Techniques of Statistical Analysis, edited by C. Eisenhart, M. W. Hastay, W. A. Wallis (McGraw-Hill Book Co., New York, N. Y.).

Eisenhart, Churchill (1947b), Planning and interpreting experiments for comparing two standard deviations, Chapter 8 of Techniques of Statistical Analysis, edited by C. Eisenhart, M. W. Hastay, W. A. Wallis (McGraw-Hill Book Co., New York, N. Y.).

Eisenhart, Churchill (1949), Probability center lines for standard deviation and range charts, Industrial Quality Control VI, pp. 24-26.

Eisenhart, Churchill (1952), The reliability of measured values--fundamental concepts, Photogrammetric Engineering XVIII, pp. 542-554 and 558-565.

Eisenhart, Churchill (1962), On the realistic measurement of precision and accuracy, ISA Proceedings of the Eight National Aero-Space Instrumentation Symposium held in Washington, May 1962, pp. 75-83.

Feller, William (1957), An Introduction to Probability Theory and its Applications, Vol. 1, 2d edititon (John Wiley & Sons, New York, N. Y.).

Galilei, Galileo (1638), Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze, Leiden.

Galilei, Galileo (1898), Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze, Le Opere di Galileo Galilei (Edizione Nazionale) VIII, pp. 39-448, Firenze.

Galilei, Galileo (1914), Dialogues Concerning Two New Sciences, translated by Henry Crew and Alfonso de Salvio, with an Introduction by Antonio Favaro (The Macmillan Co., New York, N. Y.).

Gauss, C. F. (1809), Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium, Frid. Perthes et I. H. Besser, Hamburg; reprinted in Carl Friedrich Gauss Werke, Band VII, Gotha, 1871.

Gauss, C. F. (1823), Theoria Combinationis Observationum Erroribus Minimis Obnoxiae Commentationes societatis regiae scientiarum Gottingensis recentiores, V, pp. 1-104, Gottingae; reprinted in Carl Friedrich Gauss Werke, Band IV, Gottingen, 1873.

Gauss, C. F. (1839), letter to F. W. Bessel dated February 28, 1839, reproduced in "Kritische bemerkungen zur methode der kleinsten quadrate," pp. 142-148 in Carl Friedrich Gauss Werke, Band VIII, (B. G. Teubner, Leipzig, 1900).

Gauss, C. F. (1857), Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections; English translation by Charles Henry Davis (Little, Brown and Co., Boston).

Gnedenko, B. V. (1962), The Theory of Probability (English translation by B. D. Sechler), (Chelsea Phublishing Co., New York, N. Y.).

Hermach, F. L. (1961), An analysis of errors in the calibration of electric instruments, Communication and Electronics (AIEE) 54, pp. 90-95.

Hillebrand, W. F., G. E. F. Lundell, H. A. Bright, J. I. Hoffman, Applied Inorganic Analysis, 2d ed. (1953), (John Wiley & Sons, Inc., New York, N. Y.).

Holman, Silas Whitcomb (1892), Discussion of the Precision of Measurements, (John Wiley and Sons, New York, N. Y.).

Keyser, Cassius J. (1922), Mathematical Philosophy, (E. P. Dutton and Co., New York, N. Y.).

Kline, S. J. and F. A. McClintock (1953), Describing uncertainties in single-sample experiments, Mech. Eng. 75, pp. 3-8.

Laplace, Pierre Simon (1886), Theorie Analytique Des Probabilites; 3d edition, Vol. 7 of Oeuvres Complotes de Laplace publiees sous less auspices de l'Academie des Sciences, (Gauthier-Villars, Imprimeur-Libraire de l'Ecole Polytechnique, du Bureau des Longtudes, Successeur de Mallet-Bachelier, Quai des Grands-Augustins, 55, Paris).

McNish, A. G. and J. M. Cameron (1960), Propagation of error in a chain of standards, IRE Trans. Instr. I-9, pp. 101-104.

Millikan, R. A. (1903), Mechanics, Molecular Physics, and Heat, (Ginn and Co., New York, pp. 1[illegible]-196).

Murphy, R. B. (1961), On the meaning of precision and accuracy, Materials Research and Standards 4, pp. 264-267.

NPL (1957), Calibration of temperature measuring instruments, National Physical Laboratory Notes on Applied Science, No. 12, pp. 29-30, (Her Majesty's Stationery Office, London).

Ostle, Bernard (1954), Statistics in Research, (The Iowa State College Press, Ames, Iowa).

Parzen, Emanuel (1960), Modern Probability Theory and its Applications, (John Wiley & Sons, New York, N. Y.).

Proschan, Frank (1953), Confidence and tolerance intervals for the normal distribution, J. Am. Stat. Assoc. 48, pp. 550-564.

Rossini, F. D. and W. Edwards Deming (1939), The assignment of uncertainties to the data of chemistry and physics, with specific recommendations for thermochemistry, J. Wash. Acad. Sci. 29, pp. 416-441.

Schenck, Hilbert, Jr. (1961), Theories of Engineering Experimentation, (McGraw-Hill Book Co., Inc., New York, N. Y.).

Schrock, Edward M. (1950), Quality Control and Statistical Methods, (Reinhold Publishing Corp., New York 18, N. Y.).

Shewhart, W. A. (1931), Economic Control of Quality of Manufactured Product, (D. Van Nostrand Company, Inc., New York, N. Y.).

Shewhart, Walter A. (1939), Statistical Method from the Viewpoint of Quality Control, (The Graduate School, U. S. Department of Agriculture, Washington, D. C.).

Shewhart, Walter A. (1941), Contribution of statistics to the science of engineering, University of Pennsylvania Bicentennial Conference, Volume on Fluid Mechanics and Statistical Methods in Engineering, pp. 97-124, (University of Pennsylvania Press, Philadelphia).

Simon, Leslie E. (1941), An Engineer's Manual of Statistical Methods, (John Wiley & Sons, Inc., New York, N. Y.).

Simon, Leslie E. (1942), Application of statistical methods to ordnance engineering. J. Am. Stat. Assoc, 37, pp. 313-324.

Simon, Leslie E. (1946), On the relation of instrumentation to quality control, Instruments 19, pp. 654-656 (Nov. 1946); reprinted in Photogrammetric Engineering XVIII, pp. 566-573 (June 1952).

Spinks, A. W. and T. L. Zapf (1954), Precise comparison method of testing alternating-current watthour meters. J. Research NBS 53, pp. 95-105.

Student (1908), The probable error of a mean, Biometrika VI, No. 1, pp. 1-25.

Student (1926), Mathematics and agronomy, Journal of the American Society of Agronomy XVIII, 703-719.

Student (1927), Errors of routine analysis, Biometrika, XIX, pp. 151-164.

Swindells, James F. (1959), Calibration of liquid-in-glass thermometers, NBS Circ. 600, pp. 11-12, (U. S. Government Printing Office, Washington 25, D. C.).

Tukey, J. W. (1960), Conclusions vs. decisions, Technometrics 2, No. 4, pp. 423-433.

Waidner, C. W. and H. C. Dickinson (1907), On the standard scale of temperature in the interval 0° to 100° C, Bul. Bur. Stds. 3, pp. 663-728.

Webster's Dictionary of Synonyms (1942, 1st ed.), (G. and C. Merriam Co., Springfield, Mass.).

Youden, W. J. (1950), Comparative tests in a single laboratory, ASTM Bulletin No. 166, pp. 48-51.

Youden, W. J. and J. M. Cameron (1950), Use of statistics to determine precision of test methods, Symposium on Application of Statistics, Special Technical Publication No. 103, pp. 27-34, (American Society for Testing Materials, Philadelphia).

Youden, W. J. (1951a), Statistical Methods for Chemists. (John Wiley & Sons, New York, N. Y.).

Youden, W. J. (1951b), Locating sources of variability in a process, Ind. Eng. Chem. 43, pp. 2059-2062.

Youden, W. J. (1953), Sets of three measurements: The Scientific Monthly LXXVII, pp. 143-147.

Youden, W. J. (1954-1959), Statistical Design, Industrial and Engineering Chemistry, Feb. 1954 to Dec. 1959, Bimonthly articles collected in a single booklet available from Reprint Department, ACS Applied Publications, 1155 Sixteenth St., Washington 6, D. C.

Youden, W. J., W. S. Connor, and N. C. Severo (1959), Measurements made by matching with known standards, Technometrics 1, pp. 101-109.

Youden, W. J. (1960), The sample, the procedure, and the laboratory, Anal. Chem. 32, pp. 23A-37A.

Youden, W. J. (1961a), How to evaluate accuracy, Mat. Res. & Std. 1, pp. 268-271.

Youden, W. J. (1961b), What is the best value? J. Wash. Acad. of Sci. 51, pp. 95-97.

Youden, W. J. (1961c), Statistical problems arising in the establishment of physical standards, Proceedings Fourth Berkeley Symposium on Mathematical Statistics and Probability III, pp. 321-335 (University of California Press, Berkeley and Los Angeles.).

Youden, W. J. (1961d), Systematic errors in physical constants, Phys. Today 14, pp. 32-42 (1961); also in Technometrics 4, pp 111-123 (1962).

Youden, W. J. (1962a), Experimentation and Measurement, National Science Teachers Association Vistas of Science Series No. 2, (Scholastic Book Services, New York 36, N. Y.).

Youden, W. J. (1962b), Measurement Agreement Comparisons, presented at the Standards Laboratory Conference, National Bureau of Standards, Boulder, Colo., August 8-10, 1962.

Youden, W. J. (1962c), Uncertainties in calibration, IRE Trans. Instr. I-11, pp. 133-138 (1962).

Youden, W. J. (1962d), Realistic estimates of errors in measurements, ISA Journal 9, No. 10, pp. 57-58.

STATISTICAL STUDY OF RELIABILITY AND ACCURACY OF SURFACE-TO-AIR MISSILES

Bruce W. Sterner
Surveillance Group, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland

INTRODUCTION. The original paper presented at the Ninth Conference on the Design of Experiments in Army Research, Development and Testing consisted of the presentation of a program for the statistical study of data derived from troop training flights of Nike Hercules and Hawk missiles, followed by a discussion and interpretation of the results of this study with particular emphasis on reliability, failure modes, and the analysis of miss distance data. The major portion of the paper contained classified information which cannot be presented here. For access to the classified material, which consists primarily of the results of the statistical study, the reader is referred to Ballistic Research Laboratories Memorandum Report No. 1487, July 1963, Reliability and Performance of the Nike Hercules Missile System; and Ballistic Research Laboratories Memorandum Report No. 1513, October 1963, Reliability and Performance of the Hawk Missile System; both of which are classified confidential. Some classified information which was presented at the conference on the analysis of miss distance data for Nike Hercules has not been published but will appear in a forthcoming BRL Memorandum Report. The material which follows is an abridged paper including the unclassified portions of the original paper.

PURPOSE. The statistical study of reliability and accuracy of surface-to-air missiles is part of the overall surveillance program to evaluate the stockpile of Army surface-to-air missiles. The data used in this portion of the program is derived from the Army's Package Training Program and Annual Service Practice Program. The results of this study will be integrated with other phases of the surveillance program for the accomplishment of the following objectives:

1. To provide a continuing evaluation of the safety, reliability, and performance of the stockpile of missiles.

2. To provide advance information concerning any degradation in the safety, reliability, or performance of these missiles so that timely action can be taken to maintain the required inventory levels and performance standards by either repair or replacement.

Other benefits which can be realized from the analysis of troop training firings are such things as:

1. The accumulation of basic knowledge of the characteristics of current surface-to-air missiles,

2. An increase in our general fund of knowledge about missile systems reliability, and

3. A basis for the prediction of expected reliability of future missile systems.

All of these benefits will be valuable in formulating test programs and analysis programs for future missile systems.

PROGRAM FOR STATISTICAL ANALYSIS. The first step in the analysis of the data consists of establishing standards for classifying each missile firing as either successful or unsuccessful. With such a definition in hand, an overall reliability figure for the entire time period under consideration can be obtained. However, in order to study reliability as a function of time, it is helpful to construct a control chart for the percent of successful flights. The data is divided into groups, ordered in time, of n missiles each and the percent success is computed for each group. The number of missiles to include in each group is arbitrary, but should be large enough so that random runs of successes or failures will not adversely affect the group reliability. Group sizes of 20 to 40 missiles with 2σ control limits have been satisfactory for this application. When this data is plotted on a control chart it is relatively easy to observe whether any significant changes in reliability have occurred with respect to time. Groups out of control or trending up or down, will indicate a need for a closer examination of the data to determine the cause. As additional data becomes available it may be necessary to recompute the control limits or to compute different control limits for different groups depending on the factors involved for each particular group. Differences suggested by the control chart can be tested for independence by the use of contingency tables and the χ^2 distribution.

In the analysis of failure modes for a missile system, it is generally satisfactory to classify failures, as a percent of total missiles launched, in a two-way table of failure categories and time intervals. With such a classification it may be possible to show significant changes in failure modes corresponding to changes in reliability suggested by the control chart for percent success. This table will also show the particular failure categories or sub-categories which contribute the most failures to the missile system in any given time interval. The magnitude of the time intervals will depend upon the quantity of data available for analysis but the intervals do not necessarily have to be equal. The existance of significant differences in failure rates between two time intervals or between several time intervals can be tested by the use of contingency tables and the χ^2 distribution.

The analysis of miss distance data for a missile system generally consists of presenting a frequency distribution of miss distance and a control chart for the mean miss distance and for the standard deviation of miss distance. For the control chart the miss distances are ordered in time and divided into groups of n missiles each. Group sizes of 10 to 25 missiles with 3σ control limits have been satisfactory for this application. When group means and standard deviations are plotted on the control chart, it will be readily observable if any of the groups are out of control. Differences between groups or among several groups can be tested for significance through the use of a t-test or one-way classification analysis of variance.

In addition to the above analysis of miss distance, it is desirable to examine the frequency distribution of miss distance to determine if it can be characterized by some theoretical frequency distribution. Several typical distributions are fitted to the observed data. The "goodness of fit" test is used in the comparison of the observed distribution with the theoretical distribution to test the hypothesis that a particular distribution fits the observed data.

Another point recently considered in connection with the analysis of miss distance data was the possible existance of a relationship between miss distance and the target parameters at the time of intercept. This problem is investigated through a linear regression analysis of miss distance as a function of the range, altitude, and velocity of the target at intercept. The equation used in this study is of the form

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

where y is miss distance, X_1 is range, X_2 is altitude, and X_3 is velocity. The parameters α, β_1, β_2, and β_3 are true values, to be estimated by a, b_1, b_2, and b_3 respectively. The method of least squares is employed which minimizes the sum of squares of the deviations of the y_k from the hyperplane

$$y = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

The b values are then tested for significant effects through the use of a t-test.

RESULTS OF THE STATISTICAL ANALYSIS. As mentioned earlier in this document, the results of the statistical analysis are classified and cannot be presented here. However, it should be pointed out at this time that the statistical techniques employed in this program made a substantial contribution toward the accomplishment of the objectives of the overall surveillance program to evaluate the stockpile of Army surface-to-air missiles. Furthermore, the results of the statistical analysis were most valuable in providing several clues for improving the analysis program.

CONCLUSION. In conclusion, it should be emphasized that the current program for the analysis of data derived from troop training flights of Nike Hercules and Hawk missiles, and the projected improvements in the program, will not necessarily become the final program to be followed for these or any other surface-to-air missile systems. No doubt there are some factors due to aging, design improvements, etc., which have not yet occurred and which will affect the program. All of these factors will have to be integrated into the program. Constantly changing data and improved methods of analysis make it necessary that the analysis program remain flexible, so that the maximum amount of information can be generated from the basic input data.

STATISTICAL DECISION THEORY

Lionel Weiss
Mathematics Research Center, U. S. Army
The University of Wisconsin

Statistical decision theory can be applied to the following type of problem. $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are jointly distributed random variables, whose joint distribution is not completely known, but is known to be one of a given set of possible joint distributions. After observing $X_1, \ldots, X_m$, but before observing $Y_1, \ldots, Y_n$, somebody has to make a decision. After the decision is made, $Y_1, \ldots, Y_n$ are observed, and a loss is incurred which depends on the decision chosen and on the observed values of $X_1, \ldots, X_m, Y_1, \ldots, Y_n$. The problem is how the decision is to be based on the observed values of $X_1, \ldots, X_m$. As an example of such a problem, suppose we have to decide how large a factory to build to produce a certain commodity. The X's might be the results of a survey taken to examine potential demand, and the Y's are the demands that will be observed in the various accounting periods after the factory is built.

A decision rule attaches a particular decision to each possible set of X's that might be observed. For example, in the case of the factory, we might use the decision rule that makes weekly capacity equal to 1000 multiplied by the sum of the X's . More generally, a decision rule may use randomization: that is, it assigns a set of probabilities of choosing the various decisions to each possible set of X's that might be observed.

We introduce the following notation. We abbreviate $X_1, \ldots, X_m$ by X, and $Y_1, \ldots, Y_n$ by Y. We use the symbol Θ as an index for the possible joint distributions of X,Y: that is, a particular value of Θ picks out a particular distribution. We use the symbol D as an index for the possible decisions: a particular value of D picks out a particular decision. The loss we incur when X is observed, D is chosen, and the Y is observed, is denoted by $W(Y;D;X)$. When X,Y are discrete variables, $f(x,y;\Theta)$ denotes the probability that $X = x$ and $Y = y$ when the joint distribution is given by Θ. When X,Y are continuous variables, $f(x,y;\Theta)$ denotes the joint density function. A decision rule s is defined by nonnegative numbers $s(D;X)$, where $s(D;X)$ denotes the probability assigned by the decision rule s to choosing the decision D when X is observed, and therefore $\sum_D s(D;X) = 1$ for each X. We denote by $r(\Theta;s)$ the expected value of the loss when the decision rule s is used and the true joint distribution is given by Θ. Then it is easily seen that when X,Y are discrete variables, $r(\Theta;s) = \sum_X \sum_Y \sum_D W(Y,D;X) f(X,Y,\Theta) s(D;X)$. When X,Y are continuous, we replace summation with respect to X and Y by integration.

As a numerical illustration of the above, suppose we have to decide whether or not to buy a nonguaranteed device for \$500. If we buy the device and it fails, we must then buy a guaranteed device for \$1000. If we do not buy the nonguaranteed device, we buy the guaranteed device for \$1000. The proportion of defectives turned out by the factory which produced the nonguaranteed device is unknown, but before making our decision we can observe two similar devices from the same factory to see whether or not each failed. We introduce the following notation. Y is defined to be 1 if the nonguaranteed device fails, and 0 otherwise. X_1 is defined to be 1 if the first of the two similar devices to be observed fails, and 0 otherwise, with a similar definition for X_2 in terms of the second device to be observed. Θ denotes the unknown proportion of defectives turned out by the factory producing the nonguaranteed devices. Then X_1, X_2, Y are independent and identically distributed random variables, each with probability Θ of being equal to 1 and probability $(1 - \Theta)$ of being equal to 0. We label the decision to buy the nonguaranteed device as decision number 1, and the decision not to buy it as decision number 2. Then the loss function is seen to be given as follows. $W(0;1;X) = \$500$, $W(1;1;X) = \$1500$, $W(0;2;X) = W(1;2;X) = \$1000$ no matter what X is. Then if the decision rule s_1 is given by $s_1(1;0,0) = 1$, $s_1(1;0,1) = s_1(1;1,0) = \frac{1}{2}$, $s_1(1;1,1) = 0$, we find that $r(\Theta; s_1) = 500 + 1500\,\Theta - 1000\,\Theta^2$. If the decision rule s_2 if given by $s_2(1;0,0) = 0$, $s_2(1;0,1) = s_2(1;1,0) = \frac{1}{2}$, $s_2(1;1,1) = 1$, we find that $r(\Theta; s_2) = 1000 - 500\,\Theta + 1000\,\Theta^2$.

From the above, it is clear that in each decision problem there are infinitely many decision rules, and we need some way of comparing the goodness of one decision rule to another. If s_1 and s_2 are two decision rules, where $r(\Theta;s_1) \leq r(\Theta;s_2)$ for all possible values of Θ, and $r(\Theta;s_1) < r(\Theta;s_2)$ for at least one possible value of Θ then we say that "s_1 is better than s_2" and we would not use the decision rule s_2. For example, in the numerical illustration of the preceding paragraph, it can be verified that s_1 is better than s_2. We say that a decision rule s is "inadmissible" if there is a better decision rule than s. If a decision rule is not inadmissible, we call it "admissible". How can we find admissible decision rules? For simplicity, let us limit ourselves to the case where there is a finite number h of possible joint distributions of X, Y so that we may assume that the possible values of Θ are the integers from 1 to h inclusive. If $b_1, \ldots, b_h$ are nonnegative quantities adding to unity, the decision rule s is called a "Bayes decision rule relative to $b_1, \ldots, b_h$" if $b_1 r(1;s)+\ldots+b_h r(h;s) \leq b_1 r(1;t)+\ldots+b_h r(h;t)$ for every decision rule t. If s is called simply a Bayes decision rule, it means that there is some set of nonnegative quantities $b_1, \ldots, b_h$ adding to unity, such that s is a Bayes decision rule relative to $b_1, \ldots, b_h$. A basic theorem states that every admissible decision rule is a Bayes decision rule. Therefore the class of Bayes decision rules contains the class of admissible decision rules. Since it is computationally easy to find Bayes decision rules, it is easy to find admissible decision rules.

In most problems there are infinitely many admissible decision rules, only one of which will be chosen for use. What principles can be used for selecting one particular decision rule from all the admissible decision rules? There is much disagreement about this, and we merely mention two of the various principles that have been proposed. One principle calls for the assignment of "subjective probabilities" or "degrees of belief" to the various possible values of Θ, and then using a Bayes decision rule relative to $b_1, \ldots, b_h$, where b_Θ is set equal to the subjective probability assigned to the value Θ. Another principle defines $M(s)$ as $\max_\Theta r(\Theta; s)$, and then proposes using a decision rule which minimizes $M(s)$, which decision rule is called "minimax".

WHAT TYPE OF STATISTICIANS ARE NEEDED IN RESEARCH AND DEVELOPMENT LABORATORIES

Chairman: Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute

Panelists: Dr. E. L. Cox, Biometrical Services ARS, Plant Industry Station
Dr. Churchill Eisenhart, National Bureau of Standards
Dr. Frank E. Grubbs, Army Ballistics Research Laboratories Aberdeen, Md.
Mr. John L. McDaniel, Directorate of R and D, Army Missile Command
Dr. Paul R. Rider, Office of Aero-Space Research, Wright-Patterson AFB
Dr. William Wolman, Goddard Space Flight Center, NASA
Dr. Donald A. Gardiner, Mathematics Division, Oakridge National Laboratories

INTRODUCTORY REMARKS BY THE CHAIRMAN

Boyd Harshbarger

Mathematical statistics, including the statistical design of experiments, is a new discipline. It was in the late 20's that R. A. Fisher began his work in the design of experiments. In 1933 he moved to the University of London where the field of design was greatly expanded. Industry was slow to adopt many of the ideas developed by Fisher, Yates, and others.

Dr. Walter Shewhart of the Bell Laboratories might be considered the "Father of Statistics in Industry" in this country, but it was the organization and promotion ability of General Leslie Simon that finally caused the proper recognition of statistics. Many of us remenber the ten-day wonder courses of World War II which were directed by Dr. Holbrook Working and Dr. Edward Olds.

Industry and government installations moved quickly from quality control to the design of experiments. In this area government installations appear to have developed their own pattern for the organization and application of statistics. On the other hand, companies like Bell Laboratories developed large statistical departments where consulting is concentrated and where statistical research is highly developed. As an example of an industrial company engaged in government activities, Hercules Powder Company has been active in the application of statistics to propellant research. As early as World War II, classes were set up at the Radford Arsenal and soon after full-time statisticians were employed.

Steel companies, like U. S. Steel, oil companies, chemical industries, automobile corporations, and most of the defense industries now employ statisticians. Some companies in certain industries have been more successful in the introduction of statistical methods than others.

It is fair to point out that statistics would not have developed as rapidly or as successfully if it had not been for the interests of government agencies. In this connection it is well to note that places like Aberdeen Proving Grounds have been most successful in the use of statistics. The Army Research Center at Durham gave some of their earliest grants in support for research in the design of experiments. Redstone Arsenal and White Sands Proving Grounds have used statistical methods since their inception. The success of these two installations in statistics has been limited somewhat by their ability to secure competent statisticians.

What I have said about the Army is true of other military installations, including the Navy and Air Force, as well as space programs.

I am hoping that the panel will outline the most efficient method for the use of statistics and statistical designs. I am suggesting that they outline the best method of approach as well as to where statistics should be located in an organization. This will vary from organization to organization and from installation to installation. I do feel, however, that the panel should tell us what they are doing in their particular activities and what they might think would be best for them to do.

I have asked each of the panelists to give us a ten minute, and only ten minutes, summary of their suggestions. When this is completed, we will then throw the program open for discussion.

STATISTICAL PERSONNEL ASSOCIATED WITH AGRICULTURAL RESEARCH

Edwin L. Cox

Statisticians are active in many phases of the work of the U. S. Department of Agriculture. The remarks which follow will be concerned with statistics and statisticians principally from Statistical Reporting Service (SRS), Forest Service (FS), and Agricultural Research Service, (ARS). It is probably correct to say that the main involvement of statisticians with R and D problems will be located in these above mentioned organizational units.

Before commenting specifically on some of the differences in needs and practices as exemplified by these different organizational units some overall generalizations will be presented. The statisticians presently employed will be found in the main to have had education and training in a field of specialization other than mathematics or statistics. Of necessity through experience or further training this prior background has been added to in the pursuit of professional activity in statistics. The reason for this finding is not entirely due to shortages of graduates with statistical majors. It has bearing on various other factors; historical, personal, etc.

The Statistical Reporting Service conducts and supervises a program of Sample Surveys. Communication with the Census Bureau has been frequent. The Research and Development Branch of the Standards and Research Division has a program of interpreting and developing the practice of sampling methods for production and other estimates. SRS has a definite program of sending statistical personnel where they can pursue a program of graduate training in statistics. As might be expected from the mission of the organization, the training received is mainly in theory and methodology of sampling surveys and censuses.

The Forest Service has placed or intends to place statistical personnel at all regional laboratores. The policy has generally been to expect that such statistical personnel have training in forestry. Additional training in statistics has in the past been accomplished in part by the provision of special courses by forest service statisticians. The statistical contributions from the Forest Service have been typically in the areas of technique and methodology. Rather specialized techniques of sampling have received statistical evaluation.

In the Agricultural Research Service most of the statistical personnel are organizationally located in a unit called Biometrical Services. Except for those assigned to the four utilization laboratories, the members of Biometrical Services are stationed in the Washington, D. C. area. Their statistical consulting service to field stations is provided by correspondence. In addition to statistical aid to ARS scientists in Washington and Beltsville and in the field a computing laboratory is maintained as an integral part of the unit. Biometrical services personnel in the main have not received their primary graduate training in statistics. Some have pursued additional study in statistics or have gained competence by experience. The scope of the subject matter which may come under consideration with respect to problems of design and analysis of experiments covers the complete range of agricultural experimentation. Statistical contributions from this group have come principally in association with subject matter specialists. Several bulletins of specially statistical interest have been produced.

It would seem from the experience of these working organisations that a firm acquaintance with subject matter is important for statistical personnel associated with agricultural research. A coordination of statistical techniques with this foundation seems to produce a person who has rapport with the scientists and can aid and advise them on numerical problems. It may be that some areas deserving statistical tools have been ignored but until there is a felt need for attention, these areas are likely to remain uninvestigated.

THE NEEDS OF THE STATISTICAL ENGINEERING PROGRAMS OF THE NATIONAL BUREAU OF STANDARDS

Churchill Eisenhart

I welcome this opportunity to review the staffing needs and recruitment difficulties of the statistical engineering programs of the National Bureau of Standards because I feel that they are symptomatic of the needs and recruitment problems of many of the statistical advisory and consulting groups in Government laboratories. To provide the setting for my story, let me begin with a few words about the programs and functions of the three statistical groups at the National Bureau of Standards.

The major statistical program of the National Bureau of Standards is that of the Statistical Engineering Laboratory of the Applied Mathematics Division, on the main "campus" of the Bureau, in Washington, D. C. This group, of which I was the Chief from its formal establishment on July 1, 1947 until June 30, 1963, is now under the direction of Joseph M. Cameron, Chief, and Joan R. Rosenblatt, Assistant Chief. The principal function of the Statistical Engineering Laboratory is to serve in an advisory and supporting capacity to the Bureau's scientific and technical personnel on the application of modern probability and statistical methods in the physical sciences and engineering. Its staff also conducts research on the theory and techniques of statistical inference and statistical design of experiments, with special reference to problems that arise in physical science and engineering experimentation; and to a lesser extent, on mathematical and statistical aspects of the definition, measurement, and specification of the reliability of components and assemblies, with particular attention to the evaluation of the reliability of electronic components and systems.

There are two other, much smaller, statistical groups at the National Bureau of Standards. One of these is located in the Polymers Division, on the Washington campus; the other, in the Office of the Manager of the Boulder Laboratories of the National Bureau of Standards, in Boulder, Colorado. Both of these groups operate independently of the Statistical Engineering Laboratory. The Polymers Division group concentrates on development of statistical techniques for the design, analysis, and interpretation of physical and chemical data pertinent to the evaluation of the measurement techniques in the field of polymers; and, in recent years, has been actively engaged in studies of measurement processes dealing with polymeric materials from the viewpoints of interlaboratory standardization and the need for standard reference materials. The Boulder Laboratories group specializes in adapting existing general statistical methods to the experimental work of the Boulder Laboratories, and to developing mathematical models of stochastic phenomena of interest in scientific and engineering programs of the Boulder Laboratories.

Both the service and the research activities of all three of these groups would benefit from additional personnel. I shall, however, restrict my comments principally to the needs of the Statistical Engineering Laboratory, as this is the group with which I am most familiar.

To begin with, the Statistical Engineering Laboratory's service activities -- its provision of advice and assistance to NBS staff members on statistical aspects of <u>their</u> calibrations, research, and development programs -- is continually hampered by shortage of personnel. The available "consultants" are frequently involved in four or five major consulting projects at one time, and therefore forced to reduce the time spent on any one project. The Laboratory has continual need for additional specialists in experimental statistics to enable it to give some projects the detailed thorough study they warrant. The background of these specialists should include as a minimum a strong MS in statistics with some first-hand experience in the design and analysis of experiments. Second, both the service and the research activities of the Statistical Engineering

Laboratory would benefit if the Laboratory were able to attract a succession of topnotch young Ph. D's in mathematical statistics who desire to gain some firsthand contact with applications in the physical sciences and engineering before devoting the remainder of their careers to teaching in a university or college, or to working in industry or in one of the ever-increasing number of non-profit research institutes. The scientific and technical programs of the National Bureau of Standards, embracing almost every area of the physical sciences, afford unusual opportunities for gaining experience in the application of modern statistical tools.

Unfortunately, both from the Bureau's viewpoint and from the long-range view of application of statistical methods in the physical sciences and engineering in this country, the number of qualified applicants for positions in the Statistical Engineering Laboratory at all levels is almost nil. There are several reasons for this situation. Let us consider first the Laboratory's "red carpet" positions, its Postdoctoral Resident Research Associateships in probability and mathematical statistics.*

A recipient of one of these Research Associateships is expected to devote his efforts entirely to advanced training and research in some aspect of probability and mathematical statistics related to the work of the Statistical Engineering Laboratory. In recent years, research activities of the Laboratory have been directed toward the evaluation of existing and the development of new statistical techniques of value for the applications in physical science and engineering experimentation, with special attention to "practical properties" of statistical techniques, including questions of power, efficiency, robustness, consequences of misapplications. The "advanced training" may consist merely of passively observing nature, and the manner in which we carry out, our consultative and advisory services in probability and mathematical statistics. The Research Associate may elect to participate in one of these consultative and advisory

---

*These awards are part of a broad program sponsored by the National Bureau of Standards, in cooperation with the National Academy of Sciences -- National Research Council. Their purpose is "to provide young scientists of unusual ability and promise an opportunity for fundamental research in various branches of the physical and mathematical sciences". Intended principally for recent Ph. D's and Sc. D's, the awards carry an annual gross stipend equal to the GS-12 entrance salary in the U. S. Civil Service. Applicants' qualifications are evaluated by a board of selection appointed by the National Academy of Sciences -- National Research Council.

activities to gain insight, inspiration, and some practical experience -- but he is under no obligation to do so. In spite of the honor accruing to the recipients of these Research Associateships by virtue of the administration of the program by the National Academy of Sciences -- National Research Council, and of the definitely preferential status accorded them in the Statistical Engineering Laboratory, in 4 out of the 7 years of the program's existence there were no applicants for the Research Associateship in probability and mathematical statistics in the Statistical Engineering Laboratory, the present year being one of the "dry" years, and in only one year were there as many as 2 candidates. In contrast, each year there have been ample applicants in "pure" mathematics, and in many branches of physics and chemistry.

There are, I believe, at least four reasons for this dearth of applicants for these Postdoctoral Resident Research Associateships in probability and mathematical statistics: First, there are far fewer Ph.D's specializing in probability and mathematical statistics than in the other branches of mathematics -- and, *a fortiori*, than in the more popular branches of physics and chemistry. Second, although the Statistical Engineering Laboratory of the National Bureau of Standards is one of the principal centers of research in the application of statistical theory and methods in the physical and engineering sciences in this country, its activities are primarily of interest to groups engaged in measurement, calibration, and testing in the physical and engineering sciences, and are of very little interest to -- and carry no special authority in -- the departments of mathematics, where the majority of courses in mathematical statistics are taught. In contrast, the work of many divisions of the National Bureau of Standards is vital to research and development programs in the physical-science and engineering departments of colleges and universities. Third, the present-day training in statistics at most of the universities or colleges in this country is so strongly oriented toward theory, with little attention to applications, that the majority of the successful candidates for MA and PhD degrees in probability and mathematical statistics have already developed a strong distaste for applications, partly from the expressed contempt of "mathematician's mathematicians" towards anything "applied" in contrast to "pure", and partly from lack of experience with applications and resulting fear of the unknown. Consequently, even explicitly research-oriented "positions" in a government laboratory are shied away from, being regarded as having an applied "taint". Fourth, the small number of individuals who do have the requisite training and interest are, for the most part, either lured into industrial applications by the much higher salaries offered by industry and the ever-growing number of "non-profit" private research institutes, or take positions with universities where they can engage simultaneously in teaching and research under the contract arrangements now generally available, for the most part with Government funds. These latter are able to offer the usual inducements of university appointments and titles, less stringent working hours, opportunity for teaching and private consulting, for extra money from the "outside", and now also in many cases equal or higher salaries as well.

Let us now consider the situation at the lower professional levels in the Statistical Engineering Laboratory. In comparison with other sections of the NBS, and in relation to customary management norms, the Statistical Engineering Laboratory has always been "top heavy", that is, has always had an unusually large fraction of its staff in the upper professional brackets. This, I believe, is characteristic of most consulting groups. In the 1950's ("the good old days") when our recruiting at the PhD-level was more successful, the Laboratory was especially top-heavy with PhD's; and the median Civil Service grade of the 11 professional members of the year-around staff was GS-12, i.e., one grade above the entrance level for brand-new PhD's in Mathematics and Mathematical Statistics. Today (October 1963) the corresponding median grade of the year-round staff, including myself as an "honorary member", is GS-13; i.e. one grade above the entrance level for "quality" PhD's in mathematics and mathematical statistics -- and only "quality" PhD's seem to be available.

The Statistical Engineering Laboratory has always managed to get along with far fewer personnel in the lower professional levels than most statistical consulting and advisory groups because it has never needed a sizeable computing group within the Laboratory itself on account of the existence and general availability within the Bureau of the computing services of the NBS Computation Laboratory. The senior staff of the Statistical Engineering Laboratory, on the other hand, have been obliged for many years to do for themselves various "chores" that normally would be the assigned work of "graduate assistants" at a large university and of GS-5 and GS-7 level professional assistants in many other laboratories. Our shortage of support staff at these levels has stemmed from the almost complete lack of qualified candidates for lower-and-middle rank professional positions in analytical and mathematical statistics, a situation which, in turn, stems from the fact that in the United States statistics is largely a post-graduate field of specialization. As a result there are very few applicants for Statistician positions at the bachelor's-degree (GS-5 or GS-7) entrance levels who have training in statistics that is comparable in scope and depth to the training in physics or chemistry that is required of applicants for Physicist and Chemist positions at the same entrance levels.

In other sections of the Bureau, where the work involves the traditional fields of chemistry, physics, and engineering, there is a continual influx of "new young blood" directly out of colleges and universities, at the bachelor's and master's degree levels. These junior professionals assist the higher-ups with the more straightforward phases of the day-to-day operations of the groups concerned, and a few, through on-the-job-experience and in-service training, rise in due course to higher positions in the organization. To date, the Statistical Engineering Laboratory has been obliged to get along without the "normal" complement of lower level assistants on account of the peculiar circumstances of the statistical profession. It may be, however, that in time the current trends in the undergraduate mathematics curriculum will produce an increased number of qualified applicants for junior positions.

The staff of a consulting group like the Statistical Engineering Laboratory will, nevertheless, always tend to be top-heavy because of the nature of its work. A statistical consultant has to be prepared to handle a wide variety of problems: when a physicist telephones to say he has a problem in least-squares analysis, it may turn out that his real problem (from the professional statistician's point of view) is in experiment design, or numerical analysis, or components of variance. The consultant who is assigned to provide advice must have a sufficiently broad background to identify the problem correctly and to be able to apply the relevant theory and methods. An especially important part of the requisite background is the thorough advanced training in probability and distribution theory that enables the statistician to steer a safe course through the adaptation of standard techniques for application in the almost universally non-standard circumstances that arise in physical science investigations.

While it is not necessary that each consultant be an expert in every field of statistics (this would be absurd), it is also true that consulting work will not suit the tastes of a narrow specialist who doesn't care to know about any field except his own.

In summary, our efforts to recruit brand-new college graduates, MA's, and PhD's has been discouragingly unsuccessful in recent years. I have summarized some of the factors contributing to this situation. Fortunately we have been a bit more successful in recruiting MA's and PhD's who have been "out" for a year or two; and some of our MA's have grown into full PhD's right on the job, through on-the-job experience and "outside" job-related course work; and at least two of our present staff are headed in the same direction. We just wish that more promising young men and women would "give us a try" enroute to their ultimate careers.

SOME THINGS THE STATISTICIAN SHOULD KNOW

Frank E. Grubbs

In continuing this completely unrehearsed program, perhaps the first remark that should be made concerning what types of statisticians are needed in R & D laboratories is that they should be good statisticians! The next remark is that the statistician should know quite well or learn as thoroughly as possible the field of application. This is especially desirable or necessary in the engineering and scientific fields, or other fields, where physical, mathematical, biological, etc. models are in many cases already available. To speak only in statistical terms is a lost cause indeed! Also, the statistician in many R & D laboratories is a member of a team -- his methods are well worthwhile and needed, although it seems to be a rather natural fact that the statistician is more prone to criticism than any of his colleagues. But this is all in the game of things, and we must remember that the non-statistical investigator is often a bit jealous of the power and usefulness of statistical methods. In some problems, on the other hand, statistics cannot help.

The types of statisticians needed depend very much on what they actually do in R & D laboratories, so we must examine this momentarily. There is the obvious need for occasional statistical analysis of data (including the common tests of significance, regression analysis, etc.) and the statistical design and analysis of scientific experiments. Factorial experimentation, incomplete block designs, Latin and Graeco-Latin Squares, lattices, etc. are all of considerable importance, as is also the search for optimum operating conditions utilizing Box techniques. Interpretations of the data, however, must always be in terms of the physical picture involved, as this is the basic requirement originally desired.

At the Ballistic Research Laboratories (and no doubt R & D laboratories generally), curve fitting is of importance. I distinguish between the fitting of "physical" laws or curves in general and the fitting of frequency distributions. In ordinary curve fitting, the physicist or engineer is oftentimes trying to linearize his data in addition to trying to get the "best" fit. He plots data on a logarithmic scale, etc. The statistician could be especially helpful (if he can get his hands on the data!) since he is usually adept in the use of transformations, least squares and testing the significance of parameters. (Iteration is often required to estimate parameters.)

There is also the need for fitting frequency distributions. The Pearson system of frequency curves is quite general (including even the Weibull distribution of much current interest) and the method of moments very useful in practice as compared to maximum likelihood estimation, which although theoretically elegant and efficient, is often not so satisfying practically. (I must remark that statisticians coming out of the graduate schools now have most likely not studied the Pearson system of frequency curves or the Gram-Charlier Series, etc.!)

The big, new field of endeavor in Army R & D work is that of Operations Research, and in this I include the very important field of Weapon Systems evaluation. In the latter case, as in much OR work, we often get somewhat away from significance testing, experimental designs, etc. and come face-to-face with the requirement for probabilistic models. We predominantly speak of "hit probabilities," "kill probabilities," coverage problems concerning the overlap of weapon "lethal areas" on the target area, distribution of the range of engagement, chance of winning an engagement, etc. But again, the statistician is badly needed. The calculation of many probabilities of hitting, for example, is closely tied in with the distribution of quadratic forms in normal variables, the distribution of range of engagement is often of the Pearson Type III or Gamma class of frequency distribution, etc. Also, for weapon systems evaluation problems, there is ever-present the requirement for combining several frequency distributions to obtain the single or overall characterization of weapon system performance, if possible. Reliability studies represent a very important facet of the field of weapons evaluation and the statistician can be very valuable here too. Some reliability theory should be in the present curricula for statistics. Thus, the properly qualified statistician can do much for this new field and he is a necessary part of it.

Finally, my time is getting short, but I must mention that a very special, highly qualified type of statistician is very often needed. He is the type that must be right up-to-date with the most recent advances in the state of the art theoretically. He is the one who is pushing the frontiers forward in many areas, combining developments from many areas and, for example, including time series studies, variance-component analyses, multi-variate normal theory, the treatment of errors of measurement, precision and accuracy, etc. The practical types of problems I speak of are already with us and the demand is most urgent for accurate estimation procedures. I refer to the tracking data analysis problem for missile trajectories, which Dr. David Duncan will discuss at the general session tomorrow. For such problems, we see the need for the most highly trained statistician and the requirement is likely to grow considerably.

I remark and observe that many of the papers presented in this conference indicate the types of statisticians needed in R & D laboratories.

In summary, therefore, in R & D laboratories we need both the applied statistician who will analyse data statistically, carry out design of experiments, etc. and help others do their job better. Also we need the theoretical statistician for many very complex problems cropping up on the frontiers of knowledge. In either case, however, the statistician must know the physical or engineering fields quite well in order that he may make the best contribution possible.

DESIRABLE ATTRIBUTES OF STATISTICIANS IN R & D LABORATORIES

John L. McDaniel

The word "laboratory" in the Defense context has many meanings. Used loosely, as we all tend to do, it encompasses the full spectrum of weapon technology from basic research through production. Further, each service has problems unique to their missions, and thus have discreet requirements for its laboratories. Naturally, my remarks will be aligned toward our Army laboratories in general and our missile laboratories of the Army Missile Command in particular. I believe however my remarks will be sufficiently general to be useful in our broad discussion of Research & Development Laboratories.

The R & D laboratories of the Army Missile Command are an organic part of a Commodity Command. The commodity is missiles, therefore it is an intrinsic responsibility of our laboratories to concern themselves with a mission which covers first, research; second, the generation of new missile concepts to fulfill the requirements of the future Army; and third the technical support required in the fielding of missile systems now in development. In carrying out this broad mission the laboratories must be constantly mindful of the system characteristics which assure that the weapon will function reliably in the environment anywhere in the world that the Army is required to fight.

The formula for discharging this responsibility contains many terms. One of the terms is men and it is this term that we will discuss. The successful operation of R & D laboratories requires teamwork of the men. The team is made up of persons trained in the physical sciences and in the various fields of engineering. The statistician is now a major, rather than a casual partner of this team.

My first desirable attribute then for the statistician for R & D laboratories is an individual to work as a team member in provisioning reliable weapon systems. To operate effectively as a member of this team the statistician must have a fair amount of formal education in his field and a great amount of appreciation for the physical sciences and for engineering.

The statistician plays a key role in the area of reliability. Reliability is a function of the Army's decision from beginning to end. Men experienced in this unique business can look objectively at a development program, almost before it begins, and identify the reliability problems we can expect to encounter. Reliability must be built into all the decisions which are made through an entire program, through design, tests of successive stages of design, through corrections made during design, through further testing, through production and the assurance of quality production, through supply and protection of quality in the supply network.

My second desirable characteristic of a statistician in the R & D laboratories then is an individual trained in the techniques of reliability as this reliability relates to a weapon system throughout its life cycle.

Decisions regarding the missile systems generated by the R & D laboratories must be made by men who have to stand-up and be counted. Stand up and be counted by the men who must answer to the combat soldier who takes the missile system into the battlefield with his life in his hands. We cannot have those decisions made by other people for us. These decisions must be made on the basis of our knowledge, reinforced by the knowledge of others in equally objective positions. In dealing with the competitive commercial world we must be in a position truly to control our missile business. We must have technically competent people in our laboratories who have no conflicting interests and who therefore can afford to be objective.

Third, then the statistician in the R & D laboratories must be technically competent in missile technology and must have a full measure of objectivity.

There are those who believe that reliable missile systems can be provided to the user without government competence in in-house R & D laboratories. This belief ignores some very basic principles.

First, the competitive commercial industry engaged in missile system engineering requires control. To apply this control a capability to control is required, not just by suggestion but by direction when necessary.

Second, the capability to control requires knowledge and experience. Without this experience one must rely on faith and hopeful trust. The engineer cannot know whether his faith is well placed unless he has first hand knowledge.

Engineering and scientific knowledge deteriorates in a vacuum. Unless the engineer has the responsibility for actual hardware work, for getting his hands dirty, he will become the kind of engineer who looks over drawings and hopes he is right, but has a measure of doubt as to his wisdom since he is not current with the state-of-the-art.

My fourth premise then is that the statistician in the R & D laboratories must have confidence in his own knowledge. This knowledge must come from working with the hardware, not just sitting in a staff position and waiting for a problem to be brought to him.

In order to afford to the scientists and engineers the required knowledge a small amount of our business must be done in-house. Enough to allow us to control the rest of our missile business. I have heard that a good many years ago when Henry Ford started to build a small steel plant a committee of the Congress called him and questioned him about monoply. "Isn't there enough steel in the country for you", he was asked. "yes Sir" Ford answered, "well then", the committee asked, "why do you build a steel plant?" Henry Ford's answer is said to have been -- "Gentlemen, to learn enough about making steel to know what I'm buying from the steelmakers."

Gentlemen, I submit finally we require statisticians in the R & D laboratories as a key member of the team to help us remain leaders in missiles so we can influence what we're buying from the missile-makers.

STATISTICIANS IN AIR FORCE ORGANIZATIONS

Paul R. Rider

In the first place let me tell you about some of the Air Force organizations that are concerned with statistical methods. How these organizations originated and developed is ably described by my colleague and officemate H. Leon Harter, in an article, "Statistics in the Air Force Research Program," which appeared in the American Statistician for October, 1962 (Vol. 16, no. 4, pp. 23-24) and was reprinted as Appendix B in Harry J. Eisenman's History of Mathematical Statistics Research at the Aeronautical Research Laboratories, Office of Aerospace Research, United States Air Force, Washington D. C., 1962. Consequently I shall not go into historical matters here.

Two elements of the Office of Aerospace Research handle practically all of that organization's basic research program in statistics. One of these elements is the Air Force Office of Scientific Research, also located in Washington. It has as its principal function the awarding and monitoring of contracts and grants for basic research. The subject matter of the statistics contracts includes nonparametric methods, order statistics, estimation theory, distributions of statistical functions, design and analysis of experiments, decision theory, foundations of probability, limit theorems, stochastic processes, Markov processes, and combinatorial analysis, as well as applications of probability and statistics.

The other element is the statistics group of the Applied Mathematics Research Laboratory, a component of the Aerospace Research Laboratories, located at Wright-Patterson Air Force Base, Ohio. This element performs most of the in-house basic research in statistics; it also does a small amount of consulting work and monitors a few contracts in areas closely related to the internal work. These areas include design and analysis of experiments, estimation of parameters, and probability theory. Specific topics being studied are transformations in the analysis of variance, multiple comparisons tests, multivariate analysis, the use of order statistics and functions of order statistics (e.g., range and quasi-range) to obtain estimates of parameters, mixed and truncated distributions, information theory, circular error probabilities, and probabilistic models for applied mathematical problems. Members of this group have conducted courses in statistics and sponsored masters' theses by students in the Air Force Institute of Technology, although they are not at present engaged in any formal instructional activities.

The Operations Analysis Office at Headquarters USAF uses statistical methods extensively and performs a certain amount of research in applied statistics.

The Air Force Systems Command makes extensive use of statistics as a research tool and performs a fair amount of research in applied statistics. Among the various divisions of AFSC, probably the one having the most extensive program of statistical research is the recently formed Bioastronautics Division; particular mention should be made of the work of the School of Aviation Medicine located at Brooks Air Force Base, Texas, and the Aerospace Medical Research Laboratories at Wright-Patterson Air Force Base, Ohio. At the School of Aviation Medicine, research is being performed or monitored on the statistical theory of epidemics, the sampling distribution of the characteristic roots of Wishart matrices, the analysis of repeated measurements, classification statistics, and non-parametric several-sample tests. At the Aerospace Medical Research Laboratories research has been performed on distribution-free tests and considerable use has been made of the design and analysis of experiments in their testing of materials, equipment, rations, physical endurance, etc. In the Ballistic Missile Division of AFSC some applied statistical research has been administered by the Air Force Missile Development Center at Holloman Air Force Base, New Mexico, and by the Air Force Missile Test Center at Patrick Air Force Base, Florida. Naturally, much of the work at these two centers is concerned with reliability.

The Air Force Logistics Command employs statistical quality control and acceptance sampling techniques.

Contract groups are the Rand Corporation at Santa Monica, California and the Institute for Air Weapons Research at the University of Chicago.

I have tried to sketch a picture of the present state of statistics in the Air Force research and development program, and now I shall attempt to say how I think the situation can be improved. I would not urge any change in the overall organization. I do believe that many more statisticians could be used in all of our fields of endeavor. We could use more research statisticians. Of course, it is impossible, or at least difficult, to force the direction of basic research, but some effort could be made to assemble groups of statisticians with common interests in problems that would likely have some useful application to Air Force problems some time in the future. I would not, however, throw out any project simply because no possible application can be imagined for it, even in the far-distant future. It is well known that some theories which at one time were considered of academic interest only have subsequently been found to be of great practical utility. For example, group theory, once regarded as of no conceivable use has become extremely important in the design of experiments.

But mainly I should like to see more statisticians employed as consultants in various fields. These persons should be familiar with the particular fields in which their services are employed, whether this be the physical sciences, the biological, logistic, or other. Ideally, a strong group of such persons with practical knowledge and sound theoretical training should be attached to every important research or development organization. This would probably not be feasible because of the difficulty of getting a sufficient number of qualified statisticians. The next-best plan would doubtless be to have a central bureau of consultants whose services would be available to those organizations desiring them. I do not want to bring up the question of unifying the armed services, but I see no reason why the advice and assistance of such a bureau should not be available to any organization within the Department of Defense.

In summary, I think we need more statisticians of all kinds, and although I am connected with a group whose primary mission is basic research, I believe that the greatest need is for statistical consultants, those who can give practical aid on important research and development problems needing immediate solution.

DESCRIBABLE CHARACTERISTICS IN A STATISTICIAN

William Wolman

"What type of physicist is needed in a laboratory" or "What type of astronomer is needed for an observatory" are questions which are rarely asked. Why? Because the prime and overriding requirement is for a scientist with the highest technical qualifications. In considering the question as to what type of statistician is needed, we are dealing with somewhat of a different problem. What makes it a different problem? To enumerate a few characteristics regarding a statistician, the following points come to mind:

1. He must deal with other scientists and engineers.

2. The field is new and is a relatively unknown scientific discipline.

3. There are many people in the profession who got in through the back door and who are technically not up to par.

4. He must be willing to communicate and listen to other people's technical problems.

As a basic requirement the statistician must have adequate technical qualifications. I would include a sound foundation in the design and analysis of experiments -- the Analysis of Variance, Multiple Comparison Methods and knowledge of some of the more exotic designs, such as, "split plot", "fractional replication" and response surface techniques. The level of understanding should be that which is described in such excellent books as Kempthorne, Brownlee, Cochran and Cox, and Scheffé, references 1, 2, 3, and 4. A knowledge of the theory of estimation including confidence intervals and the determination of sources of variation as measured by variance components is another technical area of importance for the laboratory statistician. Furthermore, some knowledge of the various types of fixed and random models for the Analysis of Variance is important. I would like to recommend that if a laboratory desires a statistician that they hire a professional. Today, there are at least a dozen first-class universities which have programs in modern statistical theory and its application. It is not sufficient to transform a member of another discipline into a statistician. The common practice of using a "warmed over" engineer as a statistician can only lead to undesirable consequences. I would like to reiterate that I am not speaking about many fine statisticians whose original training was in some other discipline and who later in their career drifted into statistics. I have frequently encountered individuals who make the statement, "I am not a statistician" and then enter into a lengthy discourse on some statistical questions and consider their opening disclaimer a license to commit every conceivable kind of technical and logical error. In closing this paragraph I would like to point out that the technical requirements which I have enumerated above are only examples and should be supplemented with a sound foundation in statistical inference and probability. Furthermore, I am referring to a consultant statistician and not a research statistician, namely, one who does research in the theory and methods of mathematical statistics.

Having sound technical qualifications is not sufficient for a consultant statistician in the laboratory. He must be able to communicate and speak the language of his fellow scientists and engineers. Some understanding of the subject matter is also mandatory. He must certainly find out what the investigator whom he is trying to help wants to uncover or determine. For example, if it is a problem of determining which of two processes are superior, then he must recommend to the experimenter a procedure which will provide at least a 50:50 chance of detecting differences of a magnitude which are considered to be of consequence. A knowledge and understanding of laboratory techniques and metrology is certainly a most desirable knowledge for the statistician in the laboratory. Wilson's book, reference 5, gives an account of some of the functions and requirements for the experimenter and statistician in the laboratory.

As a consultant, the statistician must also have certain personal characteristics if he is to perform his job successfully. He must have a desire to help his fellow scientists and he must be able to communicate with them freely. His attitude should not be antagonistic or domineering. He should realize that he is performing a technical support function which is usually not in the limelight of an operation. I have found that if the experimenters in the laboratory feel that you have something to sell, namely, that you are able to help them in their work that you will soon be in a position of having more to do than you can possibly handle.

Some closing remarks as to the organizational location for a statistician in a laboratory follow: The statistician should have easy access to the experimenters and should not be hindered by any organizational barriers. It is usually desirable to deal with the experimenters in an informal manner. Constraints in terms of reporting and approval procedures should be held to an absolute minimum. For instance, keeping track as to the number of contracts a statistician has with his colleagues is virtually meaningless. The contributions of a consultant statistician are often very difficult to measure and may consist of an acknowledgment at the end of an experimenter's paper in a technical journal, or a statement of appreciation made at a meeting. I do not consider the formal organizational location of the statistician to be particularly important if he can perform his functions in a manner as I have outlined above. It is only if the organizational location is concerned with the remuneration that the individual is to receive that it may be an important question.

REFERENCES

1. Kempthorne, O. (1952). "The Design and Analysis of Experiments." Wiley, New York.

2. Brownlee, K. A. (1960). "Statistical Theory and Methodology in Science and Engineering." Wiley, New York.

3. Cochran, W. G. and Cox, G. M. (1957). "Experimental Designs." Wiley, New York.

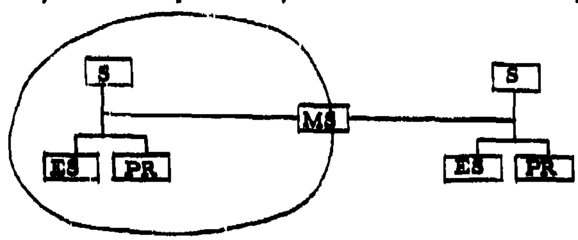
4. Scheffé, H. (1959). "The Analysis of Variance." Wiley, New York.

5. Wilson, E. B., Jr. (1952). "An Introduction to Scientific Research." McGraw-Hill, New York.

TYPES OF STATISTICIANS NEEDED

Donald A. Gardiner

The research and development laboratory needs two types of statistician to occupy three kinds of position on the statistical consulting staff. I would call the two types of statistician the external consultant and the internal consultant. The external consultant is the statistician who talks with the scientist or subject matter specialist about his problem and attempts to work out a solution. The internal consultant is the statistician who consults with statisticians and mathematicians. The three kinds of position may best be explained by reference to the accompanying diagram.



The diagram shows what I think the structure of a statistical consulting staff should look like. The circled portion represents a basic "module" or building block of the organization. I have shown two modules, but a large organization may need more. A smaller laboratory may need one module or only a part of one. The letter symbols are interpreted as follows:

MS	Mathematical or theoretical statistician
ES	Experimental or applied statistician
S	Statistician, either mathematical or experimental
PR	Programmer

Although I consider the programmer a basic and very important part of the organization, a discussion of his role and his characteristics is not pertinent to this panel.

The external consultants are the S and ES statisticians. Both of these people treat directly with the customer. Therefore, they should be personable, they should meet people well, and have the ability to express themselves clearly in a non-mathematical manner. These qualities are desirable in the internal consultant, too, but they are not absolutely necessary. While we are considering the internal consultant, it might be well to point out how his position could be filled in the case of the laboratory which does not have a full statistical staff. I think that in this case the services of an MS statistician might be obtained on a part-time basis--perhaps by hiring a consultant from a university for a few days each month.

The S and ES statisticians should have a science background. This is probably best obtained by taking a Bachelor of Science degree at a university. The mathematical scientistician would need mathematics training equivalent to a masters degree in mathematics. More important, however, is their education in statistics. Both the S and MS statistician should have statistical training at the Ph. D. level and the ES statistician at the masters level. I do not want to say that these degrees should be required. However, the levels of training I have in mind can be obtained today only in the graduate schools by courses of study leading to the Ph. D. and masters degrees.

I do not suscribe to the idea that a statistician working with chemists should have intensive training as a chemist nor that a statistician working with engineers should have training as a engineer. To paraphase George Box * in this regard, there would then be little point in training chemists or engineers -- only statisticians. But I would require that the statisticians be willing to show their ignorance in the field of application. To be able to help the scientist they must be inquisitive even to the point of asking stupid questions.

---

*G. E. P. Box, "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples, " p. 16-60 in Biometrics, vol 10, (1954).

AN ANALYSIS OF FACTORIAL EXPERIMENTAL DESIGNS

L. W. Kesting
U. S. Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT. A simple technique involving the analysis of variance is proposed for the initial evaluation of data obtained from a factorial experimental design. The application is presented for a general two-factor rectangle that could result from any factorial design. Equations were developed that permit the use of a uniform, simplified work sheet. An experimental design using the technique is proposed for studying the effects of heat on the insulation barriers separating auxiliary rockets from other missile components. A factorial design is suggested with four factors at levels 4, 3, 3, and 2, "confounded" in the multiple interactions. The type, thickness, and positions of insulation materials and the type of structural material supporting the insulation are considered. More test information and greater precision for less test time and equipment are the advantages of factorial designs of the type suggested. A numerical example illustrates the application of the analysis of the variance technique at marginal conditions of significance.

SYMBOLS.

A, B, C, D	Positions
F, R	Insulation type
o, a, b	Insulation thickness
k, m, n, p	Levels
Q	Combinations
S	Sum of squares
T	Total
V	Variance
X	Data point or measurement
x, y, z	Support material type of thickness

SYMBOLS (CONT.)

d. f.	Degrees of freedom
I, II, . . .	Factors
(1), (2), . . .	Material sets
σ	Estimated standard deviation of a population

SUBSCRIPTS

C	Column
E	Error
e	Combination measurement
i	Column measurement
j	Row measurement
R	Rows
S	Subtotal
T	Total
W	Within groups
.	Variable i, j, or e

INTRODUCTION. The use of a factorial design with confounded multiple interactions can improve precision and reduce the cost of testing. The various combinations, when collected into groups, form natural partitions of the whole program. Tests on the parts can often be performed independently. Thus, different batches of raw materials, pieces of test equipment, and experimental environments can be assigned to the parts. As factors are shown to be insignificant, the associated results and degrees of freedom can be combined with other factors to more precisely determine their effect. The investigator, in consequence, can achieve a systematic examination of the data available.

Factorial designs are applicable to many fields of study in which observations are made. The principles are used extensively in biology, medicine, psychology, economics, and sociology. In certain fields chemistry and engineering the use is more limited. To quote Dr. Youden (Reference 1): "Among chemists, the reception so far accorded this experimental device is reminiscent of the dilemma of the young man seeking employment. No one will hire him until he has experience, and he cannot get this experience until someone gives him a job."

The conflict between the use of formal experimental designs and the "shotgun" approach to research problems is well known. Formal designs will usually be attempted if time is available to select or originate them and properly treat the results. For the average scientist or engineer, who has only a nodding acquaintance with statistical principles, the digging from text books for a method to analyse a specialised factorial design is no small task. It boils down to a matter of cost, time, and ability, but not necessarily naiveness on the part of the investigator.

The purpose of this paper is to show in detail a simple application of the analysis of variance to a general factorial experimental design. This paper is intended for the scientist or engineer who can see advantages in confounding but has limited time and resources for an analysis of this type. Detailed steps are presented including a work sheet and presentation form. The method of application is based upon fundamental expressions derived in several texts (References 2, 3, 4, and 5) and upon the solution of numerical examples in Reference 1. An example is given that illustrates the method for a specific research problem.

METHOD. In some scientific investigations an attempt is made to hold all factors constant except the one under immediate study. Tests are run on the variable under immediate study, which is then held constant while a second factor is varied. The procedure is continued until all pertinent variables have been investigated. Each factor (I, II, III, and IV) may be varied at two or more levels (k, m, n, and p). The total number of different experimental combinations, Q, is

$$Q = k\,m\,n\,p.$$

Tests on every possible combination are usually prohibitively expensive and time consuming or possible outside the scope of the investigation.

An investigation that includes the study of the effects produced by varying two or more factors simultaneously provides the setting for a factorial experimental design. These designs often have advantages over holding constant all variables except one. For example:

1. The number of test runs can usually be reduced without sacrificing important data.
2. The precision of the investigation can be improved.
3. The cost and complexity of test equipment can be reduced.
4. Investigations can be undertaken for which it is impossible to hold constant several important variables.

The advantages accrue from confounding factors and levels and eliminating certain combinations on the basis of prior knowledge or engineering judgment.

A general factorial experimental design is shown in Table I. Since the design is presented primarily to illustrate the application of a method, no combinations have yet been eliminated. Under the heading Example, to be found a few pages later one, the design is expanded and certain combinations are eliminated.

The analysis of variance is usually employed initially to evaluate the data taken in a factorial experiment. The variance, as used here, is the square of the estimate of the standard deviation of the population (σ^2) represented by the experimental data. Variances are calculated of groups of data associated with the factors and levels. If the groups are all from the same population, the results will be essentially the same. The groups showing a separate influence on the data will have higher variances. F tables indicate the magnitude of the differences at various confidence or probability levels. The theory underlying the analysis of variance and the F tables is presented in detail in References 1 and 2, and a development will not be attempted here. However, formulas and procedures are given to simplify the computations. Fundamental expressions are given in Table II.

Table I

VARIANCE RECTANGLE

ROW FACTOR II		COLUMN FACTOR I: LEVEL K_1	LEVEL K_2	LEVEL K_3	
	LEVEL N_2	X_{111}	X_{211}	X_{311}	
		X_{112}	X_{212}	X_{312}	
		X_{113}	X_{213}	X_{313}	
		X_{114}	X_{214}	X_{314}	
		$T_{11.}$	$T_{21.}$	$T_{31.}$	$T_{.1.}$
		$\bar{X}_{11.}$	$\bar{X}_{21.}$	$\bar{X}_{31.}$	$\bar{X}_{.1.}$
	LEVEL N_1	X_{121}	X_{221}	X_{321}	
		X_{122}	X_{222}	X_{322}	
		X_{123}	X_{223}	X_{323}	
		X_{124}	X_{224}	X_{324}	
		$T_{12.}$	$T_{22.}$	$T_{32.}$	$T_{.2.}$
		$\bar{X}_{12.}$	$\bar{X}_{22.}$	$\bar{X}_{32.}$	$\bar{X}_{.2.}$
		$T_{1..}$	$T_{2..}$	$T_{3..}$	$T_{...}$
		$\bar{X}_{1..}$	$\bar{X}_{2..}$	$\bar{X}_{3..}$	$\bar{X}_{...}$

k = NUMBER OF LEVELS FOR COLUMNS

n = NUMBER OF LEVELS FOR ROWS

m = NUMBER OF MEASUREMENTS IN EACH CATEGORY

X_{ije} = THE ETH MEASUREMENT IN THE ITH COLUMN OF THE JTH ROW

$$T_{ij.} = \sum_{e=1}^{m} X_{ije}$$

$$\bar{X}_{ij.} = \frac{T_{ij.}}{m}$$

$$T_{.j.} = \sum_{i=1}^{k} T_{ij.}$$

$$\bar{X}_{.j.} = \frac{T_{.j.}}{k}$$

$$T_{i..} = \sum_{j=1}^{n} T_{ij.}$$

$$\bar{X}_{i..} = \frac{T_{i..}}{nm}$$

$$T_{...} = \sum_{j=1}^{n} T_{.j.} = \sum_{i=1}^{k} T_{i..} = \sum_{i=1}^{k} \sum_{j=1}^{n} \sum_{e=1}^{m} X_{ije}$$

$$\bar{X}_{...} = \frac{T_{...}}{knm}$$

Table II

BASIC EQUATIONS FOR THE ANALYSIS OF VARIANCE

SOURCE	SUMS OF SQUARES	d.f.
COLUMN MEANS	$S_C = nm\sum(\bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2$	$k-1$
ROW MEANS	$S_R = km\sum(\bar{x}_{\cdot j\cdot} - \bar{x}_{\cdot\cdot\cdot})^2$	$n-1$
INTERACTION	$S_I = S_S - S_C - S_R$	$(k-1)(n-1)$
SUBTOTAL	$S_S = m\sum(\bar{x}_{ij\cdot} - \bar{x}_{\cdot\cdot\cdot})^2$	$nk-1$
WITHIN GROUPS	$S_W = S_T - S_S$	$nk(m-1)$
TOTAL	$S_T = \sum(x_{ije} - \bar{x}_{\cdot\cdot\cdot})^2$	$nkm-1$

THE EXPRESSION FOR THE SUM OF SQUARES OF THE COLUMN MEANS CAN BE REARRANGED TO SIMPLIFY THE COMPUTATIONS.

$$S_C = nm\sum(\bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2 = nm\left[(\bar{x}_{1\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2 + (\bar{x}_{2\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2 + \ldots + (\bar{x}_{k\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2\right] \quad (1)$$

WHICH IS SQUARED TO OBTAIN

$$\begin{aligned} S_C = {} & nm\left[\bar{x}^2_{1\cdot\cdot} - 2\bar{x}_{1\cdot\cdot}\bar{x}_{\cdot\cdot\cdot} + \bar{x}^2_{\cdot\cdot\cdot}\right] + \\ & nm\left[\bar{x}^2_{2\cdot\cdot} - 2\bar{x}_{2\cdot\cdot}\bar{x}_{\cdot\cdot\cdot} + \bar{x}^2_{\cdot\cdot\cdot}\right] + \ldots \\ & nm\left[\bar{x}^2_{k\cdot\cdot} - 2\bar{x}_{k\cdot\cdot}\bar{x}_{\cdot\cdot\cdot} + \bar{x}^2_{\cdot\cdot\cdot}\right]. \end{aligned} \quad (2)$$

SUMMING,

$$S_C = nm\sum_{i=1}^{k}\bar{x}^2_{i\cdot\cdot} - 2nm\bar{x}_{\cdot\cdot\cdot}\sum_{i=1}^{k}\bar{x}_{i\cdot\cdot} + knm\,\bar{x}^2_{\cdot\cdot\cdot} \quad (3)$$

FROM TABLE I.

$$nm\sum_{i=1}\bar{x}_{i\cdot\cdot} = \sum_{i=1}^{k}T_{i\cdot\cdot} = T_{\cdot\cdot\cdot} = knm\,\bar{x}_{\cdot\cdot\cdot} \quad (4)$$

SQUARING BOTH SIDES AND REARRANGING,

$$nm\sum_{i=1}^{k}\bar{x}^2_{i\cdot\cdot} = \sum_{i=1}^{k}\frac{T^2_{i\cdot\cdot}}{nm} \quad \text{AND} \quad knm\,\bar{x}^2 = \frac{T^2_{\cdot\cdot\cdot}}{knm}. \quad (5)$$

SUBSTITUTING EQUATIONS (4) AND (5) IN EQUATION (3) GIVES

$$S_C = \sum_{i=1}^{k}\frac{T^2_{i\cdot\cdot}}{nm} - \frac{T^2_{\cdot\cdot\cdot}}{knm},$$

WHICH IS MORE CONVENIENT FOR USE IN THE WORK SHEET FORM.

Sums of squares for rows, subtotals, and totals can be treated in the same manner to obtain the expressions in Tables III and IV.

Table III

WORK SHEET FORM

	SINGLE MEASUREMENTS	REPLICATIONS	ROW TOTALS	COLUMN TOTALS	GRAND TOTAL
SQUARE AND SUM	$\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{e=1}^{m} X_{ije}^2$	$\sum_{i=1}^{k}\sum_{j=1}^{n} T_{ij\cdot}^2$	$\sum_{j=1}^{n} T_{\cdot j\cdot}^2$	$\sum_{i=1}^{k} T_{i\cdot\cdot}^2$	$T_{\cdot\cdot\cdot}^2$
DIVIDE BY	1	m	km	nm	knm
QUOTIENT	$\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{e=1}^{m} X_{ije}^2$	$\sum_{i=1}^{k}\sum_{j=1}^{n} \frac{T_{ij\cdot}^2}{m}$	$\sum_{j=1}^{n} \frac{T_{\cdot j\cdot}^2}{km}$	$\sum_{i=1}^{k} \frac{T_{i\cdot\cdot}^2}{nm}$	$\frac{T_{\cdot\cdot\cdot}^2}{knm}$
SUBTRACT	$\frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\frac{T_{\cdot\cdot\cdot}^2}{knm}$	
SUM OF SQUARES	$\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{e=1}^{m} X_{ije}^2 - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\sum_{i=1}^{k}\sum_{j=1}^{n} \frac{T_{ij\cdot}^2}{m} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\sum_{j=1}^{n} \frac{T_{\cdot j\cdot}^2}{km} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$\sum_{i=1}^{k} \frac{T_{i\cdot\cdot}^2}{nm} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	

Table IV

ANALYSIS OF VARIANCE PRESENTATION FORM

SOURCE	SUM OF SQUARES	d.f.	VARIANCE	F RATIO
COLUMN MEANS	$S_C = \sum_{i=1}^{k} \frac{T_{i\cdot\cdot}}{nm} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$k-1$	$V_C = \frac{S_C}{k-1}$	$\frac{V_C}{V_E}$
ROW MEANS	$S_R \sum_{j=1}^{n} \frac{T_{\cdot j}^2}{km} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$n-1$	$V_R = \frac{S_R}{n-1}$	$\frac{V_R}{V_E}$
INTERACTION	$S_I = S_S - S_C - S_R$	$(k-1)(n-1)$	$V_I = \frac{S_I}{(k-1)(n-1)}$	$\frac{V_I}{V_W}$
SUBTOTAL	$S_S = \sum_{i=1}^{k} \sum_{j=1}^{n} \frac{T_{ij}^2}{m} - \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$nk-1$		
WITHIN GROUPS	$S_W = S_T - S_S$	$nk(m-1)$	$V_W = \frac{S_W}{nk(m-1)}$	
TOTAL	$S_T = \sum_{i=1}^{k} \sum_{j=1}^{n} \sum_{e=1}^{m} X^2{}_{ije} \frac{T_{\cdot\cdot\cdot}^2}{knm}$	$nKm-1$		

Table IV is the conventional form for presenting the results of an analysis of variance. The F ratios are the final computations and are compared with critical values found in F tables in many references on statistics (References 1 and 2). If an F ratio exceeds the critical value, a significant effect from the factor is proved at the confidence level selected.

The analysis of variance is applicable to any design for experiments at two or more levels provided the following assumptions are acceptable:

1. There is no interaction between row and column factors at the levels investigated.

2. The data relative to the factors have a homogeneous variance and are from populations with normal distributions.

However, moderate violations of these assumptions change the analysis very little and the validity of the assumptions can be checked from the data to be treated.

The first ratio to test is "interaction" over the "within groups" variance. The presence of significant interaction invalidates further tratment by the analysis of variance, and other methods must be used. When the critical F value is exceeded indicating interaction, the following conditions exist separately or in combination:

1. The row or column factors are producing effects when tested together that do not occur when the factors receive the same test separately.

2. An additional factor is of sufficient importance to be included in the analysis.

3. The items in the subgroups are not drawn at random.

If no interaction is present, the sums of squares for the "interaction" and "within groups" sources can be added and divided by their total degrees of freedom to provide a more inclusive error term, V_E, where

$$V_E = \frac{S_T - S_C - S_R}{nkm - k - n + 1} .$$

The F ratios for rows and columns, V_C/V_E, are now more sensitive. A result higher than the critical value for F tables indicates a significant influence from the row or column factor. The presence or absence of interaction is in itself an important result of many investigations.

The test for homogeneity of variance requires the computation of the ratio of S_C and S_R to S_W. A critical F value is computed, and the results are compared (Reference 2).

An example illustrating this method of applying the analysis of variance to factorial designs follows. One case of interaction between two factors is given.

EXAMPLE. A factorial design is presented with four factors: one at four levels, two at three levels, and one at two levels. The analysis of variance techniques are used to compute the desired probabilities. Hypothetical results are used to illustrate the analysis at marginal conditions of significance.

Present missile designs usually include small auxiliary rockets that perform various functions within the overall mission. Gas generators exhausting through turbines are used to rotate propellant pumps and electrical generators. In some designs, the generators directly pressurize propellant tanks and hydraulic systems. Small rocket motors are used to spin certain stages of the missile for axial stability. Vernier motors are fired to adjust speed and direction. Retrorockets used on planet probes decrease the speed of the satellite when entering the desired orbit. The auxiliary rockets, located at many points on the missile, vary in size, thrust, and burning time and operate from solid, liquid, and gaseous propellants.

Firing auxiliary rockets often causes severe environmental changes around adjacent missile components. High temperatures result from exhaust gases and hot motor surfaces. Corrosion and erosion are sometimes caused by the discharged gases.

This example is a proposal for calculating the probability of an influence from materials of adjacent components in controlling the heat from auxiliary rockets. The influence could originate from either type, thickness, or position of insulating materials or the type of supporting structure.

The proposed design may not be optimum when a free choice of factors and levels is permitted. In this case, the number and arrangement of the factors and levels depend upon the scope, precision, method of testing, and future test programs. Factors and levels are unique for a particular experiment. In general, designs with all factors at the same level are simpler. Other possible combinations are enumerated in References 6 and 7.

<u>ASSUMPTIONS</u>. The proposed design applies to a particular investigation characterized by the following assumptions:

1. Eight sets of insulating and supporting material samples are the maximum that can be placed around a single rocket motor. (The arrangement may be similar to that shown in Figures 1 and 2.)

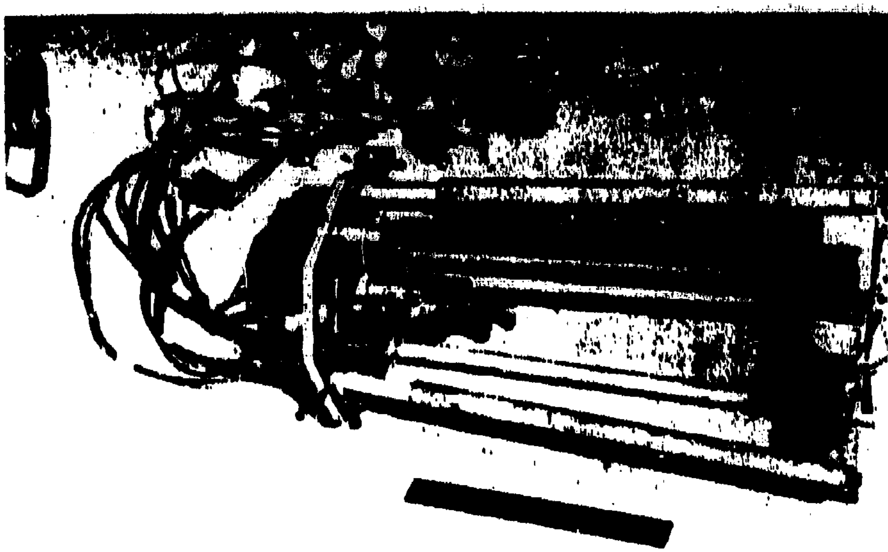


Figure 1. SATELLITE ANTENNA RODS GROUPED AROUND A VERNIER ROCKET MOTOR

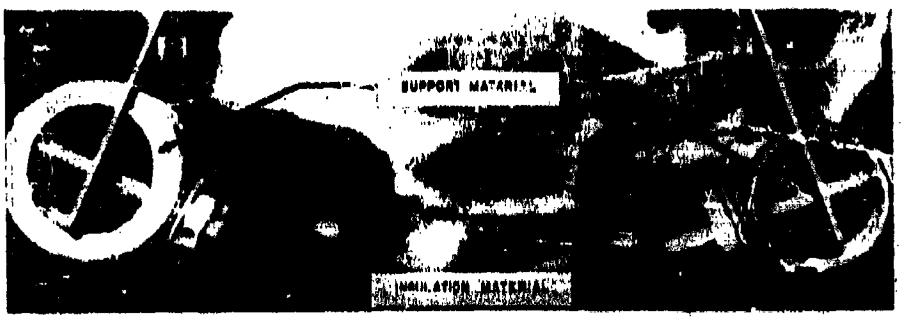


Figure 2. EXAMPLES OF MATERIAL SETS

2. Thermocouples are uniformly installed on the surface of the supporting material on the side away from the motor (Figure 3).

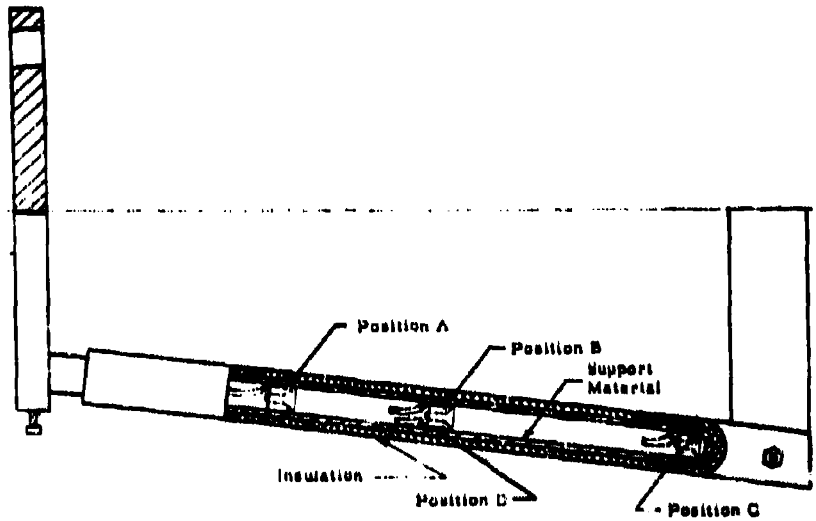


Figure 3. THERMOCOUPLE LOCATION ON A MATERIAL SET

3. Thermocouples are located at the same four positions on each set of material samples.

4. Only one rocket motor will be fired with a given set of material samples. Thermocouple outputs will be recorded until the samples reach thermal stability.

5. Four factors at levels 4, 3, 3, and 2, as given in Table V, will be investigated.

ARRANGEMENT

For an investigation based upon the above assumptions, the design in Table V is proposed for the arrangement of factors and levels.

Table V

PROPOSED FACTORIAL DESIGN

FACTORS AND LEVELS ASSUMED

FACTOR	SYMBOL	LEVEL
POSITION - INSULATION AND SUPPORT MATERIAL	A, B, C, D	4
TYPE - INSULATION MATERIAL	F, R	2
THICKNESS - INSULATION MATERIAL	o, a, b	3
THICKNESS OR TYPE - SUPPORT MATERIAL	x, y, z	3

		x			y			z		
		o	a	b	o	a	b	o	a	b
F	A	$X_{(1)}$		$X_{(2)}$		$X_{(3)}$				$X_{(4)}$
	B	$X_{(1)}$		$X_{(2)}$		$X_{(3)}$				$X_{(4)}$
	C	$X_{(1)}$		$X_{(2)}$		$X_{(3)}$				$X_{(4)}$
	D	$X_{(1)}$		$X_{(2)}$		$X_{(3)}$				$X_{(4)}$
R	A		$X_{(5)}$		$X_{(6)}$		$X_{(7)}$		$X_{(8)}$	
	B		$X_{(5)}$		$X_{(6)}$		$X_{(7)}$		$X_{(8)}$	
	C		$X_{(5)}$		$X_{(6)}$		$X_{(7)}$		$X_{(8)}$	
	D		$X_{(5)}$		$X_{(6)}$		$X_{(7)}$		$X_{(8)}$	

The numbers in parenthesis refer to the eight sets of insulation and support material samples grouped around the rocket motor (Figures 1 and 2). Four thermocouples are uniformly located on each set representing the four positions, A, B, C, and D (Figure 3).

The analysis of variance for the factorial design shown in Table V is summarized in Table VI.

Table VI

ANALYSIS OF VARIANCE

SOURCE	d.f.
A B C D - POSITION OF INSULATION AND SUPPORT MATERIAL	3
x y z - TYPE OF INSULATION MATERIAL	2
o a b - THICKNESS OF INSULATION MATERIAL	2
F R - THICKNESS OR TYPE OF SUPPORT MATERIAL	1
TWO-FACTOR INTERACTIONS	
A B C D vs x y z	6
A B C D vs o a b	6
A B C D vs F R	3
x y vs o a x y vs a b y z vs a b	3
F R vs x y F R vs y z	1
F R vs o a F R vs a b	1
ERROR	3
TOTAL	31

NUMERICAL COMPUTATIONS

The following numerical example (Tables VII through XIII) is given to illustrate the initial method of analysis for a factorial experiment. The composite design is broken down into simple two-factor rectangles suitable for the analysis of variance technique. The hypothetical data are thermocouple readings in degrees F.

Table VII

WORK SHEET FOR ABCD VS xyz

PROPOSED FACTORIAL EXPERIMENT
4 x 3 x 3 x 2 NUMERICAL EXAMPLE

		x			y			z		
		o	a	b	o	a	b	o	a	b
F	A	330		361		361				373
	B	337		355		375				372
	C	333		320		359				359
	D	320		357		310				370
R	A		329		370		375		379	
	B		351		330		361		360	
	C		315		340		358		359	
	D		327		300		329		347	

		A	B	C	D	TOTAL
x	F_b	361	355	320	357	
	R_a	329	351	315	327	
		690	706	635	684	= 2715
y	F_a	361	375	359	310	
	R_b	375	361	358	329	
		736	736	717	639	= 2828
z	F_b	373	372	359	370	
	R_a	379	360	359	347	
		752	732	718	717	= 2919
		752	732	718	717	
TOTALS		2178	2174	2070	2040	= 8462

Table VIII

ANALYSIS OF VARIANCE FOR ABCD VS xyz

	24 READINGS	12 TOTALS OF DUPLICATES	ROW TOTALS	COLUMN TOTALS	GRAND TOTAL
SQUARE AND SUM	2,992,894	5,982,360	23,889,370	17,916,460	71,605,444
DIVIDE BY	1	2	8	6	24
QUOTIENT	2,992,894	2,991,180	2,986,171	2,986,076	2,983,560
SUBTRACT	2,983,560	2,983,560	2,983,560	2,983,560	
SUM OF SQUARES	9,334	7,620	2,611	2,516	

SOURCE	SUM OF SQUARES	d.f.	VARIANCE	F
COLUMNS ABCD	2516	3	838	3.59
ROWS xyz	2611	2	1305	5.60
INTERACTION	2493	6	415	2.92
SUBTOTAL	7620	11		
WITHIN GROUPS	1714	12	142	
TOTAL	9334	23		

$$V_E = \frac{2493 + 1714}{6 + 12} = \frac{4207}{18} = 233$$

CRITICAL F VALUES (REF. 2)

d.f	CONFIDENCE LEVEL	
	95 PERCENT	99 PERCENT
6-12	3.00	4.92
3-18	3.16	5.09
2-18	3.55	6.01

THE EFFECTS OF ABCD AND xyz ARE SIGNIFICANT AT THE CONFIDENCE LEVEL OF 0.95 BUT NOT AT 0.99 SINCE $3.16 < 3.59 < 5.09$ AND $3.55 < 5.60 < 6.01$.

Table IX

EXAMPLE OF INTERACTION BETWEEN ABCD AND xyz

	A	B	C	D	
x	361 329 690	372 360 732	320 315 635	357 327 684	2741
y	361 375 736	359 358 717	375 361 736	310 329 639	2828
z	373 379 753	355 351 706	359 359 718	370 347 717	2893
	2178	2155	2089	2040	

	24 READINGS	12 TOTALS	ROWS	COLUMN	GRAND TOTAL
SQUARE AND SUM	2,992,894	5,982,360	23,880,114	17,913,230	71,605,444
DIVIDE	1	2	8	6	24
QUOTIENT	2,992,894	2,991,180	2,985,014	2,985,538	2,983,560
SUBTRACT	2,983,560	2,983,560	2,983,560	2,963,560	
SUM AND SQUARE	9,334	7,620	1,454	1,978	

ANALYSIS OF VARIANCE ILLUSTRATING INTERACTION

SOURCE	SUM OF SQUARES	d.f.	VARIANCE
COLUMNS	1978	3	659
ROWS	1454	2	727
INTERACTION	4138	6	690
SUBTOTAL	7620	11	
WITHIN GROUPS	1714	12	142
TOTAL	9334	23	

INTERACTION IS SIGNIFICANT ABOVE THE 99-PERCENT CONFIDENCE LEVEL SINCE 4.91 > 4.82 FOR 6 AND 12 DEGREES OF FREEDOM.

Table X

WORK SHEET ABCD VS oab

PROPOSED FACTORIAL EXPERIMENT
4 x 3 x 3 x 2 NUMERICAL EXAMPLE

		x			y			z		
		o	a	b	o	a	b	o	a	b
F	A	330		361		361				373
	B	337		355		375				372
	C	333		320		359				359
	D	320		357		310				370
R	A		329		370		375		379	
	B		351		330		361		360	
	C		315		340		358		359	
	D		327		300		329		347	

		A	B	C	D	
o	F_x	330	337	333	320	
	R_y	370	330	340	300	
		700	667	673	620	2660
a	F_y	361	375	359	310	
	R_z	329	351	315	327	
		690	726	674	637	2727
b	F_x	361	355	320	357	
	R_y	375	361	358	329	
		736	716	678	686	2816
		2126	2109	2025	1943	8203

Table XI

ANALYSIS OF VARIANCE FOR ABCD VS oab

	24 READINGS	12 TOTALS OF DUPLICATES	ROW TOTALS	COLUMN TOTALS	GRAND TOTAL
SQUARE AND SUM	2,814,227	5,620,071	22,441,985	16,843,631	67,289,209
DIVIDE BY	1	2	8	6	24
QUOTIENT	2,814,227	2,810,035	2,805,248	2,807,271	2,803,718
SUBTRACT	2,803,718	2,803,718	2,803,718	2,803,718	
SUM OF SQUARES	10,509	6,317	1,530	3,553	

SOURCE	SUM OF SQUARES	d.f.	VARIANCE	F
COLUMNS ABCD	3553	3	1184	3.93
ROWS oab	1530	2	765	2.54
INTERACTION	1234	6	206	0.59
SUBTOTAL	6317	11		
WITHIN GROUPS	4192	12	349	
TOTAL	10509	23		

$$V_E = \frac{1234 + 4192}{18} = 301$$

THE FACTOR ABCD IS SIGNIFICANT AT THE 95- TO 99-PERCENT CONFIDENCE RANGE.
ROWS AND INTERACTION ARE NOT SIGNIFICANT.

Table XII

PROPOSED FACTORIAL EXPERIMENT

4 x 3 x 3 x 2 NUMERICAL EXAMPLE

		x			y			z		
		o	a	b	o	a	b	o	a	b
F	A	330		361		361				373
	B	337		355		375				372
	C	333		320		359				[illegible]
	D	320		357		310				370
R	A		329		370		375		379	
	B		351		330		361		360	
	C		315		340		358		359	
	D		327		300		329		[illegible]	

		A	B	C	D	
F	x_b	361	355	320	357	
	y_a	361	375	359	310	
		722	730	679	667	2798
R	x_a	329	351	315	327	
	y_b	375	361	358	329	
		704	712	673	656	2745
		1426	1442	1352	1323	5543

Table XIII

ANALYSIS OF VARIANCE FOR ABCD VS FR

	16 READINGS	8 DUPLICATES	ROW TOTALS	COLUMN TOTALS	GRAND TOTAL
SQUARE AND SUM	1,927,069	3,845,939	15,363,829	7,691,073	30,724,849
DIVIDE BY	1	2	8	4	16
QUOTIENT	1,927,069	1,922,969	1,920,478	1,922,768	1,920,303
SUBTRACT	1,920,303	1,920,303	1,920,303	1,920,303	
SUM OF SQUARES	6,766	2,666	175	2,465	

SOURCE	SUM OF SQ	d.f.	VARIANCE	F
COLUMNS	2465	3	822	2.19
ROWS	175	1	175	0.47
INTERACTION	26	3	8	.02
SUBTOTAL	2666	7		
WITHIN GROUPS	4100	8	512	
TOTAL	6766	15		

V_E = 375

THE CRITICAL F VALUE FOR 3 AND 11 d.f. IS 3.59 AT 95 PERCENT.

NO SIGNIFICANCE OF ANY FACTOR IS SHOWN BY THE F RATIO.

In the proposed design, two to four data points are available at each combination of the factors. The points can be used as replications to check for interaction. However, the points are not true replications until the secondary factors in the combination are proved insignificant (for instance, F, R, o, a, and b in the analysis of the primary factors ABCD vs xyz). When secondary factors are significant, the test for interaction is valid but less sensitive.

As factors are shown to be insignificant, the associated results and degrees of freedom can be combined with other factors to more precisely determine their effect. For example, in the proposed design, if the insulation thickness factor at levels o, a, and b is shown to be insignificant, three data points for each combination instead of two become available for analyzing the remaining factors. Insignificance of x, y, and z must also exist to provide extra data points for ABCD vs FR.

The probability of an influence from materials of adjacent components in controlling heat from auxiliary rocket motors can be determined from the F tables. Values selected from the tables that bound the F ratio are computed for the material factor. The probability is then obtained by interpolation of the corresponding headings of the F table.

CONCLUSION. Since the use of formal experimental designs and other statistical principles is limited largely by funds, time, and capabilities of the investigator, simplified methods of analysis should be developed for those not thoroughly acquainted with statistics. One approach has been attempted by orienting a computational procedure with fundamental expressions for the analysis of variance and applying the technique to factorial designs. An example is presented illustrating the technique for a research problem.

The use of auxiliary rocket motors on missiles often creates space and weight problems in insulating adjacent components from hot motor surfaces and exhaust gases. An experimental design is proposed for computing the probability of an effect from material factors, such as insulation type, thickness, position, and support material type.

The arrangement of factors and levels is based upon limited test conditions, including eight sets of materials, four thermocouple positions, two types and three thicknesses of insulation, and three types of support material. The arrangement provides the following characteristics:

1. The four factors can be varied simultaneously.

2. Two to four data points at each combination of primary factors are available as replications to test for interaction.

3. As factors are proved insignificant, the associated results can be combined with other factors to increase the sensitivity of the remaining analysis.

4. The inherent advantages of general factorial designs apply, such as a natural partition of the whole test program, reduction in the number of test runs, and an improvement in scope and precision.

REFERENCES

1. W. J. Youden: Statistical Methods for Chemists. John Wiley and Sons, Inc., New York, 1951.

2. Wilfrid J. Dixon and Frank J. Massey: Introduction to Statistical Analysis. (First Edition), McGraw-Hill Book Company, New York, 1951.

3. F. E. Croxton and D. J. Cowden: Applied General Statistics. Prentice Hall, N. W., 1939.

4. S. S. Wilks: Elementary Statistical Analysis, Princeton University Press, 1951.

5. William Volk: Applied Statistics for Engineers. McGraw-Hill, New York, 1958.

6. Ronald A. Fisher: The Design of Experiments. Oliver and Boyd, Edinburgh, 1949.

7. William G. Cochran and Gertrude M. Cox: Experimental Designs. John Wiley and Sons, Inc., New York, 1959.

RELIABILITY ESTIMATION FOR MULTI-COMPONENT SYSTEMS

J. R. Kniss
Ballistic Research Laboratories
Aberdeen Proving Ground, Md.

INTRODUCTION. The U. S. Army must maintain a world-wide stockpile of war reserve ammunition for both combat and training. This ammunition must be kept in a constant state of readiness; and it is, therefore, important that the quality and reliability of the stockpile be estimated periodically.

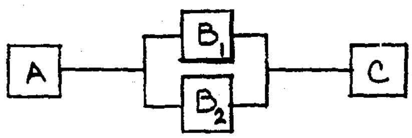
The Surveillance Group of the Ballistic Research Laboratories periodically analyzes the results of tests conducted on samples taken from the stockpile. In the case of conventional ammunition the items are usually tested ballistically by means of some type of statistical design, and the samples are destroyed or consumed in the tests. In the case of nuclear ammunition, it is usually impractical to evaluate the quality and reliability of the stockpile through ballistic tests. Therefore, laboratory test data are obtained, where the major components of the particular item are tested statically and usually non-destructively. In these tests each component is tested as an individual item, and it is assumed that the functioning of one component is independent of the functioning of the remaining components, once it has been removed from the system for testing (i.e., although it is true that the functioning of one component will generally depend upon the functioning of other components, the inherent capability of its functioning as a component within itself still exists).

The results of the individual component tests must then be analyzed and combined in order to estimate the reliability of the system and also to place an interval about this estimate which will yield some specified degree of confidence that the true reliability lies within this interval. A number of solutions for reliability estimation have been developed and appear in the literature. However, none of these solutions is satisfactory for the type of problem with which this organization is confronted. It was, therefore, necessary to devise a system for estimating reliability that is highly flexible, fast and simple; and it is felt that the method outlined in this paper satisfies these conditions. Although this system depends upon access to a high speed digital computer, it does have the advantage of flexibility, speed and simplicity once the original problem has been programmed.

DISCUSSION. Determination of the reliability of each type of component[1] is relatively straightforward. It is first required that a decision, based on the test results, be made on each item tested as to whether it is a success or a failure. Once a decision has been reached on each item, the number of successful components of a particular type (x) follows a binomial distribution. The point estimate of reliability ($\hat{p}$) for each type of component is $\hat{p} = \frac{x}{n}$ where n is the number tested. A confidence interval may then be placed on p using statistical procedures.

However, for stockpile reliability testing it is of primary interest to determine the reliability of the entire system. This relationship can be expressed in the form of a block diagram where each block represents a specific component or part tested. (Two or more blocks could represent the same type of component where redundancy is built into the system.)

For instance, a system of components might be arranged in the following manner

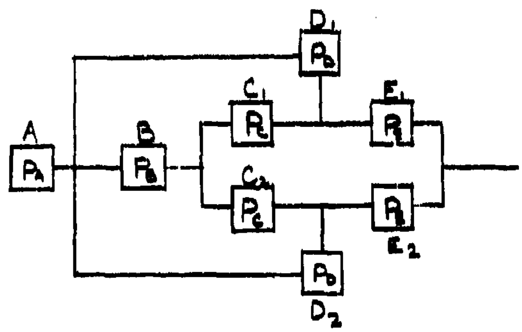


where B_1 and B_2 are the same types of components. If the reliabilities of these types of components were established at p_1, p_2, and p_3 respectively, then the reliability of (R) of the system would simply be

$$R = p_1 (2p_2 - p_2^2) p_3$$

1 - A type of component could be a timer, radar fuze, etc.

Nuclear warheads, of course, are much more complicated than the above example; and it is not always possible to write the reliability equation through inspection of the block diagram. The following is a sample of a block diagram whose reliability equation cannot be readily obtained.



In order to compute the reliability of this system of components it is necessary to enumerate all combinations of success-failure possibilities for components B through E. Letting a success be represented by a 1 and a failure by a 0 the reliability of the system may be computed in the following manner:

Components							Probability	System Success or Failure
B	C_1	C_2	D_1	D_2	E_1	E_2		
1	1	1	1	1	1	1	$P_B\ P_C^2\ P_D^2\ P_E^2$	S
1	1	1	1	1	1	0	$P_B\ P_C^2\ P_D^2\ P_E\ (1-PE)$	S
1	1	1	1	1	0	1	$P_B\ P_C^2\ P_D^2\ (1-P_E)\ P_E$	S
			.				.	.
			.				.	.
			.				.	.
1	1	1	1	1	0	0	$P_B\ P_C^2\ P_D^2\ (1-P_E)^2$	F
1	1	1	1	0	1	0	$P_B\ P_C^2\ P_D\ (1-P_D)\ P_E(1-P_E)$	S
			.				.	.
			.				.	.
			.				.	.
1	0	0	0	0	0	0	$P_B\ (1-P_C)^2\ (1-P_D)^2\ (1-P_E)^2$	F
0	0	0	0	0	0	0	$(1-P_D)\ (1-P_C)^2\ (1-P_D)^2(1-P_E)^2$	F

All probabilities that yield successes should be added, and the reliability of the system can then be written

$$R = P_A P_E \left\{ P_D (2 - P_D P_E) + P_B P_C (2 - P_C) + P_B P_C P_D \left[2P_E (P_D - 1) + P_C (2 - P_D) - 2 \right] \right\}.$$

The reliability of the system can now be computed from the above equation. This example is one of the least complex systems that has been encountered in practice. Most systems require thousands of enumerations resulting in reliability equations covering several pages.

The problem of placing a confidence interval on R is difficult. The number of successes ($\hat{x}$) associated with each value of $\hat{p}$ has a binomial distribution and a confidence interval may be placed on each true p using this distribution. In order to place a confidence interval on R, however, it is also necessary to utilize the distribution of the individual $\hat{x}$'s, since R is a function of x's.

If it can be assumed that $\hat{p}$ is a random variable and that the distribution of $\hat{p}$ is known, then random samples can be drawn from these distributions and substituted into the reliability equation and random values of $\hat{R}$ can be obtained. The distribution of $\hat{R}$ can be estimated, and a confidence interval can be constructed which will produce some measure of assurance that the true reliability, R, lies within the interval.

The distribution of $\hat{p}$ (given p) may be represented as

$$g(\hat{p} ; p) = \binom{n}{n\hat{p}} p^{n\hat{p}} (1-p)^{n(1-\hat{p})}$$

where $p = 0, 1/n, 2/n, \ldots, 1$. However, it is not possible or desirable to sample the above distribution for values of $\hat{p}$ since $\hat{p}' = x/n$ must be substituted for p and, for the case where $\hat{x} = 0$ or $\hat{x} = n$, $g(\hat{p} ; p) = 0$. This is a result of the assumption that $p = \hat{x}/n$, and, of course, if p equals either 0 or 1 there can be no sampling error. In fact, for any value of $\hat{x}$ the variation introduced would be entirely due to sampling, and any confidence intervals generated would not be realistic since the error involved in estimating p would be ignored.

Confidence intervals may be placed on an individual p at the level of confidence by using the binomial distribution in the following manner.

$$(1) \quad \sum_{x=0}^{\hat{x}} \binom{n}{x} \hat{p}_1^{\,x} (1 - \hat{p}_1)^{n-x} = (1 - \alpha)/2 \qquad \text{(upper tail)}$$

$$(2) \quad \sum_{x=\hat{x}}^{n} \binom{n}{x} \hat{p}_2^{\,x} (1 - \hat{p}_2)^{n-x} = (1 - \alpha)/2 \qquad \text{(lower tail)}$$

$x = 0, 1, \ldots, n.$

By solving the above equation at specified levels of α, it is possible to construct an interval such that $100\alpha\%$ of all intervals constructed in such a manner will contain p. It is also possible to generate values of $\hat{p}$ over a continuous range from the above equation by varying α. However, it is not possible to construct intervals in the vicinity of $\hat{x}/n$ where α is small or where $(1-\alpha)/2$ is near or equal to 0.5. In fact, for $(1 - \alpha)/2 = 0.5$, where one would expect the interval to be a point, two values of $\hat{p}$ are obtained depending upon the equation used. This result becomes intuitively apparent when it is considered that equation (1) gives a value of $\hat{p}_1$ for which $(1-\alpha)/2$ (100)% of the distribution will be $\hat{x}$ or less and equation (2) gives a value of $\hat{p}$ for which $(1-\alpha)/2$ (100)% of the distribution will be $\hat{x}$ or more. Therefore, where $(1-\alpha)/2 = 0.5$, 50% of the distribution must be $\hat{x}/n$ or more in one case and 50% must by $\hat{x}/n$ or less in the other. This, of course, is a contradiction caused by the fact that the binomial distribution is descrete with respect to x, but $\hat{p}$ takes on continuous values.

Since equations (1) and (2) are accurate at the tails, it was decided to construct a distribution function using these tails and completing the center with a rectangular distribution. This distribution function is in effect a function of a "created" random variable and has the property that the probability of selecting a sample between certain limits from this distribution is equal to the probability that p lies within that interval, or

$$(5) \qquad P(\hat{p}_1 \leq p \leq \hat{p}_2) = \int_{\hat{p}_1}^{\hat{p}_2} f(\hat{p})\, d\hat{p}.$$

We shall call this function $f(\hat{\beta})$, and the variable, of course, is $\hat{\beta}$ which is now a continuous variable ranging from 0 to 1. It is now possible to generate values of $\hat{\beta}$ and substitute them into the reliability equation.

Once specific values of $\hat{\beta}_1$ and $\hat{\beta}_2$ are substituted into equation (5), the relationship, of course, becomes nonsense since obviously p either lies within specific limits, or it does not. However, although (5) is not in this sense a true probability, it is a measure of confidence in the truth of the statement on the left of (5); and, therefore, the relationship is considered to be sufficient justification for the use of the distribution described above.

Actually equation (2) may be used to select samples of $\hat{\beta}$ ranging from 0 to $(\hat{R} - 1)/n$ and equation (1) may be used for samples of $\hat{\beta}$ ranging from $(\hat{R} + 1)/n$ to 1. The interval from $(\hat{R} - 1)/n$ to $(\hat{R} + 1)/n$ may be covered by a rectangular distribution whose height must be determined so that the area under the entire distribution will be 1. Figure 1 is an example of such a distribution function for a particular value of $\hat{R}$ and n.

This distribution is actually the combination of two beta distributions and a rectangular distribution. It is a result of the fact that the beta distribution integrates to the binomial.

e. g. $$\frac{(\alpha + \beta + 1)!}{\alpha!\ \beta!} \int_0^y t^{\alpha} (1 - t)^{\beta}\ dt = F(y; \alpha, \beta)$$

$$= -\sum_{i=0}^{\alpha} \binom{\alpha + \beta + 1}{i} \cdot y^i (1 - y)^{\alpha + \beta + 1 - i} + 1.$$

Using standard notation the binomial distribution may be written

(3) $$\sum_{x=0}^{\hat{R}} \binom{n}{x} p^x (1 - p)^{n-x} = 1 - F(p; x, n-x-1).$$

Similarly since

$$\sum_{\hat{x}}^{n} \binom{n}{x} p^x (1 - p)^{n-x} = 1 - \sum_{0}^{\hat{x}-1} \binom{n}{x} p^x (1 - p)^{n-x}$$

then

(4)

$$\sum_{\hat{x}}^{n} \binom{n}{x} p^x (1 - p)^{n-x} = F(p; \hat{x}-1, n-\hat{x}).$$

Therefore, equation (4) may be used for the lower tail of the distribution using only that portion of the curve from 0 to $(\hat{x} - 1)/n$, and equation (3) may be used for the upper tail using only that portion of the curve from $(\hat{x} + 1)/n$ to 1. The interval from $(\hat{x} - 1)/n$ to $(\hat{x} + 1)/n$ will be covered by a rectangular distribution whose height will be determined so that the total area under the distribution will be equal to one. (If only the lower limit on R is desired, equation (4) may be used for the entire distribution; similarly, if only the upper limit is desired, equation (3) may be used.) Therefore, a distribution can be set up for each component or part depending upon the number tested and the observed number of successes.

In the event that $\hat{x} = 0$ or that $\hat{x} = n$ for a large percentage of the components, it is possible that the point estimate of reliability ($\hat{R}$) will fall outside the confidence interval. This results from the binomial distribution where, for instance, if $\hat{p}' = 1$ any two-sided confidence interval (except where $\alpha = 1$) will not include $\hat{p}'$. Should this situation occur, only a one-sided interval should be given. (It is quite possible that a one-sided interval may in any event be of more interest than a two-sided interval.)

Through the use of high speed digital computers each of these distributions can be randomly sampled and thousands of values of $\hat{p}_i$ can be generated for the i^th component. If, for instance, 1,000 values of $\hat{p}_i$ are randomly selected for each $\hat{p}_i$ in the reliability equation, and if these values are randomly placed in that equation, then 1,000 values of $\hat{R}$ can be generated.

A function of $\hat{R}$ can be approximated by plotting these 1,000 values on a histogram. Confidence limits can then be placed on R either by fitting a curve to the histogram or by using the generated values of $\hat{R}$ directly. The sampling error involved can be decreased by increasing the number of values generated. It is no problem for the BRLESC (Ballistic Research Laboratories Electronic Scientific Computer) to generate even 10,000 values of $\hat{p}_i$ for each distribution.

A number of trial runs for the previous example have been generated by the computer and Figure 2 shows sample histograms that have resulted for both individual $\hat{p}$'s and for $\hat{R}$ using $\hat{x} = 90$ and $n = 100$. Figure IIIa and IIIb are histograms generated where $\hat{x} = 12$ and $n = 20$ and where R represents the reliability of a simple system of three components in parallel. It is anticipated that the system can be improved in the future by programming the machine to fit curves to the histogram and to integrate under these curves in order to produce confidence intervals.

CONCLUSION.

1. The objective of flexibility has been satisfied in that a point estimate of reliability and a confidence interval about that estimate can be obtained for almost any system regardless of the number of tests performed on the individual components or of the success ratios obtained.

2. The objective of speed and simplicity have been satisfied since it is possible for the machine to construct an interval in a few minutes. The only manual operation required is to punch the estimates of p on cards and place these cards in the machine.

3. It is understood that certain liberties have been taken in creating a random variable such as $\hat{p}$. However, it is felt that this assumption is justified by the results and by equation (5).

4. Obviously the distribution used for these estimates is not the best, particularly in the vicinity of $\hat{x}/n$. As is generally the case, this difficulty is most pronounced for smaller sample sizes. However, the important portions of these distributions are instrumental in determining the tails of the distribution of $\hat{R}$. Therefore, in only rare cases would the rectangular portion of the distribution be important. (Special instructions to the machine are anticipated where $\hat{x} - 1 = n$ or where $\hat{x} = n$. These instructions will eliminate the rectangular distribution for these cases.)

ACKNOWLEDGEMENT. The author wishes to express his appreciation to Mr. Warren Wenger of the Surveillance Group and to Mr. Glen Beck and Mr. John Clouse of the Computing Laboratory for their valuable assistance in completing this study.

Mr. Wenger acted as a liaison between the statistical and machine considerations of the problem and offered many helpful suggestions. Mr. Beck and Mr. Clouse programmed the problem for the machine computations.

FIGURE I

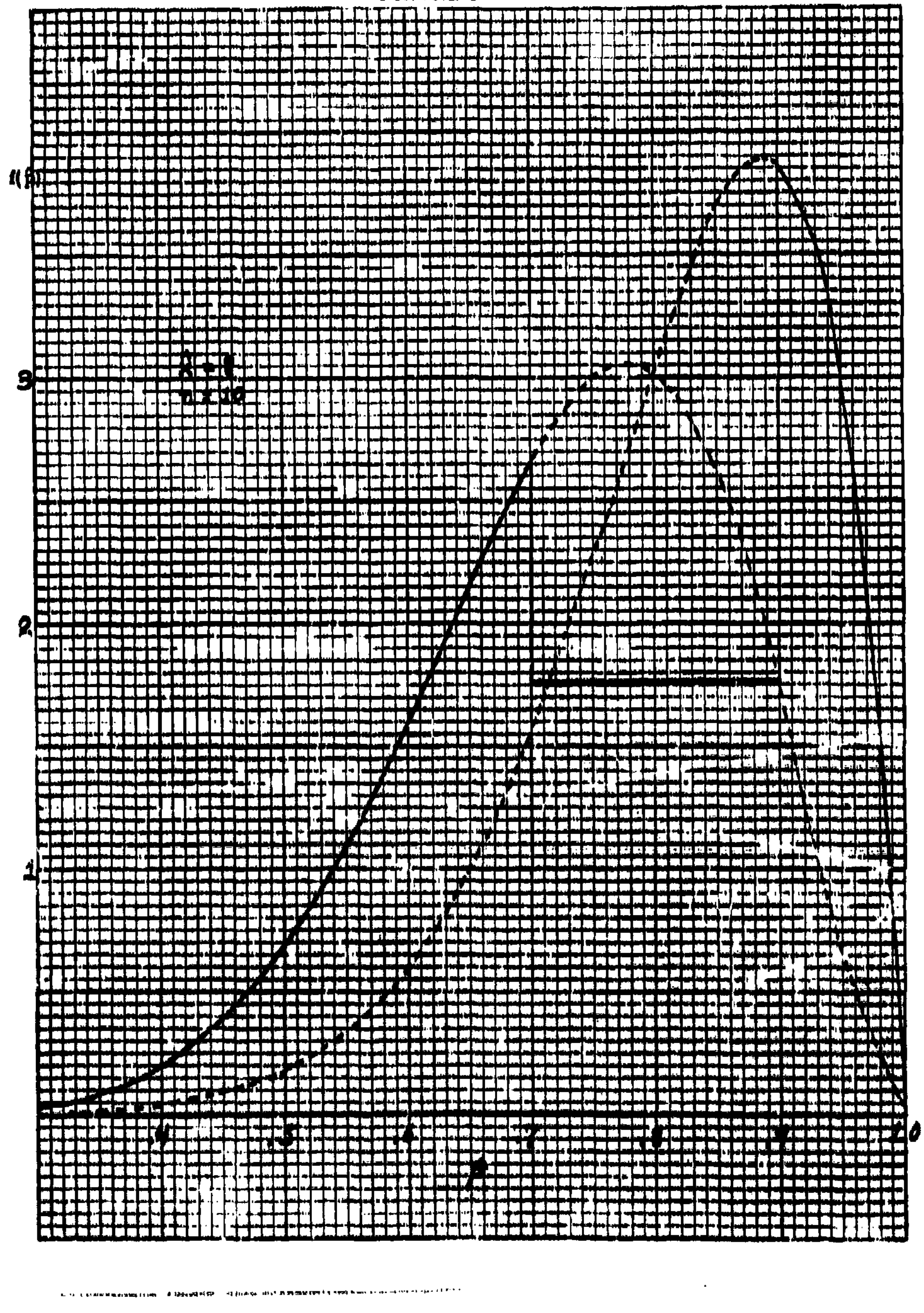


FIGURE II

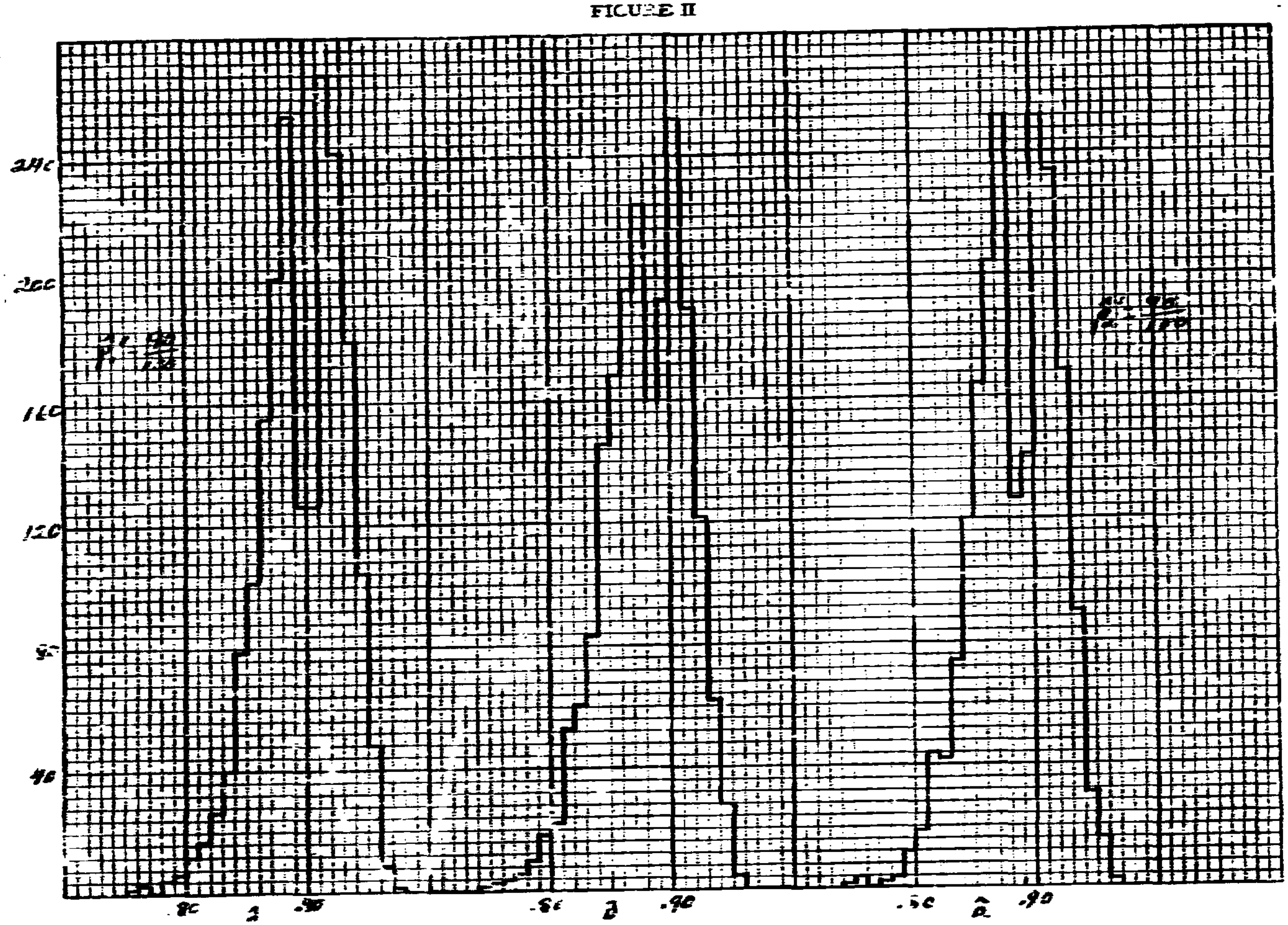


FIGURE IIIa

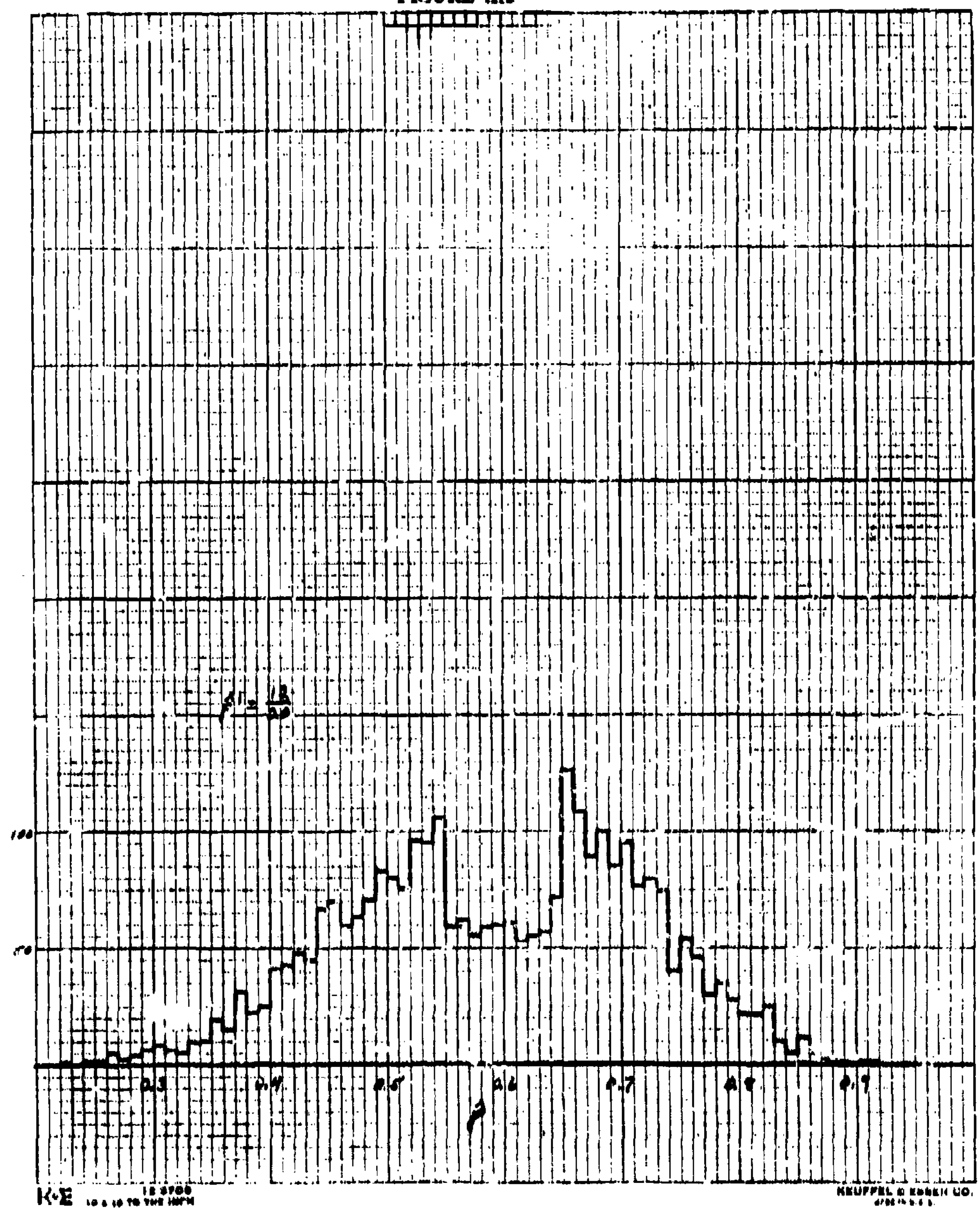
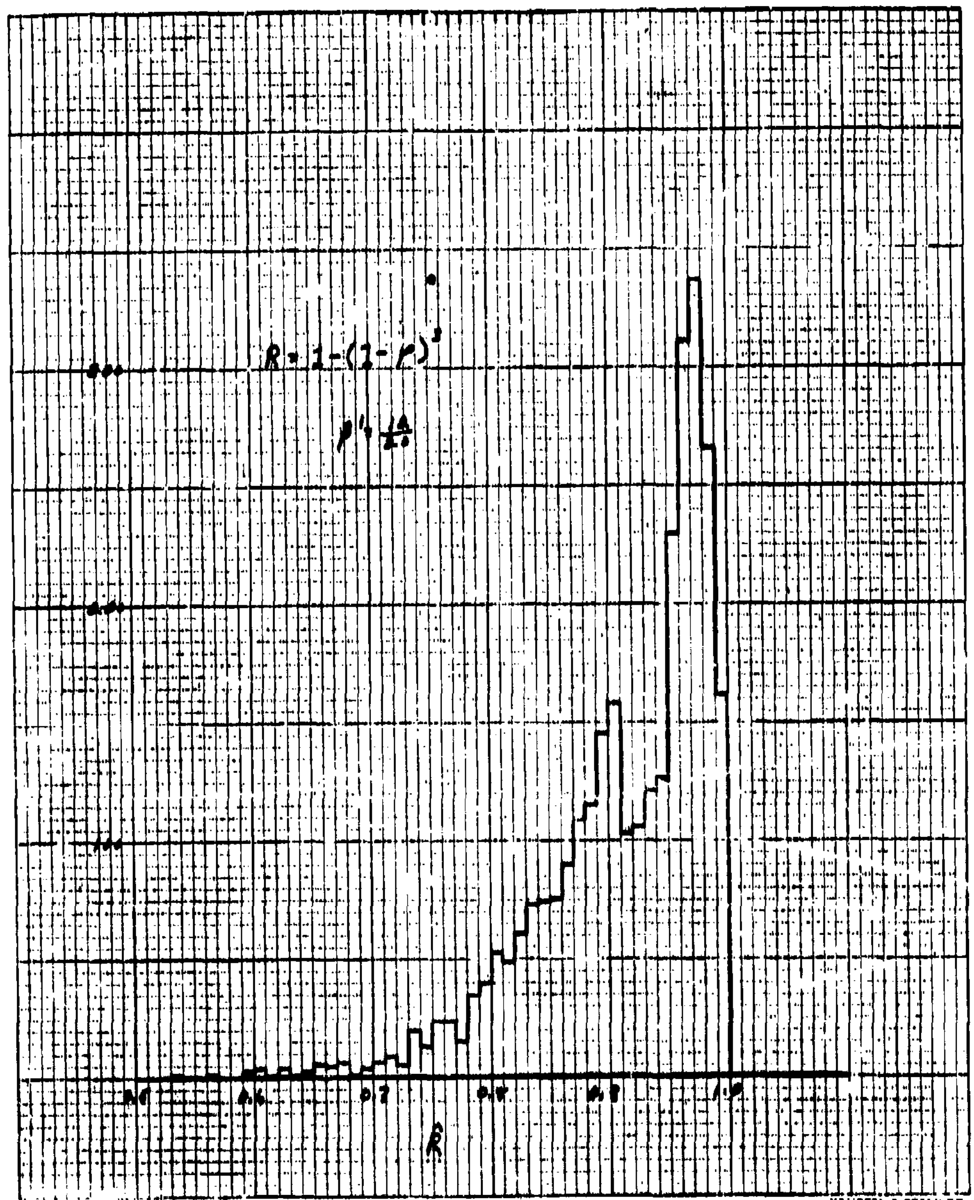


FIGURE IIIb



STATISTICAL STUDY OF AGING CHARACTERISTICS OF ARTILLERY MISSILES

Raymond Bell
Surveillance Group, Ballistic Research Laboratories,
Aberdeen Proving Ground

INTRODUCTION. Among the various methods employed by the Surveillance Group in investigating the aging characteristics of Artillery Missiles are rocket motor static tests, chemical and physical test of propellant, fully instrumented proving ground flight tests and evaluation of service practice firings.

As indicated, evaluation of service practice firings is one of the methods employed by the Surveillance Group in investigating aging characteristics. These firings are conducted by troop units for training purposes. Although these firings are not fully instrumented flights, sufficient data is obtained for performance analysis purposes.

This paper presents a typical analysis of service practice firing results and indicates the significance of these results in the Surveillance Program. An example of the evaluation of the annual service practice firings for the Honest John Rocket will be presented.

DISCUSSION. In this evaluation 934 firings of Rocket 762MM: M31 Series, conducted for troop training and other purposes by both United States and NATO firing units have been considered. The purpose of this study was to investigate the overall accuracy performance of the M31 rocket system when fired by troop units and to establish if there is any indication of a deterioration of this accuracy performance with increasing age of the M6 series rocket motors of these M31 series rockets.

In investigating the effect of age on accuracy performance, two possible types of changes were considered: first, changes in the center of impact (C.I.) of the rockets as a function of age and second, changes in the dispersion of the rockets about the C.I. as a function of age. For example, if the total impulse of the rocket motor which is a measurement of the total thrust of the rocket decreased with increasing age, then the mean range of the rocket should also decrease with increasing age. It is also possible that as the age of the rocket motor increases, its performance could become more erratic thus increasing the range dispersion of the Honest John Rocket. Although the effect of age on both range and deflection miss-distances was investigated, in general it is expected that deterioration of rocket motor performance will be reflected primarily in changes in the range performance of the M31 series rocket. In order to investigate the effect of age, simple and multiple linear regression techniques and analysis of variance techniques were used.

In investigating the effect of age on mean range and deflection performance, all firings were grouped according to the type of burst and type of launcher system employed. Separate analyses were carried out for each of these groups for both range and deflection miss-distances. The regression model used in these analyses assumes range and deflection miss-distances for each firing are linear functions of the age of the rocket motor and the launcher to target range. The launcher to target range was included in the regression model since any age effects would probably be range dependent.

In the division of the rockets according to type of burst and type of launcher system, the majority of the rockets (848 of the 924 considered) fell into three main groups. These were firings with the M289 launcher for air burst (211 rockets fired), M386 launcher for air burst (473 rockets fired) and the M289 launcher for ground burst (164 rockets fired). In addition, there were 49 rockets fired with the M33 launcher for air burst, 21 firings with the M386 launcher for ground burst and 6 firings with M33 launcher for ground burst. The regression analyses which were carried out with all these groups except for the firings with the M33 launcher for ground burst due to only six rockets being fired in this group indicated that after 7 1/2 years of shelf life there appears to be a significant age effect. This is indicated by a significant age effect (.01 level) in the analysis of range miss-distance in the M386 launcher firings for air burst. Although none of the other groups indicated a significant age effect at the .01 level, the other two large groups of rockets were near significance at this level. For example, for the group containing rockets fired with the M289 launcher for ground burst $t = -2.25$ which is not significant at the .01 level as $t_{.995} = 2.61$ but is significant at the .05 level where $t_{.975} = 1.97$. With the group containing rockets fired from M289 launcher for air burst $t = -1.95$ which is just short of the .05 level of significance. Thus, it appears that there is an indication of a decrease in the mean range performance as a function of age of the rocket motor of the M31 rocket. No significant (.01 level) age effect was found in the analysis of deflection miss-distance as a function of age.

In the analysis of the M386 launcher firings for air burst in which a significant (.01 level) age effect was indicated, a significant (.01 level) launcher to target range effect was also noted. To further investigate these significant results the rockets within this group were further subdivided into four launcher-to-target range groups. These groups were rockets with a launcher-to-target range of $<$10,000 meters, 10,000 to 14,000 meters, 14,001 to 18,000 meters and $>$18,000 meters. A regression analysis of range miss-distance as a linear function of rocket motor age was performed on each of the four launcher-to-target range groups. The regression analysis for the 10,000 to 14,000 meter group which contained the largest number of rockets fired indicated a significant (.01 level) age effect. A scatter diagram of the range miss-distance for the four launcher-to-target range groups is presented on Figures 1 thru 4. Figures 1, 3 and 4 indicate the range miss-distances versus the M6 series motor age for the three range groups, $<$10,000 meters, 14,001 to 18,000 meters and $>$18,000 meters in which no significant (.01 level) age effect is indicated. Since there was no significant age effect for these three groups the mean miss-distance for each group is also indicated on these figures. A scatter diagram of the range miss-distances versus the M6 series motor age for the rockets fired in the 10,000 to 14,000 range group is presented on Figure 2. As previously indicated, a significant (.01 level) age effect was noted in this group. Indicated on this figure is the regression line derived from the regression analysis performed with this group. This regression line indicates that the average range of each rocket for the period covered has been decreasing at the rate of 18.6 meters/year such that the average range miss-distance for 7 1/2 year old motors was -80 meters (80 meters short of the target).

As previously indicated, in addition to an age effect with the M386 launcher firings for air burst, a significant launcher-to-target range effect was noted. As indicated on Figures 1, 3 and 4 the average range miss-distance was -50 meters for the $<$10,000 meter range group, + 27 meters for the 14,001 to 18,000 meter group and + 104 meters for the $>$18,000 meter group. These results seem to indicate that for the shorter range targets the rockets are falling short of the target and as the target range increases the rockets begin going over the target so that at the long ranges ($>$18,000 meters) the rockets are falling on the average 104 meters beyond the target. This occurrence is probably due to a small bias in the firing table. Table I contains a summary of the reported miss-distances for the M386 launcher firings for air burst. These firings are grouped according to age and according to the range groups previously mentioned. On Table I, it is noted that for rockets fired at ranges $<$ 10,000 meters the reported mean miss-distances for all ages are negative except the 6 to 7 year old rockets. This is also indicated in the 10,000 to 14,000 meter group except in this group the 3 to 4 year old rockets mean miss-distances are positive. Whereas in the 14,001 to 18,000 meter group all the mean miss-distances are positive with the exception of the 4 to 5 year old rockets. Similarly in the $>$ 18,000 meter group all the mean miss-distances are positive.

In investigating the possible effect of age on range and deflection dispersion regression techniques were also employed. In order to remove the dependence of the variance of cells on the true population variance of the cell a logarithmic transformation of the range and deflection sample variance was made prior to carrying out the regression analyses. Due to the limited number of ground impact firings, these analyses were limited to the air burst rounds. Separate analyses were carried out for each calendar year's firings with each launcher as well as a combined analysis over all calendar years for each launcher. These analyses disclosed no significant (.01 level) age effect for either range or deflection dispersion.

Tables 2 and 3 present a summary of the means and standard deviations of range ($\bar{R}$, S_R), deflection ($\bar{D}$, S_D) and height of burst ($\bar{H}$, S_H) miss-distances. Table 2 summarizes those rounds fired for air burst and Table 3 summarizes those rounds fired for ground impact. Each table contains a separate summary for the firings conducted during each calendar year and a combined summary over all firings. The firings for each calendar year are grouped according to the age of the M6 series motor in one-year intervals. These groupings are further subdivided according to the type of launcher system employed, i.e., M289, M386 and M33. Prior to computing the sample means and standard deviations in these tables, the consistency of the data was investigated by testing the extreme values using a criterion for outlying observations. Any outlying values (at the .01 level of significance) were not used in computing the means and standard deviations of the miss-distances, but were tabulated individually as footnotes to the tables.

SUMMARY. In summarizing, the results of these analyses have indicated that there appears to be a significant age effect on the functioning results of the M31 Series Honest John Rocket. This is indicated by the regression analyses carried out with the M386 launcher firings for air burst and with the M289 launcher firings for ground and air burst. With the M386 launcher firings for air burst, the average range appears to be decreasing at the rate of 18.6 meters/yr such that for 7 1/2 year old motors the average range miss-distance is -80 meters. It should be pointed out that the decrease in range of 18.6 meters/yr is considered as the rate of decrease for rockets 1 to 7 1/2 years old and should not be construed as indicative of a rate of decrease for rockets older than 7 1/2 years. Extrapolation of this rate of decrease to older motors may be hazardous as the rate may change as the motors age. It might also be pointed out that the decrease in range observed for older motors appears to be substantiated by rocket motor static tests. It was also noted with this group of firings that there was a significant launcher-to-target range effect. This effect appears to indicate that the shorter range rockets (< 14,000 meters) are impacting short of the target while the longer range rockets (> 14,000 meters) are impacting beyond the target. This appears to be due to a small bias in the firing table.

CONCLUSION. In conclusion, we might consider the effect of the significant age and launcher to target range effect on the current status of the M31 stockpile. First, it should be pointed out that the bulk of the stockpile consists of rockets that vary in ages from 3 to 6 years. For rockets of these ages, with the age effect indicated by the M386 launcher firings for air burst an average miss-distance of 0 to -50 meters due to age may be expected. Thus, the bulk of the stockpile does not as yet appear to be adversely affected by this age effect.

If at some later date, it is determined that the age effect is reducing the effectiveness of the Honest John Rocket System it would be possible to correct for this age effect by establishing a propellant weight correction factor to be used with these rockets.

In relation to future analyses, the Surveillance Group is continually reviewing all service practice firings for inclusion in these analyses. As mentioned earlier, the evaluation of service practice firings is only a part of an integrated program which has been initiated to determine the reliability and ballistic characteristics of the artillery missile stockpile.

FIGURE 1

Rocket, 762MM: M31 Series - M386 Launcher
Range-Miss Vs. M6 Series Motor Age
Range < 10,000 Meters

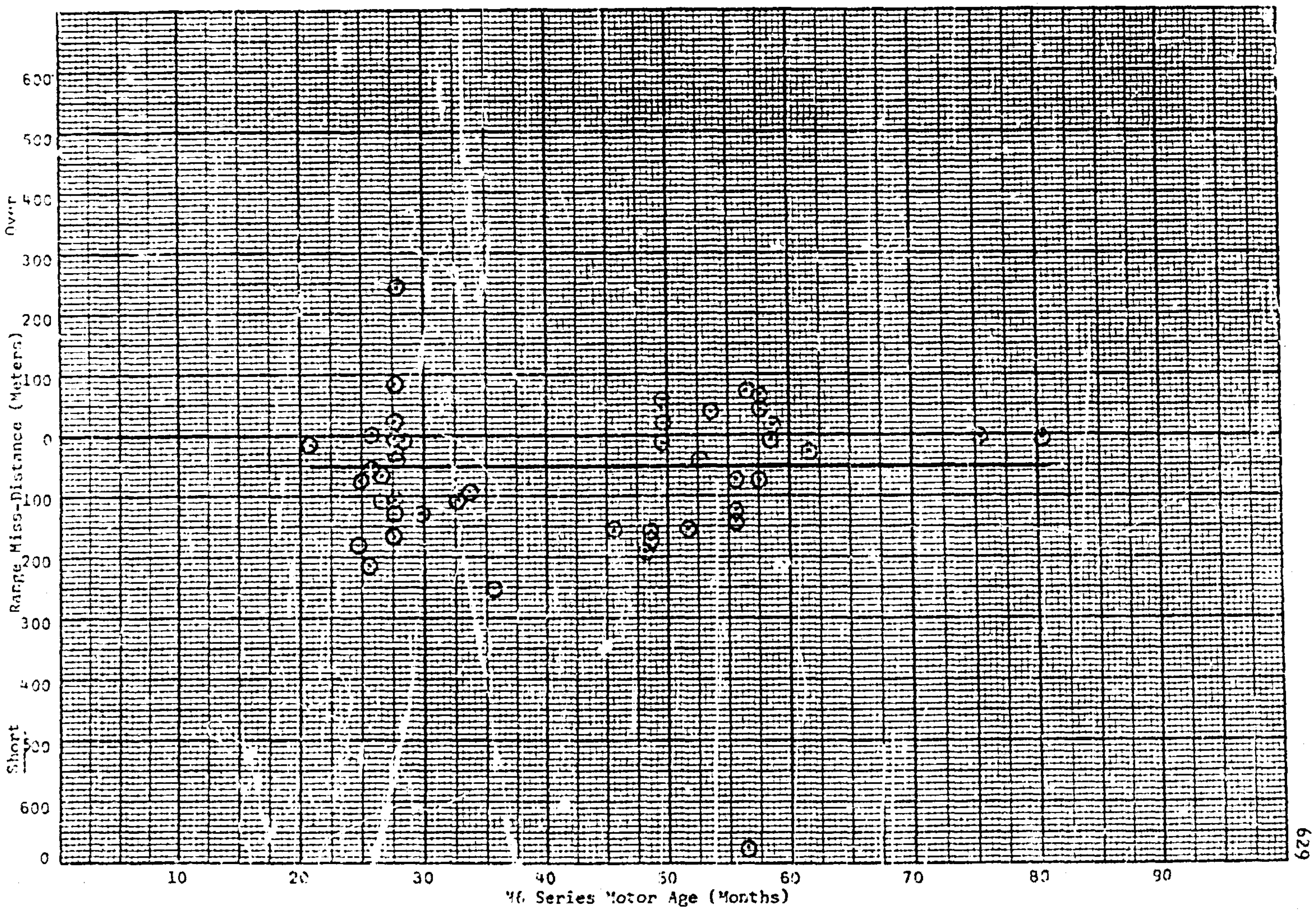


FIGURE 2

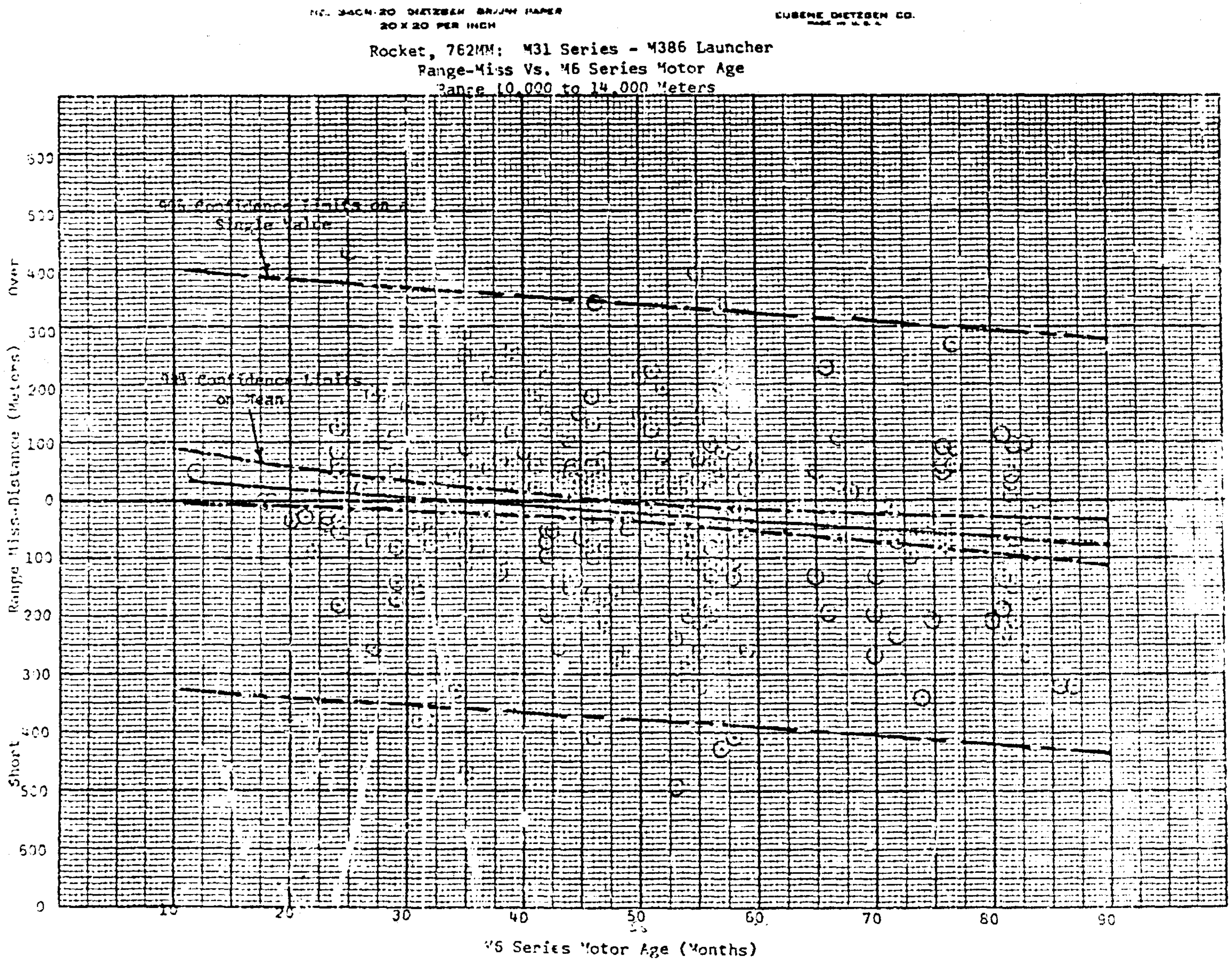

NO. 340N-20 DIETZGEN GRAPH PAPER
20 X 20 PER INCH
EUGENE DIETZGEN CO.
Rocket, 762MM: M31 Series - M386 Launcher
Range-Miss Vs. M6 Series Motor Age
Range 10,000 to 14,000 Meters
Single Value
on Mean
Over
Range Miss-Distance (Meters)
Short
600
500
400
300
200
100
0
100
200
300
400
500
600
0
10
20
30
40
50
60
70
80
90
M6 Series Motor Age (Months)

FIGURE 3

Rocket, 762MM: M31 Series - M386 Launcher
Range-Miss Distance Vs. M6 Series Motor Age
Range 14,001 to 18,000 Meters

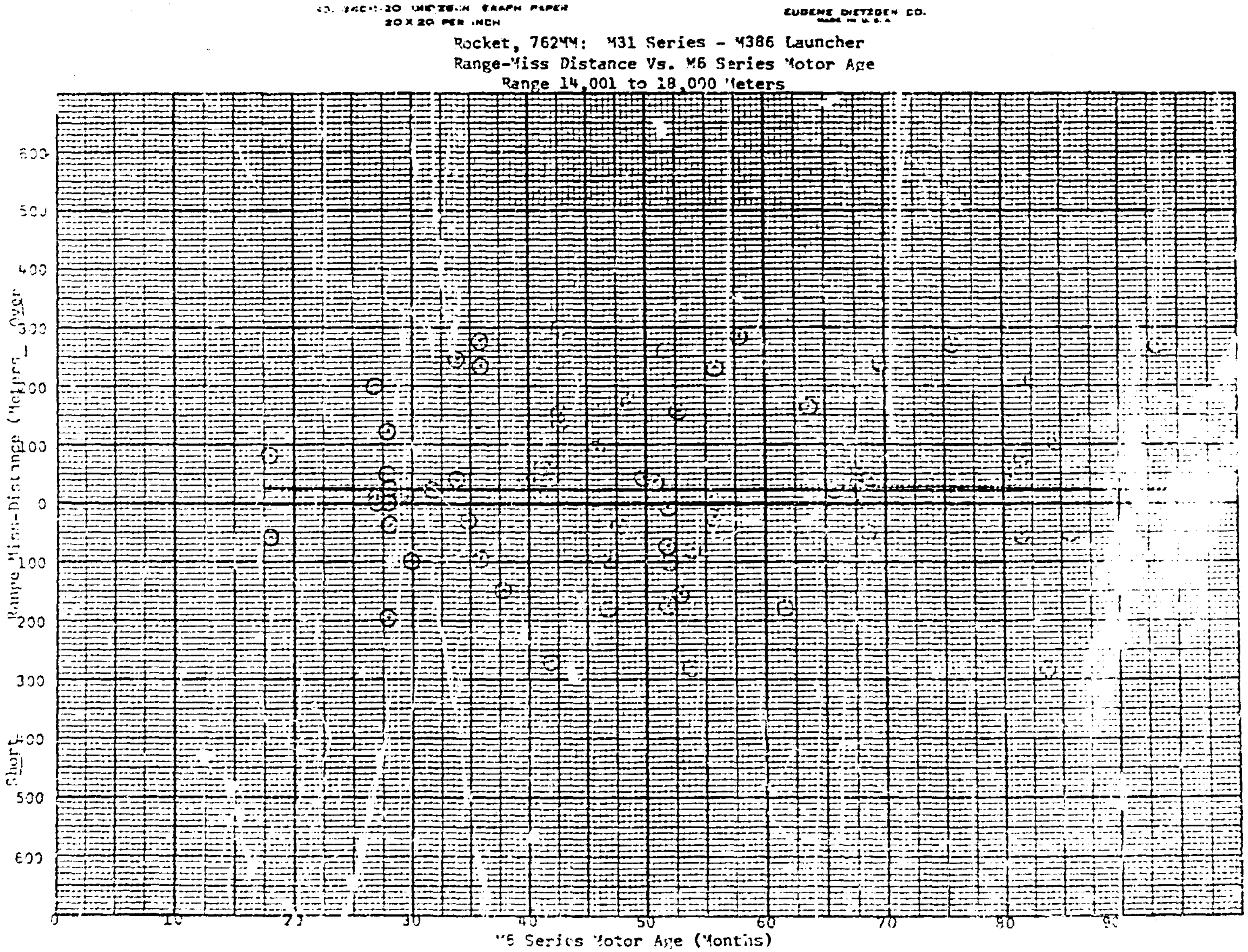


FIGURE 4

Rocket, 762MM: M31 Series - M386 Launcher
Range Miss Vs. M6 Series Motor Age
Range > 18,000 Meters

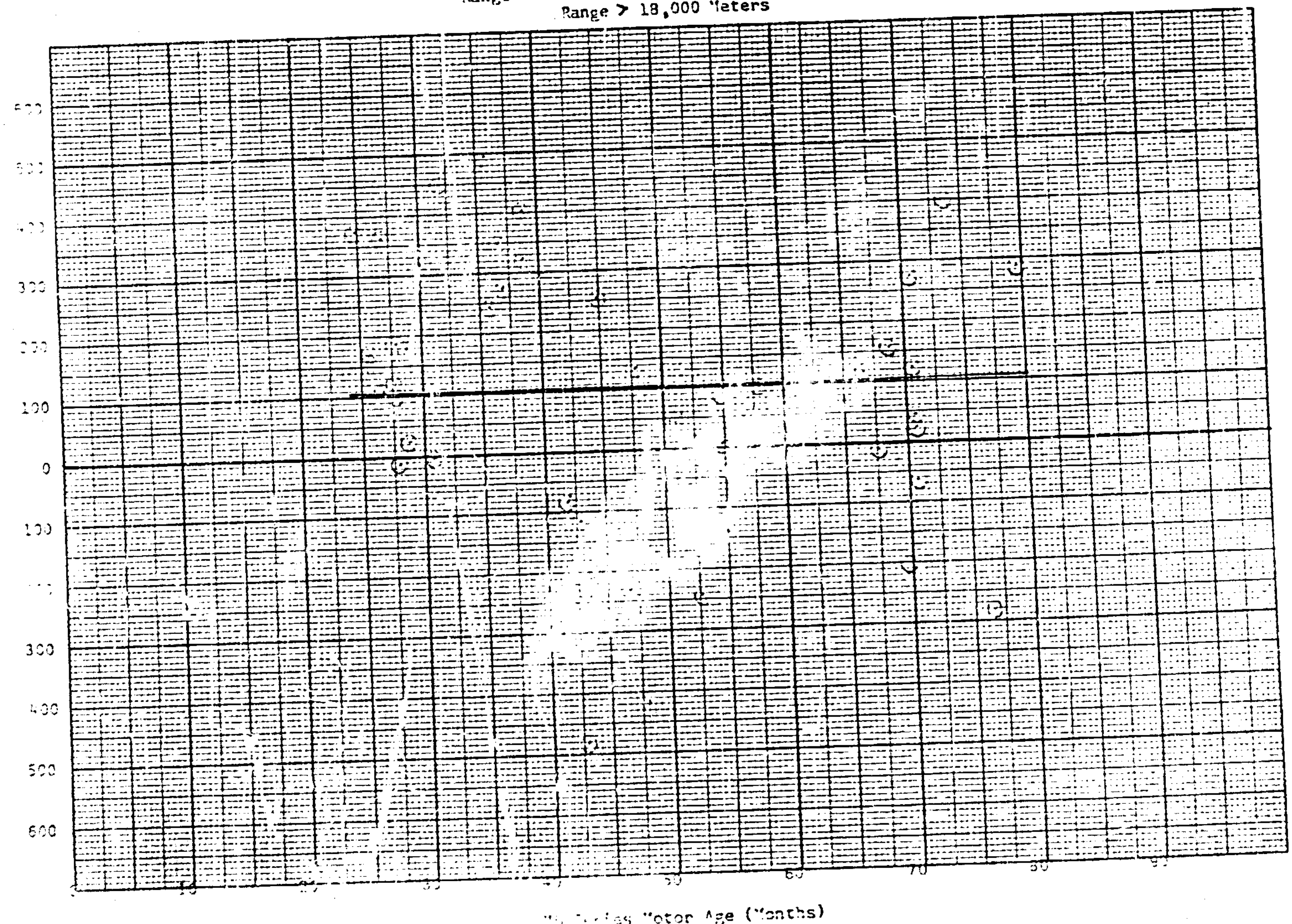


TABLE E

Summary of Reported Miss-Distances (meters) for Troop Firings of Rocket, 762MM: M31 Series

Air Bursts - M386 Launcher

Launcher to Target Range (Meters)	Parameter	Age of M6 Series Motor at Time of Firing: 1 yr. but under 2 yr.	2 yr. but under 3 yr.	3 yr. but under 4 yr.	4 yr. but under 5 yr.	5 yr. but under 6 yr.	6 yr. but under 7 yr.	7 yr. but under 8 yr.
≤ 10,000	n	1	18	2	18	1	2	
	$\bar{R}$	-10	-54	-199	-40	-20	+3	
	S_R		107	69	92		4	
	$\bar{D}$	L160	L22	R451	R16	L70	R151	
	S_D		140	380	123		69	
	$\bar{H}$	H156	H16	L42	L31	L2	L7	
	S_H		45	10	71		55	
	No. outliers	-	3[a]	0	1[b]	-	0	
10,000 to 14,000	n	8	47[d]	77[f]	114	20	39	5
	$\bar{R}$	-17	-19	+21	-28	-25	-54	-182
	S_R	37	170	121	144	115	135	136
	$\bar{D}$	R15	R13	L10	L31	L81	L17	R151
	S_D	267	179	170	198	190	161	253
	$\bar{H}$	H25	L12	H19	L1	L19	0	L45
	S_H	87	89	64	82	112	74	59
	No. outliers	1[c]	2[e]	1[g]	7[h]	0	0	0
14,001 to 18,000	n	2	14	16	18	9	5	4
	$\bar{R}$	+10	+23	+36	-16	+73	+103	+8
	S_R	99	113	165	160	171	129	243
	$\bar{D}$	R88	L66	L56	R21	R98	L80	R8
	S_D	39	121	233	255	249	86	427
	$\bar{H}$	H90	L13	L38	L48	H15	H44	L1
	S_H	6	115	130	113	55	145	226
	No. outliers	0	0	0	0	0	0	0
> 18,000	n		13	12	7	9	3	
	$\bar{R}$		+171	+106	+26	+55	+138	
	S_R		140	311	126	139	366	
	$\bar{D}$		L8	R225	R54	R50	R43	
	S_D		214	462	232	358	209	
	$\bar{H}$		L26	H87	L5	L34	L7	
	S_H		198	196	132	104	290	
	No. outliers		1[f]	0	0	0	0	

TABLE 1 CONT'D

n = number of rockets
$\bar{R}$ = average range miss-distance
S_R = standard deviation of range miss-distances
$\bar{D}$ = average deflection miss-distance
S_D = standard deviation of deflection miss-distances
$\bar{H}$ = average height of burst miss-distance
S_H = standard deviation of height of burst miss-distances
No. outliers-number of outliers based on W. J. Dixon's Criteria (.01 level of significance)

a. Range Outlier (+477 meters)
   Deflection Outlier (R670 meters)
   Height-of-burst outlier (High 408 meters)
b. Range Outlier (-670 meters)
c. Range Outlier (+1099)
d. Mean height-of-burst miss-distance based on a sample size of 43. Height-of-burst miss-distances were not reported for four rockets
e. Range Outlier (+1356 meters)
   Deflection Outlier (L809 meters)
f. Mean height-of-burst miss-distance based on a sample size of 76. The height-of-burst miss-distance was not reported for one rocket.
g. Height-of-burst outlier (High 352 meters).
h. Deflection outlier (R1323 meters)
   Six Range Outliers (-1500 meters, +820 meters, +658 meters, +840 meters, +823 meters, +860 meters)
i. Deflection Outlier (R1269 meters)

TABLE II

Summary of Reported Miss-Distances (meters) for Troop Firings of Rocket, 762MM: M31 Series

AIR BURSTS

Year Fired	Parameter	Age of M6 Series Motor at Time of Firing: Under 1 yr.			1 yr. but under 2 yr.			2 yr. but under 3 yr.			3 yr. but under 4 yr.			4 yr. but under 5 yr.			5 yr. but under 6 yr.			6 yr. but under 7 yr.			7 yr. but under 8 yr.		
		M289	M386	M33	M289	M386	M33	M289	M116	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33
1958	n				1		2	7		17	2														
	$\bar{R}$				-365		+276	-157		-45	+98														
	S_R						329	285		257	224														
	$\bar{D}$				R45		L16	L140		L108	L27?														
	S_D						497	217		333	756														
	$\bar{H}$				H147		L75	H10?		L13	L164														
	S_H						339	205		226	197														
	No. Out				-		0	0		0	0														
1959	n	8			11	5		14	32	2	12	11	5	8	10										
	$\bar{R}$	-5			+12	-22		-50	+11	+430	-73	-62	+275	-150	-39										
	S_R	205			281	13		151	135	13	146	196	260	167	125										
	$\bar{D}$	L41			L8	L83		L35	L6	R134	R83	R37	L123	L124	L158										
	S_D	224			179	159		186	189	278	167	261	218	145	246										
	$\bar{H}$	L14			H61	L9		H25	H37	H78	H6	L53	H5	L19	L61										
	S_H	87			150	41		111	75	165	113	108	36	69	109										
	No. Out	0			0	1[a]		0	1[b]	0	0	0	0	0	2[c]										
1960	n	7			17	3			36	1	1	19	3	1	76	10	1	11							
	$\bar{R}$	-66			-78	+23			+46	-50	-117	+59	+470	-154	-8	+131	-137	+39							
	S_R	113			104	74			150			167	252		207	157		91							
	$\bar{D}$	L131			L39	L28			L81	R90	R143	R9	R77	L153	L6	R143	L18	R7							
	S_D	139			134	106			174			238	137		182	118		269							
	$\bar{H}$	L49			H12	H14			147	L19	H68	H65	L70	L23	L27	H103	L24	L29							
	S_H	39			71	179			143			115	186		88	51		95							
	No. Out	0			0	0			1[d]	-	-	0	0	-	0	0	-	0							
1961	n				24			35	4			31	2	2	40	1		19	1		17				
	$\bar{R}$				+12			-98	+15			-36	-526	-102	-12	+55		-24	+110		+52				
	S_R				203			274	139			141	217	139	225			144			139				
	$\bar{D}$				R7			L53	L90			R12	L136	R42	L24	R850		R22	L768		R10				
	S_D				122			228	268			200	448	32	231			257			167				
	$\bar{H}$				L2			L13	L67			H14	H360	H7	H11	H16		H1	L25		L10				
	S_H				95			111	72			82	410	13	[illegible]			105			86				
	No. Out				0			0	0			1[e]	0	0	0	-		0	-		0				

TABLE II (Cont'd)

Summary of Reported Miss-Distances (meters) for Troop Firings of Rocket, 762MM: M31 Series

AIR BURSTS

Age of M6 Series Motor at Time of Firing

Year Fired	Parameter	Under 1 year M289	Under 1 year M386	Under 1 year M33	1 year but under 2 years M289	1 year but under 2 years M386	1 year but under 2 years M33	2 years but under 3 years M289	2 years but under 3 years M386	2 years but under 3 years M33	3 years but under 4 years M289	3 years but under 4 years M386	3 years but under 4 years M33
1962	n				1	3		31	13		18	45	
	$\bar{R}$				+120	-30		-106	-54		-79	+25	
	S_R					44		217	226		103	159	
	$\bar{D}$				L30	R148		L15	R118		L42	L35	
	S_D					401		203	158		121	214	
	$\bar{H}$				H30	H58		H55	[illegible]		H34	H16	
	S_H					93		[illegible]	[illegible]		81	103	
	No. Out				-	0		[illegible]	0		0	0	
1963	n								5		1	1	
	$\bar{R}$								-46		-73	+280	
	S_R								151				
	$\bar{D}$								L186		R43	R797	
	S_D								375				
	$\bar{H}$								H115		H54	0	
	S_H								93				
	No. Out								0		-	-	
Years Combined	n	15			54	11	2	87	96	20	34	107	10
	$\bar{R}$	-34			-21	-12	+276	-99	+11	+2	-67	+26	+173
	S_R	167			201	45	329	241	165	278	126	161	440
	$\bar{D}$	L83			L11	L5	L16	L44	L14	L74	L4	R2	L16
	S_D	188			136	234	497	210	203	323	202	224	248
	$\bar{H}$	L30			H18	H9	L75	H27	L2	L4	H10	H17	H51
	S_H	69			105	92	339	127	114	213	106	103	234
	No. Out	0			0	1	0	0	2	0	0	1	0

Year Fired	Parameter	4 years but under 5 years M289	4 years but under 5 years M386	4 years but under 5 years M33	5 years but under 6 years M289	5 years but under 6 years M386	5 years but under 6 years M33	6 years but under 7 years M289	6 years but under 7 years M386	6 years but under 7 years M33	7 years but under 8 years M289	7 years but under 8 years M386	7 years but under 8 years M33
1962	n		37			5	2	5	30	3		7	
	$\bar{R}$		+17			+6	+582	-64	-54	-53		-61	
	S_R		175			115	575	117	153	123		212	
	$\bar{D}$		R1			L150	L486	R70	R7	L95		R23	
	S_D		194			189	105	322	155	213		325	
	$\bar{H}$		H2			L18	H11	L132	L6	L3		L16	
	S_H		87			41	298	163	110	21		163	
	No. Out		0			1f	0	0	0	0		0	
1963	n	4				3			2			2	
	$\bar{R}$	-155				+179			-223			-224	
	S_R	66				211			166			98	
	$\bar{D}$	L65				R95			R155			R313	
	S_D	118				214			121			289	
	$\bar{H}$	H60				L14			L9			L52	
	S_H	72				12			6			100	
	No. Out	0				0			0			0	
Years Combined	n	15	163	11	1	38	3	5	49	3		9	
	$\bar{R}$	-145	-5	+124	-137	+16	+424	-64	-24	-53		-98	
	S_R	129	201	151		137	489	117	160	123		204	
	$\bar{D}$	L88	L18	R207	L18	L9	L580	R70	R14	L95		R87	
	S_D	130	204	241		253	179	102	158	213		326	
	$\bar{H}$	H5	L13	H95	L24	L19	L1	L132	H4	L3		L26	
	S_H	69	90	55		102	212	163	99	21		147	
	No. Out	0	2	0	-	1	0	0	0	0		0	

TABLE II CONT'D

a. Range Outlier (+1099 meters).
b. Deflection Outlier (R1269 meters).
c. Range Outlier (-1500 meters).
   Deflection Outlier (R1323 meters).
d. Range Outlier (+1356 meters).
e. Deflection Outlier (R1184 meters).
f. Height-of-Burst Outlier (Low 325 meters).
g. Mean height-of-Burst Miss-Distance based on a sample size of 14. Height-of-burst miss-distances were not reported for four rockets.
h. Mean height-of-burst miss-distance based on a sample size of 44. The height-of-burst miss-distance was not reported for one rocket.
i. Mean height-of-burst miss-distance based on a sample size of 92. Height-of-burst miss-distances were not reported for four rockets.
j. Mean height-of-burst miss-distance based on a sample size of 106. The height-of-burst miss-distance for one rocket was not reported.

n = number of rockets
$\bar{R}$ = average range miss-distance
S_R = standard deviation of range miss-distances
$\bar{D}$ = average deflection miss-distance
S_D = standard deviation of deflection miss-distances
$\bar{H}$ = average height of burst miss-distance
S_H = standard deviation of height of burst miss-distances
No. outliers-number of outliers based on F. E. Grubbs' Criteria (.01 level of significance)

TABLE III

Summary of Reported Miss-Distances (Meters) for Troop Firings of Rocket, 762MM: M31 Series

GROUND IMPACTS

Year Fired	Parameter	Age of M6 Series Motor at Time of Firing																	
		1 yr. but under 2 yr.			2 yr. but under 3 yr.			3 yr. but under 4 yr.			4 yr. but under 5 yr.			5 yr. but under 6 yr.			6 yr. but under 7 yr.		
		M289	M386	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33	M289	M386	M33
1958	n	27		2	73		3	14			1								
	$\overline{R}$	+8		-366	-65		+82	-262			-346								
	S_R	276		531	310		482	164											
	$\overline{D}$	R3		L157	L9		L262	L17			L292								
	S_D	211		75	295		409	376											
	No. Outliers	1[a]		0	1[b]		0	0			-								
1959	n	4			35	6	1	4	2		2								
	$\overline{R}$	-181			-30	+36	-286	-21	-30		+70								
	S_R	201			128	158		315	424		56								
	$\overline{D}$	R200			L45	R35	L125	R44	L190		R48								
	S_D	272			332	260		89	808		24								
	No. Outliers	0			0	0	-	1[c]	0		0								
1950	n					3		1				7							
	$\overline{R}$					+145		-191				-11							
	S_R					380		45				206							
	$\overline{D}$					L112		L23				L96							
	S_D					68		103				111							
	No. Outliers					0		0				0							
1961	n	2										4			2			2	
	$\overline{R}$	+75										-111			-364			+138	
	S_R	177										260			77			80	
	$\overline{D}$	L130										L22			R276			R8	
	S_D	42										98			22			95	
	No. Outliers	0										0			0			0	
1962	n					2			2			1							
	$\overline{R}$					+505			-126			-30							
	S_R					205			310										
	$\overline{D}$					L160			L115			R310							
	S_D					198			160										
	No. Outliers					0			0										
Years Combined	n	33		2	108	11	4	20	4		3	12			2			2	
	$\overline{R}$	-11		-366	-53	+131	-9	-207	-78		-68	-46			-364			+138	
	S_R	2[illegible]7		531	315	302	434	209	308		244	210			77			80	
	$\overline{D}$	R19		L157	L20	L41	L223	L6	L153		L65	L37			R276			R8	
	S_D	22[illegible]		75	307	215	341	315	478		197	144			22			95	
	No. Outliers	1		0	1	[illegible]	0	1	0		0	0			0			0	

a. Range Outlier (+1578 meters). b. Deflection Outlier (L1860 meters). c. Deflection Outlier (L1541 meters).

THE MEASUREMENT OF THE MORALE AND SUPPRESSIVE EFFECTS OF WEAPONS*

D. H. Chaddock
C.B.E., M.Sc., M.I. Mech E.

In the last decade considerable progress has been made in assessing at an early stage the probable performance of proposed weapon systems in terms of their chance of a hit, and given a hit, the extent of the damage they will inflict. Information of this kind is frequently used to compare one weapon system with another or to decide whether a particular project is worth pursuing further. The methods used are mainly statistical and are based upon measurements made during controlled trials, or extrapolation from such data.

In so far as physical damage to equipment or men is concerned the basic data, although not all that could be desired, is sufficient for a reasonably satisfactory assessment to be made of the overall effectiveness of the weapon system. In cases where this is high (for example, over 50%) it is not perhaps an unreasonable assumption that the system must possess some value, and other things being equal, should receive further development. In other cases, due either to the extreme difficulty of the task that has been proposed or limitations imposed on the assessment (for example, in weight and size) it is not possible to show that a proposed system would have more than a very small effectiveness, perhaps of the order of a few percent. In these cases it is natural to reject the proposal as valueless for military purposes.

A decision of this sort, while valid in relation to the material damage which is being discussed, admittedly completely ignores any other effect which the use of the weapon might produce. Typical examples, well known to Service personnel are --

1. "Raising the morale" from the point of view of the man firing the weapon.
2. "Deterring the attacker" from the point of view of the man being fired at.
3. "Keeping their heads down" from both points of view.

A brief examination of three typical cases might serve to illustrate the type of situation in which, although the probability of material damage

---

* This paper by D. H. Chaddock, Director of Artillery Research and Development, The War Office, United Kingdom was presented at the conference by G. F. Komlosy.

might be small, the "Morale" value of delivering an attack could be quite high.

INFANTRY ANTI-TANK WEAPONS. Because of limitations in weight and size most infantry anti-tank weapons, such as the Bazooka, the Panzerfaust and the Carl Gustav, deliver their attack on impact by means of a hollow charge warhead. The hollow charge produces a high velocity jet which, although it can penetrate a considerable thickness of steel, produces only a relatively small hole. Inside the target the remnants of the jet and the fragments from the hole are confined to a relatively narrow cone having an axis in prolongation of the light path. Within the cone the fragments have considerable damaging power, which can be readily assessed in terms of material and men. Outside the cone the results appear to be very disappointing, even when anesthetised animals are subjected to attack.

Therefore, scoring only on the measurable physical damage, assessments lead to a rather low probability of "killing" a tank by this means. However, such assessments are not universally acceptable. French experts, for example, consider that when a tank is penetrated by a hollow charge, the crew will be so demoralized that they will be unable to continue to fight. Direct evidence of this is hard to come by. Some British crews report being unaware of having been hit until after the battle when they have discovered a hole in their tank.

ATTACK OF LOW FLYING AIRCRAFT. Ever since aircraft flew in battle there has been a problem of attacking them from ground based weapons. Owing to the fundamental difficulties of detecting, sighting and aiming at a moving target, and in spite of a great deal of ingenuity expended in the development of sighting systems and weapons, the problem has never been satisfactorily solved. Current assessments, for example, show that for a simple 20mm machine gun, eye sighted and manually laid, the chances of successfully engaging a low-flying target are very small even under favorable circumstances. Technically sophisticated solutions involving the use of radar for early warning, range finding, etc. lead to weapon systems too complicated to envisage as an all-arms anti-aircraft weapon.

The current range of armoured fighting vehicles are therefore being produced without any means of defending themselves from low flying aircraft, on the grounds that even if they had a weapon it would be ineffective. This situation, although accepted, is causing some disquiet. Users with battle experience remember the times when they fired their weapons at inquisitive enemy aircraft and believe they deterred them from carrying out the purpose of their sortie by doing so. The reaction of aircraft pilots to being fired at no doubt varied, but the general impression seems to be that although "flak" from heavy anti-aircraft guns was frightening to unseasoned crews, and a nuisance, it was never a real deterrent. Small arms fired from the ground seems to be ignored by most pilots. This might not be the case, however, with relatively unprotected and slow-moving helicopters.

ARTILLERY SUPPORTING FIRE. One of the classical roles of Artillery is to provide supporting fire for other arms in both defence and attack. In general this involves bringing indirect fire to bear upon an area target. Except in certain specific cases, for example counter battery fire, no attempt is made to engage individual targets within the area, the volume, rate, and weight of fire being judged as likely to produce some casualties and general disruption of the enemy's activities. In the past, when Infantry could only move by exposing themselves, they were vulnerable to attack by high explosive shell fragments, particularly when VT fuses were used, and the efficacy of this sort of fire is unquestioned. Its very efficacy has lead to the development of the armoured personnel carrier and self propelled and armoured guns, in whose military specification there is a requirement that they shall protect their occupants from the effects of near burst, high explosive shells. Assessment therefore of the material damage likely to be inflicted on a highly mechanised army, such as the Russian Army, by indirect fire with conventional weapons is depressingly low. The user, however, is not prepared to accept the conclusion that his fire is therefore ineffective and points to the morale effect of subjecting the troops to fire of any sort. In this it is a widely held belief that a single large explosion is very much more effective than the same weight of fire delivered in a number of smaller explosions--the philosophy of the "big crump".

These are typical cases, and in them and others it is quite obvious that experienced users believe that their weapons can and do produce an effect over and above the material damage that can be observed. So far, no means of measurement or assessment has been found for this so-called deterrent or neutralization effect. Assuming that such an effect does in fact exist, if it is small then no great harm will be done if, as now, weapons are assessed solely on the material damage they can inflict. If, on the other hand, the deterrent effect is large, then it might materially affect the choice of weapon system and armament with which the Forces are provided.

To investigate and measure these effects is obviously going to be difficult, if not impossible by direct experiment, because unless there is real danger the subject's reaction may not be representative. Simulated battle as provided in "Battle Indoctrination" courses is unlikely to be a satisfactory substitute for the real thing since however well done it is, the participants retain confidence that their instructors and safety officers would allow very little real danger. An analysis of past battle experience might produce some usable data, although debriefing now on the memory of events more than a decade ago could be very unreliable. A possibility that suggests itself is the use of experiments using drugs or hypnosis to pursuade the participants that they really are in danger, and that, for example, the simulated enemy is really shooting to kill, then their reactions to a "real" situation can perhaps be assessed.

Personally, as a layman, I believe that troops in battle very rapidly make their own assessment of the lethality, imagined or real, of the situation in which they find themselves, and act accordingly. Below a certain threshold the risk is acceptable and they proceed, if with some caution, in the normal manner. Above this threshold the risk is intolerable and action is taken to avoid it. This suggests that a clue might be found in other dangerous activities, such as mountain climbing, working on high buildings, or even crossing the road. A city dweller will cross the road "jay-walking" in the fact of oncoming traffic because he believes that the risk is negligible. A country-man or elderly person will not accept the risk and either crosses at lights or waits until the road is clear. In the construction of high buildings normal rates are paid to workers below a certain height, above it they receive a bonus of danger money. If conditions are bad with high winds, ice and snow etc., all work ceases, the risk is intolerable to employer and employee alike. Many similar examples could be drawn from other activities; in all of them we find the acceptance of a certain degree of risk but a threshold above which the risk is intolerable. If a threshold value really does exist and can be found, then it should be possible to relate it to the probability of material or physical damage.

As to what the threshold value might turn out to be, I believe that it will be found to be much smaller than generally supposed. The Romans, who knew a great deal about the conquest and subjugation of foreign countries, gave us the word "decimate" which according to the dictionary means to kill a tenth. I believe that now, as then, if one tenth of a group of infantry, tanks, aeroplanes or what have you are killed and seen to be killed, the morale of the survivors will deteriorate very rapidly. This is in sharp contrast to much current military thinking and might well revolutionize some of our current military requirements, and make much more attractive proposals for weapons systems that would now be thrown into the waste paper basket. For this reason I think it important to direct scientific thinking and research into this area. I have given an account of some of the difficulties that will be met with, and made a few suggestions regarding possible avenues of approach. I would now be most interested to hear the views of any member of the audience on this subject, particularly with regard to any actual experimentation that may have been performed.

COMMENTS ON THE PAPER BY D. H. CHADDOCK

Clinical Session C

As a member of this panel, Boyd Harshbarger suggested that under the topic "The Measurement of the Morale and Suppressive Effects of Weapons" many discoveries in this area could possibly be derived by having psychological studies made of the reactions of the personnel employed at munition installations and manufacturers in which they have had explosions. For instance, at the Allegheny Ballistic Laboratories a group of psychologists might have been able to evaluate many of the characteristics that affect morale by observing the action of the personnel after a major explosion that happened there about a year ago. Psychological designs could be set up for activities at various areas in dangerous sections to study human reactions to emergencies. These designs should be set up prior to any accidents for various plants and installations, and should be ready to be executed immediately upon a catastrophe.

SOME SMALL SAMPLE THEORY FOR NONLINEAR REGRESSION ESTIMATION

H. O. Hartley
Institute of Statistics, Agricultural and Mechanical College of Texas, College Station, Texas

NOTE. The paper read at the Ninth Conference by H. O. Hartley under the title, "Non-Linear Estimation" presented results, both published and unpublished, on the general topic of least squares estimation of the parameters on non-linear regression laws. For material already published or about to be published [in this area] reference is made to:

H. O. Hartley (1961), "The Modified Gauss-Newton Method for the fitting of Non-Linear Regression Functions by Least Squares," Technometrics 3, 269.

H. O. Hartley, "Exact Confidence Regions for the Parameters in Non-Linear Regression Laws," (shortly to be published in Biometrika.)

H. O. Hartley and A. Booker, "Non-Linear Least Squares Estimation". Accepted for publication to the Annals of Mathematical Statistics.

The note reproduced below represents as yet unpublished material.

1. INTRODUCTION AND SUMMARY. We are given a set of N responses y_t which have arisen from a nonlinear regression model

$$(1) \qquad y_t = f(x_t, \theta) + e_t; \; t = 1, 2, \ldots, N$$

Here x_t denotes the t^{th} "fixed" input vector of k elements giving rise to y_t, whilst θ is an m-element unknown parameter vector with elements θ_i and the e_t are a set of N independent error residuals from $N(0, \sigma^2)$ where σ^2 will be assumed either known or unknown as indicated in the various sections. The expectations of the y_t are therefore the k + m variable functions $f(x_t, \theta)$ which will be assumed to satisfy certain regularity conditions. The problem is to estimate θ notably by "least squares" which in the present case is identical with "maximum likelihood". Whilst computational procedures to obtain the nonlinear least squares estimators and their large sample properties have received some consideration in the literature, we are in this note concerned with deriving some small sample results for the above estimation problem.

2. A THEOREM ON THE SUFFICIENCY IN NONLINEAR REGRESSION ESTIMATION. In what follows we shall be predominantly using vector notation and shall denote by y, $f(\theta)$, e and θ vectors having elements y_t, $f(x_t, \theta)$, e_t and θ_t, respectively. The first three of these are N-vectors, the last an m-vector. The N x k matrix of input values (x_t) will not be specifically referred to in this section and these matrix elements merely enter as fixed arguments into the functions $f(x_t, \theta)$. The nonlinear model (1) can then be written as

$$(2) \qquad y = f(\theta) + e$$

and the joint distribution of y; i.e. the sample likelihood as

$$(3) \qquad L(\theta, y) = (2\pi\sigma^2)^{-\frac{1}{2}N} \exp\left\{-\frac{1}{2\sigma^2}(y'y - 2y'f(\theta) + f'(\theta)f(\theta)\right\}$$

where the ' denotes transposition of column vectors and where we assume that, for the time being, σ^2 is known. Adopting classical* definitions of statistical sufficiency we say the $L(\theta, y)$" admits a set of p-statistics s(y) jointly sufficient for the estimation of θ" if

$$(4) \qquad L(\theta, y) \equiv F(S(y), \theta)\, H(y)$$

where the p-vector s(y) has as its elements the p-statistics $s_j(y)$ which are mathematical N-variable functions of the N elements $y_1 \ldots y_N$ of the sample vector y so that F is a p + m variable function of the elements

* Distribution functions and statistics are here assumed to be at least piecewise continuous functions of their argments, i.e. for all y_t and all θ in a convex n dimensional space Ω.

of $s(y)$ and θ, whilst $H(y)$ is an N-variable function of y. Comparing (4) with (3) we note that the condition for joint sufficiency can be written as

(5) $$y'f(\theta) - \frac{1}{2} f'(\theta) f(\theta) = J(s(y), \theta)$$

where we have introduced the p + m variable function J by the equation

(6) $$F(s(y), \theta) = \exp\left\{\sigma^{-2} J(s(y), \theta)\right\}.$$

We now state the following theorem which, under certain regularity conditions, is more or less obvious: -

Theorem 1:-

For the existence of a set of p statistics which are jointly sufficient for a set of p parameter functions $g_j(\theta)$ ($j = 1, 2, \ldots, p$) it is sufficient and, under regularity condition C, necessary that

(7) $$f(\theta) = U\, g(\theta)$$

where U is an n x p matrix of fixed numbers and $g(\theta)$ is a p-vector with elements $g_j(\theta)$.

The regularity condition C mentioned in Theorem 1 is as follows:- There exists at least one N-vector y(say $y = \eta$) such that all $s_j(y)$ have first partial differentials at $y = \eta$ and the function J defined by (6) has first partial differentials with regard to the s_j at $T = s(\eta)$ for all θ in Ω.

Proof of Theorem 1

The sufficiency of (7) is trivial. We have from (7)

(8) $$2y' f(\theta) = 2y' U g(\theta) = 2g'(\theta) U' y$$

and

(9) $$f'(\theta) f(\theta) = g'(\theta) U' U g(\theta)$$

Substituting (8) and (9) in (3) we have the form (4) for the parameter functions $g(\theta)$ and the p statistics $s_j(y)$ which are the elements of the p-vector

$$s(y) = U'y, \tag{10}$$

To prove that (7) is necessary we are given a likelihood function which satisfies (4), (5) and (6) and which we restrict to satisfy the regularity condition C. In view of the latter condition we are able to apply to the function $J(s(y), \theta)$ the standard differentiation rules at the point $y = \eta$ and $s(y) = s(\eta) = T$. We therefore obtain by differentiating (5) with regard to y_t the relation

$$f(x_t, \theta) = \sum_{j=1}^{p} \frac{\partial J(T, \theta)}{\partial s_j} \frac{\partial s_j(\eta)}{\partial y_t} \tag{11}$$

or the vector form (7) if we define the N x p matrix U by

$$U = \frac{\partial s_j(\eta)}{\partial y_t} \tag{12}$$

and the p-vector $g(\theta)$ by

$$g(\theta) = \frac{\partial J(T, \theta)}{\partial s_j} \tag{13}$$

Theorem 1 shows that we cannot expect to obtain a sufficient set of statistics unless the regression function $f(\theta)$ is of the "essentially linear" form. By essentially linear we mean that it is linear in a set of p parameter functions $g_j(\theta)$, $j = 1, 2, \ldots, p$.

The estimability of the linear parameters $g_j(\theta)$ depends on the rank of the matrix U, in accordance with linear estimation theory. The special case of p = m and rank (U) = m is of interest as the number of $g_j(\theta)$ is thus equal to the number of the original parameters θ_i. In such a situation it may be asked under which circumstances the mapping of the Ω space of θ_j to the Γ-space of the $g_j(\theta)$ is a one-to-one mapping. Whilst the classical condition of a full rank Jacobian only assures this one-to-one mapping "in the small" we prove the following unique-mapping theorem "in the large":-

Theorem 2:-

If for a set of m functions $g_j(\theta)$ depending on m variables θ_j we assume continuity of the first partial derivatives and definiteness of the quadratic form

$$(14) \qquad \sum_{i,j=1}^{m} \frac{\partial g_j}{\partial \theta_i} u_i u_j \gtrless 0 \qquad \text{for } u \neq 0$$

for all θ in the convex Ω space then the mapping of Ω to Γ by

$$(15) \qquad \gamma_j = g_j(\theta)$$

is a one-to-one mapping.

Proof:-

We have to show that two different θ vectors always generate two different γ-vectors. Suppose that, on the contrary, for θ^* and θ^+ in Ω we find

$$(16) \qquad g(\theta^*) = g(\theta^+) \text{ for } \theta^* \neq \theta^+$$

then the function $q(z)$ of the scalar argument z given by

$$(17) \qquad q(z) = \sum_{j=1}^{m} (\theta_j^* - \theta_j^+) \cdot g_j(\theta^+ + z(\theta^* - \theta^+))$$

would have identical values at $z = 0$ and $z = 1$, i.e.

(18) $q(0) = q(1)$.

Therefore for some value of $\hat{s}$ with $0 < \hat{s} < 1$ we have $\frac{dq}{ds}(\hat{s}) = 0$ and hence

(19) $$\frac{dq}{ds}(\hat{s}) = \sum_{i,j=1}^{m} (\theta_j^* - \theta_j^+)(\theta_i^* - \theta_i^+) \frac{\partial g_i}{\partial \theta_j} (\theta^+ + \hat{s}(\theta^* - \theta^+)) = 0$$

which would contradict (14).

ON THE SIMULTANEOUS ESTIMATION OF A MISSILE TRAJECTORY AND ERROR COMPONENTS INCLUDING THE ERROR POWER SPECTRA OF SEVERAL TRACKING SYSTEMS
(Preliminary Report)

David B. Duncan
Johns Hopkins University
and
Pan American World Airways
Guided Missiles Range Division
Patrick Air Force Base, Florida

1. INTRODUCTION. The problem under consideration is that of estimating the trajectory of a missile in powered flight using the data of several tracking systems observing it simultaneously. At the same time, it is also desired to estimate error components principally in the tracking systems and, in addition, the spectra of the autocorrelated errors of the observations. The discussion divides into two parts: The development of a linear mixed model up through section 4 and the development of models with stationary processes in sections 5 through 8.

Many of the ideas discussed in the earlier part are basic ones being used in the current reductions of trajectory data in programs included under the names of "best estimate of trajectory" and "Glotrac". In the second part, similar ideas are extended to include the additional presence and estimation of stationary autocorrelated errors in the tracking observations and, briefly in section 8, in the trajectory itself.

The early part is discussed in the notation and style of regression theory and methods as developed in statistical books of recent years such as, for example, Anderson and Bancroft [1] and Graybill [10]. The latter part involving spectral analyses and filtering is developed, more than is usual, in the regression-theory-and-methods style of the first part. This is done in the hopes of better bringing out how it all can fit together with the methods of the first part in complete unified analyses. Many references have been of help here, including for example Whittle [15] and [16], Hannan [11], Jenkins [14] and Durbin [9], no to mention Blackman and Tukey [2] to which further references are given. This brief list, of course, does not even begin to mention all of the many articles pertinent in this area.

No attempt is made here to discuss many details of methodology involved in writing computing programs. These will be written up in reports by the author, E. E. McGehee, and S. B. Burkett and by L. B Collins and A. Rinaldi who have been cooperating in the programming work. (As currently planned, these reports will appear as Pan American Technical Staff Memos and RCA Data Reduction Programming Memos respectively.)

2. LINEAR MODEL WITH PARAMETERS ALL FIXED. A simple first-approach model for trajectory data may be written as

$$(2.1) \qquad y = X\beta + e$$

where $y = (y_1, \ldots, y_n)'$ is a vector of n observations, X is an $n \times r$ matrix of known constants, β is an $r \times 1$ vector of fixed parameters (regression coefficients) to be estimated, and e is an $n \times 1$ vector of errors to be estimated with, as the name implies, all zero expectations.

The normal equations for estimating β are

$$(2.2) \qquad Ab = g$$

or, more fully,

$$(X'WX)\,b = X'Wy$$

where the solution b is the $r \times 1$ vector estimating β, and $W = V^{-1}$ where V is the variance (variance-covariance matrix) of e, $V = \text{var}(e)$.

Actually, in trajectory applications, the problem is initially of a non-linear form, such as,

$$(2.3) \qquad y = f(\beta) + e$$

which is handled iteratively in the familiar form

$$(2.3) \qquad y^* = X\beta^* + e^*$$

where β^* represents deviations $(\beta - \beta_0)$ of the parameters of interest β from initial values β_0, similarly $y^* = (y - y_0)$ where $y_0 = f(\beta_0)$ and X

is the matrix of partial derivatives $(x_{ij}) = (\partial f_i / \partial \beta_j)$ of the elements f_i of $f(\beta)$ with respect to the elements β_j of β, taken at $\beta = \beta_o$. In most of what follows the distinction between (2.3) and (2.1) is relatively unimportant. Except where distinctions are important y^* and β^* will be referred to as the observations and the trajectory parameters respectively and the star notation will not be carried.

In a typical trajectory estimation problem the elements of y, β and ϵ will be subvectors and submatrices. For a system with q observables the elements y_t of y will be $q \times 1$ subvectors for each set of observations at each point t in time, $t = 1, \ldots, n$. Thus with Mistram, say with the observations P_1, Q_1, P_2, Q_2, R we have $q = 5$; with an Azusa, say with the observations l, m, r we have $q = 3$, and so on. For a trajectory in d dimensions, the elements β_i and b_i of β and b will be $d \times 1$ vectors, and of course in powered missile flight $d = 3$. In conformance with these, each element x_{ti} of X is a $q \times d$ submatrix (usually of partial derivatives). In passing it will be noted that we have referred to the dimensions of y, X, β and ϵ in counts of those subvector and submatrix elements and not in terms of the ultimate scalar elements. It will be convenient to continue in these terms unless otherwise indicated.

The most popular forms of models in current use for the reduction of position data have square diagonal X matrices

$$(2.4) \qquad X = \begin{bmatrix} x_1 & & & & 0 \\ & x_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & x_n \end{bmatrix},$$

with as many $(r = n)$ trajectory points to be estimated as there are observations. (The submatrices comprising the elements of course are not necessarily square or diagonal). These we will refer to as single-point-in-time models because they estimate mathematically independent parameters at each point in time and employ no constraints for the trajectory.

In the first-approach model the errors are assumed uncorrelated over time making the variance and weight matrix diagonal

$$(2.5)\qquad V = \begin{bmatrix} v_1 & & & & 0 \\ & v_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & v_n \end{bmatrix}, \quad W = \begin{bmatrix} w_1 & & & & 0 \\ & w_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & w_n \end{bmatrix}$$

In each of these the elements are $q \times q$ submatrices which are often assumed diagonal.

In a single-point-in-time problem with time-uncorrelated errors, the normal equations $Ab = g$ are also diagonal, being of the form

$$(2.6)\qquad \begin{bmatrix} a_1 & & & & 0 \\ & a_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \cdot \\ \cdot \\ \cdot \\ g_n \end{bmatrix}$$

in which the solution for each unknown b_t can be made separately in $a_t b_t = g_t$ and in which

$$(2.7)\qquad \underset{\substack{3\times 3}}{a_t} = \underset{\substack{3\times q}}{x_t'}\,\underset{\substack{q\times q}}{W_t}\,\underset{\substack{q\times 3}}{x_t}, \qquad \underset{\substack{3\times 1}}{g_t} = \underset{\substack{3\times q}}{x_t'}\,\underset{\substack{q\times q}}{W}\,\underset{\substack{q\times 1}}{y_t}, \qquad t = 1, \ldots, n.$$

In a case with p tracking systems working simultaneously, the complete data can be thought of as fitting the same model (2.1) in the form, with $p = 3$ say,

$$(2.8)\qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where y_i, X_i, e_i, and correspondingly V_i and W_i are for the i^{th} system $i = 1, \ldots, p$. If the errors are taken to be uncorrelated between systems, the normal equations are readily seen to reduce to the same diagonal form (2.6), in which now,

$$a_t = \sum_{i=1}^{p} a_{it}, \quad g_t = \sum_{i=1}^{p} g_{it}$$

and a_{it} and g_{it} are the submatrices and subvectors defined as in (2.7) for the i^{th} system at time t, $i = 1, \ldots, p$, $t = 1, \ldots, n$.

All of the above are well-known and discussed in a multitude of references, a typical example for a single system being Davis's development of the reduction of Cinetheodolite data [4]. Here, it is hoped that the review has served primarily as a relatively simple introduction to the notation and background context of work in the following sections.

3. A LINEAR MIXED MODEL. A big weakness of the classic model of section 2 is that it makes no allowances for autocorrelation of errors within systems, i. e., time correlations in the errors at one point in the trajectory to another. A big step forward in this regard can be made with the use of a mixed model of the form

$$(3.1)\qquad y = X\beta + Z\gamma + e$$

where y, X, β, e are as already defined, and $\delta = Z\gamma$, or more fully,

$$(3.2)\qquad \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix} \gamma$$

is a $q \times 1$ subvector δ_t of errors in y_t which can be expressed as a known linear function $z_t\gamma$, (z_t being $q \times k$) of k error components $\gamma = (\gamma_1, \ldots, \gamma_k)'$ common to all points in time $t = 1, \ldots, n$. For example, for a given system with say $q = 3$ observables it might be desired to make allowance for a set of "zero set" errors a_{10}, a_{20}, a_{30} and a set of linear drift coefficients a_{11}, a_{21}, a_{31}. In such an example $z_t\gamma$ would be

$$(3.3)\qquad \begin{bmatrix} 1 & 0 & 0 & t & 0 & 0 \\ 0 & 1 & 0 & 0 & t & 0 \\ 0 & 0 & 1 & 0 & 0 & t \end{bmatrix} \begin{bmatrix} a_{10} \\ a_{20} \\ a_{30} \\ a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}$$

Better models, if available, are ones which express the errors δ_t in y_t as physically identifiable error components y_i in the tracking system, such as, tilt errors, refraction errors, etc.

The important point that distinguishes the elements of γ from the trajectory parameters in β is that the γ's are regarded as random variables (which vary from flight to flight) with a vector of means $\bar{\gamma}$, and a variance (variance-covariance) matrix V_γ. (The error components γ and the remaining errors e are uncorrelated.)

For p systems working simultaneously the complete data fits the same model (3.1) in the form, with $p = 3$ say,

$$(3.4)\qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \beta + \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where, in addition to the terms already discussed for (2.8), Z_i and γ_i are the Z's and γ's for the i^{th} system.

In connection with the general linear mixed model (3.1) a very useful result can be readily obtained which may be stated as follows: The maximum-likelihood estimates b, c of β and γ, (assuming that γ and ϵ are Gaussian) are the same as though we assume the linear-with-all-fixed parameters model

$$(3.5) \qquad y = X_1 \beta_1 + X_2 \beta_2 + \epsilon$$

with terms defined as in

$$\begin{bmatrix} y_1 \\ y_0 \end{bmatrix} = \begin{bmatrix} X \\ O_k \end{bmatrix} \beta + \begin{bmatrix} Z \\ I_k \end{bmatrix} \gamma + \begin{bmatrix} \epsilon_1 \\ \epsilon_0 \end{bmatrix}$$

where y_1 and ϵ_1 are the given $n \times 1$ vectors y and ϵ, O_k and I_k are the null and identity matrices of order k, the additional observations y_0 are $y_0 = \bar{\gamma}$ and ϵ_0 contains the k deviations $\bar{\gamma} - \gamma$. In other words we can treat the means $\bar{\gamma}$ as being observations about γ fixed as mean. The same variance $V_0 = V_\gamma$ is taken for ϵ_0 as was given for γ. The full variance and weight matrices for ϵ in (3.5) are thus

$$(3.6) \qquad V = \begin{bmatrix} V_1 & 0 \\ 0 & V_0 \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & 0 \\ 0 & W_0 \end{bmatrix}$$

where V_1 and W_1 are the V and W of the original ϵ in (3.1) and $W_0 = V_0^{-1}$.

From this pseudo model with pseudo-observations and all fixed parameters (or, in other ways) the estimates $b = b_1$ and $c = b_2$ are seen to be given by

$$(3.7) \qquad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

with terms as defined in

$$(3.7a)\qquad \begin{bmatrix} X_1' W X_1 & X_1' W X_2 \\ X_2' W X_1 & X_2' W X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1' W y \\ X_2' W y \end{bmatrix}$$

which, in terms of the given mixed model (3.1) reduce to

$$(3.7b)\qquad \begin{bmatrix} X' W_1 X & X' W_1 Z \\ Z' W_1 X & Z' W_1 Z + W_0 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} X_1' W_1 y \\ Z' W_1 y + W_0 y_0 \end{bmatrix}$$

If the underlying model is non-linear in the error components γ, the matrix Z will consist of partial derivatives, the γ in the model equations will consist of increments $\gamma^* = (\gamma - \gamma_0)$ to be made to γ_0, a set of values reached in a previous iteration. The pseudo-observations y_0 in the equations will consist of deviations $y_\gamma^* = (y_0 - \gamma_0) = (\bar{\gamma} - \gamma_0)$. In the first iteration it is natural to take γ_0 as $\bar{\gamma}$ thus y_0^* is zero and the term $W_0 y_\gamma^*$ drops out on the right hand side of (3.7b). In all subsequent iterations it will appear as $W_0(\bar{\gamma} - \gamma_0)$.

In most work a vector of expected fixed errors as is represented by $\bar{\gamma}$, can be calibrated out and this is usually done. Thus more often than not the effective $\bar{\gamma}$ is 0. Thus in a problem linear in γ we have $y_0 = \bar{\gamma} = 0$ and $W y_0$ never appears, in the right hand side of (3.7b). In a non-linear problem $W y_\gamma^*$ will only appear as before after the first iteration and then as $W_0(\bar{\gamma} - \gamma_0) = -W\gamma_0$. Details like these can be reworked and checked very simply starting from the pseudo-model (3.5) or its non-linear problem equivalent.

Many trajectory estimation problems are solved using the mixed linear model (3.1) assuming that the variance matrices $V_1 = \text{var}(\epsilon)$ and $V_0 = \text{var}(\omega)$ are diagonal. The $n \times n$ matrix A_{11} and the corresponding $n \times 1$ vector g_1 are the same as the A and g of the standard model of section 2. (A_{11} is diagonal with 3×3 submatrices a_i as its $n \times n$ elements, g_1 has $n \times 1$ subvectors). With p systems, the elements of A_{11} and g_1 from each system add together to give an overall A_{11} and g_1 of the same $n \times n$ and $n \times 1$ dimensions.

For a single system $A_{22} = X'_2 W X_2 = Z' W_1 Z + W_0$ is $k \times k$ and symmetric though not diagonal. $A_{12} = X'_1 W X_2 = X' W_1 Z$ is correspondingly $n \times k$ and is non-null. $A_{21} = A'_{12}$. For p systems, because of the additional error component vectors γ_i entering with each system A_{22} is $k \times k$ where now $k = \sum_{i=1}^{p} k_i$ and k_i is the size of γ_i, $i = 1, \ldots, p$. A_{12} widens correspondingly.

To take an example, for 300 seconds of data at say 10 points per second, $n = 3{,}000$. With $p = 6$ systems say with $\sum k_i = 60$ error components in all, $k = 60$. Thus A_{11} would be $3{,}000 \times 3{,}000$ and diagonal with 3×3 submatrices as elements. A_{22} would be symmetric and 60×60 with scalar elements. A_{12} would be non-null $3{,}000 \times 60$, with 3×1 subvectors as elements and $A_{21} = A'_{12}$.

Further discussions and applications of this model are available in many references including Brown [3], Duncan [5] and [6] and Henderson et al [12].

4. PARTITIONING OF THE NORMAL EQUATIONS. TWO-STEP REGRESSION. Because A_{11} above in (3.7) is diagonal, it is easily invertible; large though may be its size. As a result of this it is convenient to solve for b_2 first in the "reduced" normal equations

$$(4.1) \qquad (A_{22} - A_{21} A_{11}^{-1} A_{12}) b_2 = g_2 - A_{21} A_{11}^{-1} g_1$$

and then for b_1 by substituting b_2 in

$$(4.2) \qquad A_{11} b_1 = g_1 - A_{12} b_2 \quad .$$

Both of these equations follow directly from (3.7). In this way the largest non-diagonal matrix to be inverted is the left hand side one $(A_{22} - A_{21} A_{11}^{-1} A_{12})$ in (4.1) which is only $k \times k$.

In this connection it is often useful to note the following development of (4.1) and the subsequent use of (4.2). Substituting for the A terms from (3.7a) we get

$$(X_2' W^{1/2} D W^{1/2} X_2) b_2 = X_2' W^{1/2} D W^{1/2} y \tag{4.3}$$

where $D = (I - W^{1/2} X_1 A_{11}^{-1} X_1' W^{1/2})$ is readily seen to be idempotent and symmetric. From this we can write the reduced normal equations as

$$\tilde{A} b_2 = \tilde{g} \tag{4.4}$$

where $\tilde{A} = \tilde{X}'\tilde{X}$, $\tilde{g} = \tilde{X}'\tilde{y}$ and $\tilde{X} = DW^{1/2}X_2$, $\tilde{y} = DW^{1/2}y$.

In this way b_2 may be regarded as the "fit" (or the result of "fitting") a regression (unweighted) of $\tilde{y}$ on $\tilde{X}$ where $\tilde{y}$ is the weighted residual vector

$$\begin{aligned} \tilde{y} &= DW^{1/2}y = W^{1/2}y - W^{1/2}X_1 A_{11}^{-1} X_1' Wy \\ &= W^{1/2}(y - X_1 b_1^*) = W^{1/2}(y - \hat{y}) \end{aligned} \tag{4.4}$$

from fitting a regression of y on X_1 alone and $\tilde{X}$ comprises similar weighted residual vectors from fitting regressions of the columns of X_2 on X_1.

The substitution step (4.2) may be rewritten as $b_1 = A_{11}^{-1} g_1 - A_{11}^{-1} A_{12} b_2$ and then

$$b_1 = b_1^* - K b_2 \tag{4.5}$$

and may be regarded as one of "correcting" $b_1^* = A_{11}^{-1} g_1$ to give b_1 by subtracting the "correction" $K b_2 = A_{11}^{-1} A_{12} b_2$. In the sequel this will be referred to as the two-step regression method, the first step comprising the solution giving $\tilde{y}$ and $\tilde{X}$ in which b_1^* is obtained, the second the solution for b_2 and the use of this to correct b_1^* to b_1.

5. LINEAR ALL-FIXED-PARAMETERS MODEL WITH STATIONARY ERRORS. A further improved model for trajectory estimation would appear to be the same as the linear mixed model of section 3, but with allowance for stationary autocorrelations among the errors within systems. The linear terms $Z\gamma$ would take care of large non-stationary effects. An allowance for stationary autocorrelations among the subvectors (e_t ; t = 1, . . . , n) of the errors e would take care of a lot of the remaining correlations. Before passing on to consideration of a model of this type in the next section, we will first deal here with stationary autocorrelations among the errors in an all-fixed parameters model of the type considered in section 2.

In the notation

$$(5.1) \qquad y = x\beta + e$$

Or, more fully for p = 3 systems, say,

$$(5.1a) \qquad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \beta + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

(cf (3.4) before), the variance submatrices in

$$(5.1b) \qquad \text{var}(e) = V \cdot \begin{pmatrix} V_1 & & 0 \\ & V_2 & \\ 0 & & V_3 \end{pmatrix}$$

are now of the stationary Laurent form

$$
(5.1\,c) \qquad \underset{n \times n}{V_i} = \begin{pmatrix} v_{i0} & v_{i1} & v_{i2} & \cdots \\ v'_{i1} & v_{i0} & v_{i1} & \cdots \\ v'_{i2} & v'_{i1} & v_{i0} & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}
$$

Each element v_{ih} is a lag-h autocovariance matrix for the errors at time t and time $t+h$, $t = 1, 2, \ldots$ and $h = 1, 2, \ldots$

In getting simultaneous estimates of a trajectory over several systems it comes out that a more simple solution is provided if we assume that appropriately weighted forms of trajectory estimates obtained separately for each system are the observations with the stationary errors, rather than the original system observations themselves.

Thus, for the i^{th} system we may choose to take

$$
(5.2) \qquad \underset{3 \times 1}{y_{it}} = \underset{3 \times 3}{D_{it}} \, \underset{3 \times 1}{b_{it}}
$$

as the t^{th} element of the observation vector, where b_{it} is a single-point-in-time estimate (without assuming autocorrelated errors) of the individual trajectory deviations $\beta_t^* = (\beta_t - \beta_{to})$, $i = 1, 2, 3$; $t = 1, \ldots, n$; and D_{it} is a diagonal matrix with d_{jj} put equal to the reciprocal of the standard deviation of the j^{th} element of b_{it}. (In this way d_{jj} is positive and d_{jj}^{-2} is the j^{th} diagonal element of

$$
a_{it}^{-1} = (x'_{it} w_{it} x_{it})^{-1}
$$

for the i^{th} system at time t in (2.9).

With the observations defined in this way, the matrix X_i becomes a $3n \times 3n$ identity matrix, or in keeping with our previous notation, an $n \times n$ identity matrix in which each diagonal element is a 3×3 identity submatrix I_3, $i = 1, 2, 3$. From this the normal equations for $p = 3$, say systems again, are

$$(5.3) \qquad (W_1 + W_2 + W_3)b = W_1 y_1 + W_2 y_2 + W_3 y_3$$

where $W_i = V_i^{-1}$ for each system and is of the approximate Laurent form

$$(5.4) \qquad W_i = \begin{pmatrix} w_{i0} & w_{i1} & w_{i2} & \cdots \\ w_{i1} & w_{i0} & w_{i1} & \cdots \\ w_{i2} & w_{i1} & w_{i0} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}$$

We can then write

$$(5.5) \qquad b = K_1 y_1 + K_2 y_2 + K_3 y_3$$

where $K_i = (W_1 + W_2 + W_3)^{-1} W_i$ is approximately Laurent, or, in the style of filtering,

$$(5.5a) \qquad b_t = \sum_{h=-m}^{m} k_{1h} y_{1(t+h)} + \sum_{h=-m}^{m} k_{2h} y_{2(t+h)} + \sum_{h=-m}^{m} k_{3h} y_{3(t+h)}$$

$$t = 1, \ldots, n$$

where

$$(5.6) \qquad k_i = (k_{i(-m)}, \ldots, k_{i0}, \ldots, k_{im})'$$

is a typical column, except for zeros fore and aft, of K_i, with $k_{i(-h)} = k_{ih}$, $h = 1, \ldots, m$; $i = 1, 2, 3$.

A simple matrix-and-regression-theory approach for finding K_i which is being tried and which may be useful in extending to more complex models may be put as follows.

Suppose for the moment that the number of Euclidian coordinates is only $d = 1$ instead of 3, and that the elements $w_{i0}, w_{i1}, \ldots$ of the weight matrices W_i are thus scalars instead of 3×3 submatrices. Assume n is odd, say, and introduce a Fourier matrix $F(n \times n) = ((f_{tj}))$ comprised of the Fourier terms

$$(5.7) \qquad f_{tj} = \frac{1}{\sqrt{n}} e^{i \frac{2\pi}{n} (t - \bar{t})(j - \bar{j})}$$

where $\bar{t} = \bar{j} = (n+1)/2$. It is readily established that this matrix is orthogonal, i.e. $F^{-1} = \overline{F'}$, and that $F V_i F^{-1} \overset{\text{approx}}{=} S_i$ where S_i is diagonal with diagonal elements comprised of the most-detailed symmetric power spectrum (the symmetric expected periodogram) of the errors e_{it}, i.e. $e_i = (e_{i1}, \ldots, e_{in})'$.

Thus we may write

$$F K_i F^{-1} = F(W_1 + W_2 + W_3)^{-1} F^{-1} F W_i F^{-1}$$

$$= (F W_1 F^{-1} + F W_2 F^{-1} + F W_3 F^{-1})^{-1} F W_i F^{-1} .$$

But

$$(5.8) \qquad F W_i F^{-1} = (F V_i F^{-1})^{-1} \overset{\text{approx}}{=} S_i^{-1}$$

which we shall write as $1/S_i$ to emphasize the point that the algebra of diagonal matrices is the same as that of scalars.

Hence

(5.9) $$F K_i F^{-1} \text{ approx } \frac{1/S_i}{1/S_1 + 1/S_2 + 1/S_3}$$

where R_i is a diagonal matrix with diagonal terms of the form

$1/s_i(f)/(1/s_1(f) + 1/s_2(f) + 1/s_3(f))$, $f = -(n+1)/2, \dots, 0, \dots, (n+1)/2$,

and from this

(5.10) $$K_i = F^{-1} R_i F.$$

By assuming a certain degree of smoothness in each spectra, the R_i can be obtained as much smaller matrices with $f = -m, \dots, 0, \dots, m$, the Fourier matrices can be correspondingly reduced, and the required filter elements given by (5.10) are found by the simple finite Fourier transforms of the form

(5.11) $$k_{ih} = \frac{1}{N} \sum_{j=-m}^{m} e^{i \frac{2\pi}{N} hj} r_{ij}$$

where $r_{i(-m)}, \dots, r_{im}$ are the diagonal elements of the reduced R_i in which $r_{ij} = (1/s_{ij})/(1/s_{1j} + 1/s_{2j} + 1/s_{3j})$ and the s_{ij} are spectral estimates (except for a constant) from, say, a Blackman-Tukey (2) method, $N = 2m + 1$ and the imaginary i in the exponent of e is not to be confused with the other i's for the i^{th} system. Because of symmetry (5.11) reduces to the simple cosine expansions of the form

$$k_{ih} = \frac{1}{N}\left[r_{io} + 2 \sum_{j=1}^{m} r_{ij} \cos \frac{2\pi}{N} hj \right]$$

In the case of $d = 3$ dimensions the same approach extends to the construction of three-dimensional filters. The elements f_{tj} of F are replaced by 3×3 scalar matrices $f_{tj} I_3$, the diagonal scalar elements s_{ij} of the spectral matrices S_i are replaced by 3×3 complex spectral submatrices $s_{ij} = c_{ij} + i q_{ij}$ and the elements k_{ih} of the filters are 3×3 submatrices

6. ESTIMATION OF ERROR SPECTRA. Trajectory-free estimates of the spectra required for the construction of the filters in section 5 can be obtained by the cross-spectrum-of-differences method already discussed by Dr. Wells in an earlier paper on this program, (Duncan and Wells [8]). Once a process $\{b_t\}$ of trajectory estimates obtained by the method of section 5 is available, the initial spectral estimates can be improved by iterations of the same form of procedure.

To explain this, the method based on differences can be identically re-represented as a method based on residuals $\{u_{it} = y_{it} - \bar{y}_t\}$, $i = 1, 2, 3$ where

$$\left\{ \sum_{i=1}^{3} y_{it}, \quad t = 1, \ldots, n \right\}.$$

(cf. Duncan and Carroll [7]). The desired spectral estimates $\tilde{s}_1(f)$, $\tilde{s}_2(f)$, $\tilde{s}_3(f)$ can be obtained (in the case of $d = 1$ dimension) by solving the normal equations.

$$\begin{aligned}
& a_{11}\tilde{s}_1(f) + a_{12}\tilde{s}_2(f) + a_{13}\tilde{s}_3(f) = \hat{s}_{u_1}(f) \\
(6.1) \qquad & a_{21}\tilde{s}_1(f) + a_{22}\tilde{s}_2(f) + a_{23}\tilde{s}_3(f) = \hat{s}_{u_2}(f) \\
& a_{31}\hat{s}_1(f) + a_{32}\tilde{s}_2(f) + a_{33}\tilde{s}_4(f) = \hat{s}_{u_3}(f)
\end{aligned}$$

where the right-hand side values are, say, estimates of the spectra $s_{u_1}(f)$, $s_{u_2}(f)$, $s_{u_3}(f)$ of the residual processes, and where

$$(6.2) \qquad (a_{ij}) = \begin{pmatrix} 4/9 & 1/9 & 1/9 \\ 1/9 & 4/9 & 1/9 \\ 1/9 & 1/9 & 4/9 \end{pmatrix}$$

The solution is carried out at each frequency f.

After an estimate $\{b_t\}$ of the trajectory has been made with filters K_1, K_2, K_3 the residuals can now be recomputed as $\{u_{it} = y_{it} - b_t\}$. The equations (6.1) are then solved again with

$$(6.2) \qquad a_{ij} = (\delta_{ij} - r_{1j} - r_{2j} - r_{3j})^2$$

where r_{ij} is the j^{th} frequency element of $R_i = FK_iF^{-1}$. Steps like these interspersed with the iterations for getting $\{b_t\}$ ultimately yield good estimates of both the spectra and the trajectory. In the case of $d = 3$ dimensions, the same approach extends to the estimation of 3×3 complex spectral matrices. (Note: In the discussions of sections 5 and 6, $\{b_t\}$ sometimes refers to deviations $b_t^* = b_t - \beta_{to}$ as previously indicated, and sometimes to $\{b_t\}$. If this is kept in mind, no trouble should arise in distinguishing which is which.)

7. LINEAR MIXED MODEL WITH AUTOCORRELATED ERRORS. The same ideas can be adapted to the mixed model

$$(7.1) \qquad y = X\beta + Z\gamma + \epsilon$$

in which the errors ϵ are autocorrelated as in section 5.

By defining the observations as in section 5 the system submatrices X_i of X can each be reduced to $n \times n$ (or $3n \times 3n$) identity matrices. If the model for the system observations is linear in γ, the modified model will be non-linear in γ. The vector γ will thus represent deviations $\gamma^* = (\gamma - \gamma_0)$, and Z will consist of partial derivatives.

Rewriting the model in the pseudo all-fixed parameters form

$$(7.2) \qquad y = X_1 \beta_1 + X_2 \beta_2 + \epsilon$$

which, in more detail represents

$$\begin{pmatrix} y_1 \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ O_k \end{pmatrix}\beta + \begin{pmatrix} Z \\ I_k \end{pmatrix}\gamma + \begin{pmatrix} \epsilon_1 \\ \epsilon_0 \end{pmatrix}$$

with terms defined as before (cf (3.5)), the first steps, talking in terms of the two-step regression approach (section (4)), is to fit regressions of y on X_1 and of X_2 (i.e., of the columns of X_2) on X_1 forming the residuals

(7.3) $$\tilde{y} = DW^{1/2}y \quad \text{and} \quad \tilde{X} = DW^{1/2}X_2$$

where

$$D = I - W^{1/2}X_1(X_1'WX_1)^{-1}X_1'W^{1/2} \quad .$$

This will reduce to simple filtering operations on y and on Z the details of which can be readily developed by the method of section 5. In this step a first-step trajectory estimate b_1^* is also obtained by filtering.

In the second step the equations

(7.4) $$\tilde{X}'\tilde{X}b_2 = \tilde{X}\tilde{y}$$

are solved for b_2 and b_1^* is corrected using a

(7.5) $$\left\{ u_{it} = y_{it} - b_{1t} - z_{it}b_{2t} \right\}, \quad i = 1, 2, 3.$$

The desired estimates of the respective error processes could then be obtained by solving equations of the form (6.1) where the right hand side terms are based on these residuals. This would involve ignoring the effect in the residuals of $z_{it}(b_{2t} - \beta_{2t})$ which would not cause any problem since b_{2t} is fitted over the whole time span of data and $b_{2t} - \beta_{2t}$ would have a relatively low variance. To start the process the initial trajectory estimates in the first two-step solution would be obtained using filters from previous experience, or at the very worst, using identity filters.

The estimation procedure, in summary, would consist of several three-step iterations, two giving closer estimates of the trajectory parameters in β and the error components in γ, the third giving closer estimates of the error spectra.

8. CONCLUDING REMARKS.

(1) In connection with the spectral estimates, the emphasis here has been placed on the role these play in leading to better trajectory estimates. The error spectra however, and especially those of the initial tracking system observations, are of considerable interest in themselves. With this in mind, when a set of estimates in the trajectory coordinate system have converged, it will then be desirable to project the trajectory estimates into the coordinates of each of the tracking systems. The residuals of the initial observations from these will then provide a basis for getting good estimates of the spectra of the errors in the initial observations. Each will consist of a linear combination of the spectral estimate of the residual concerned and the spectral estimate of the error in the trajectory estimate.

In the event that the variance matrix of the initial observations turns out at this point to be much different from that initially assumed in transforming to the trajectory coordinates, it may be considered desirable to make another iteration starting with the latest weights.

(2) A direction for improvement is that of allowing for spectra and filters which can change with some flexibility over time. It is hoped to develop methods for handling this along with the development of a promising model to which we now refer very briefly.

(3) Perhaps the most interesting and useful features of the foregoing mixed-model-plus-filtering methods are their potential usefulness in the analysis of data in a more advanced model. This model is the same as the linear mixed model with stationary errors in section 7 except that the specification for the trajectory itself is no longer of the single-point-in-time form (with a diagonal X matrix for each system). The trajectory is now represented at the centers of progressive medium-length intervals by a low degree polynomial (say a quadratic) plus a deviation δ_t, $t = 1, \ldots, n$ for successive intervals are assumed to form a stationary time series.

Similar step-by-step-regression-plus-filtering methods can be used to give estimates of (i) the polynomial component and (ii) the stationary component and thus the combined components of the trajectory, (iii) the error components in the tracking systems, (iv) the spectra of the error processes and (v) the spectrum of the stationary trajectory components. A model with polynomial plus stationary components in a signal is suggested in similar contexts by several writers, e.g., Zadeh and Ragazzini [17], and Hunziger [13].

The prospect here of also simultaneously estimating the spectrum of the stationary trajectory component has special appeal. Roughly speaking, the trajectory data, in this approach, will be fitted under constraints as tight as the data themselves will condone. The estimates will have an appreciable part of the low-variance advantages that can be added by smoothing without having lost much of the responsiveness which is shown by the data to be needed. Current work is in progress on these aspects.

REFERENCES

1. Anderson, R. L. and T. A. Bancroft, "Statistical Theory in Research," McGraw Hill, New York, 1954.

2. Blackman, R. B. and J. W. Tukey, "The Measurement of Power Spectra", Dover, New York, 1958.

3. Brown, Duane C., "A Treatment of Analytical Photogrammetry with Emphasis on Ballistic Camera Applications," RCA Data Reduction Technical Report No. 34, RCA AFMTC-TR-57-22, ASTIA Document No. 124144, 1957.

4. Davis, R. C., "Techniques for the Statistical Analysis of Cinetheodolite Data", Navord Report 1599, NOTS 369, China Lake, California, 22 March 1951.

5. Duncan, David B., "On the Optimum Estimation of a Missile Trajectory from Several Tracking Systems". RCA Systems Analysis Technical Report No. 12, RCA AFMTC-TR-60-16, ASTIA Document.

6. Duncan, David B., "On Some Theory of Mixed Linear Models and Its Application in the Estimation of a Missile Trajectory", Paper No. 332 from the Dept. of Biostatistics, The Johns Hopkins University, December, 1961.

7. Duncan, David B. and C. L. Carroll, "Studies in the Simultaneous Estimation of a Trajectory form Several Tracking Systems", PAA Technical Staff Technical Memo No. 4, MTV-TM-26-13, 7 September 1962.

8. Duncan, David B. and W. T. Wells, "Estimation of Error Spectra from the Cross-Autocovariance Functions of Differences", PAA Technical Staff Technical Memo No. 11, MTV-TM-64-2, 30 January 1964.

9. Durbin, J., "The Fitting of Time-Series Models", Invited Review Paper, Econometric Society Mtg., Washington, D. C., 1959.

10. Graybill, Franklin A., "An Introduction to Linear Statistical Models", Volume I, McGraw Hill Book Company, Inc., New York, 1961.

11. Hannan, E. J., "Time Series Analysis", Methuen, London, 1960.

12. Henderson, C. R., Oscar Kempthorne, S. R. Searle, and C. M. Krosigk, "The Estimation of Environmental and Genetic Trends from Records Subject to Culling", Biometrics 15 pp 192-218.

13. Hunziger, R. R., "Trajectory Estimation from Radar and Inertial Guidance Data", RCA Systems Analysis TN 20 AFMTC, 31 August 1962.

14. Jenkins, G. M., "General Considerations in the Analysis of Spectra", Technometrics 3, 11, 133-166, 1961.

15. Whittle, P., "Hypothesis Testing in Time Series Analysis", Almquest and Wiksells, Uppsala, 1951.

16. Whittle, P., Prediction and Regulation, English Universities Press, London and Van Nostrand, New York, 1963.

17. Zadeh, L. A. and John R. Ragazzini, "An Extension of Wiener's Theory of Prediction", J. Appl. Phys., Volume 21, No. 7, July, 1950.

ATTENDEES TO THE NINTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

23-25 Oct. 1963

OFF-POST ATTENDEES

Abbott, K. H.	Army Materials Research Agency, Watertown, Mass.
Addelman, Sidney	RTI, Durham, N. C.
Bailey, Robert J.	BRL, Aberdeen PG, Md.
Bell, Raymond	BRL, Aberdeen PG, Md.
Bengston, Nils M.	Army Research Office, Durham, N. C.
Biser, Erwin	USAELR&D Labs., Ft. Monmouth, N. J.
Bombara, E. L.	NASA Marshall Space Flight Center, Huntsville, Ala.
Breaux, Harold J.	BRL, Aberdeen PG, Md.
Bruno, O. P.	BRL, Aberdeen PG, Md.
Bulfinch, A.	Picatinny Arsenal, N. J.
Carr, Frank L.	Army Materials Research Agency, Watertown, Mass.
Carter, Fredrick L., Jr.	U. S. Army Biological Labs, Ft. Detrick, Md.
Cogan, Eugene A.	Human Resources Research Office, Alexandria, Va.
Cohen, A. C., Jr.	University of Georgia
Cox, Edwin L.	Biometrical Services ARS, Plant Industry
Cox, Paul C.	WSMR, New Mexico
Cowie, William S.	Army Natick Labs, Mass.
Crenshaw, Craig M.	Army Material Command, Washington, D. C.
Crews, George C.	BRL, Aberdeen PG, Md.
Del Priore, Francis R.	COMPTEVFOR, Norfolk, Va.
Doverspike, Charles A.	LTV Div., Warren, Mich.
Dressel, Francis	Army Research Office, Durham, N. C.
Dressler, Frederick	BRL, Aberdeen PG, Md.
Duncan, David B.	Johns Hopkins Univ., Baltimore, Md.
Dwyer, W. A.	Boeing Company, Seattle, Washington
Eckard, Cecil O.	Ft. Douglas, Utah
Eckles, Andrew J., III	RAC, McLean, Va.
Eisenhart, Churchill	National Bureau of Standards, Washington, D. C.

Ellner, Henry	Edgewood Arsenal, Md.
Foster, W. D.	Ft. Detrick, Md.
Free, Bernie R.	WSMR, New Mexico
Fursa, Alex	Hughes Aircraft Co., Fullerton, Calif.
Gabriel, Richard F.	Douglas Aircraft Co., Sacramento, Calif.
Galbraith, A. S.	Army Research Office, Durham, N. C.
Gardiner, Donald A.	Oak Ridge National Lab., Tenn.
Gayle, J. B.	NASA Marshall Space Flight Center, Huntsville, Ala.
Grant, Roy L.	Bureau of Mines, Pittsburg, Pa.
Grim, Walter H.	U. S. Army Materiel Command, Washington, D. C.
Grubbs, Frank E.	BRL, Aberdeen PG, Md.
Hafle, Ralph S.	Baldwin Electronics Inc., Little Rock, Ark.
Haines, B. W.	U. S. Army Biological Labs, Ft. Detrick, Md.
Hanamoto, Ben	USATAC, Warren, Mich.
Harshbarger, Boyd	Virginia Polytechnic Institute, Blacksburg, Va.
Hartley, H. O.	Texas A-M University
Heathcock, Ray	NASA Marshall Space Flight Center, Huntsville, Ala.
Helvig, Thomas Noble	Minneapolis-Honeywell Regulator Co., Hopkins, Minn.
Holmes, Frederick B.	West Sacramento, Calif.
Hoppes, H. N.	RAC, McLean, Va.
Isaac, Gerhard J.	U. S. Army Medical Research & Nutrition Lab, Fitzsimmons General Hospital, Denver, Colorado
Jessup, Gordon L., Jr.	U. S. Army Biological Labs, Ft. Detrick, Md.
Johnson, Jerome	BRL, Aberdeen PG, Md.
Johnston, Fred J.	U. S. Army Electronic Research [illegible] View, California
Katsanis, David J.	Frankford Arsenal, Philadelphia, Pa.
Kearms, James L.	Hercules Powder Co.
Kniss, James	BRL, Aberdeen PG, Md.
Kirkland, Jessee L.	U. S. Army Engr Waterways Experiment Station, Vicksburg, Miss.
Komlosy, C. F.	War Office, United Kingdom
Krane, Scott A.	Dugway Proving Ground, Dugway, Utah
Krimmer, Manfred W.	Picatinny Arsenal, Dover, N. J.
Kullback, Solomon	George Washington Univ, Washington, D. C.
Kurkjian, Badrig M.	Harry Diamond Labs, Washington, D. C.
Langlie, H. J.	Philco Corp. Newport, Calif.
Lavin, George I.	BRL, Aberdeen PG, Md.

Leins, Willard C.	Bechtel Corp., Huntsville, Ala.
LeMaster, Vaughn	APSA, Joliet, Ill.
Long, Louis J.	U. S. Army Engr. Res. & Dev. Labs, Ft. Belvoir, Va.
Lubin, Ardie	WRAMC, Washington, D. C.
Lucas, H. L.	North Carolina State College
Lundy, Hazel E.	Springfield Armory, Springfield, Mass.
Mallios, W. S.	Booz, Allen Applied Research, Inc., Bethesda, Md.
Maloney, Clifford J.	National Institutes of Health, Bethesda, Md.
Mandelson, Joseph	Edgewood Arsenal, Md.
Marsaglia, George	Boeing Co., Seattle, Washington
Marshall, Jack A.	Univ. of Chicago, Chicago, Ill.
Mellinger, J. U.	U. S. Army Personnel Research & Consulting Branch, Washington, D. C.
Mountjoy, Garrard	Baldwin Electronics Inc., Little Rock, Ark.
Nielson, Howard C.	Dugway Proving Grounds, Dugway, Utah
Noreiko, Herbert E.	USAADEA, Ft. Meade, Md.
Olkin, Ingram	Stanford University, Stanford, Calif.
Peterson, Donald	Sandia Corp., Albuquerque, N. M.
Pierce, Bruce	Lake City Army Ammunition Plant, Independence, Mo.
Podolsky, Benjamin	Frankford Arsenal, Philadelphia, Pa.
Polanecsly, Robert	Frankford Arsenal, Philadelphia, Pa.
Prairie, Richard R.	Sandia Corp., Albuquerque, N. M.
Proschan, Frank	Boeing Scientific Research Labs, Seattle Washington
Purtell, John C.	Watervliet Arsenal, Watervliet, N. Y.
Randolph, Mrs. Lila K.	Food & Drug Adm, LDA, Wash., D. C.
Rechtschaffner, R. L.	Rocketdyne Div., NAA, Inc., Calif.
Resnic, Burton	USACIRDL, Ft. Monmouth, N. J.
Rider, Paul L.	Wright-Patterson Air Force Base, Ohio
Riggs, Charles W.	Ft. Detrick, Md.
Rolle, Sylvan D.	Frankford Arsenal, Philadelphia, Pa.
Ryan, John	LTV Div., Warren, Mich.
Saboe, John C.	International Resistance Co., Philadelphia, Pa.
Scamurra, Robert	Cornell Aeronautical Lab., Buffalo, N. Y.
Schlenker, George	U. S. Army Weapons Command, Rock Island, Illinois
Schmidt, T. W.	U. S. Army Research Office, Durham, N. C.

Schnell, Raymond B.	Edgewood Arsenal, Md.
Selig, Seymour	Office of Naval Research, Washington D. C.
Selman, Jerome H. N.	U. S. Army Munitions Command, Dover, N. J.
Sloane, Harry S.	Dugway Proving Grounds, Dugway, Utah
Smith, R. S	RAC, McLean, Va.
Sterner, Bruce	BRL, Aberdeen PG, Md.
Sumner, N. R., Jr.	RAC, McLean, Va.
Suta, Benjamin E.	Hercules Powder Co., Salt Lake City, Utah
Tiedemann, Harold J.	Pratt & Whitney Aircraft, Fla.
Tietjen, Gary L.	Hercules Powder Co., Salt Lake City, Utah
Tingey, Henry B.	BRL, Aberdeen PG, Md.
Ungar, Andrew	IIT Research Institute, Chicago, Ill.
Vivona, Stefano	WRAMC, Washington, D. C.
Wampler, Roy H.	National Bureau of Standards, Washington D. C.
Weaver, L. A.	Honeywell, St. Petersburg, Fla.
Weinert, Miss Aria	RAC, McLean, Va.
Weinstein, Joseph	USAERDL, Ft. Monmouth, N. J.
Weintraub, Gertrude	Picatinny Arsenal, Dover, N. J.
Weiss, Lionel	Wisconsin
Wells, William T.	Patrick AFB, Fla.
Wilburn, Nicholas	USACIRDL, Ft. Monmouth, N. J.
Wilks S. S.	Princeton Univ., Princeton, N. J.
Williamson, Arthur C.	Applied Physics Lab., Silver Spring, Md.
Wolman, William	NASA Goddard Space Flight Center, Greenbelt, Md.
Wright, Mrs. Julia B.	U. S. Naval Propellant Plant, Md
Zimmer, W. J.	Sandia, Albuquerque, N. M.
Zundel, Evan B.	Dugway Proving Grounds, Dugway, Utah

ON-POST ATTENDEES

Alley, Bernard J	Prop. Lab, R & D
Anderson, James B.	P&P
Anderson, M. B.	S&M Dir.
Atchison, Martin W.	MSFC
Bayliss, James R.	Dir. of Msl Intel
Beckoel, Samuel E.	Dir. of Msl Intel
Black, Robert O.	P&P Dir.
Bowers, Daniel E.	Zeus Project Office
Brockway, Stanley J.	Zeus Project Office
Burrows, Dale L.	MSFC

Camp, Dennis W.	MSFC
Carter, A. L.	R&D
Carter, William W.	MICOM
Cason, F. A.	Zeus Project Office
Clinton, Rankin	Dir. of Msl Intel.
Clugston, Paul J.	R&D
Collier, B. D.	R&D
Colbert, Eleanor	R&D
Comus, N. L.	R&D
Conard, Robert G.	R&D
Cox, Charles D.	R&D
Cremin, John D.	Zeus Project Office
Crowson, Herbert	Zeus Project Office
Cummings, Robert E.	MSFC
Daniels, Glenn E.	MSFC
Davis, Alonzo J.	MSFC
Dean, H. A.	R&D
DeBerry, Deanna	Brown Engineering
DeJonge, Clark	Zeus Project Office
Dickey, Lyle R.	MSFC
Dihm, Henry A.	R&D
Dillard, Richard A.	R&D
Essick, Martin	Rohm-Haas
Eulitz, Werner R.	MSFC
Ewart, Wade	R&D
Fagan, James J.	R&D
Fisher, Perry	AMSC Calibration Center
Foshee, Charles E.	Brown Engineering
Fulton, Don	R&D
Fultz, James M.	Zeus Project Office
Gale, John B.	MSFC
Garrard, Phillip H.	R&D
Graham, Donald	R&D
Gudzent, Dietrich E.	R&D
Hadaway, William M.	R&D
Hammett, C. E.	Lance Project Office
Hanks, William R.	Zeus Project Office
Hardenburgh, Miles R.	P&P Dir.
Harrell, William B.	MSFC
Harris, Robert A.	P&P Dir.
Hasseltine, Carroll L.	MSFC
Heithecker, H. A.	R&D
Henricksen, Howard H.	R&D

Hewitt, H. H.	R&D
Hill, Harry L.	MSFC
Hoard, Earl G.	MSFC
Howerton, John W.	R&D
Hussey, C. W.	R&D
James Robert E.	R&D
Johnson, Lloyd J.	R&D
Johnson, W. B.	R&D
Katz, Lester	MSFC
Kesting, Lawrence W.	Prop Lab, R&D
Keyserling, Ben H.	Lance Project Office
Kittleberger, Kenneth	R&D
Langston, Russell L.	R&D
Lehnigk, Siegfried H.	R&D
Lewis, Will A.	R&D
Lifsey, J. David	MSFC
Link, D. T., Jr.	Zeus Project Office
Lishman, Sidney	Thiokol
Lucas, Kirk S., Jr.	MSFC
Lynch, Johnnie A.	Dir. of Msl Intel
Mayfield, W. K.	Lance Project Office
McClay, LeRoy	Mauler Project Office
McCombs, Guy C.	Lance Project Office
McDaniel, John L.	R&D
Miller, James C.	R&D
Miller, James R., III	AMSC Calibration Center
Miltewede, Edgar	Thiokol
Moore, Thomas M.	MICOM
Morgan, Michael T.	R&D
Morrow, Miian B.	R&D
Mulder, John H.	S&M Dir
Murray, Hubert, Jr.	R&D
Nordstrom, A. H.	MSFC
Olszewski, Joseph V.	Zeus Project Office
O'Neil, Thomas D.	R&D
Onkst, James R.	Mauler Project Office
Persons, Harold P.	R&D
Rees, Daniel M.	R&D
Rhodes, S. E.	Dir. of Msl. Intel.
Ross, J. M.	Zeus Project Office
Shannon, R. E.	MSFC
Shepherd, Daniel F.	R&D
Shriver, Edward L.	MSFC

Design of Experiments

Smack, R. B.	AMSC Calibration Center
Smith, Norman H.	P&P Dir.
Smith, Thomas D.	Zeus Project Office
Spears, L. E.	MSFC
Steele, William H.	Thiokol
Steinberg, Alvin	MSFC
Storey, Billy M.	R&D
Sullivan, H. D.	Zeus Project Office
Taylor, James M.	R&D
Thoman, Charles A.	Dir. of Msl Intel.
Thomson, John D.	R&D
Vincent, Harry F.	R&D
Ward, Clyde	R&D
White, Charles E.	Zeus Project Office
White, Nils	R&D