AROD R 61-2

U. S. ARMY RESEARCH OFFICE-DURHAM

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

U. S. ARMY RESEARCH OFFICE-DURHAM
BOX CM, DUKE STATION
DURHAM, NORTH CAROLINA

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 61-2
December 1961

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

Sponsored by the Army Mathematics Steering Committee

conducted at

The Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland
19-21 October 1960

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

TABLE OF CONTENTS

---

*This paper was presented at the Conference. It carries a security classification of CONFIDENTIAL. It is being published in a classified appendix to this report.

**This paper was presented at the Conference. It does not appear in these proceedings.

TABLE OF CONTENTS (Cont'd)

---

*This article is being issued in a classified security (SECRET) appendix of this technical manual.

**This paper was not presented at the Conference. It is not published in these proceedings.

*This paper was presented at the Conference. It does not appear in these proceedings.

TABLE OF CONTENTS (Cont'd)

*This paper presented by title. It does not appear in these proceedings.

TABLE OF CONTENTS (Cont'd)

Page

*This paper presented by title. It does not appear in these proceedings.

FORWORD

The Army Mathematics Steering Committee initiated the present series of conferences in October 1955. It is the intent of this Committee that these Design Conferences afford an opportunity to statistical design specialists and Army research, development and testing personnel to get together and exchange views and experiences in this rapidly growing field. It is also the intent that through invited addresses and special panel discussions many of the new developments in the theory of Statistical Design and Analysis of Experiments be brought to the attention of Army scientists; they can then make use of these new theories to help solve some of their complicated design problems.

It is of interest to note that the host of this Conference, the Ballistic Research Laboratories, has for many years recognized the importance of mathematical statistics and actively applied the methods in Army research and development. In 1926 Dr. L. S. Dederick derived in an unpublished manuscript the probability distribution of the sample range. In 1936 General (then Captain) Leslie E. Simon formed within the Ballistic Research Laboratories a group essentially concerned with scientific sampling. At about this same time, he, with Colonel H. H. Zornig, published a paper entitled "The Proposed System of Surveillance of War Reserve Ammunition." Among the many BRL reports in the field of statistics two papers by R. H. Kent, one issued in 1938 on "The Most Economical Sample Size" and the other in 1940 on "The Estimation of the Probable Error from Successive Differences" are representative of early work and served to stimulate additional research in this field. Other early Ordnance contributions include the Ordnance Sampling Inspection Tables. These were very important during World War II and were the forerunners of the tables standardized by the Department of Defense as Military Standard 105A. Mathematical Statistics has been, and is, playing an important role in the continuing research, development and testing activities at the Aberdeen Proving Ground, and the above-mentioned papers are but a few of the many contributions that have been made, and are being made, to this field by the scientists at the Ballistic Research Laboratories.

The five invited hour addresses at the Sixth Design Conference were delivered by F. J. Anscombe, W. S. Conner, J. R. Duffett, J. E. Jackson, and W. J. Youden. Residuals, experimental designs, reliability, and multivariate analysis were the topics discussed by the first four of these speakers. W. J. Youden, the banquet speaker, talked on "The Enduring Values". A panel discussion on "Common Pitfalls in the Design and Analysis of Experiments" was organized and chairmanned by G. E. P. Box. The members of his panel were C. Daniel, J. S. Hunter, W. J. Youden, and M. Zelen. In addition to these addresses, ten papers were presented in the Clinical Sessions, fifteen in the Technical Sessions, and eight by title. Specialists in the Clinical Sessions were asked to discuss experimental designs in the areas of tolerance and calibration problems, optics, bomb salvos, fatigue limits, missile safety, and multivariate analysis. In the Technical Sessions, personnel management, simulation, trajectory analysis, aerosol chamber data, nerve gas experiments, reliability of weapon systems, and response surface analysis were but a few of the topics that were considered.

The Sixth Conference was attended by 115 registrants and participants from 58 organizations outside of the Ballistic Research Laboratories. In addition, 71 staff members and other personnel of the host organization were present. Speakers and panelists came from Booz, Allen Applied Research, Inc., Canadian Army Operational Research Establishment, Cornell University, Defence Research Medical Laboratories (Canada), Eastman Kodak Company, General Analysis Corporation, Hercules Powder Company, National Bureau of Standards, Princeton University, Research Triangle Institute, Space Technology Laboratories, Inc., University of California, University of Chicago, University of Georgia, University of Maryland, University of Michigan, University of Wisconsin, Virginia Polytechnic Institute, and 15 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 H. Weber, the Commanding General of the Aberdeen Proving Ground, and Colonel J. P. Hamill, Director of the Ballistic Research Laboratories, for making such excellent facilities available for the Conference; and to Dr. Frank E. Grubbs who served as Chairman on Local Arrangements. Thanks are due many others at the Laboratories for the time and the help they gave the participants. Of these, Mr. O. P. Bruno and Major Joseph E. Sowa deserve special mention. They handled many of the local details for Dr. Grubbs and organized the interesting tour of the local facilities.

Finally, the Chairman wishes to express his appreciation to the Advisory Committee: G. E. P. Box, F. G. Dressel (Secretary), Frank E. Grubbs, Boyd Harshbarger, Clifford J. Maloney, J. S. Hunter, and Marvin Zelen for their help in organizing the program of the Conference, and especially to Dr. Dressel for coordinating the Conference program and steering these Proceedings through publication.

S. S. Wilks
Professor of Mathematics
Princeton University

PROGRAM

SIXTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

19 - 21 October 1960

Wednesday, 19 October

REGISTRATION: 0845 - 0945 (Eastern Daylight Saving Time)

Theater No. 1, Aberdeen Proving Ground

GENERAL SESSION I: 0945 - 1215 - Theater No. 1

Calling of Conference to Order:

Dr. F. E. Grubbs, Local Chairman

Welcome:

Brigadier General J. H. Weber, Commanding General, Aberdeen Proving Ground

Introduction:

Colonel J. P. Hamill, Director of the Ballistic Research Laboratories

Chairman:

Professor S. S. Wilks, Princeton University

Invited Papers:

Reliability

Dr. James R. Duffett, Space Technology Laboratories, Inc.

Examination of Residuals

Professor F. J. Anscombe, Princeton University

At 1215 buses leave Theater No. 1 for the Chesapeake

LUNCH: 1230 - 1400 - Chesapeake

Wednesday Afternoon

There will be three Technical Sessions and one Clinical Session conducted Wednesday afternoon. Technical Sessions I and II will be held concurrently from 1400 to 1500. From 1515 to 1645 Technical Session III and Clinical Session A will be running concurrently. The security classification of the first paper in Technical Session III is SECRET. No clearances are required for the other papers given on Wednesday.

TECHNICAL SESSION I: 1400 - 1500 - Chesapeake - Room A

Chairman: Joseph Weinstein, U. S. Army Signal Research and Development Laboratory

A Simulation Error-Model for an Airborne Target Location System - E. Biser and John Beckmann, U. S. Army Signal Research and Development Laboratory

Analysis of Some Trajectory Measuring Instrumentation Systems - O. L. Kingsley, Range Instrumentation Division, White Sands Missile Range

TECHNICAL SESSION II: 1400 - 1500 - Chesapeake - Room B

Chairman: Clifford J. Maloney, U. S. Army Biological Warfare Laboratories

A Trial Comparing Certain Side Effects of Two Nerve Gas Antidotes, Using Human Subjects - C. A. de Candole, Defence Research Medical Laboratories, Downsview, Ontario, and B. A. Richardson, Canadian Army Operational Research Establishment, Ottawa, Canada

An Application of the Exponential Hazard Model to Aerosol Chamber Trial Data - Theodore W. Horner, Booz, Allen Applied Research, Inc.

COFFEE: 1500 - 1515 - Chesapeake

CLINICAL SESSION A: 1515 - 1645 - Chesapeake - Room A

Chairman: Ralph E. Brown, Frankford Arsenal

Panelists: G.E.P. Box, The University of Wisconsin

A. C. Cohen, Jr., The University of Georgia

W. S. Connor, The Research Triangle Institute

H. A. David, Virginia Polytechnic Institute

J. R. Duffett, Space Technology Laboratories, Inc.

Calibration of a Zinc Sulfide Particle Detector - John E. Malligo, Methods Research Section, MR & AE Branch, Technical Evaluation Division, Chemical Corps Biological Laboratories

Effects of Aiming Point Patterns on Bomb Salvo Target Coverage - R. D. Doner, Systems Analysis Laboratory, OML Division, Army Rocket and Guided Missile Agency, Redstone Arsenal

The Tolerance Structure of Complex Systems - William S. Agee, Flight Simulation Laboratory, White Sands Missile Range

TECHNICAL SESSION III: 1515 - 1645

The first paper in this session carries a security classification of SECRET and will be held in Room 259, BRL Bldg. 328. Transportation from the Chesapeake to BRL Bldg. 328 will be provided at 1505 hours.

The second paper will be given in Room B, the Chesapeake, beginning at 1610 hours. Transportation from BRL Bldg. 328 back to the Chesapeake will be provided at 1600 hours.

Chairman: F. Howard Forsyth, Office, Chief of Ordnance, Department of the Army

BRL Bldg. 328

Allocation of Resources and Military Worth - Walter E. Cushen, Operations Research Office, The Johns Hopkins University

TECHNICAL SESSION III (Cont'd)

Room B, Chesapeake

An Experiment in Personnel Management Evaluation - Richard R. Blough, Statistical Research Center, The University of Chicago

Wednesday Evening: The cocktail lounges at the Chesapeake and the Main Club are open from 1630 to 2300 hours. The dining room at the Main Club is open from 1800 to 2000 hours.

Buses will take conferees to motels or Main Club.

Thursday, 20 October

Clinical Session B carries a security classification of SECRET. It and Technical Session IV will run from 0900 to 1015. Clinical Session C and Technical Session V scheduled from 1030 to 1230 complete the morning phase of the program. In the afternoon Technical Sessions VI and VII run concurrently from 1400 to 1440. General Session II will be a panel discussion and is timed from 1500 to 1645.

TECHNICAL SESSION IV: 0900 - 1015 - Chesapeake - Room A

Chairman: Gertrude Weintraub, Missile Warhead and Special Projects Laboratory, Picatinny Arsenal

Reliability of Weapon Systems Estimated from Component Test Data Alone - Henry DeCicco, U. S. Army Ordnance Special Weapons-Ammunition Command

Performance of Propellants Evaluated by Tensile and Ballistic Tests - Niles White, Propellant Branch, Propellant Laboratory, ARGMA, and Boyd Harshbarger, Virginia Polytechnic Institute

CLINICAL SESSION B: 0900 - 1015 - BRL Bldg. 328, Room 259

Security Classification - SECRET

(Transportation from the Chesapeake to BRL Bldg. 328 will be available [illegible] hours. Bus from BRL Bldg. 328 to Chesapeake at 1015 hours)

CLINICAL SESSION B: (Cont'd)

Chairman: Edward W. Chittenden, Diamond Ordnance Fuze Laboratories

Panelists: R. M. Eissner, Ballistic Research Laboratories

Walter Foster, U. S. Army Biological Warfare Laboratories

J. R. Johnson, Ballistic Research Laboratories

Clifford J. Maloney, U. S. Army Biological Warfare Laboratories

S. S. Wilks, Princeton University

Marvin Zelen, University of Maryland

Multivariate Analysis for Project Michigan Experiments - Emil H. Jebe, The University of Michigan, Willow Run Laboratories, Operations Research Department

Problems in the Analysis and Interpretation of Project Michigan - William A. Brown and Emil H. Jebe, The University of Michigan, Willow Run Laboratories, Operations Research Department

COFFEE: 1015 - 1030 - Chesapeake

CLINICAL SESSION C: 1030 - 1230 - Chesapeake - Room A

Chairman: Elizabeth Scott, University of California, Berkeley

Panelists: R. J. Anscombe, Princeton University

Robert E. Bechhofer, Cornell University

H. A. David, Virginia Polytechnic Institute

J. Edward Jackson, Eastman Kodak Company

Jerzy Neyman, University of California, Berkeley

CLINICAL SESSION C: (Cont'd)

Computation of Expected Resolution Improvement Factor - Chandler Stewart, Mine Detection Branch, Engineering Research and Development Laboratories

Panoramic Viewing Utilizing Hyperbolic Ellipsoidal Reflecting Optics - Donald W. Rees, Physical Sciences Laboratory, U. S. Ordnance Tank - Automotive Command, Detroit Arsenal

Some Statistical Problems Related to Missile Safety - Paul C. Cox, Reliability and Statistics Office, Ordnance Mission, White Sands Missile Range

TECHNICAL SESSION V: 1030 - 1230 - Chesapeake - Room B

Chairman: Lawrence Langwell, Warhead and Special Projects Laboratory, Picatinny Arsenal

Design for Weighing Calibrations - Neilson, Hercules Powder Company, Magna, Utah

Response Surface Analysis as Related to Repellant Research - D. G. Boyle and E. A. Perlman, Hercules Power Company, Magna, Utah

Application of Factorial Experiment and Box Technique to Ballistic Devices - D. J. Katsanis and C. L. Fulton, Frankford Arsenal

LUNCH: 1230 - 1400 - Chesapeake

TECHNICAL SESSION VI: 1400 - 1440 - Chesapeake - Room A

Chairman: A. Bulfinch, Quality Assurance Division, Picatinny Arsenal

On the Problem of Negative Estimates of Variance - W. S. Thompson, Jr., University of Delaware, and J. R. Moore, Surveillance Branch, Weapon Systems Laboratory, Ballistic Research Laboratories

TECHNICAL SESSION VII: 1400 - 1440 - Chesapeake - Room B

Chairman: S. A. Krane, General Analysis Corporation, Dugway Proving Ground Office

"Build-Up" of Single Point Source Data - R. F. White, General Analysis Corporation, Dugway, Utah

COFFEE: 1440 - 1500 - Chesapeake

GENERAL SESSION II: 1500 - 1645 - Chesapeake - Room A

Panel Discussion on Common Pitfalls in the Design and Analysis of Experiments.

Chairman: G.E.P. Box, The University of Wisconsin

Panel Members: Cuthbert Daniel, Private Consultant

J. S. Hunter, Mathematics Research Center, The University of Wisconsin

W. J. Youden, National Bureau of Standards

Marvin Zelen, The University of Maryland

After General Session II, buses will take conferees to motels or Main Club for cocktails and dinner.

SOCIAL HOUR: 1730 - 1830 - Main Club, Officers' Open Mess

DINNER: 1830 - Main Club

Chairman: Frank E. Grubbs, Ballistic Research Laboratories

Speaker: W. J. Youden, National Bureau of Standards - "The Enduring Values."

Friday, 21 October

Technical Session VIII and Clinical Session D are scheduled for 0900 - 1015. General Session III is called for 1030 and will run until 1230. After lunch there will be conducted tours of the Ballistic Research Laboratories.

TECHNICAL SESSION VIII: 0900 - 1015 - Chesapeake - Room A

Chairman: Badrig M. Kurkjian, Diamond Ordnance Fuze Laboratories

Some Tests for Outliers - C. P. Quesenberry and H. A. David, Virginia Polytechnic Institute

Note on Precision of Graded vs. All-or-None Response in Bioassay - Francis M. Wadley, U. S. Army Chemical Corps Biological Laboratories

CLINICAL SESSION D: 0900 - 1015 - Chesapeake - Room B

Chairman: T.N.E. Grenville, Research and Engineering Division, Department of the Army, Office of the Quartermaster General

Panelists: R. E. Bechhofer, Cornell University

O. P. Bruno, Ballistic Research Laboratories

A. C. Cohen, Jr., The University of Georgia

Boyd Harshbarger, Virginia Polytechnic Institute

J. S. Hunter, Mathematics Research Center

Comparison of Field Wear and Laboratory Testing of Fabrics for Military Garments - William S. Cowie, Textile Clothing and Footwear Division, QM R & E Center Laboratories, Quartermaster Research and Engineering Command

Estimation of Condemnation Limits from Limited Fatigue Runout Data on Full Scale Components - J. P. Purtell and C. W. Egan, Research Branch, Watervliet Arsenal

COFFEE: 1015 - 1030 - Chesapeake

GENERAL SESSION III: 1030 - 1230 - Chesapeake - Room A

Chairman: Boyd Harshbarger, Virginia Polytechnic Institute

Development in the Design of Experiments - W. S. Connor, The Research Triangle Institute

GENERAL SESSION III: (Cont'd)

Multivariate Analysis Illustrated by Nike-Hercules
I. Separation of Product and Measurement Variability
II. Acceptance Sampling - J. Edward Jackson, Eastman Kodak Company

LUNCH: 1230 - 1400 - Chesapeake

TOURS: 1330 - Conducted tour will be initiated at the Chesapeake

SUPPLEMENTARY PROGRAM

The following papers were received too late to be considered for places on the agenda. We hope that the manuscripts of these papers will be submitted for publication in the Proceedings of this Conference. (Papers are listed in order of receipt in the Office of Ordnance Research).

A Virulence Measure for Minute Organisms - S. A. Krane, General Analysis Corporation, Dugway Proving Ground Office

Design of an Experiment for the Most Efficient Conduct of Safety, Reliability and Performance Tests of Fuzes in the Design and Development Stages - Gertrude Weintraub, Missile Warhead and Special Projects Laboratory, Picatinny Arsenal

Design of an Experiment to Evaluate the Effects of Various Factors Affecting the Acceleration of Unconventional Fragments - Gertrude Weintraub, Missile Warhead and Special Projects Laboratory, Picatinny Arsenal

Design of the Laboratory Statistical Reliability Program for the T46E1 Warhead - Alfred Fiorentiono, Warhead and Special Projects Laboratory, Picatinny Arsenal

Design of a Laboratory Reliability Program for the XM44 Shillelagh Missile Warhead and the XM805 Fuzing System - Lawrence Langweil, Warhead and Special Projects Laboratory, Picatinny Arsenal

Reliability Prediction - A. Bulfinch, Quality Assurance Division, Picatinny Arsenal

SUPPLEMENTARY PROGRAM (Cont'd)

On the Use of Monotone Functions in Multi-Dimensional Environmental Testing - Edward W. Chittenden, Diamond Ordnance Fuze Laboratories

Asymptotically Locally Most Powerful Test for the Identity of Regressions of Variables Requiring Transformations - Jerzy Neyman and Elizabeth L. Scott, Statistical Laboratory, University of California, Berkeley

RELIABILITY*

James R. Duffett
Space Technology Laboratories, Inc.

A. BACKGROUND. Those personnel who have been involved in the flight testing of complex guided missiles are, in general, aware of the phenomenon that a much higher system reliability (R) is obtained for an essentially series arrangement of the components than would prevail if estimates of the component reliabilities (r_i's, $i = 1, 2, \ldots, n$) were substituted into the mathematical model

$$R = r_1 r_2 \ldots r_n .$$

This phenomenon has been eloquently exposed, and elucidated upon, by Frank A. Fleck of United Electro Dynamics. The model

$$R = r_1 r_2 \ldots r_n$$

is usually referred to as the "series" model; it is also known as the "cascade" or "tandem" model. By contrast, those personnel who have been concerned with the use of redundancy (i.e., the paralleling of components) in order to lower the probability of a dud (i.e., the unreliability of the payload) have observed a phenomenon which is opposite in effect to that which has been observed in the series case. Namely, a system of components arranged in parallel evidences, in general, a lower reliability (R) than would be obtained if the estimates of the component reliabilities (r_i's, $i = 1, 2, \ldots, m$) were substituted into the mathematical model

$$R = 1 - (1 - r_1)(1 - r_2) \ldots (1 - r_m).$$

This last model is usually referred to as the "redundant" or "parallel" model.

A plausible explanation for these two phenomena, which have opposite effects for the series and redundant systems, is that the variation of the environmental stresses from component to component within the same flight is small relative to the variation of the environmental stresses from flight to flight. It is believed that, in many redundant systems, the stresses associated with the m parallel components within the same flight possess such a small standard deviation that this standard deviation can be neglected; in such instances the variation attributed to the stresses is essentially the

*This is an abstract. The paper itself is being submitted for publication in TECHNOMETRICS.

standard deviation of the stresses from flight to flight. This belief is attributed to the following facts:

(1) The m parallel components are usually mounted in close physical proximity to each other and hence tend to have the same quantitative values of the stresses.

(2) The m parallel components are usually of the same type and thus are subject to the same mode of failure and hence are susceptible to he same kind of stresses.

(3) Many kinds of stresses vary considerably from flight to flight.

These same facts can prevail in the case of some series systems, e.g., the successive amplifier stages in an equipment, the electron tubes in a guidance and control package, and the relays in a black box.

W. J. Howard[1] has considered the series situation for a specific numerical value of the component reliability r and for Gaussian distributions of strengths and flight-to-flight stresses, whereas J. R. Duffett[2] has considered the parallel situation for a large range of values of the component reliability r and rectangular distributions of strengths and flight-to-flight stresses plus certain generalizations of these assumptions.

An interesting pathological example of the complete breakdown of redundancy is afforded by the following example which is presented in (2):

> The m parallel components incorporated in the same missile flight (i.e., the same end item) are subjected to exactly the same stress; the probability distribution of the stresses from flight-to-flight consists of two isolated portions; and the probability distribution of strengths is sandwiched in between the two isolated portions of the stress distribution.
>
> It is clear that the reliability R of the system is only r, i.e., $R = r$, and thus no gain in reliability is achieved by using redundancy.

---

(1) W. J. Howard, "Chain Reliability: A Simple Failure Model for Complex Mechanisms," The Rand Corporation, RM-1058, 27 March 1953.

(2) J. R. Duffett, "Some Mathematical Considerations of Redundancy", Radioplane Company, Operations Analysis Memorandum Report Number 12, 25 October 1956.

In such a situation as that described in the foregoing example, the reliability R of a system would be r, i.e., R = r, regardless of the design arrangement of the components which compose the system.

An extremely pathological example which illustrates a situation in which the opposite effect (as that discussed above) is obtained has been given by C. R. Gates and is stated as follows:

> A system is composed of 2 components. One component fails if and only if the temperature is greater than or equal 0° F, whereas the other component fails if and only if the temperature is less than 0° F. If the components are arranged in series, the reliability of the system will be zero. However, if the components are arranged in parallel, the reliability of the system will be one.

In order to simultaneously control, at acceptable levels, both the probability of a dud (or a late detonation) and the probability of a "premature", some technical personnel have proffered, as a desirable solution, the use of a "matrix" of components, arranged both in parallel and in series; i.e., the system is to consist of m parallel circuits (or branches), where each circuit has n components arranged in series(3,4). Such a system will be referred to as a parallel-series system. (Apparently for analytical simplification, m and n are, in general, set equal to each other so that the design matrix is square.) A special case of a parallel-series system is the quad. The quad is of considerable engineering interest and consists of 2 parallel circuits, each possessing 2 components arranged in series.

Another design arrangement, which tends to decrease the percentage of duds while increasing the percentage of prematures, is a series-parallel system. This arrangement consists of n circuits (or links) in series, where each circuit has m components in parallel.

B. <u>SYSTEMS STUDIED AND ASSUMPTIONS MADE</u>. In this document, the reliability of the four types of systems--(1) a simple parallel system, (2) a simple series system, (3) a parallel-series system, and (4) a series-parallel system—are derived under the following assumptions:

(3) Burton, E. B., The Martin Company. Unpublished paper on redundancy, quads, and crossing circuitry.

(4) Creveling, C. J., "Increasing the Reliability of Electronic Equipment by Use of Redundant Circuits," NRL Report 4631, 5 December 1955, Naval Research Laboratory, Washington, D. C.

(a) The strengths of the components, regardless of the flight in which they are incorporated, are independently selected from a fixed rectangular distribution.

(b) All components which are incorporated in the same flight experience exactly the same stress.

(c) The common stress, which is applied to all of the components within the same flight, is independently selected from a fixed rectangular distribution.

(d) Failure of a component occurs if, and only if, the stress imposed on it exceeds its strength.

(e) Failure of a simple series arrangement of the components occurs if, and only if, at least one of the components fails.

(f) Failure of a simple parallel arrangement of the components occurs if, and only if, all of the components fail.

(g) Failure of a parallel-series system occurs if, and only if, all of the parallel branches (circuits) fail.

(h) Failure of a series-parallel system occurs if, and only if, at least one of the links (circuits) in series fails.

I. CONCLUSIONS.

(1) It is concluded that the reliability of a system composed of m parallel components can approach, as an upper limit, the value given by the independence model, viz.,

$$R = 1 - (1 - r_1)(1 - r_2) \ldots (1 - r_m),$$

as the variation of the stresses between flights decreases relative to the variation of the component strengths.

(2) It is concluded that the reliability of a system composed of n series components can approach, as a lower limit, the value given by the independence model, viz.,

$$R = r_1 r_2 \ldots r_n,$$

as the variation of the stresses between flights decreases relative to the variation of the component strengths.

It is further concluded that, for a given level of component reliability, the reliability of a series system can be increased by decreasing the copy-to-copy variation of the strengths of the components.

(3) It is concluded that the reliability of a system composed of m parallel components can approach, as an upper limit, the value given by the independence model, viz.,

$$R = 1 - (1 - r_1)(1 - r_2) \ldots (1 - r_m),$$

as the variation of the stresses between flights decreases relative to the variation of the stresses within flights.

(4) It is concluded that the reliability of a system composed of n series components can approach, as a lower limit, the value given by the independence model, viz.,

$$R = r_1 r_2 \ldots r_n,$$

as the variation of the stresses between flights decreases relative to the variation of the stresses within flights.

(5) It is concluded that one may obtain unwarranted optimistic estimates of the reliability of a system composed of parallel components if the independence model is assumed, but not satisfied.

(6) It is concluded that one may obtain unwarranted pessimistic estimates of the reliability of a system composed of series components if the independence model is assumed, but not satisfied.

M. RECOMMENDATIONS.

(1) It is recommended that systems integration studies be made for the purpose of determining the effect of component failures on system effectiveness and consequently to classify the systems incorporated in the end item according to such categories as series, parallel, series-parallel, and parallel-series.

(2) If parallel components are to be used, then it is recommended that consideration be given to the following:

(a) The selection of components whose modes of failure are different.

(b) The incorporation of the components in such a way that the stresses associated with the components are independently selected from a fixed probability distribution regardless of the end item in which the component is incorporated.

It may be possible to implement (b) by physically separating the parallel components by sufficiently large distances, or otherwise isolating the parallel components from one another.

(c) The isolation of the parallel components from their environments.

(3) If series components are to be used, then it is recommended that consideration be given to the following:

(a) The selection of components with (approximately) the same modes of failure.

(b) The assembly of the components in such a manner that they will experience (approximately) the same environmental regime.

One method with which to implement (b) is to package the series components as a compact unit.

(4) It is recommended that the existing relevant transportation and in-flight environments which have been measured be evaluated. It is further recommended that consideration be given to the instrumentation of missiles for the purpose of obtaining additional information on the environmental conditions which are encountered by missile systems both (i) within flights and (ii) from flight to flight.

(5) It is recommended that the probability distributions of component strengths be determined, primarily by means of laboratory tests. It is further recommended that consideration be given to decreasing the variability of the strengths of components which are to be used in series systems and subsequently to maintaining (through Statistical Quality Control) the variation of component strengths at a satisfactory level.

(6) It is recommended that the formulas given in this paper be employed to serve as a guide in the calculation of system reliability.

EXAMINATION OF RESIDUALS*

F. J. Anscombe
Princeton University

During the last few years there has been a growing interest in examining the residuals after some parameters have been fitted by the method of least squares. The author first became interested in the subject in 1954 through some suggestions made by John W. Tukey [1]; and he has recently been concerned in several studies relating to residuals [2,3,4]. The purpose of this paper is to make a brief introductory sketch of some of these developments. The presentation will be in terms of a particular example. In Section 1 an experiment is described and the conventional type of statistical analysis is outlined. In Section 2 comments are made on the validity of the analysis. In Section 3 the residuals associated with the conventional analysis are examined, and suggestions are made for modifying the analysis.

1. A LATIN-SQUARE EXPERIMENT AND ITS STANDARD ANALYSIS. In Table 1 are shown a set of observations of depth of penetration of a blast driven earth rod. Ten different propelling charge lots (denoted by the letters A, B, C, ..., J) were compared on ten different sites or "plots" (shown as columns in the table), firings being made on ten different dates spread over a period of some months (the dates are termed "blocks" and shown as rows in the table), in accordance with a 10 x 10 Latin square pattern. In each cell of the Latin square (that is, at each date, on each plot) the appropriate propellant lot was tested in duplicate, and two holes were driven. Thus there were 200 readings in all. (These data have been kindly supplied by Dr. Frank E. Grubbs.)

For such a set of readings, arranged in a Latin square design, there is a standard method of statistical analysis, which a statistician is likely to follow almost without thinking. The sum and the difference of the pair of entries in every cell of the Latin square are calculated. The sum of squares of the differences is found, to obtain a within-cell estimate of error variance; the individual differences are then forgotten about. From the sums of pairs of cell entries, row, column and letter means are calculated. If we denote the sum of the two readings in the cell in the *i*th row and *j*th column by y_{ij}, then the various means to be calculated are the row means $\bar{y}_{i.}$, the column mean $\bar{y}_{.j}$, the overall mean $\bar{y}$, and the letter means, which may be denoted by $\bar{y}_{(A)}, \ldots, \bar{y}_{(J)}$. These row, column and letter means show, respectively, the effects of blocks

---

*Prepared in connection with research supported by the Office of Naval Research.

(dates), plots and propellant lots, and can be set out in three short tables, such as Table 2, relating to propellant lots. The entries in this table are in fact the letter means divided by 2, so that we have average penetrations per firing, rather than averages of sums of two penetrations.

We can obtain an overall picture of the amount of variation present in the readings by constructing the analysis of variance table shown in Table 3. For the purpose of comparing means of rows or of columns or of letters, one would take the residual mean square in the analysis of sums of cell pairs, namely 63.6, as the estimated residual variance. The estimated standard error shown in Table 2 is equal to one-half (because the letter means were divided by 2) of $\sqrt{63.6/10}$. It will be seen that there is no evidence that the propelling charge lots have any differential effect on depth of penetration. There is a marked seasonal effect, and there seems to be a plot effect also.

II. COMMENTS ON THE STANDARD ANALYSIS. The customary analysis of experimental data goes along the lines briefly indicated above. Is it satisfactory? The orthodox treatment of the data would be perfectly appropriate and valid if certain ideal conditions were satisfied. We have no reason to suppose that these conditions are ever satisfied exactly, but it may well be that they are often nearly enough satisfied for practical purposes. The ideal conditions are sometimes referred to as the assumptions underlying the analysis of variance. For the present Latin square design, they are as follows:

> IDEAL CONDITIONS. The observations are realizations of independent chance variables all normally distributed with the same variance and with means consisting of a row constant plus a column constant plus a letter constant.

Was it reasonable to analyze the observations as though these conditions were satisfied? One may question the standard statistical analysis of any body of data under the three main headings:

> (1) Are the observations trustworthy? Should they be taken at face value?

If the answer is yes,

> (2) Are the ideal conditions nearly enough satisfied to make the standard analysis acceptable?

If the answer is no, or doubtful,

(3) How should the standard analysis be modified or replaced?

In regard to (1), no observations are absolutely trustworthy. If the results are sufficiently at variance with expectation, a mistake in the observations will be strongly suspected. Sometimes it will be possible to verify directly that a mistake has occurred, and perhaps to rectify it. But even if it is not possible to repeat or check the observations, a verdict of "presumed mistake" may still seem the most reasonable, and that implies that observations will be discarded. In some cases the whole of the observation may be discarded, in other cases just one or two aberrant readings may be picked out as presumably spurious, the rest being accepted as reliable.

In regard to (2), the divers ways in which the ideal conditions could fail to be satisfied are unlimited in number. The means may fail to have the specified simple linear structure, and the deviations of the observations from the means could in principle have any stochastic character whatever. There are, however, a few types of departure from the ideal conditions that seem to be worth looking for explicityly, as being easily intellibible and possibly important.

On the subject of how far various kinds of departure from the ideal conditions invalidate the standard method of analysis, not as much is known as one might wish. (This topic is reviewed in the last chapter of [5] and in chapter 5 of [6] .) If the ideal conditions were exactly satisfied, the standard analysis would be the most convenient and intelligible and efficient possible. In so far as the ideal conditions are not satisfied, the standard analysis will be in some degree inappropriate and perhaps misleading. For large enough departures from the ideal conditions, it would be preferable to perform some sort of modified or alternative analysis, but that means further computation and possibly less easily intelligible results.

Examination of residuals is a valuable method (though not the only possible one) of detecting isolated aberrant readings and of measuring several sorts of systematic departures from the ideal conditions. That is what this paper is about--obtaining information concerning conformity with the ideal conditions, which is a necessary step before criticizing and possibly improving the original analysis.

III. RESIDUALS AND FITTED VALUES. Corresponding to any observation, the "fitted value" is the least-squares estimate of the mean value of the hypothetical chance distribution from which the observation was drawn, according to the ideal conditions. The "residual" is the difference between the observation and the fitted value.

Our example of the penetration data has the peculiarity that there are two observations in every cell of the Latin square. One might examine the residuals corresponding to these 200 individual readings. However, for the purpose of comparing rows or columns or letters, it is the 100 cell sums y_{ij} that are relevant, and which we should hope would satisfy the ideal conditions fairly closely. So we now consider these as the effective observations and form the corresponding 100 fitted values (Y_{ij}) and residuals (z_{ij}). Each fitted value consists of the sum of the relevant row mean, column mean and letter mean, minus twice the overall mean. For example, corresponding to y_{11}(= 136.875) we have

$$\bar{y}_{1\cdot} = 123.69,\ \bar{y}_{\cdot 1} = 128.73,\ \bar{y}_{(F)} = 129.36,\ \bar{y} = 125.30,$$

and hence the fitted value is

$$Y_{11} = \bar{y}_{1\cdot} + \bar{y}_{\cdot 1} + \bar{y}_{(F)} - 2\bar{y} = 131.18$$

and the residual is

$$z_{11} = y_{11} - Y_{11} = 5.69.$$

When the fitted values and the residuals corresponding to the one-hundred cell sums y_{ij} have been calculated, the scatter diagram shown in Figure 1 can be plotted. Each point corresponds to one of the cells of the Latin square, and has the fitted value as abscissa and the residual as ordinate.

Provided that no error has been made in the calculation, the scatter diagram must have the properties

$$\sum_{ij} z_{ij} = \sum_{ij} z_{ij} Y_{ij} = 0;$$

that is, the average of the ordinates must be zero, and the coefficient of linear regression of the residuals on the fitted values must also be zero. If the ideal conditions are exactly satisfied, the diagram should have the further properties, that the residuals appear in aggregate to be normally distributed, and that they show no dependence of any sort on the fitted values.

In the present case one peculiarity is immediately noticeable, that the residuals have a negatively skew distribution; they range from + 11 to − 19, roughly. Another peculiarity is easily perceived when one looks for it, namely, that the vertical dispersion of the points is greater on the left side of the diagram than on the right. Thus the three largest positive residuals and the six largest negative residuals are all associated with fitted values that are smaller than $\bar{y}$ (= 125.3). These features of the scatter diagram suggest, respectively, that if the observations are thought of as having a chance distribution, then the distribution must be negatively skew rather than normal, and that the variance, instead of being constant, is smaller when the cell mean is greater.

It is not the case here that any one residual is so much larger in magnitude than the others as to suggest a gross error or blunder in the corresponding y. That is, there is no clear outlier, and we are not tempted to reject any observation as spurious.

Another effect which may sometimes be seen in such a scatter diagram, but is not seen here, is a quadratic or curvilinear regression of the residuals on the fitted values. We remarked above that there is necessarily no linear regression of residuals on fitted values, but a nonlinear regression is not precluded. Such a regression can arise if the effects of rows, columns, and letters are not additive, in the way stated in the ideal conditions. In fact here only the rows have a substantial effect. Columns seem to have a rather slight effect, and letters no effect at all. There is therefore not much scope for nonadditivity, and it is not surprising that no curvilinear regression is noticable.

To supplement the visual inspection of the scatter diagram, one may calculate various measures of departure from the ideal conditions, and make significance tests and other assessments. Relevant formulas are given in [3]. For example, in the present case, one may estimate a measure of skewness ($\sqrt{\beta_1}$ in Karl Pearson's notation, γ_1 in R. A. Fisher's) of the presumed common distribution of deviations of the y's from the linear cell means. The estimate comes out at −0.96, with standard error under the full ideal conditions roughly 0.39.

To sum up, inspection of the residuals and their relation with the fitted values suggests that the deviations of the y's from the cell means have a skew distribution with nonconstant variance. The physical cause for this is no doubt that occasionally, perhaps because of stones, the ground is so hard that the penetration is considerably short of the mean. On the other hand, there is no reason why penetrations much in excess of the mean should be observed, and in fact because the rocket motor is broader than the rod below it there is an effective upper limit to the depth of earth penetration achievable - though there is no definite evidence in these observations of any piling up of frequency at such a limit.

Do these phenomena matter, and are Tables 2 and 3 misleading? The correct answer is probably no, because the violation of the ideal conditions is not extreme. But if computations are done automatically and with little personal effort, it is worth while to try transforming the observations in some way to improve their conformity with the ideal conditions. Raising the readings to a power greater than 1 is suggested. This would be particularly natural if there were theoretical or experimental evidence that the propelling charge required to achieve a given average penetration was proportional to some power of the penetration; it would then be natural to use that power here. There are too few observations to fix an appropriate power closely, from examination of the observations only – a rather high power, sixth or seventh, is suggested.

Let us consider, conservatively, raising the observations to the fourth power. That is, all 200 original readings are raised to the fourth power and divided (for convenience) by 10^6; and then the previous analysis is repeated. We find that the skewness measure calculated from the residuals is now about halved (-0.54), and most of the regression effect of variance on cell means has disappeared. In place of Table 3 we have Table 4. The variance ratios are not vastly different from those of Table 3. The block and plot effects have emerged a little more distinctly, and there is still no indication of real difference between the propellant lots. Table 4 may be judged to be a fairer summary of the effects present than Table 3, but evidently our conclusions will not be much different whichever we examine. It would be desirable to investigate penetration records from a number of other trials before venturing on a general recommendation for the statistical analysis of such data.

I am indebted to Mr. John J. Simon and Mr. Carl E. Jukkola for carrying out the computations.

TABLE II

DEPTH OF PENETRATION FOR BLAST DRIVEN EARTH ROD
(LATIN SQUARE DESIGN)

Blocks	Plots 1	2	3	4	5	6	7	8	9	10
1	F 68 7/8 68	B 58 1/4 57	J 60 1/8 67 1/2	G 59 7/8 60 1/4	E 57 1/2 60 1/2	C 55 3/4 66 3/4	H 67 3/4 59 1/4	I 60 1/4 60 3/8	A 55 73	D 61 3/4 59 1/8
2	E 48 7/8 60 1/2	C 59 49 1/2	H 64 1/4 58 1/2	I 60 3/8 61 1/8	A 57 1/4 60 1/8	D 60 1/2 60 1/8	F 60 7/8 60 7/8	B 63 1/4 60 1/2	J 60 1/8 72 7/8	G 48 3/8 48 7/8
3	A 66 3/8 67 7/8	D 55 1/8 50 1/2	F 59 1/8 66	B 65 7/8 60 1/2	J 65 1/8 67 3/8	G 63 1/4 69 1/2	E 61 3/4 72 3/4	C 70 1/4 65 3/4	H 66 5/8 68 1/2	I 63 61
4	G 68 3/4 70 1/2	E 58 5/8 56 1/8	C 70 1/8 73 7/8	H 66 3/4 63 7/8	I 72 7/8 68 1/8	A 69 5/8 69 7/8	D 70 1/8 64	F 65 3/8 69 3/8	B 73 1/8 64 7/8	J 69 7/8 68 1/8
5	I 61 1/8 59 1/4	A 62 55 1/4	D 65 1/2 65 5/8	F 62 62 1/8	B 66 1/8 63 1/4	J 61 60 1/4	G 65 1/4 62	E 64 1/8 66 5/8	C 67 1/4 70	H 60 1/4 62 1/4
6	H 66 1/8 66 5/8	I 64 5/8 61 1/2	A 61 1/4 51 1/8	D 62 7/8 64 5/8	F 63 1/2 66 5/8	B 65 1/4 52	J 56 1/8 61 3/8	G 54 1/4 56	E 61 1/8 60 1/8	C 67 1/2 62 1/2
7	J 55 1/8 65 1/4	G 59 3/8 61	E 63 1/2 62 5/8	C 51 52	H 49 1/2 52	I 63 1/4 71 3/4	A 64 1/4 54 5/8	D 67 3/4 60 1/2	F 66 1/4 58	B 64 3/8 64 7/8
8	D 62 3/4 65	F 62 68	B 64 6/8 56 3/4	J 63 1/8 55	G 67 1/8 67 1/2	E 67 1/2 61 7/8	C 65 1/8 65 11/16	H 66 3/4 65 3/4	I 67 1/4 62 1/8	A 64 3/4 66 5/8
9	B 67 3/8 62 1/4	J 63 1/2 63 1/4	G 58 3/4 61 3/8	E 60 1/2 61 3/16	C 54 3/4 55 1/8	H 68 5/8 67 7/8	I 67 1/8 66 5/8	A 63 67 3/8	D 63 1/8 64 1/4	F 66 5/8 64 3/4
10	C 68 3/4 67 7/8	H 60 1/4 54 1/4	I 54 1/4 58 1/8	A 50 3/4 66	D 51 3/8 62 3/8	F 70 1/4 65	B 58 56	J 62 1/4 61 1/4	G 65 1/2 66 7/8	E 64 7/8 66 1/4

Capital Letters - Propelling Charge lots (A-J).
Plots - Plots of ground on which test was conducted (Each plot about 12' x 15').
Blocks - Firings conducted during the same period of time.

NOTE: Two observations per cell representing two propelling charges from a lot. Holes are 36" apart.

TABLE 2

Mean penetration per propellant lot

Lot	A	B	C	D	E	F	G	H	I	J
Penetration	62.3	62.2	62.9	61.8	61.8	64.7	61.7	62.8	63.2	62.9

Estimated Standard error of each mean - 1.26.

TABLE 3

Analysis of variance of penetrations

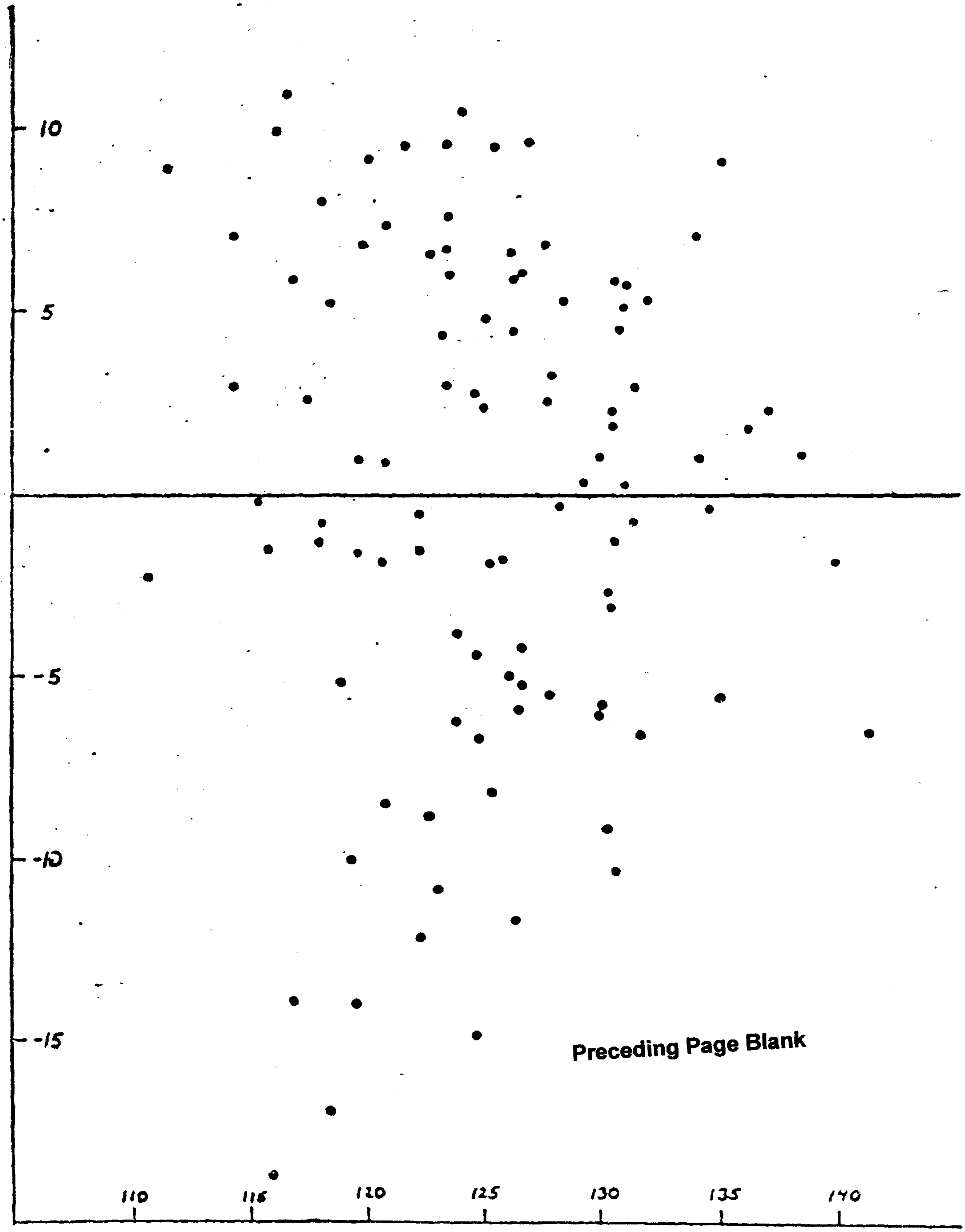
	Degrees of Freedom	Sum of squares	Mean squares
Analysis of sums of cell pairs			
Between blocks	9	2225	247.3
Between plots	9	1376	152.8
Between propellant lots	9	283	31.4
Residual	72	4579	63.6
Analysis of differences of cell pairs			
Total	100	3076	30.8

TABLE 4

Analysis of variance of penetrations after fourth-power transformation

	Degrees of freedom	Sum of squares	Mean squares
Analysis of sums of cell pairs			
Between blocks	9	2164	240.4
Between plots	9	1322	146.8
Between propellant lots	9	256	28.4
Residual	72	3787	52.6
Analysis of differences of cell pairs			
Total	100	2950	29.5

Figure 1



Preceding Page Blank

REFERENCES

[1] F. J. ANSCOMBE and J. W. TUKEY. The criticism of transformations (abstract). Journal of the American Statistical Association, 50 (1955), 566.

[2] F. J. ANSCOMBE. Rejection of outliers. Technometrics, 2 (1960), 123-147.

[3] F. J. ANSCOMBE. Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press. Vol. 1 (in the press).

[4] F. J. ANSCOMBE and J. W. TUKEY. The examination and analysis of residuals.

[5] H. SCHEFFE. The Analysis of Variance, Wiley (1959).

[6] R. L. PLACKETT. Principles of Regression Analysis, Oxford University Press (1960).

ANALYSIS OF SOME TRAJECTORY MEASURING INSTRUMENTATION SYSTEMS

Oliver Lee Kingsley
Range Instrumentation Development Division,
White Sands Missile Range

I. INTRODUCTION. The purpose of the analysis was to isolate the random and bias errors inherent in the trajectory instrumentation systems currently in use at WSMR. The preliminary analysis presented here is but the first to be made on a series of missile flights.

Data from the first flight is still undergoing further study.

The second flight of the current series was made September 1960 and the analysis will commence as the data becomes available.

The user is constantly requiring more and better data. These requirements must be met as the missile systems become more refined. It is expected that these tests can lead to improved instrumentation systems at WSMR and other ranges.

The instrumentation systems used for the initial launching were: Ballistic Camera, Askania Cine-Theodolite, DOVAP (Doppler Velocity and Position), and FPS-16 Radars. Later it is planned to include the Integrated Trajectory System (ITS) in the series of tests. The ITS is a system capable of simultaneous multiple object tracking by combining range and angle information. The range and angle measurement involves the use of electromagnetic phase-measuring systems.

Briefly, the analysis will cover the methods used to estimate the precision of each instrumentation system and the bias of each instrumentation system.

II. PRECISION AND BIAS ERROR ESTIMATES.

A. Precision Estimates by Multi-Instrument Method.

The first attempts at precision estimates were confined to the variate difference technique. Later the multi-instrument method was applied as data from other instrumentation systems became available. The latter method has become known as the "Simon-Grubbs Technique" at White Sands Missile Range because of the articles by General L. E. Simon and Dr. F. E. Grubbs illustrating this technique. This may be illustrated briefly by the assumed

Preceding Page Blank

mathematical models of simultaneous i-th paired measurements y_{i1} and y_{i2} from instruments #1 and #2 respectively:

(1) $y_{i1} = x_{i1} + b_{i1} + e_{i1}$

(2) $y_{i2} = x_{i2} + b_{i2} + e_{i2}$

where: (a) $x_{i1} = x_{i2}$ and represents the variability of the i-th quantity or characteristic being measured.

(b) b_{i1} and b_{i2} represent the measurement bias of instruments #1 and #2 respectively while measuring the i-th quantity.

(c) e_{i1} and e_{i2} represent the random error of the i-th measurement with respect to instruments 1 and 2.

Now if we have "n" of these paired measurements we may form "n" differences of the corresponding pairs. Typically:

(3) $(e_{i1} - e_{i2}) + (b_{i1} - b_{i2})$

If the bias is constant for any "n" paired measurements, we can estimate the variance of y_{i1} and y_{i2} as:

(4) $s^2_{Y1} = s^2_x + s^2_{e1}$ and

(5) $s^2_{Y2} = s^2_x + s^2_{e2}$

The estimate for the set of difference is:

(6) $s^2_d = s^2_{e1} + s^2_{e2}$.

It is now possible to estimate the instrumentation variability from the equations (4), (5) and (6). This method has a few shortcomings. Many times one achieves negative variance estimates which require some interpretation Dr. W. A. Thompson has worked on the problem of negative components of variance and I note that he is scheduled for a paper on the subject at this meeting.

The two instrument problem, in general, cannot be applied to trajectory data because the characteristic measured is extremely large and variable compared with the instrumentation system errors. The variance of the estimated instrumentation variance contains the estimated variance of the characteristic i.e σ^2_x for est. (σ^2_{e1}) we have

(7) $\frac{2}{n-1} \cdot \sigma_{e1}^{4} + \frac{1}{n-1} \left(\sigma_{x}^{2} \cdot \sigma_{e1}^{2} + \sigma_{x}^{2} \cdot \sigma_{e2}^{2} + \sigma_{e1}^{2} \cdot \sigma_{e2}^{2} \right)$

If one knows the ratio of precision for the two instrument case, then, the two instrument case could be solved for precision estimates. This is uaually not the case.

Estimates of the error variance components were obtained for the four instrumentation systems: Ballistic Camera, DOVAP, Askania and Radars. The square root of the variance estimates are presented in table 2, with the exception of the DOVAP x-component. The variance estimate for the x-component was small and negative: thus, the variance component was equated to zero.

Table 2: Standard Deviation Estimates by the Multi-Instrumentation Method.

Coordinate Component Estimated	Standard Deviation Estimate * Instrumentation System: Ballistic	DOVAP	Askania	Radar
x (feet)	2	0	11	15
y (feet)	6	4	11	21
z (feet)	10	8	8	12

* Based on 28 consecutive trajectory data points.

Other estimates can be obtained from the analysis of variance tables where the Ballistic Camera is considered as a standard for comparison.

3. Precision Estimates by the Variate Difference Method.

The variate difference method was applied to data from theDOVAP and FPS-16 radar systems for the trajectory segment covered by Ballistic Camera. Data sampling rates for the DOVAP and the FPS-16 systems were much higher than for the Askania and Ballistic Camera systems which were dependent on a flashing light at one per second. Thus, the DOVAP and Radar systems were more suited to this technique.

Table 3 presents the standard deviation estimates for the DOVAP system The estimates are based on second difference.

Table 3: Standard Deviation Estimates by Variate Difference Method.

Nominal time along trajectory segment based on missile liftoff.	DOVAP Standard Deviation Estimate* Coordinate for Components x (ft.)	y (ft.)	z (ft.)
40-50 seconds	0.17	0.31	0.24
50-55 seconds	0.17	0.36	0.26
60-65 seconds	0.23	0.40	0.31

*Each is based upon 50 consecutive trajectory data point.

These estimates filter out linear noise from the data and hence are much smaller than the Simon-Grubbs estimates of the previous section which do not filter the linear noise.

The variate difference technique was also applied to trajectory data available from the FPS-16 radars. Each radar was analyzed separately and in its natural coordinate system: range, azimuth, and elevation. The radars, for the most part, exhibit estimates close to the design intent: range ± 4 yards, azimuth ± 0.1 mils, and elevation ± 0.1 mils (these are rms. values) when evaluated by this method. Data from the three FPS-16 radars that tracked most of the trajectory are shown in Table 4. These data cover essentially the same trajectory segment as data in the preceding section.

Table 4: Precision Estimates by Variate Difference Method.

Tracking FPS-16 Radar	Nominal Time in Seconds	Standard Deviation Estimate* Range (Yds)	Azimuth (Mils)	Elevation (Mils)
112	40-50	1.62	0.33	0.16
112	55-70	2.90	0.32	0.25
114	40-55	1.86	0.14	0.12
114	55-70	2.34	0.15	0.15
122	40-55	0.75	0.22	0.18
122	55-70	0.59	0.29	0.15

*Trajectory sample of 150 points

Table 5 shows the same set of radars but with sampling interval five times as long. Most of the estimates were based on the third set of differences, and they include almost the entire missile trajectory.

Table 5: Precision Estimates by Variate Difference Method.

Tracking FPS-16 Radar	Time Segment In Seconds	Standard Deviation Estimate		
		Range (Yds)	Azimuth (Mils)	Elevation (Mils)
112	8-100	2.68	0.31	0.21
114	10-100	2.78	0.15	0.15
122	10-100	1.27	0.35	0.20

C. System Estimates of the Bias Error.

Two disjoint trajectory segments were expected for the Ballistic camera coverage. Each of these segments were to be divided into a first portion and a last portion. Actual missile trajectory segment was covered in one continuous segment from approximately 39 seconds of flight to 66 seconds of flight from missile lift-off. Thus, four sets of seven trajectory data points each were formed.

At simultaneous times, the reduced trajectory data from Askanias, DOVAP and radars were each differenced with respect to the Ballistic camera data. The set of error difference data were used in the analysis of variance. The Ballistic Camera data were considered as the reference standard.

The analysis of variance of the DOVAP difference data indicated a significant shift in the bias for the X and Z component segment means. The Y component of DOVAP data indicated no shift in the means for segments. However, a significant bias is indicated in each of the overall mean coordinates when compared with the expectation of zero. Table 6 shows the estimated means for each trajectory segment and coordinate. To compare directly with the Ballistic Camera data, these data need a nominal adjustments in each coordinate; the largest adjustment is approximately 8 feet for the X component. These adjustments do not change any of the above conclusions.

Table 6: DOVAP Mean Bias Error Estimates.

Component Coordinate	Trajectory Segment				Over-all Mean Bias
	1	2	3	4	
X (ft)*	54	49	44	42	46
Y (ft)	-17	+24	-20	-16	-20
Z (ft)	-64	-81	-80	-92	-80
	*X needs a nominal 8 foot adjustment				

The analysis of variance for the Askania errors shows a significant shift in the bias between the trajectory segments for each coordinate studied. In addition, there is a significant bias in the overall mean for each of the coordinates. These biases are undergoing further study at the present time. The estimated mean error for each segment is shown in Table 7.

Table 7: Askania Mean Bias Error Estimates.

Component Coordinate	Trajectory Segment				Over-all Mean
	1	2	3	4	
X (ft)	13	10	8	-5	6.4
Y (ft)	3	-11	-2	-20	-7.4
Z (ft)	-22	-35	-18	-44	-29.4

The FPS-16 radars exhibit a significant shifting mean along the trajectory in the Z coordinate. However, the overall mean error for the Z coordinate is not significantly different from zero. For the "X" and "Y" coordinate, the overall means exhibit a significant bias. This is not being investigated further because the tracking was not a point source such as a beacon. A Beacon track was intended for the shoot but was not attained. Table 8 below exhibits the mean data.

Table 8: Radar Mean Bias Error Estimates.

Component Coordinate	Trajectory Segment				Over-all Mean
	1	2	3	4	
X (ft)	-10	-16	-17	-3	-11.5
Y (ft)	13	20	8	25	16.8
Z (ft)	2	1	20	-9	3.6

The analysis of variance tables are shown in tables 9, 10, and 11.

Table 9: DOVAP Analysis of Variance Tables for Trajectory Segments.

Sources of Variance	d.f.	s.s.	m.s.	F
X -Trajectory Segment	3	389	129	7.58
X - Error	24	406	17	
Totals	27	795		
Y -Trajectory Segment	3	284	95	3.16
Y - Error	24	713	30	
Totals	27	997		
Z -Trajectory Segment	3	2691	897	28.9
Z - Error	24	748	31	
Totals	27	3439		

Table 10: Askania Analysis of Variance Tables for Trajectory Segments.

Sources of Variation	d.f.	s.s.	m.s.	F
X -Trajectory Segment	3	1414	471.3	7.01
X - Error	24	1614	67.2	
Totals	27	3028		
Y - Trajectory Segment	3	2048	682.5	7.21
Y - Error	24	2275	94.7	
Totals	27	4323		
Z -Trajectory Segment	3	3035	1011.7	10.10
Z - Error	24	2406	100.2	
Totals	27	5441		

Table 11: Radars Analysis of Variance Tables for Trajectory Segments.

Sources of Variation	d.f.	s.s.	m.s.	F
X -Trajectory Segment	3	880	293	1.71
X -Error	24	4093	171	
Totals	27	4973		
Y -Trajectory Segment	3	1149	383	
Y - Error	24	11980	499	
Totals	27	13129		
Z -Trajectory Segment	3	3212	1071	8.74
Z - Error	24	2957	123	
Totals	27	6159		

III. <u>CINE-THEODOLITE FILM READING PRECISION</u>. An example of sub-system study is given by this tracking correction digression. Each cine-theodolite record was read by three different reading personnel. The set of readings with the lowest reader variance was used in the data reduction process. Table 12 gives the precision estimates of the tracking correction readings.

Table 12: Precision Estimates for Eight Cinetheodolite Records.

Film Coordinate System	Standard Deviations of Tracking Corrections Cinetheodolite							
	1	2	3	4	5	6	7	8
X (arc sec)	3.0	1.1	0.4	0.4	2.0	3.1	3.8	2.8
Y (arc sec)	3.6	2.1	0.9	1.4	2.4	2.9	5.1	2.7

The estimates were derived by the three-instrument method (Simon-Grubbs). The sample sizes ranged from a low of 34 X-Y pairs to the maximum of 99 X-Y pairs. There is a trend indicated for low estimates in X to be paired with low estimates in Y. This is to be expected because a good film record could be read well in either coordinate.

IV. <u>SUMMARY AND CONCLUSIONS.</u>

A. The DOVAP and Ballistic Camera are among the most precise systems in use at WSMR. The DOVAP has the shortcoming that, in general, the trajectory data are biased from the true trajectory.

B. The shift in bias along the trajectory was significant in all coordinates for Askania data; significant in the X and Z coordinate for DOVAP data; and significant in the Z coordinate for Radars.

C. There was a significant overall bias in all coordinates except for the Radars in the Z coordinate.

D. The significant biases in the DOVAP and Askania Instrumentation systems should be studied and mathematical or physical methods developed to remove them.

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. Thompson, W. A., Moore, J. A., "On the Problem of Negative Estimates of Variance". Paper at Sixth Conference on the Design of Experiments in Army Research, Development and Testing, BRL, Aberdeen Proving Ground, Maryland (Oct. 1960).

4. Brown, D., and Patton, R. B. Jr., "A Comparison of Optical and Electronic Trajectory Measuring Methods", BRL Rpt # 965 (C). 1956. Confidential.

5. Sibol, J. L., "Askania Cine-Theodolite Accuracy Studies Conducted Under OD-039", RCA Data Processing Tech. Report # 52, Sept. 11, 1959.

6. Schmid, H., "Systematic Errors of Cine-Theodolites", BRL Rpt # 764, Aug. 1951 (U).

7. Davis, R. C., "Techniques for the Statistical Analysis of Cine-Theodolite Data", NAVORD Report 1299, China Lake, Calif., (March 22, 1951).

8. Davis, R. C., "Techniques for the Statistical Analysis of Continuous Wave Doppler Data", NAVORD Report 1312 NOTS 383, April 1951.

9. Kendall, M. G., The Advanced Theory of Statistics, Vol. II, Third Edition, C. Griffin and Co., Ltd., London (1951).

10. Cochran, W. G., Cox, G.M. "Experimental Designs", Second Ed., John Wiley & Sons, Inc., New York, 1957.

11. Bargman, R., "Separation of Random Errors of System and of Instrumentation," Proceedings of the Statistical Techniques in Missile Evaluation held at Virginia Polytechnic, Blacksburg, Va., Aug. 5-8, 1958.

12. Snedecor, G. W., Statistical Methods, 4th Edition, The Iowa State College, Press, Ames, Iowa, 1946.

13. "Final Data Report No. 10,200, Ballistic Camera Data for Nike-Hercules for Precise Tracking", Flight 71 HE Missile 11214", (U). Launched March 29, 1960, IRM-DRD, WSMR, N.M. (Sept 15, 1960). Classified Confidential.

14. "Final Data Rpt No. 9624, Askania Trajectory Data for Nike Hercules Flight 71 HE Missile 11214", Launched March 29, 1960, IRM-DRD, WSMR, N. M., May 16, 1960 (U), Classified Confidential.

15. "Final Data Report No. 10045, DOVAP Trajectory Data for Nike Hercules Flight 71 HE Missile 11214," Launched March 29, 1960, IRM-DRD, WSMR, N. M., Sept. 15, 1960 (U), Classified Confidential.

16. "Final Data Report No. 9590, Radar Trajectory Data for Nike Hercules Flight 71 HE Missile 11214". Launched March 29, 1960, IRM-DRD, WSMR, N. M., (U). Classified Confidential.

17. Bush, N., "Evaluation of Reading Error of Theodolite Readers," RCA Data Reduction Tech. Memo #2., 19 July 1961.

18. "Electronic Trajectory Systems Catalog", Vol. I, prepared by Electronic Trajectory Measurements Working Group, Inter-Range Instrumentation Group, 13 Oct 1958.

AN APPLICATION OF THE EXPONENTIAL HAZARD MODEL TO AEROSOL CHAMBER TRIAL DATA

Theodore W. Horner
Booz, Allen Applied Research, Inc.

In an aerosol cloud release, a bacterium may be ineffective because of death before some critical time. This bacterial decay, as it is called, has been studied by means of chamber trials by releasing a cloud into the chamber and then sampling the chamber at periodic intervals. One of the unique sets of data in this regard, from the standpoint of precision and amount of replication, is that supplied by Dr. T. L. Snyder and Hugh Lee, of Fort Detrick.

In these trials, an aerosol cloud of particles containing bacteria and tracer material was generated inside a chamber. Ten pairs of samples were withdrawn from the chamber at half-minute intervals starting at the first half minute. A bacteria count was obtained on one sample of each pair and a tracer measurement on the other. Estimates of initial tracer and bacteria counts were also obtained using a knowledge of the composition of the spray material, spray rate and duration. Tracer material was included in the cloud release because the particles on which the bacteria were located were continually falling to the bottom of the chamber and the tracer data were used to correct for this fallout loss. The corrected data, which are described as <u>biological recovery percentages</u>, were computed for each half-minute interval as,

$$R = 100 \times \frac{B_s}{B_I} \div \frac{T_s}{T_I}$$

where

B_s is the sample bacteria count;

B_I is the initial bacteria count;

T_s is the sample tracer measurement;

T_I is the initial tracer measurement.

Chamber trials were run at relative humidities of 12, 36, 62, and 87% with a medium 1, and at 12, 19, 36, 49, 65, and 86% with a medium 2. About twelve trials were conducted at each relative humidity, medium combination.

One of the unique features of the Snyder-Lee data is shown in Slide 1, which shows plots of viable recovery percentages versus time on log-log paper. Both of these curves are associated with the data of medium 1; the lower corresponding to a chamber relative humidity of 36% and the upper to a chamber relative humidity of 12%. The plotted points are average biological recovery percentages taken over all similar trials. The upper curve is concave downward and is typical of the type of curve that is observed for all of the relative humidity, medium combinations except the one associated with the lower curve, which is concave upward.

Several models have been proposed by Dr. Snyder and others for these biological recovery curves. The Weibull model was found to give an excellent fit in all cases except the data for medium 1, relative humidity 36%. The Weibull model will not give concave upward curves in log-log space.

A model which did give a good fit to the data for all of the medium, relative humidity combinations was the exponential hazard model. This model is defined as,

$$R = \exp\left[a + b \exp(-ct)\right]$$

where R is biological recovery percentage at time t, and a, b, and c are constants. Hazard rate $H(t)$ is defined as the chance that a bacterium will die in the interval dt given that it has lived to time t. The hazard rate for the exponential hazard model is

$$H(t) = -(1/R)\frac{dR}{dt} = -\frac{d \ln R}{dt} = bc \exp(-ct)$$

which will plot as a straight line on semi-log paper. The model was, in fact, suggested by observing the computed point-by-point hazard rate plots on semi-log paper. This hazard rate is to be contrasted to a constant hazard rate for exponential decay for biological recovery percentages and a hazard rate of

$$H(t) = \left(\frac{\beta}{\alpha}\right) t^{\beta - 1}$$

for the Weibull model.

When the exponential hazard model is plotted on log-log paper, the biological recovery percentage curve is concave downward for $t < 1/c$ and concave upward for $t > 1/c$. The initial recovery percentage at $t = 0$ is $\exp(a + b)$ and the recovery percentage at $t = \infty$ is $\exp(a)$. This model was fitted to the Snyder-Lee data by computing the regression of $Y = \ln \bar{R}$ on $X = \exp(-ct)$, choosing c so as to minimize the sum of squares of the deviations from regression.

The results of fitting the exponential hazard model to averages over similar trials of the viable recovery percentages are shown in Slide 2. The model was not fitted to the data of medium 1 at the relative humidities of 62 and 87%, because there was no evidence of bacterial decay within the time span covered by the data apart from an initial decay of 5 to 10%. In the right-hand column of the slide are the percentages of the Y sum of squares accounted for by linear regression of Y on X. These percentages, with two exceptions, are above 99%. The exceptions occur at the high relative humidities for medium 2 and are due to the fact that the slope of the regression line at these higher relative humidities is so gradual that the regression sum of squares becomes small relative to the noise which is present. This fact is pointed up perhaps better by Slide 3, which shows the plots of $Y = \ln \bar{R}$ versus $X = \exp(-ct)$ for the different relative humidities and mediums. Looking at the upper line for medium 2, relative humidity 96%, the noise did not appear greater here than at other relative humidities, but the shallow slope of the line materially reduced the sum of squares accounted for by regression.

Slide 4 shows other cases. Here again, the exponential hazard model did not appear to contradict the data.

The variation of parameters of the model with relative humidity is shown in Slide 5. The horizontal axis is associated with relative humidity and the vertical axis with parameter value. In the case of the a and b parameters, a discontinuity appears to occur in the neighborhood of a relative humidity of 45%. The a and b curves appear to be almost mirror images of each other. This is probably because the sum of a and b is associated with initial recovery and hence is a more fundamental parameter. For a particular relative humidity, the sum of a and b can be read off the graph and used as an entry in the lower right table in the graph to find the initial recovery percentage. Thus, the sum for a relative humidity of 21% is 3.6, which is associated with an initial biological recovery of 37%.

The variation of the initial biological recovery percentage with relative humidity is not too surprising. However, the nonzero recovery percentages at time equal infinity are somewhat more suspect. Thus, when the relative humidity is 58%, the predicted value for the a parameter is 2, and using the lower left table on the graph, recovery percentage at $t = \infty$ is estimated at 7.39%. This nonzero final recovery percentage is a matter which is subject to experimental verification, although such verification is not possible with present sets of data, because of the limited time span and relative humidity levels covered by the data.

Although it is not too clear in Slide 5, c declines from a value of about 0.33 at 10% relative humidity to about 0.20 at 65% relative humidity. The odd value of $c = 0.80$ for the relative humidity level of 86% was probably a poor estimate due to the noise occurring at this humidity level. In this case, the sum of squares would not be substantially changed regardless of the c value used.

Using the graphs which relate parameter values to relative humidity level and the exponential hazard model, it was possible to construct graphs relating the three quantities: relative humidity level, per cent recovery, and time, and from these graphs make predictions that could be used for further testing of the model.

Slide 6 indicates the results of fitting the exponential hazard model to individual trials. The trials are those for medium 2, relative humidity 36%. For these particular trials, the exponential hazard model explained substantial portions of the variability. The manner in which estimated initial recovery percentages varied from trial to trial is shown in the column labeled $t = 0$. Apart from trial #33, which indicates an initial biological recovery percentage of 81%, the percentages vary from a low of 14% to a high of 30%. Part of this variation may be due to imprecise control of the relative humidity of the chamber.

As regards the exponential hazard model, there are a number of areas that require further investigation; namely:

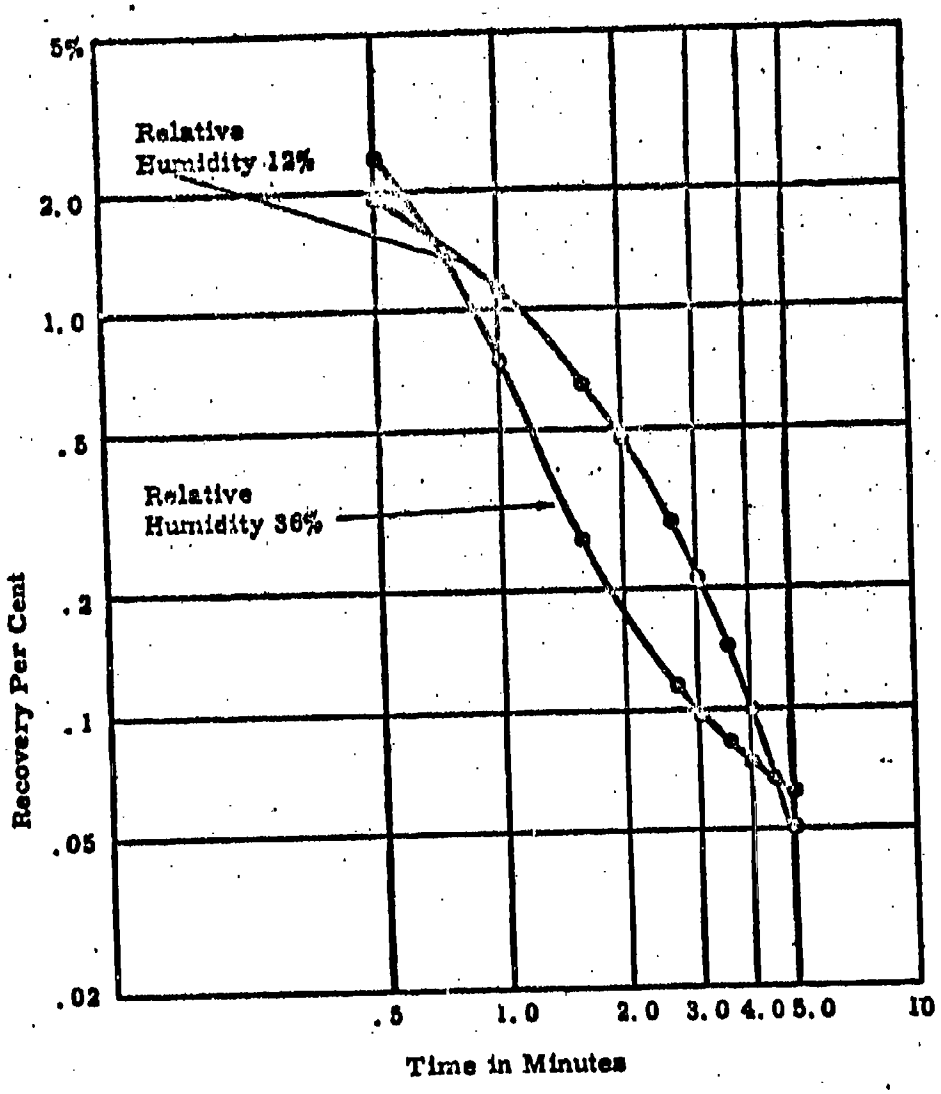
1. Validity of the model,

2. Theoretical implications of the model,

3. Application of the model as a research tool.

Slide 1

Graphs of Average Recovery Percentages on Log-Log Paper for Organism 1, Medium 1

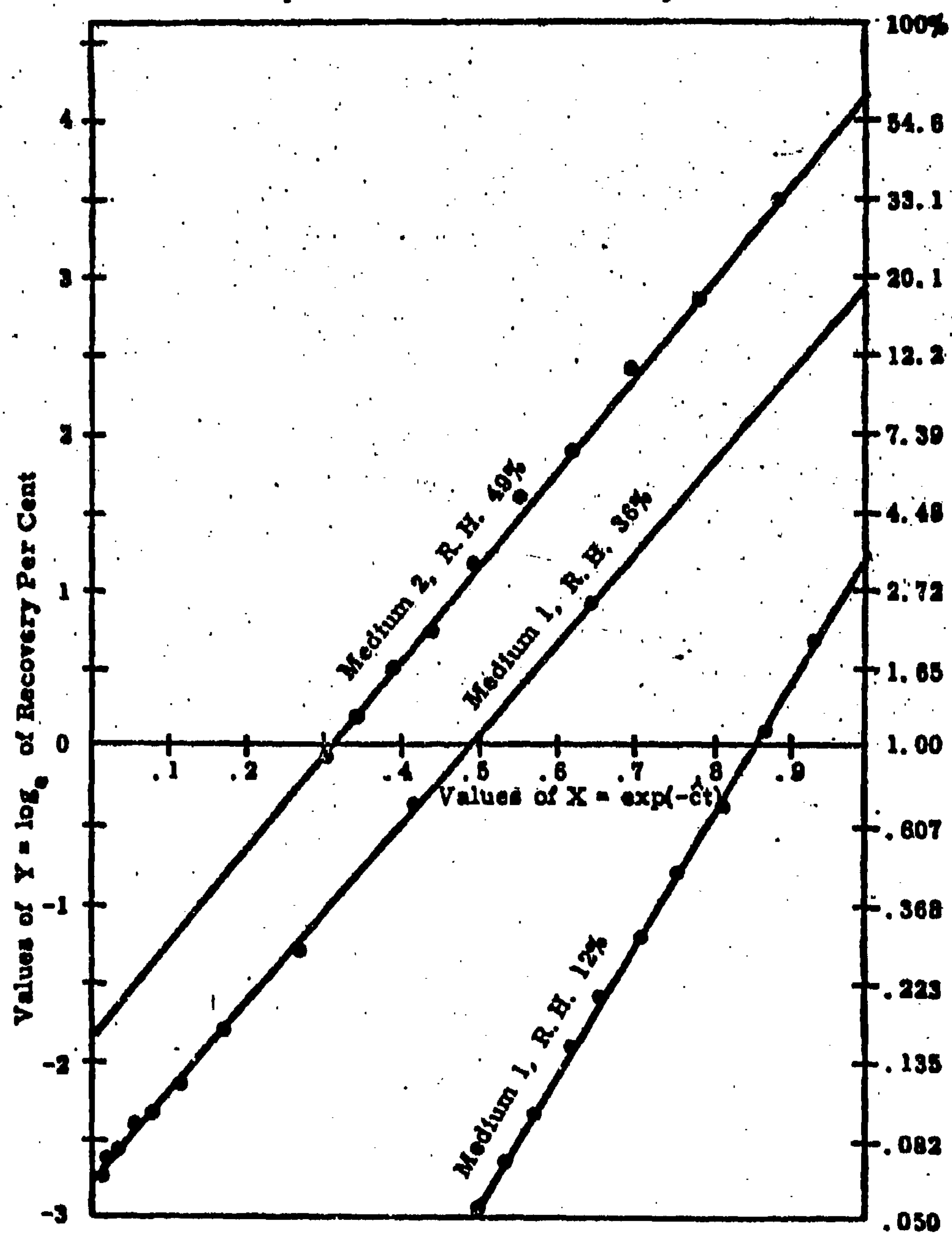


Slide 2

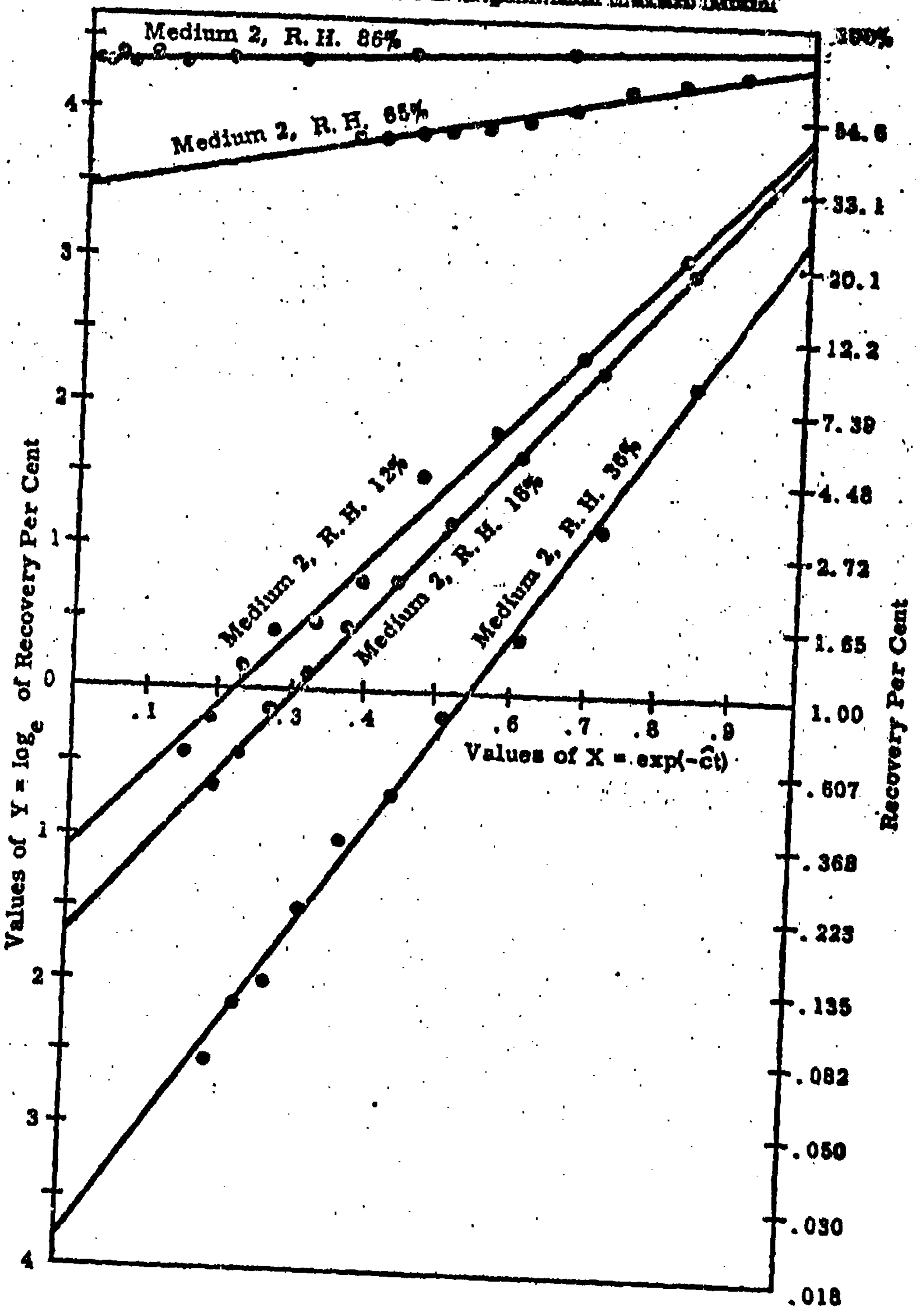
Results of Fitting the Exponential Hazard Model to Snyder-Lee Data on Average Recovery Percentages Over Similar Trials

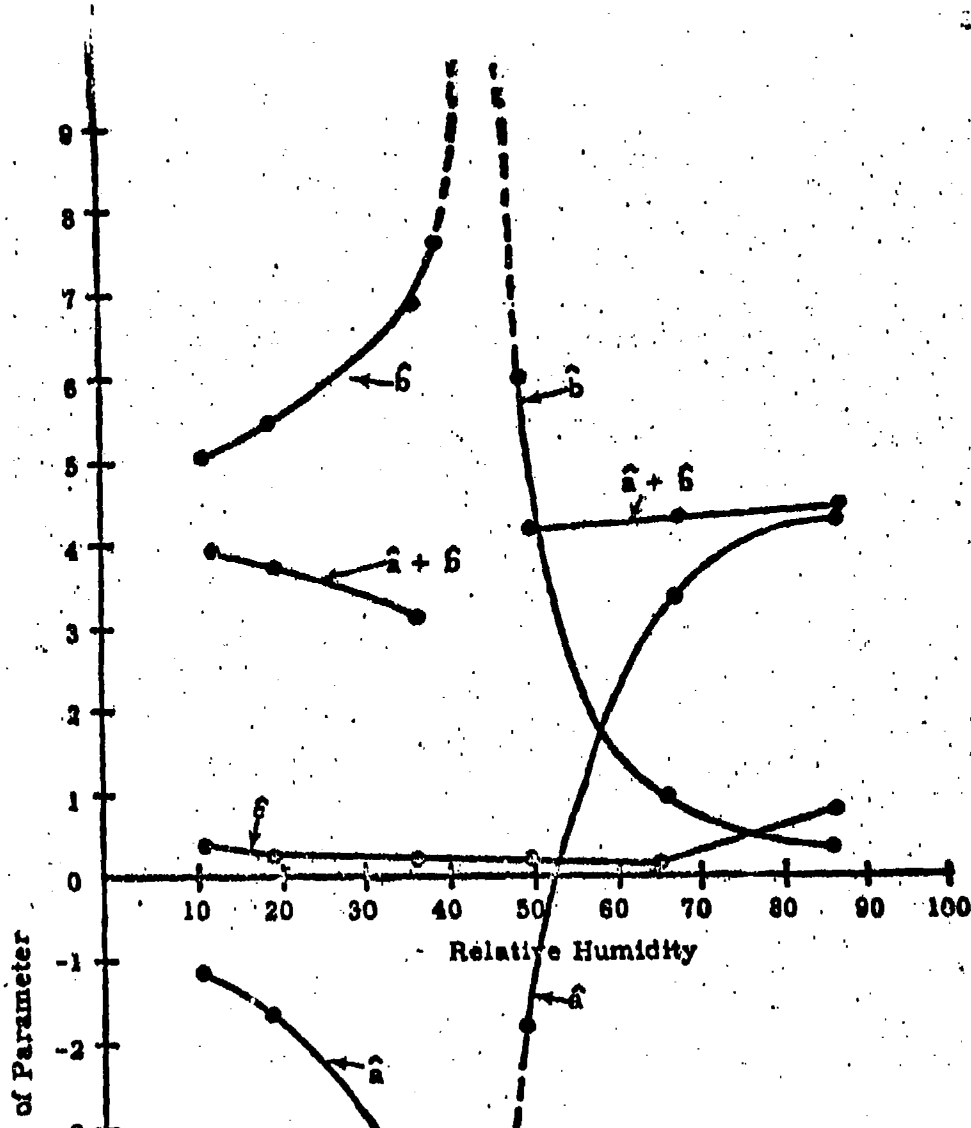
Relative Humidity	Average Biological Recovery at 1/2 Minute	Per Cent Recovery Estimates		Percentage of Y Sum of Squares Accounted for by Linear Regression of Y on X
		Initial t = 0	Final t = ∞	
Medium 1				
12	2%	3%	.00%	99.91
36	2.5	19	.06	99.89
62	87			
87	90			
Medium 2				
12	20	49	.33	99.41
18	19	44	.19	99.81
36	8	24	.02	99.59
49	32	69	.17	99.75
65	71	78	30.6	97.09
86	83	87	75.2	79.12

Fits of the Exponential Hazard Model to Snyder-Lee Data



Fits of Snyder–Lee Data to Exponential Hazard Model





$\hat{a}$	Final Recovery Per Cent = $e^{\hat{a}}$	$\hat{a} + \hat{b}$	Initial Recovery Per Cent = $e^{\hat{a}+\hat{b}}$
4	54.6	3.2	24.5
3	20.1	3.4	30.0
2	7.39	3.6	36.6
1	2.72	3.8	44.7
0	1.00	4.0	54.6
-1	.368	4.2	66.7
-2	.135	4.4	81.5
-3	.050	4.6	99.5

Slide 5

Relationship of Parameters to Relative Humidity Level for Snyder-Lee Data for Organism 1, Medium 2

Slide 6

Results of Fitting the Exponential Hazard Model to the Trials of Medium 2, Relative Humidity 36%

Trial	Estimates of Recovery Percentages: Initial $t = 0$	Estimates of Recovery Percentages: Final $t = \infty$	Percentage of Y Sum of Squares Accounted for by Linear Regression of Y on X
75	15.69	.029	99.66
76	14.56	.033	98.59
77	17.35	.030	99.61
78	30.51	.022	99.01
79	21.17	.034	99.11
80	25.09	.022	99.53
81	10.51	.009	99.20
82	26.00	.060	98.39
83	81.19	.012	99.19
84	23.23	.016	99.73
85	24.24	.005	87.20

CALIBRATION OF A ZINC SULFIDE PARTICLE DETECTOR

John E. Malligo
U. S. Army Chemical Corps Biological Laboratories
Fort Detrick, Frederick, Maryland

In the study of atmospheric currents and the behavior of particulates in aerosol, a variety of finely divided materials is used as tracers. An important class of tracer is the fluorescent particulates. Among the fluorescing mineral compounds, the sulfides of zinc and cadmium are particularly useful. Capable of being produced in extremely well controlled ranges of particle size and detectable in minute quantities, these sulfides have been used widely in aerosol studies.

A typical test using these tracers involves their aerosolization in an atmospheric system of interest and subsequent sampling at a time and place dictated by the test objectives. Sampling is achieved by filtering a metered quantity of the aerosol through a membrane filter which retains virtually all of the particles contained therein. After suitable preparation, the filters are assayed visually with a low power light microscope using ultra-violet illumination to induce fluorescence in the particles. This process, of course, entails counting all or a sample of the particles on the filter, which is at best a tedious job. On heavily laden filters, the errors due to distribution of the particles are further augmented by those from human fatigue and confusion of the point light sources. On filters with relatively low particle densities, the time consumed in counting an adequate number of particles, say 300, which is considered a reasonable sample, may be as much as 20 or 30 minutes. This is due to the necessity of observing a large number of microscopic fields, each of which contains only a few particles. In a large scale test, with possibly hundreds of filter samples, the labor involved becomes excessive. For this reason, much thought has been given to development of an automatic fluorescent particle counter.

No machine has been developed to count these particles, as such, but the General Electric Company at Hanford Atomic Works, Richland, Washington has developed a device* which, by detecting scintillation induced in zinc and cadmium sulfides by an alpha emitting source, can give a quantitative estimate of the mass of the material present. This device gives data in scaler readings of the number of nuclear disintegrations per minute.

---

*Discussed by M. O. Rankin at the Meeting of the American Meteorological Society, San Diego, California, June 15-19, 1959.

The problem to be presented here concerns the attempt to calibrate this device in terms of particles per filter. A series of filters was prepared with graded loadings of particles to cover a range considered of practical value (i.e., from 10 particles per filter to one million particles per filter in ten-fold increments).

Typical examples of the machine's response to filters of known particle count are given in slide 1, which indicates that its threshold is at about 100 particles per filter and that no apparent upper limit of useful response was reached. It is background radiation, in the form of cosmic rays, etc., which essentially determines the threshold of machine sensitivity, of course.

Slide No. 2 presents a typical set of data in which visual counts are plotted against machine response on logarithmic paper. (Consider points of machine 1 only.)

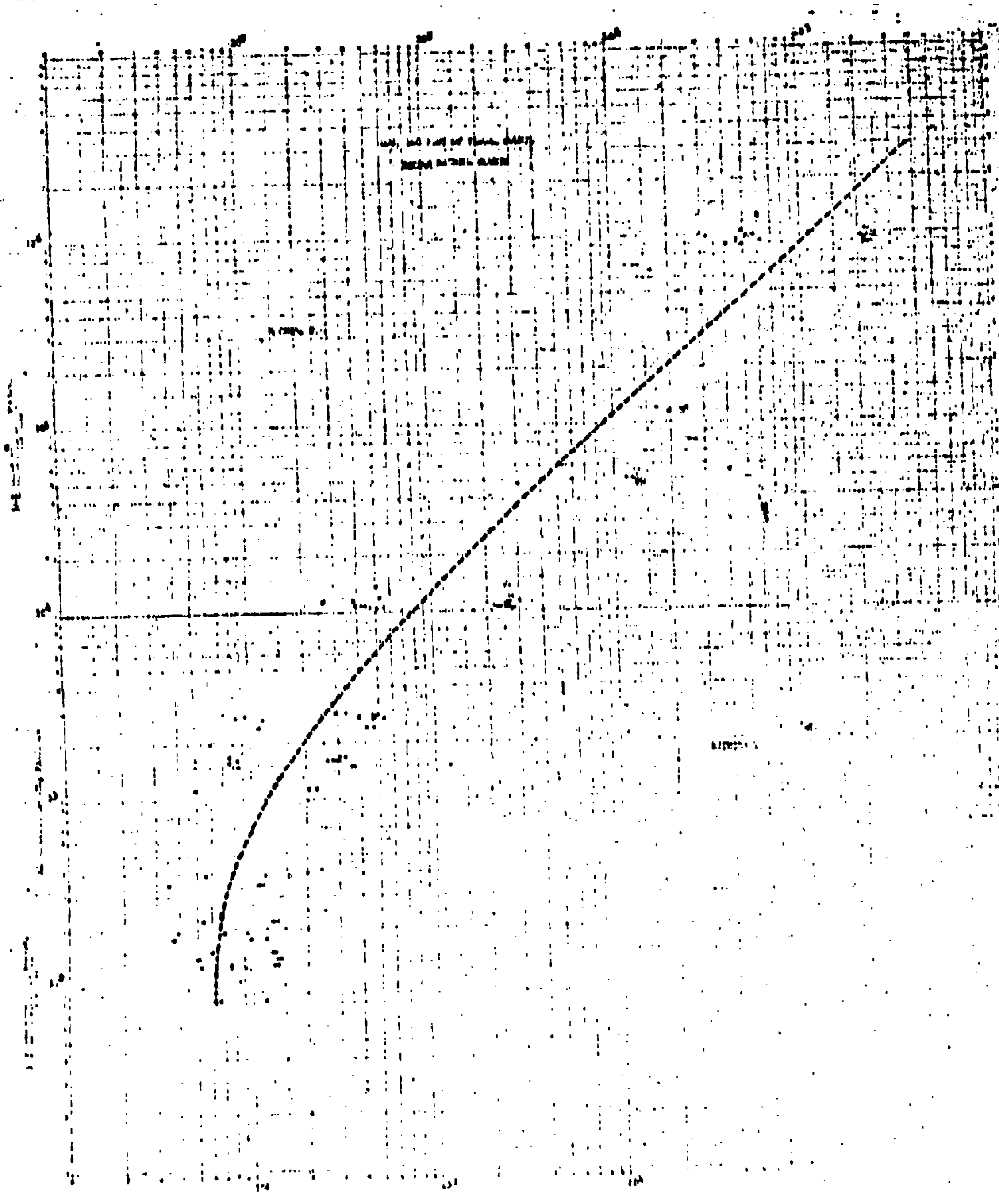
Our approach has been to fit a linear regression to the data between the limits of 1000 to one million particles per filter. This equation is given in the next slide. (Slide 3.)

The question to be asked here is, do the data warrant fitting a curvilinear regression starting at the 100 particles per filter level? A corollary question is what would be the statistical validity of such a curvilinear regression, considering that the visual counting is extremely precise and accurate at low counting levels while the machine response is just the opposite.

SLIDE 1

Visual Count (No. of particles/filter)	Machine Response (counts per minute)
138	129
2,775	418
10,236	2,826
43,298	13,737
110,193	22,211
947,971	212,621

SLIDE 2



The dashed line has no significance other than to separate the points of Machine 1 from those of Machine 2.

SLIDE 3

$$Y = -0.7992 + 1.03059\,X$$

standard error of b = 0.0090

$$r^2 = 0.993$$

Note: b = slope

r = correlation coefficient

EFFECTS OF AIMING POINT PATTERNS ON BOMB SALVO TARGET COVERAGE

Ralph D. Doner

Systems Analysis Laboratory, Requirements and Plans Division, Research and Development Operations, Army Rocket and Guided Missile Agency, Redstone Arsenal, Alabama

1.0 PROBLEM STATEMENT.

1.1 PROBLEM I. To obtain, on a given confidence level, the most nearly uniform specified fractional coverage of a large homogenous target area by varying the geometry and bomb allocation of multiple aiming points for a salvo of bombs having a fixed lethal radius and circular error probability.

1.2 PROBLEM II. To find the simplest computational techniques for obtaining acceptable solutions for Problem I on manual, analogue and digital levels.

2.0 ANALYSIS.

2.1 CENTRAL PATTERN OPTIMIZATION. Bombs aimed at the central points of the pattern will provide most of the coverage for the central portion of the large homogenous target area. Uniformity in this coverage calls for uniformity in the geometry of the pattern and in the allocation of bombs to individual points. In seeking the optimum characteristics of the central portion of the pattern the analysis will progress from consideration of a single aiming point to a row of points and finally to two dimensional arrays.

2.1.1 ONE AIMING POINT. Traditionally, the coverage for this case is determined in terms of the probabilities of hits and overlaps and the ratio of the bomb's lethal area to the area of the target. Central symmetry about this one aiming point makes coverage a function only of the radial distance from the center of the bomb burst to the aiming point. Double integration is not inherently unavoidable in empirically determining this functional relationship. A discrete set of concentric circles is in order, each to serve as an isohap* curve, that is, as a contour of constant probability. Fig. 1

*in "Handbook of Probability and Statistics with Tables", by Burrington and May, Handbook Publishers, Inc., 1958, on page 98, the expression "equiprobability curve" is used. A shorter term such as "isoaleatory", or better, "isohap", is desirable if this concept should gain wide usage.

illustrates this for ten bombs. A table of Gaussian deviates was used to locate the centers of bomb impacts. The lethal radius of the bombs is 0.5σ and the isohap circles are spaced 0.5σ apart. The resulting arc coverage on each isohap is accumulated on the centrally pivoted dividers, shown in the figure in the act of evaluating the coverage bomb number 6 has given the 2.0σ isohap. Totals are read in decimal form on the peripheral scale. In figure 1 are given the coordinates of impact centers and the coverage profile.

2.1.2 A ROW OF AIMING POINTS. Figure 2 shows several of a row of aiming points. The row is sufficiently long to fully account for the hit probabilities of all intermediate points on this line. What is needed here is a simple index of the resultant hit probability on a specific point on this continuous line, in terms of its distance from contributing aiming points. Bombs have a bivariate distribution about their aiming points. In the case of circular error probability, the ordinates of a bivariate normal surface are approximately 0.4 times the ordinates of a univariate normal curve at equal distances from centers of symmetry and for equal standard deviations σ. Both therefore have their inflection points one unit from their maximum points. As a consequence, ordinates of the univariate normal curve may be used as indices of hit probability. Addition of these ordinates, as shown in Figure 2, (where curves are spaced to intersect at their inflection points), results in a near isohap that fluctuates between 0.493 and 0.507. Such near uniformity in hit probability implies near uniformity in overlap probability. The occurance of overlap on a bomb drop exercises a negative feedback on the probability of overlap for subsequent drops, thereby tending to make coverage expectancy more nearly uniform than hit probability.

2.1.3 RECTANGULAR GRID OF AIMING POINTS. In Figure 3 part of a rectangular grid of aiming points is given. Enough points are shown to account for the total hit probability indices of a central point X and of two other centers of symmetry Y and Z. Contributions made by sets of aiming points symmetrical to each of these three points X, Y and Z are tabulated in the figure for separations S from 1.8 to 2.2. A separation of .9 appears to be near optimum. In each case X is a maximum, Y a saddle point, and Z a minimum. Repetition of these values at similar positions throughout the central portion of the pattern provides a measure of the degree of uniformity in expected coverage associated with this particular geometry with equal allocation of bombs to the aiming points.

2.1.4 HEXAGONAL GRID OF AIMING POINTS. Treatment similar to that given the rectangular grid is indicated in Figure 4 for the hexagonal array. The separation is 2.0 units, and again X is a maximum. However, Y and Z have exchanged roles. The degree of uniformity is excellent, indicating that the hexagonal grid is the optimum pattern.

2.2 PATTERN PERIPHERY OPTIMIZATION. The peripheral characteristics of the target obviously determine the boundary properties of the pattern. The circle with a six unit radius in Figure 4 is meant to represent a homogenous target area that is to be given uniform fractional coverage by a hexagonal pattern. Assume that the 19 aiming points shown have been given equal quotas of bombs to optimize the specified coverage for the central portion of the target. Obviously, this whole target area can be given coverage as indicated by the tabulation by enlarging the pattern sufficiently. What is needed here to complete the requirements is a statement limiting wastage, or peripheral coverage loss.

2.2.1 The art of tailoring aiming point patterns for maximum economy in obtaining a desired coverage must be based on verifiable principles. One such principle deserving consideration is an existence theorem such as:

EXISTENCE THEOREM. For every distinct target area with specified fractional coverage by a definite weapon system, an optimum pattern of aiming points exists, and a discrete set of isohap curves can be constructed with sufficient accuracy to make feasible a Monte Carlo method of predicting the resulting coverage.

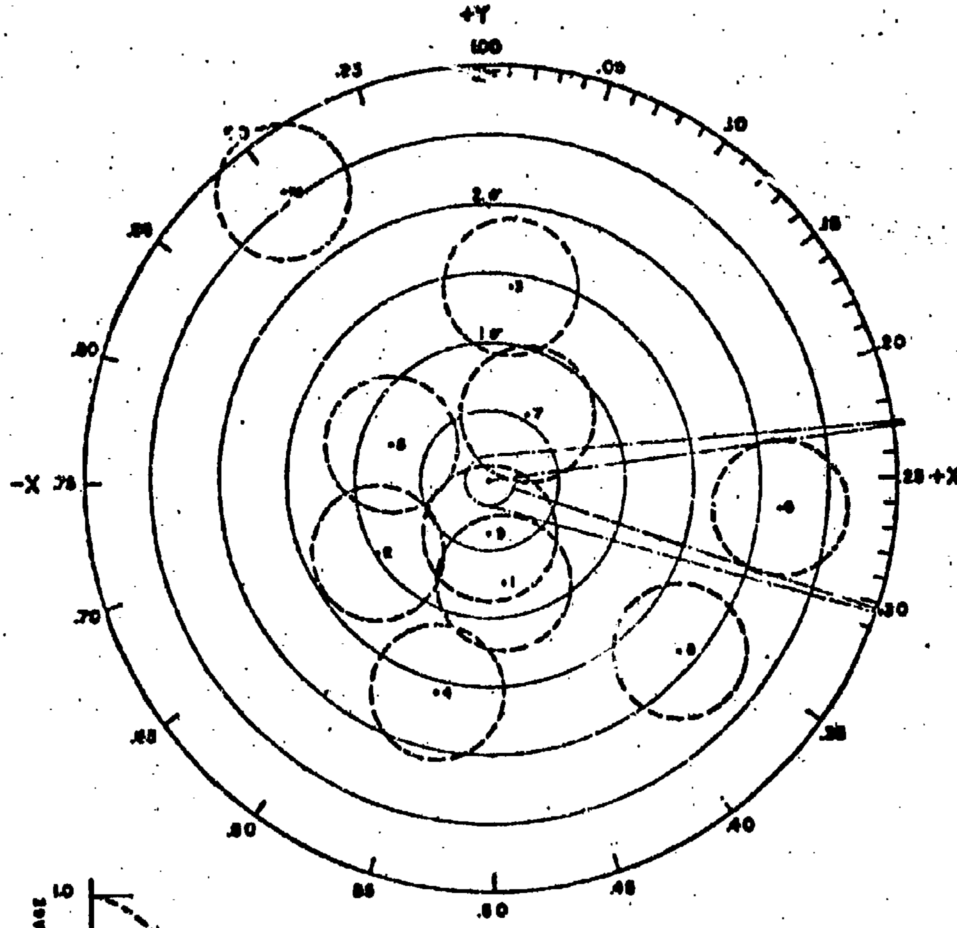
2.2.2 A second principle deserving consideration for verification or rejection has to do with the progressive modification of aiming point patterns associated with a graduated series of targets. For example, a target that has outgrown a single aiming point must accept a three point pattern if circular symmetry is to be even approximately maintained. Thereafter, the next size can be accomodated by four points, and perhaps by five. But a six point pattern, either as a pentagon with a center or a centerless hexagon, may have to give way to seven points. Such center points need not have the same allocation of bombs as the others. What is needed is more than just a continuity principle, since a formulation of this discrete continuity is also desirable. The principle may be tentatively stated thus:

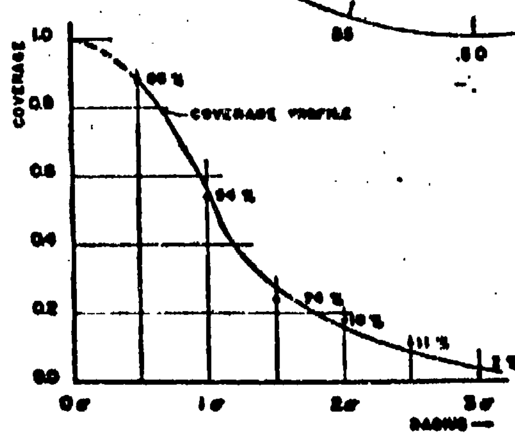
CONTINUITY THEOREM. Targets that differ slightly in size and shape, and in specified coverage by a particular weapon system, will have aiming point patterns differing moderately in configuration and bomb allocation, and isohap curves bearing strong resemblances in all characteristics.

3.0 COMPUTATIONAL SIMPLIFICATION.

3.1 ARC COVERAGE. In paragraph 2.1.1 the use of dividers in accumulating arc coverage of an isohap curve was shown. Such a curve can be replaced by a discrete set of points, thereby allowing the simple act of counting to replace continuous arc measurement. Such points when hit change only an attribute, and when hit a second time retain that same attribute, thus making unnecessary any special consideration of overlap. If a more realistic treatment of the damaging effects of a bomb hit are desired, one may assign full kill to an appropriate circular area around the center of impact, and diminishing fractions to points in surrounding rings. The attribute "hit" would give way to the accumulation of fractions at each point on an isohap, with unity representing saturation or full kill.

4. SUMMARY. A Salvo of bombs with constant lethal areas, aimed at a single point with circular error probability, give equal hit probability to points equidistant from the aiming point. Advantage is taken of the resultant central symmetry in target coverage around this point in building aiming point patterns that provide the area within the pattern maximum uniformity in coverage. Submitted for clinical consideration are suggestions for modifying these optimum patterns to accomodate irregular and nonhomogenous targets, and for simple techniques in evaluating coverage.

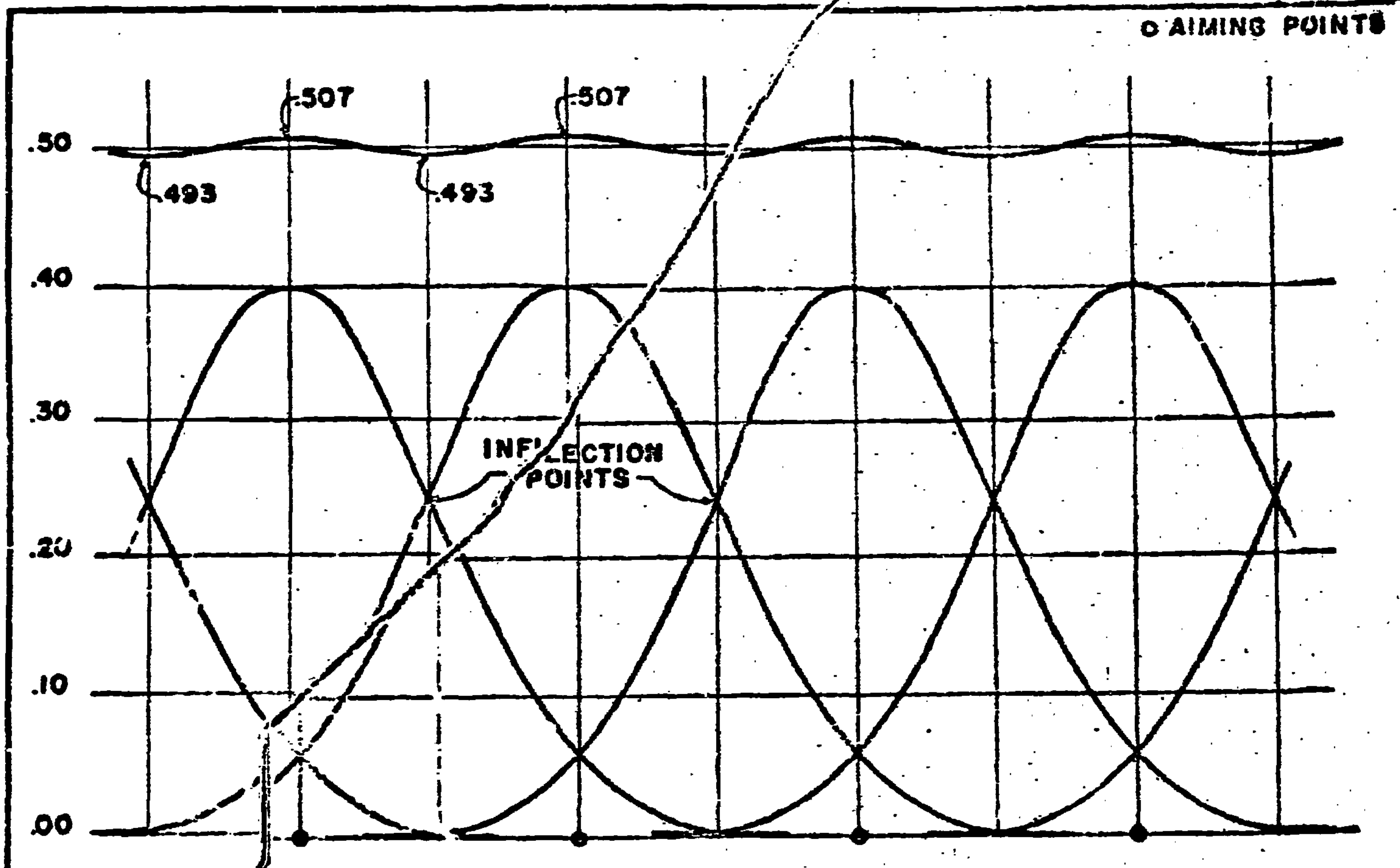




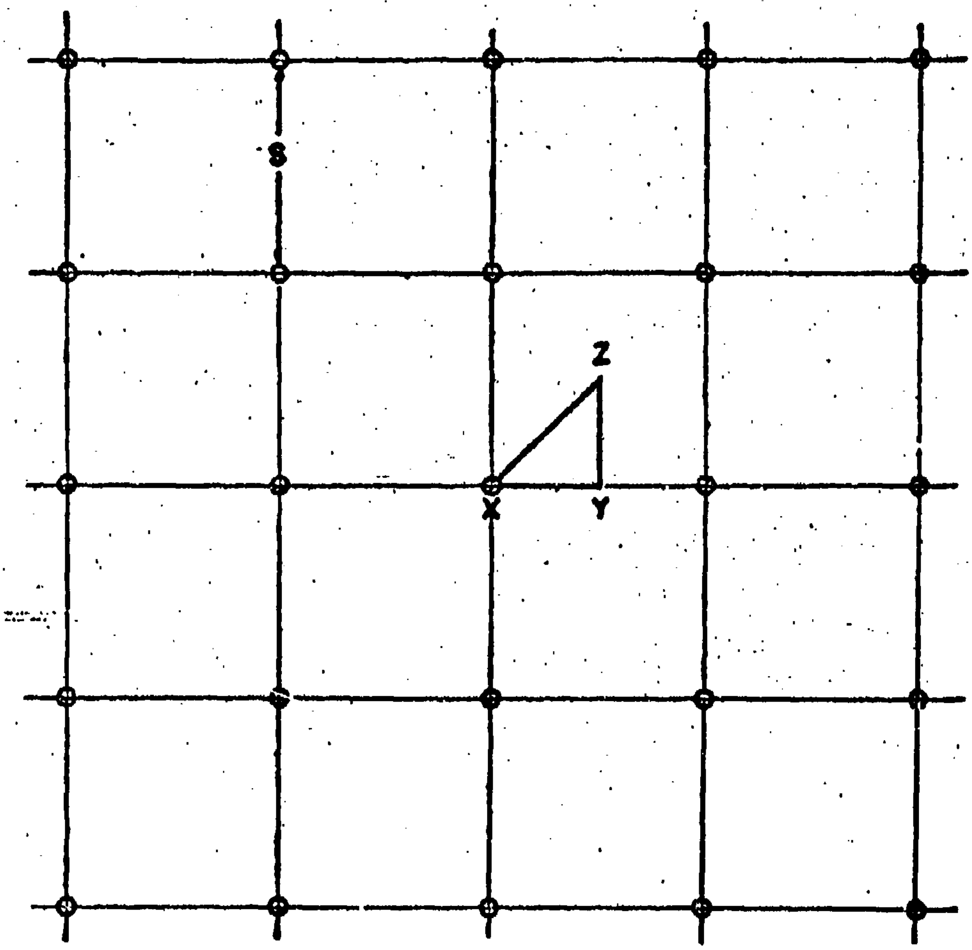
IMPACT CENTER COORDINATES

BOMB	X	Y
1	0.09	0.74
2	-0.94	0.52
3	0.17	1.48
4	-0.41	-1.53
5	-0.73	0.27
6	2.47	-0.81
7	0.58	0.49
8	1.38	-1.24
9	0.00	-0.78
10	-1.81	2.10
AVE.	0.087	-0.035

THE USE OF ISOHAPS
IN A MONTE CARLO DETERMINATION OF A COVERAGE PROFILE
FIG 1



HIT PROBABILITY INDEX
FOR A ROW OF AIMING POINTS
FIG 2

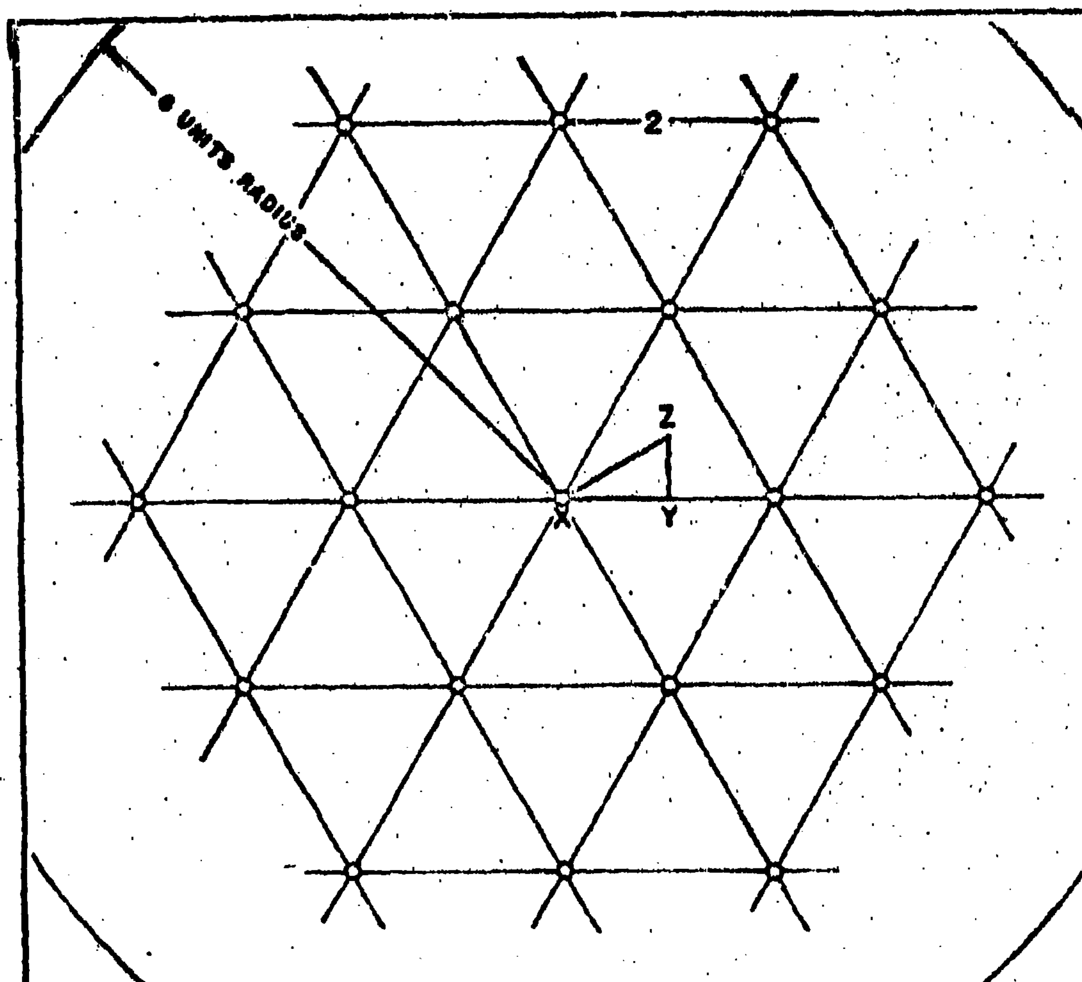


SEPARATION	S	1.8	1.9 *	2.0	2.1	2.2
HIT PROBABILITY	X	.7785	.6937	.6441	.5937	.5537
INDEX AT	Y	.7646	.6908	.6224	.5674	.5168
POINTS	Z	.7604	.6808	.6008	.5424	.4838

* OPTIMUM

COVERAGE INDICES
FOR A RECTANGULAR GRID OF AIMING POINTS

FIG 3



HIT PROBABILITY INDEX

POINT	X	Y	Z
INDEX	.7301	.7258	.7263

COVERAGE INDICES
FOR PART OF A HEXAGONAL GRID OF AIMING POINTS

FIG 4

AN EXPERIMENT IN PERSONNEL MANAGEMENT EVALUATION*

R. R. Blough
Statistical Research Center
University of Chicago

BACKGROUND. A personnel management program may be subdivided on paper into classes with such titles as

- Recruitment and placement
- Job classification
- Incentives and awards
- Disciplinary actions
- etc.

Each class may be further broken down into a list of duties or actions by management in connection with employees. Without further defining the elements constituting the program, we might ask such questions as,

Does such a program do any good, or any harm, and if so, how much?

Which possible elements of a program should be retained and which discarded in order to achieve maximum benefits?

As used here, "good" might mean an increase in productivity, or in quality of production, or in employee satisfaction.

Not much is known about answers to these questions. One reason may be that few organizations are large enough to have the facilities for finding answers experimentally. In 1957, however, a project was undertaken by the Office of the Deputy Chief of Staff for Personnel at Hq 5th Army, Chicago, to be directed by Arthur Barbour and Baldwin Sears of that Office, to acquire quantitative information. The Statistical Research Center was consulted in connection with design and analysis. It seemed that the experimenters in this instance at least had adequate manpower and

---

*This paper outlines an experiment described in more detail in SRC-600624-Bg88, a report of the Statistical Research Center dated 24 June 1960.

This work was sponsored by the Army, Navy and Air Force through the Joint Services Advisory Group for Research Groups in Applied Mathematics and Statistics by Contract No. N6ori-02035. Reproduction in whole or in part is permitted for any purpose of the United States Government.

facilities for experiments and replications: 300,000 civilian employees of the Department of the Army at numerous installations throughout the country.

In their initial trials, the experimenters decided to test the best and most comprehensive personnel management program they could devise, under the most favorable conditions for observing the effects of the program. If measurable effects indeed resulted, experiments would then be devised to investigate the program elements individually.

The maximum opportunity for observing improvement might appear to exist at an installation having no formal personnel management program at all prior to the start of the experiment. But such a primitive situation would likely exist only if the local commandant or management were unsympathetic toward personnel management programs and so probably toward the proposed experiment. In any case, such an installation would lack the trained personnel specialists capable of performing the experiment. As a compromise it was necessary to choose an installation having a reasonably good personnel management program already, expanding this program to "optimal" for the experiment.

The Decatur Signal Depot, Decatur, Illinois, which was chosen for the initial investigations, had a management personnel program "level" rated by the experimenters as 70 per cent of optimal. Thus only the effect of raising the level from 70 per cent to 100 per cent could result, and this effect might not be large. However, even a small effect might be well worth achieving. The experimenters estimated the equivalent in annual wages of a 5 per cent productivity increase throughout the Department of the Army to be about $75,000,000.

The variable of most interest at this time was in fact productivity, so the experiment was designed on that basis. Employee satisfaction, of which typical indicators are assumed to be so-called "employee reactions"--sick leave use, injuries, AWOL, voluntary separations, suggestions--was to be looked at incidentally, with interest in possible correlation with productivity. Quality of product was too subjective to be reliably assessed, and was not a variable of much concern in the experiment.

DESIGN. The design envisaged a minimum of 15 independent employee groups already existing in the organization, of size at least 10, and as alike as possible. (A primary objective here was to provide good conditions for observing an effect if present.) All groups should already have in routine operation a procedure for measuring productivity. The groups would be assigned randomly to 3 categories corresponding to what were familiarly called treatments (Table 1).

Table 1

Category	Treatment to be applied to groups in category
1. Uninformed controls	No treatment at all. It is assumed that these groups operate under the usual conditions and are ignorant of the experiment.
2. Informed controls	These groups are to be informed of the experiment but otherwise will remain under the usual conditions.
3. Experimental groups	The personnel management program applied to these groups is to be increased to 100 per cent of optimum.

(The Informed controls were included to provide against and test for the so-called Hawthorne effect, the effect on the subjects of merely being part of an experiment.) Monthly data were to be collected for 1 year, or some other suitable lengthy period, before the actual start of the experiment. The treatments would then be started and data collected for a comparable period during application of the treatments. The analysis was to be performed on numbers representing, for each group, the ratio

$$\frac{\text{treatment period performance}}{\text{pretreatment period performance}}.$$

IMPLEMENTATION. The actual experimental setup fell somewhat short of the specifications. 14 employee groups were originally chosen for the experiment, of which 5 were later dropped, leaving 9 groups (3 per treatment) instead of the recommended minimum of 15 groups. The assumption of independence for these groups appeared reasonable, but their sizes ranged from 4 to 19 employees, and they differed in composition (2 were partly made up of women). While most groups worked at storage and handling of various types of electronic equipment, one did clerical work and one manufacturing. Further, productivity standards were applicable to only about 50 to 75 per cent of the groups' jobs, so measurements of productivity reflected only a fraction of each group's total work. Data were available for only 4 months, Nov '57 - Feb '58, prior to the start of the treatments, and for 16 months after start of the treatments, Mar '58 - Jun '59. During the latter "treatment period" the experimenters estimated that the personnel management program level for the "Experimental groups" rose rather gradually from 70 per cent, reaching "very nearly" 100 per cent during the period Oct '58 - Feb '59 and then falling off. Thus the experiment proceeded under a number of handicaps which had not been foreseen.

DATA. The productivity data were constructed as follows. Suppose for job ν a time study has specified H_ν hours per unit of production. If an employee actually spends A_ν hours producing U_ν units, the product $H_\nu U_\nu$ is called the "Earned Hours" and A_ν the "Actual Hours" for that amount of work, and the corresponding productivity is

$$100 \frac{H_\nu U_\nu}{A_\nu} = 100 \frac{\text{Earned hours}}{\text{Actual hours}},$$

a measure of the productive use of time. Total productivity for work done on, say, n jobs, is

$100 \sum_{\nu=1}^{n} H_\nu U_\nu / \sum_{\nu=1}^{n} A_\nu$. Once $H_\nu U_\nu$ and A_ν are obtained for each employee's work on each job each day, productivity for any combination of employees, jobs, and days may be calculated. In the routine collection of productivity data at Decatur, Earned and Actual hours were respectively summed within each group over an entire month, so the ratio reported for each group, each month, was a monthly productivity for the group, applying of course, only to that part of the group's work covered by the time study standards.

Data on the above mentioned employee reactions (considered indications of employee satisfaction) were reported in terms of index numbers apparently intended to express the various "reactions" in comparable units, as, say, percentage of a "norm", where the norm is usually some average of past experience. It is not clear that the index numbers are always more informative or easier to interpret than the actual observed quantities. For example, the index originally adopted by the experimenters for use in this study for reporting reactions whose increased frequency of occurrence indicates a decrease in employee satisfaction (e.g., sick leave usage) was computed as

$$(\text{Norm} - \text{Observed}) + 100 = \text{Index},$$

for example, $$(98 - 100) + 100 = 98,$$

or $$(.5 - 2.5) + 100 = 98.$$

The indexes finally adopted for the experiment were, for such "desirable" employee reactions as "suggestions",

$$100\,\frac{\text{Actual}}{\text{Norm}}\,,$$

and for "undesirable" reactions,

$$100\left(1 - \frac{\text{Actual} - \text{Norm}}{\text{Norm}}\right) = 100\left(2 - \frac{\text{Actual}}{\text{Norm}}\right).$$

(These latter indexes are now standard for reporting about a hundred different items in routine evaluations of the Army's Civilian Personnel Program. However, the universal usefulness of the indexes is not clear, as illustrated by an example arising in the present experiment. A Norm for voluntary separations--mainly, number of employees quitting--was computed as the average percentage of the work force being separated voluntarily per month for each month, from data collected over an earlier 2-year period. There were 2 voluntary separations from all 9 groups over the 20 month period of the experiment, one of these occurring in a group of size 22, for a rate of 4.545 per cent in the month. The norm for that month was .247 per cent, and resulting index was $100(2 - 4.545/.247) = -1640$--representing the smallest nonzero separation rate that could have been obtained for this group--to be compared with the index 200 for a zero rate.)

Because of their rather dubious meaning, the index numbers were not used in the analysis of the experiment. Actually, for the most part, there were too few occurrences among the "employee reactions" to permit analysis.

ANALYSIS. In general, let

$i = 1, 2, \ldots, I$ treatments. In this case $I = 3$.

$j = 1, 2, \ldots, J$ groups within each treatment. Here $J = 3$.

$k = 1, 2, \ldots, K_p$ pretreatment months. Here $K_p = 4$.

$k' = 1, 2, \ldots, K_T$ treatment months. Here $K_T = 16$.

Then, for group j within treatment i, and pretreatment month k, designate the productivity random variable by X_{ijk}, and let $Y_{ijk'}$ be the random variable denoting the same group's productivity in treatment month k'. Making the transformation to $\log X_{ijk}$ and $\log Y_{ijk'}$ (in order to normalize the productivities, which are ratios) and averaging, for group j within treatment i, the transformed pretreatment period data and treatment period data over desired sets of months, form the difference

$$Z_{ij} = \frac{1}{K_2} \cdot \sum_{k'}^{K_2} \log Y_{ijk'} - \frac{1}{K_1} \cdot \sum_{k}^{K_1} \log X_{ijk},$$

where K_1 = some number of pretreatment months, $0 \leq K_1 \leq K_p$,

K_2 = some number of treatment months, $0 \leq K_2 < K_T$.

Then

$$Z_{ij} = \log \left(\frac{\text{geometric mean of } K_2 \text{ treatment monthly productivities}}{\text{geometric mean of } K_1 \text{ pretreatment monthly productivities}} \right),$$

the geometric means being a consequence of the log transformation. The operations of forming Z_{ij} may be assumed to have removed the group effect, and since groups are independent of each other, to have resulted in an observation for each group to which the following simple model applies:

$$Z_{ij} = \mu + \alpha_i + e_{ij}$$

where

μ = over-all mean

α = treatment effect, $\sum \alpha_i = 0$.

e_{ij} = random error, normally distributed with mean 0 and variance σ^2, the e_{ij}'s being mutually independent.

Analyses may now be performed on observed values of the Z_{ij}'s calculated from the productivity data.

The Z_{ij}'s could be formed from averages over any months available, and in fact 5 analyses of productivities were performed using various combinations of monthly observations. The analysis which employed productivity comparison ratios for the two periods Nov '57 - Feb '58 and Nov '58 - Feb '59 appeared to be the most appropriate and indeed gave the lowest estimate of residual variance. The results of this analysis are summarized below in Tables 2 and 3. Table 2 contains point estimates $\widehat{\mu+\alpha_i}$ of $\mu+\alpha_i$, converted back (by taking antilogs) to a (treatment period)/(pretreatment period) productivity ratio. Table 3 contains antilogs of differences $(\widehat{\mu+\alpha_i}) - (\widehat{\mu+\alpha_{i'}})$, that is, the entries are ratios of the ratios in Table 2, and of 95 per cent confidence limits for the true differences, also converted back to ratios.

Table 2

i	Category or treatment	Antilog $(\widehat{\mu+\alpha_i})$
1	Uninformed controls	1.06
2	Informed controls	1.02
3	Experimental groups	.98

Table 3

i/i'	Corresponding categories	Antilog $(\widehat{\mu+\alpha_i})-(\widehat{\mu+\alpha_{i'}})$	Antilog of John Tukey 95% confidence limits
3/2	Experimental/Informed	.96	.80, 1.14
2/1	Informed/Uninformed	.96	.81, 1.14
3/2	Experimental/Uninformed	.92	.77, 1.10

The estimates of Table 2 say, for example, that the Uninformed controls improved 6 per cent during the treatment period as compared to their performance in the pretreatment period, while the Experimental groups declined 2 per cent. In Table 3 the relative improvement of the Experimental as compared with the Informed groups was .95, which superficially suggests that optimization of the personnel management program is detrimental. However, the confidence limits for the ratios of Table 3 obviously indicate such large variability that nothing is, and little could be, statistically significant. In fact, the power of this test against a real 5 per cent increase in productivity was estimated as about .1, and to raise the power to .9 would require an estimated 44 groups per treatment, or 132 groups in all, a seemingly prohibitive number.

The sketchy analyses of employee reactions which were possible also showed no statistically significant effects which could be attributed to the treatments. One of the groups did show a statistically significantly higher rate of sick leave usage than the others. This group was small, with a high proportion of women employees.

SOURCES OF ERROR. Some possible contributors to the large variability are

1. Differences between groups in composition, size, and type of work.

2. Supervisory differences and differences in the personnel management treatment received by groups within a given category.

3. Differences created by the standards of productivity. For example, time study may allot too few hours (say H_s) or too many hours (H_ℓ) per unit of production. Then for A hours actually spent producing U units,

$$\frac{H_s U}{A} < \frac{H_\ell U}{A} .$$

That is, the apparent productivity depends on the standards, and shifting from jobs with strict standards (H_s) to those with lenient standards (H_ℓ) will cause an apparent increase in productivity when the actual productivity is unchanged. Standards for the same job are often revised, but this is not believed to have happened during this particular experiment.

Also, standards may be set up, in an installation like Decatur, to apply to handling individual items. Occasionally, large orders will require handling of gross lots by lift truck with consequent remarkable temporary rises in reported productivity.

4. Fluctuating workload. For example, during this experiment the invasion of Lebanon occurred, which caused a great increase in demands made on this installation.

5. Errors in collecting, computing and reporting. The task of recording and computing H_iU_i and A_i for all employee x day x job combinations contains many opportunities for error. In one case (found in previous work where raw data were examined in detail) one employee on one day on one job was reported to have produced 1403 units. His productivity was 1822. The job number turned out to be 1403, and this apparent error, when eliminated, reduced the group's monthly figure from 106 to 102. Thus a single error had increased the group's reported productivity by an amount comparable with that of the effect looked for in this experiment.

Other basic difficulties may be inferred from the fact that after close of the experiment the experimenters said that they doubted that the optimization of the personnel management program sought for the Experimental groups had been attained, and that the level which the treatment had actually reached was not very precisely known.

COMMENTS. Improvements in experimental technique are evidently required to obtain useful results from future experiments of practical size. It is likely that variability can be reduced by such means as care in selecting groups, elimination of clerical errors, and exclusion of data arising from abnormal circumstances. However, it appears only prudent at this stage to utilize as many employee groups as possible to attempt to overcome the effect of variability still present.

There remain the basic requirements, such as, that the groups must be and must remain independent, and must receive the treatment specified. It would seem that only local management can assure that even the most general design conditions are met, and so, as essential participants, local management should have adequate understanding of the experiment and its objectives.

A NOTE ON APPROXIMATE CONFIDENCE INTERVALS FOR FUNCTIONS OF BINOMIAL PARAMETERS

Henry DeCicco
Ordnance Special Weapons-Ammunition Command
Dover, New Jersey

I. INTRODUCTION. A system is made up of a number of components in arbitrary combination, and it is required to obtain a confidence interval for the reliability of the system without testing the whole system itself. That is, we have at our disposal only the test data (let us assume in the form of binomial success ratios) on the components of the system.

Special cases of this problem have been treated by Buehler (1) and Madansky (4). The method presented herein leads to approximate confidence intervals but is general enough to cover arbitrary systems with relative ease. It is also capable of accommodating the case where the components of the system are statistically dependent, although this case is not developed here. The method involves computing moments of functions of random variables, in particular those functions of the observed binomial data described by the probability structure of systems of interest. Although it is entirely feasible to compute the first four moments of such functions and thereby settle the question of a relevant distribution function, practical work generally requires no more than the first two. The following discussion is, accordingly, so limited.

II. N INDEPENDENT COMPONENTS IN SERIES. Let $\hat{p}_i$ denote the observed k successes in m binomial tests recorded for the i-th component, and let p_i denote the associated binomial parameter. For a system of n statistically independent components in series we have

$$(2.1) \qquad F(\hat{p}_1, \hat{p}_2, \ldots \hat{p}_n) = \prod_i \hat{p}_i \qquad i = 1, 2, \ldots n$$

$$(2.2) \qquad E(F) = \prod_i p_i$$

$$(2.3) \qquad \operatorname{Var}(F) = \prod_i (p_i^2 + \operatorname{Var} \hat{p}_i) - \prod_i p_i^2$$

the last two following directly from the definitions of expected value and variance.

A more explicit form of (2.3) for computation is

$$\text{(2.3a)}\quad \mathrm{Var}(F) = \sum_{i_2}\Big(\prod_{i_1\neq i_2} p_{i_1}^2\Big)\mathrm{Var}\,\hat{p}_{i_2} + \sum_{i_2,i_3}\Big(\prod_{i_1\neq i_2,i_3} p_{i_1}^2\Big)\mathrm{Var}\,\hat{p}_{i_2}\,\mathrm{Var}\,\hat{p}_{i_3} + \dots + \sum_{i_2,i_3,\dots,i_n}\Big(\prod_{i_1\neq i_2,i_3,\dots,i_n} p_{i_1}^2\Big)\mathrm{Var}\,\hat{p}_{i_2}\,\mathrm{Var}\,\hat{p}_{i_3}\dots\mathrm{Var}\,\hat{p}_{i_n} + \prod_i \mathrm{Var}\,\hat{p}_i$$

where all distinct subscripts are summed from 1 to n to yield $2^n - 1$ terms and where

$$\mathrm{Var}\,\hat{p}_i = p_i(1 - p_i)/m_i \ .$$

While the last result is exact, it is easily seen that a serviceable approximation exists in its linearized version, obtainable directly from the classical propagation of error formula, that is

$$\text{(2.4)}\quad \mathrm{Var}(F) \approx \sum_i \left(\frac{\partial F}{\partial \hat{p}_i}\right)_o^2 \mathrm{Var}\,\hat{p}_i$$

where the partials are understood to be evaluated at the point $p_1, p_2, \dots p_n$. Applying this to (2.1) gives

$$\text{(2.5)}\quad \mathrm{Var}(F) \approx \sum_{i_2}\Big(\prod_{i_1\neq i_2} p_{i_1}^2\Big)\frac{p_{i_2}(1 - p_{i_2})}{m_{i_2}}$$

which is just the first n terms out of the total $2^n - 1$ in (2.3a). Application of (2.4) naturally assumes that F has been redefined for continuity since the $\hat{p}_i$'s take on only fractional values on the unit interval.

Finally, corresponding to (2.2) and (2.3) the relevant unbiased estimates are easily shown to be

$$(2.6) \quad \widetilde{E}(F) = \prod_i \hat{p}_i$$

$$(2.7) \quad \widetilde{\mathrm{Var}}(F) = \prod_i \hat{p}_i^2 - \prod_i \left\{ \hat{p}_i^2 - \frac{\hat{p}_i(1-\hat{p}_i)}{m_i - 1} \right\}$$

III. MORE GENERAL SYSTEMS. The probability structure of each system is of course special, and it would be pointless to attempt a catalogue of these. Even so, it might be useful to characterize a fairly general structure to suggest the flexibility of the method of linearized estimates. Such a structure might be as follows: assuming statistical independence throughout, we consider n assemblies in series where an assembly is made up of s_i identical components in parallel and where, further, at least a_i of the s_i components must function for the assembly to function. We then have

$$(3.1) \quad F = \prod_i \left\{ \sum_{y=a_i}^{s_i} \binom{s_i}{y} \hat{p}_i^y (1-\hat{p}_i)^{s_i - y} \right\} \qquad i = 1, 2, \ldots n$$

and to a first order approximation

$$(3.2) \quad \widetilde{E}(F) \approx \prod_i \left\{ \sum_{y=a_i}^{s_i} \binom{s_i}{y} \hat{p}_i^y (1-\hat{p}_i)^{s_i - y} \right\}$$

$$(3.3) \quad \widetilde{\mathrm{Var}}(F) \approx$$

$$\sum_{i_2} \left\{ \prod_{i_1 \neq i_2} \left(\sum_{y=a_{i_1}}^{s_{i_1}} \binom{s_{i_1}}{y} \hat{p}_{i_1}^y (1-\hat{p}_{i_1})^{s_{i_1}-y} \right) \cdot \right.$$

$$\left. \left(\sum_{y=a_{i_2}}^{s_{i_2}} \binom{s_{i_2}}{y} \left[y \hat{p}_{i_2}^{y-1} (1-\hat{p}_{i_2})^{s_{i_2}-y} - (s_{i_2}-y)\hat{p}_{i_2}^{y} (1-\hat{p}_{i_2})^{s_{i_2}-y-1} \right] \right) \right\}^2 \cdot$$

$$\widetilde{\mathrm{Var}}\, \hat{p}_{i_2}$$

where, as before, both subscripts are summed from 1 to n. The foregoing are usually biased in keeping with a general limitation of linearized estimates. (Refer Concluding Remarks.)

In the case of simply redundant systems where $a_i = 1$ we readily obtain the corresponding expressions

$$(3.1a)\quad F = \prod_i \left\{1 - (1-\hat{p}_i)^{a_i}\right\}$$

$$(3.2a)\quad \tilde{E}(F) \approx \prod_i \left\{1 - (1-\hat{p}_i)^{a_i}\right\}$$

$$(3.3a)\quad \widetilde{Var}(F) \approx \sum_{i_2} \left[\prod_{i_1 \neq i_2} (1 - [1-\hat{p}_{i_1}]^{a_{i_1}})\, a_{i_2} (1-\hat{p}_{i_2})^{a_{i_2}-1}\right]^2 \widetilde{Var}\ \hat{p}_{i_2}$$

IV. APPROXIMATE CONFIDENCE INTERVALS. The limited experience of the writer to date with systems of particular interest to Ordnance has indicated that Var (F) is generally very small compared to E (F). We now quantify the term "small" to determine a numerical condition under which the first two moments as discussed above are enough to give reasonably useful confidence intervals. To do this we examine the ratio $E(F)/\sqrt{Var\ F}$ in the context of Tchebycheff's inequality. We readily obtain

$$(4.1)\quad Pr\left\{\left|\frac{F - E(F)}{E(F)}\right| \leq \epsilon\right\} \geq 1 - \frac{1\ Var(F)}{\epsilon^2\, E^2(F)}$$

Which has the common sense interpretation that the larger the ratio $E(F)/\sqrt{Var(F)}$, the smaller is the probability that a particular observation of our chance quantity $F(\hat{p}_1, \hat{p}_2, \ldots \hat{p}_n)$ will deviate beyond a given distance ϵ from $E(F)$.

Figure 1 is a plot of the bound, ϵ of the relative deviation against the ratio $E(F)/\sqrt{Var(F)}$ with confidence level δ as parameter. That is, the condition

$$Pr\left\{\left|\frac{F-E(F)}{E(F)}\right| \leq \epsilon\right\} \geq 1 - \delta$$

implies the relation

$$(4.2) \qquad \epsilon = \frac{\sqrt{\text{Var }(F)}}{\sqrt{\delta}\, E\,(F)}$$

In Figure 1, δ is taken as .10 for the 90% confidence interval usually desired in engineering applications.

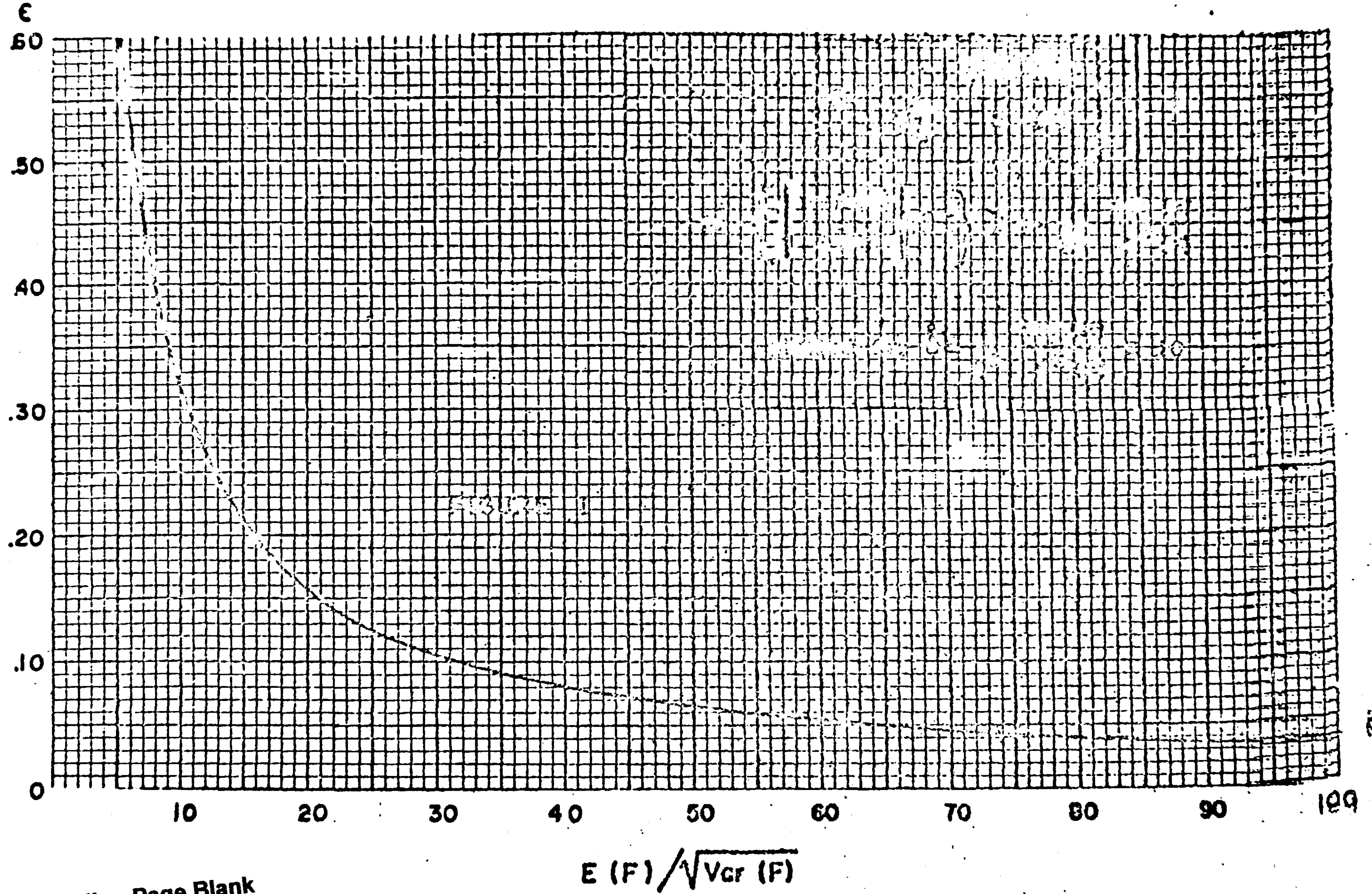
The ratio $E(F)/\sqrt{\text{Var }(F)}$ is commonly referred to as the "signal-to-noise ratio." Designating this ratio by R, we recognize the plot of Figure 1 as hyperbolic, that is,

$$\epsilon = k/R, \quad k = 1/\sqrt{\delta}\,.$$

We see that when the ratio exceeds 30, roughly, the relative deviation is not likely to exceed 10%. This is, of course, saying nothing more than that some 3 standard deviations on either side of the mean value of a distribution (unspecified save for having a finite variance) will cover about 90% of the range of values. However, the real advantage of such a plot is that it shows that after a certain point, large values of $E\,(F)/\sqrt{\text{Var }(F)}$ do not influence the bound ϵ very much. The fact that the curve in Figure 1 is relatively flat over a whole region is often useful in deciding when estimates of even the first two moments are enough to settle, in a practical sense, questions concerning whether a prescribed level of reliability for a complicated system is likely to have been satisfied. Observe further that this fact also allows for considerable imprecision in the estimates of both E(F) and Var (F). If, after such a computation is made, one requires fuller information, it would be necessary to calculate higher moments. The work of Tukey (5) provides expressions for the first four moments that go considerably beyond the level of refinement of linearized estimates. However, a major conclusion of that work is that, with particular reference to the classical propagation of error formula (that is, the formula for Var (F)), linearized estimates are often better than commonly supposed.

It might be well to emphasize that the explicit form exhibited by linearized estimates of Var (F) serves the further useful purpose of exposing those components and substructures of a system that appear as major contributors to the overall variability. A rational allocation of additional component tests, for the purpose of reducing that variability, is thereby indicated. (See example below.)

RELATIVE DEVIATION AS A FUNCTION OF SIGNAL-TO-NOISE RATIO



Preceding Page Blank

Finally, a comparison of certain calculations given by both Buehler (1) and Madansky (4) indicates good agreement with the method of this note. For example, Madansky (4) gives the following comparison with a result of Buehler for the upper limit of a 90% confidence interval for the probability of failure of a two component, parallel system, where 3 failures in 100 tests were recorded for one component and 5 failures in 100 tests were recorded for the other; Buehler (1) obtains, for the upper limit, .004.2. Madansky (4) obtains .00518.

Interpreting $\hat{p}_1 = 3/100$, $\hat{p}_2 = 5/100$ now as <u>failure</u> rates, and using equations (2.6) and (2.7), we obtain the following unbiased estimates

$$\tilde{E}(F) = \hat{p}_1\hat{p}_2 = .00150$$

$$\tilde{Var}(F) = \hat{p}_1^2\,\hat{p}_2^2 - \left\{\hat{p}_1^2 - \left(\frac{\hat{p}_1(1-\hat{p}_1)}{m_1-1}\right)\right\}\left\{\hat{p}_2^2 - \left(\frac{\hat{p}_2(1-\hat{p}_2)}{m_2-1}\right)\right\}$$

$$= 102.56 \times 10^{-8}$$

so that

$$\tilde{E}(F) + 3.162\sqrt{\tilde{Var}(F)} = .00470.$$

The corresponding linearized estimate based on (2.4) yields an upper limit of .00491 so that a positive bias in the amount .00021 is thereby incurred.

V. <u>EXAMPLE.</u> The following example indicates a simple application of (3.2a) and (3.3a) to a system of a common generic type. In addition to computing a lower 90% confidence limit for the reliability of the system, we exhibit the structure or the associated variability explicitly and also take note of the signal-to-noise ratio.

The system is made up of four assemblies in series (we shall assume statistical independence throughout). The first assembly consists of a single component. The second assembly consists of two identical components in parallel, at least one of which must function for the assembly to function. The third assembly consists of three identical components at least one of which must function. The fourth assembly consists of a single component.

Preceding Page Blank

The following test data applies (observed success ratios):

$$\hat{p}'_1 = \frac{198}{200}$$

$$\hat{p}'_2 = \frac{194}{200}$$

$$\hat{p}'_3 = \frac{190}{200}$$

$$\hat{p}'_4 = \frac{196}{200}$$

Corresponding to (3.2a) and (3.3a), respectively, we have

$$\tilde{E}(F) \approx \hat{p}_1 \left\{1 - (1-\hat{p}_2)^2\right\} \left\{1 - (1-\hat{p}_3)^3\right\} \hat{p}_4 \approx .969$$

$$\widetilde{Var}(F) \approx \left[\left\{1 - (1-\hat{p}_2)^2\right\} \left\{1 - (1-\hat{p}_3)^3\right\} \hat{p}_4\right]^2 \cdot \frac{\hat{p}_1(1-\hat{p}_1)}{199} +$$

$$+ \left[2\hat{p}_1 (1-\hat{p}_2) \left\{1 - (1-\hat{p}_3)^3\right\} \hat{p}_4\right]^2 \cdot \frac{\hat{p}_2(1-\hat{p}_2)}{199} +$$

$$+ \left[3\hat{p}_1 \left\{1 - (1-\hat{p}_2)^2\right\} (1-\hat{p}_3)^2 \hat{p}_4\right]^2 \cdot \frac{\hat{p}_3(1-\hat{p}_3)}{199} +$$

$$+ \left[\hat{p}_1 \left\{1 - (1-\hat{p}_2)^2\right\} \left\{1 - (1-\hat{p}_3)^3\right\}\right]^2 \cdot \frac{\hat{p}_4(1-\hat{p}_4)}{199}$$

$$\approx 144.5 \times 10^{-6} .$$

We obtain the desired lower limit

$$\tilde{E}(F) - 3.16\sqrt{\widetilde{Var}(F)} = .931.$$

The components of the variability are as follows:

$$\left(\frac{\partial F}{\partial \hat{\beta}_1}\right)_0^2 \widetilde{\text{Var}}\ \hat{p}_1 = 47.68 \times 10^{-6}$$

$$\left(\frac{\partial F}{\partial \hat{\beta}_2}\right)_0^2 \widetilde{\text{Var}}\ \hat{p}_2 = .4954 \times 10^{-6}$$

$$\left(\frac{\partial F}{\partial \hat{\beta}_3}\right)_0^2 \widetilde{\text{Var}}\ \hat{p}_3 = .01262 \times 10^{-6}$$

$$\left(\frac{\partial F}{\partial \hat{\beta}_4}\right)_0^2 \widetilde{\text{Var}}\ \hat{p}_4 = 93.33 \times 10^{-6}$$

We thereby observe that some two-thirds of the total variability comes from the final assembly alone, the remainder arising almost entirely from the first. Observe that this conclusion is far from obvious, since the assemblies cited are precisely those with the lowest observed failure rates and, indeed, the lowest individual variances. It is therefore clear that the most direct approach to reducing the overall variability would be to increase the number of tests on the fourth and first component types.

Finally, we observe a signal-to-noise ratio about 80 which, according to Figure 1, is within a relatively flat region of the bound ϵ. Considerable variation in the estimate of $E(F)/\sqrt{\text{Var}(F)}$ is therefore not likely to influence a practical decision based on the estimated reliability of the system.

VI. <u>CONCLUDING REMARKS.</u> It is clear that if the observed success ratios on <u>all</u> the components of a system are either zero or one, then the computed variance of $F(\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{p}_n)$ will vanish and no very useful information is obtained. This case would represent an intrinsic limitation of the method of moments, but from the standpoint of applications it does not appear to be one frequently or even occasionally encountered.

The bias associated with estimates of E (F) and Var (F) has been considered here only with respect to series systems. It is possible, in principle, to assess and remove the bias implied in linearized estimates, computed for more general systems, in terms of expressions for moments of the kind given in reference (4). The amount of work involved in this will generally be prohibitive, though for common system types it will be feasible.

ACKNOWLEDGEMENT. For discussions of the problem of this note and its ramifications, the writer gratefully thanks D. H. Evans, A. Stein, F. E. Grubbs and S. Ehrenfeld.

REFERENCES

(1) Buehler, R. J.: Confidence Intervals for the Product of Two Binomial Parameters, Journal of the American Statistical Association, December 1957.

(2) DeCicco, H.: The Reliability of Weapon Systems Estimated From Component Test Data Alone, Ordnance Special Weapons-Ammunition Command Technical Note 1, December 1959.

(3) DeCicco, H.: The Error in Linearized Estimates of the Variance of Products, Ordnance Special Weapons-Ammunition Command Technical Note 2, February 1960.

(4) Madansky, A.: Approximate Confidence Limits for the Reliability of Series and Parallel Systems, The Rand Corporation, 4 April 1960, RM 2552.

(5) Tukey, J. W.: Propagation of Errors, Fluctuations and Tolerances, Basic Generalized Formulas, Technical Report No. 10, Princeton University, 1958.

PERFORMANCE OF PROPELLANTS EVALUATED BY TENSILE AND BALLISTIC TESTS

Boyd Harshbarger
Virginia Polytechnic Institute
ARGMA Redstone Arsenal

Niles White
Propellant Branch
Propulsion Laboratory
ARGMA Redstone Arsenal

The objective of this paper is to show the functions which describe the relations existing between static test results (average motor pressure and 50% burning time) and flight test results (burning distance, burnt velocity, maximum cartridge case pressure, muzzle velocity, and burn-out time) of a spin stabilized rocket. A second objective is to list confidence limits in order that the results may be better evaluated.

The equations found relating flight test results to static test results are generally linear or of linear form, except for one logarithmic term due to flight temperatures. Equations predicting burning distance, burnt velocity and burn-out time are described. These equations express the flight test values as functions of various static test results. The equations which express flight test values are given in terms of (i) six static test values, (ii) four static test values, (iii) three values found from the six static test values, and (iv) logarithms of six static test values. Generally, the equations which involve all the static values are best for predictive purposes. All the different type equations mentioned are related to arithmetic means of flight test results and to arithmetic means of static test results.

Some relations similar to those mentioned above were found involving logarithms of variances rather than means. These equations do not predict variances with as much accuracy as the corresponding equations predict means. Thus, most of this paper will be devoted to a discussion of the best equations which use means for variables.

There are a number of flight variables but of the three flight variables, burning distance, burn-out time, and burnt velocity, the latter can be predicted with the most accuracy. The best equations predicting burnt velocity are as follows:

$$y_b = -5603.6 - 66004X_1 - .69340X_2 + .32081X_3 - 589.54X_4$$
$$-1359.2X_5 + 733.30X_6 + 3675.9 \log (X_7 + 623). \qquad (1)$$

gives the relation between burnt velocity, y_b, in feet per second and the six static variables and temperature, where X_1, X_2, X_3 represent average motor pressures in pounds per square inch at -40°F, 70°F, and 160°F, respectively; X_4, X_5, X_6 represent 50% burning time in seconds at -40°F, 70°F, 160°F, respectively; and X_7 is one of the flight temperatures -20°F, 70°F, 140°F.

$$y_b = 1952 + .07765X_2 + .1536X_3 + 543.2X_5 - 268.7X_6 + 2.329X_7 \quad (2)$$

gives burnt velocity in terms of only five variables, X_2, X_3, X_5, X_6, X_7, defined as for equation (1).

$$y_b = -7831.7 - 1.6075u_1 - 97.430u_2 + 386.49u_3 + 3387.7 \log (X_7 + 578.8) \quad (3)$$

gives burnt velocity as a function of four variables u_1, u_2, u_3, and X_7, where $u_1 = \log X_1 - \log X_4$, $u_2 = \log X_2 - \log X_5$, $u_3 = \log X_3 - \log X_6$, and X_7 is defined as above.

All three of these equations are highly predictive. Just how good equation (1) is as a predictor of the actual flight values can be seen by looking at Tables I, II, and III, where it will be observed that the predicted values are less than 4.5% in error when compared with the actual flight values. The 99% confidence intervals for predicted values at -20°F, 70°F, 140°F are respectively,

$$2464.6 \leq y \leq 2593.4,$$

$$2686.6 \leq y \leq 2815.4,$$

$$2840.8 \leq y \leq 2969.6.$$

TABLE I

Predicted Flight Measurements Using Equation (1)

Mix Number	at -20°F		
	Actual Value	Predicted Value	Percent Value
1495	2440	2486	1.88
1157	2588	2529	-2.27
1174	2580	2531	-1.89
1187	2628	2513	-4.37
1248	2550	2560	.39
1261	2502	2529	.99
1280	2480	2469	-.44
1293	2528	2538	.39
1308	2482	2525	1.73
1342	2497	2516	.76
1351	2500	2490	-.40
1399	2555	2523	-1.25
1361	2558	2490	-2.65
1372	2496	2534	1.52
1373	2475	2541	2.66
1385	2472	2563	3.68

TABLE II

Predicted Flight Measurements Using Equation (I)

Mix Number	at 70°F		
	Actual Value	Predicted Value	Percent Value
1495	2737	2708	-1.05
1157	2786	2751	-1.25
1174	2818	2753	-2 -2.03
1187	2782	2735	-1.68
1248	2775	2782	.25
1261	2708	2751	1.58
1280	2708	2691	-.62
1293	2737	2760	.84
1308	2746	2747	.03
1342	2737	2738	.03
1351	2672	2712	1.49
1399	2720	2745	.91
1361	2745	2712	-1.20
1372	2733	2756	.84
1373	2743	2763	.72
1385	2740	2785	1.45

TABLE III

Predicted Flight Measurements Using Equation (1)

Mix Number	at 140°F Actual Value	Predicted Value	Percent Value
1495	2837	2862	.88
1157	2954	2905	-1.65
1174	2893	2907	.48
1187	2980	2889	-3.05
1248	2952	2936	-.54
1261	2890	2905	.51
1280	2878	2845	-1.14
1293	2910	2914	.13
1308	2950	2901	-1.66
1342	2932	2892	-1.36
1351	2764	2866	3.69
1399	2890	2899	.31
1361	2865	2866	.03
1372	2852	2910	1.96
1373	2910	2917	.24
1385	2888	2939	1.76

A study similar to the relationship between static firings and flight firings have been considered by the Army Rocket and Guided Missile Agency involving physical test data. Propellants designated A, B, C, and D were mixed and cast in cylindrical specimens and subsequently guillotine sliced and dog bones stamped out. Propellants A, B, and C specimens had two vertically aligned dots spaced one inch apart, and a photographic technique was used to measure the longitudinal extension. This technique should eliminate minor dimensional changes. Propellant D did not utilize the photographic technique. In all cases, however, the propellant was obtained from a cylindrical carton and no attempt was made to number these propellant samples to designate them from adjacent samples. All propellants are within batch data except propellant D which is five batches.

The standard deviations expressed as a percent of ballistic and physical property data are summarized below:

Propellant

	A			B			C		
	-40	78	143	-40	78	143	-40	78	143
Strain at Max. Stress Std. Dev. %	4.1	6.3	3.4	39.6	7.6	14.6	31.2	7.5	6.8
Total Variation Impulse-%		0.48 N=18			1.03 N=10			0.96 N=10	

Propellant D

	-25°F	78°F	125°F
Strain at Maximum Stress Standard Deviation	12.08 13.51 15.25 14.93 15.85	8.35 27.50 21.92 30.66 31.41	5.30 9.70 8.27 12.58 9.46
AVERAGE	16.85	25.6	9.1
Total Impulse Variation			0.70 (N-26)

This data is considered tentative; however, the wide variation in standard deviation for propellants B, C, and D is considered signigicant, and attempts will be made to explain this phenomenon. Certainly with these wide within batch variations quality control at the mix site does not appear to be the answer. Perhaps polymer control and/or better dispension of the liquid solid phase and subsequent polymerization is in order. It may be significant that the propellant designated A is a solution process and the variation of total impulse might be even lower if the formulation were out of the research stage as in propellant D.

Statistically designed experiments on a rather large scale have been proposed for this physical test study. When these experiments involve stamped out dog bones the analysis should show whether the new techniques can be used to describe the variable in flight tests for the several types of solid propellants and for environmental effects.

PROBLEMS IN THE ANALYSIS AND INTERPRETATION OF INFORMATION PROCESSING EXPERIMENTS*

Emil H. Jebe and William A. Brown**
Operations Research Department
Institute of Science and Technology
The University of Michigan

Studies in information processing are being conducted by the Information Processing Task (IPT) of Project MICHIGAN in the Institute of Science and Technology of The University of Michigan. The experiments utilize "real world" elements and simulated elements.

The "real world" part consists of a processing station manned and operated by a six member crew. On the other hand, the non-real parts comprise a simulated tactical situation and simulated military sensors whose purpose is to furnish information on enemy task force movements. A crew is assigned the task of keeping up with the locations of enemy elements in the simulated tactical situation. Differences between the "postulated" locations and the true positions in a specific problem or run provide a score which is the measurement of the performance of a crew. This score is the response variable which is being analyzed.

This general description of what is being done needs to be expanded in terms of:

1. The tactical situtation being studied,

2. The operation of the processing station,

3. The factors which have been studied,

4. The inputs to the station,

5. The outputs of the station.

*This work was conducted by the Information Processing Task (IPT) of Project MICHIGAN under Department of the Army Contract DA-36-039 SC-78801, administered by the U. S. Army Signal Corps.

**Mr. Brown is now with the Physics Department.

Understanding of these five items will be aided by a flow chart diagram of the operations, see Figure 1.

The simulated tactical maneuvers take place within a military reservation in the western United States. The area of interest is roughly a square measuring 20 miles on a side. It is assumed that it is daylight with clear weather and the Blue forces have air superiority. Red forces consist of parts of two divisions which have moved into the reservation as an attacking force. Blue forces prepare to repel the attacking Reds. It is expected that Red Forces will be dispatched to counter the Blue movements. These Red Forces may move about 20-25 miles usually on roads through the reservation area between 1300 and 1700 hours. A normal amount of miscellaneous traffic in the area (called tactical noise) takes place due to Red movements not directly associated with the moving task forces.

It is the function of the surveillance information processing station to receive and process reports on the movements of these Red Task Forces, i.e., to track these concentrations and report where they are at selected times. "Where they are", of course, means where they are estimated to be. Thus, the purpose of these experiments is to investigate the performance of combat surveillance processing system concepts as implemented in a laboratory station. Specifically, measurements are made of the ability of station personnel to locate military concentrations as a function of selected sensor characteristics, the given tactical situation and the modes of operation of the station itself.

In a number of experiments five crews have been used to operate the station with each crew repeating the same problem five times. The latter effect is referred to as "Repetitions." Other factors that have been varied are the scan rates for the two sensors used and the detection probabilities of these sensors. The number of moving Red Task Forces has usually been held fixed for a single experiment but has varied from one to five among experiments. In one experiment the number of moving task forces was varied with levels 1, 2, 3, 4, or 5.

An IBM 709 computer is used to prepare the inputs to the processing station. A set of computer programs store the terrain information and the Red Task Force movements in the target area. In addition, the computer programs simulate the output of the sensors and messages and overlays are prepared for the surveillance station to process. The content and nature of these inputs to the station are modified from the actual task force movements by:

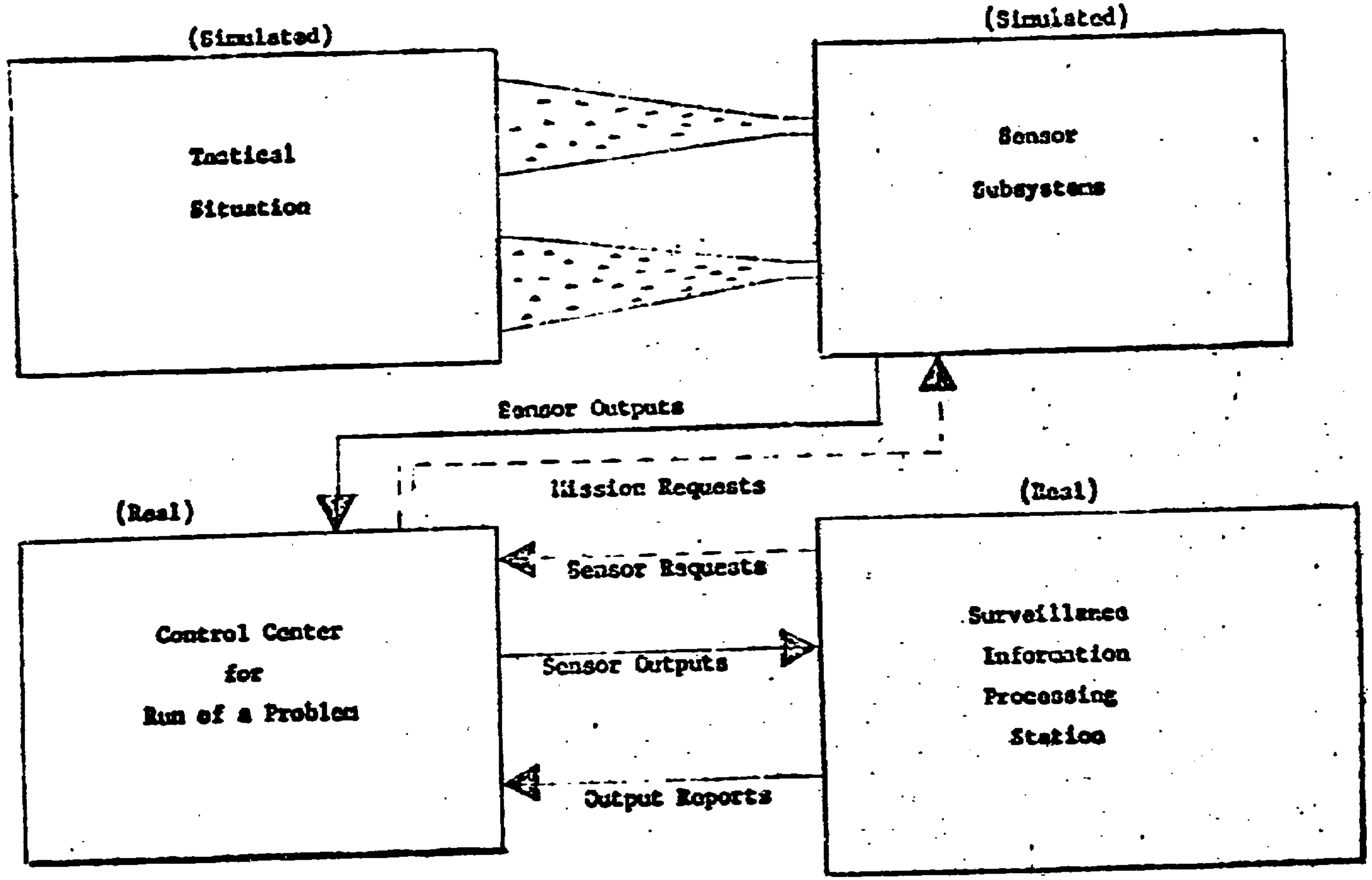


FIGURE 1. OVER-ALL INFORMATION FLOW

1. Line of sight considerations which are checked by the computer program.

2. The particular treatment combinations applied for the run (experimental unit), which introduce stochastic elements through the detection probabilities of the sensors simulated.

For a single experimental unit (a run of the station for one afternoon) the information received by the station is based upon these computer simulations and this information is used by the station personnel to track the Red Task Forces.

The output of the processing station consists of reports on the location of a Red Task Force. The chief of the station crew is designated as the Postulator. He is expected to give the locations of the task forces when requested to do so by the control section. These locations are specified in terms of map coordinates that delineate the perimeters of the terrain occupied by a task force. Such locations may be either an area or a route or a combination of areas and routes. The response for any one task force is limited to eight map points for each prediction. A scoring program using Monte Carlo techniques converts the station outputs into error distance scores.

In designing the first experiments, the Latin-Square configuration seemed most adaptable to the investigation of effects of interest. Performance or employment variations of the sensors have provided the treatment levels for the experiments. Since it was desirable to investigate several factors at a number of levels in one experiment, the 5 x 5 orthogonal square was chosen as the basic design. Rows and columns of a square have been designated as "repetitions" and "crews," respectively. That is, five crews have been used with each crew being presented the same problem five times but with a different set of treatment combinations imposed for each time. From some points of view it would be desirable to present a new tactical situation to each crew for each repetition or run of a problem.

In order to grasp more readily the layout of the experiments, Tables 1 and 2 are presented. For the designation of the treatment levels in the cells of the orthogonal square, the numbers 1, 2, 3, 4, and 5 are used. For example, crew 3 on its first repetition was presented with level 2 of Factor X, level 5 of Factor Y and Level 4 of Factor Z.

Within each cell of the orthogonal square, i.e., for a single run, usually eight reports are made by the crew on the location of task forces. These reports are spaced in time at 20-minute intervals during the development of the

Table 1

RANGE OF EXPERIMENTAL CONDITIONS AND LEVELS OF FACTORS CHOSEN FOR STUDY IN INFORMATION PROCESSING EXPERIMENTS

Symbol	Description of Experimental Conditions or Factors	Designation of Factor Levels				
		1	2	3	4	5
C	Crew	1	2	3	4	5
R	Repetition	1	2	3	4	5
	<u>Block I Factors</u>					
X	Factor X	0/hr	1/hr	2/hr	3/hr	4/hr
Y	Factor Y	0.3	0.4	0.5	0.6	0.7
Z	Factor Z	1.7/hr	2/hr	2.4/hr	3/hr	4/hr
	<u>Block II Factors</u>					
U	Factor U	45 mi	40 mi	35 mi	30 mi	25 mi
V	Factor V	25 min	20 min	15 min	10 min	5 min
W	Factor W	37 min	27 min	18 min	8 min	1 min
T	Situation Time	(1 through 8, or 8 reports at 20 minute intervals)				

Table 2

RANDOMIZATION OF THE DESIGN FOR BLOCK 1*

Crews

Repititions	1	2	3	4	5
1	3,3,2	1,4,3	2,5,4	4,1,5	5,2,1
2	2,4,5	4,5,1	5,1,1	2, 1,2	1,2,3
3	4,2,4	5,3,5	3,4,1	1,5,2	2,1,3
4	5,5,3	3,1,4	1,2,5	2,3,1	4,4,2
5	1, 1, 1	2,2,2	4,3,3	5,4,4	3,5,5

*For Block I the factors varied were X, Y, and Z (refer Table 1 above). Thus, the numbers 3,3, and 2, for Crew 1, Repetition 1 refers to levels 3,3, and 2, respectively for X, Y, and Z as described in Table 1. The factors U, V, and W were fixed at levels 1, 5, and 1, respectively, for each of the 25 cells of the Block I experiment (again refer to Table 1).

simulated tactical situation. Thus, the structure of an experiment might be described as orthogonal square with split-plot features provided by the time spacing and the targets. The term split-plot is used because of the similarity with the standard split-plot experiment whether the main plot structure be a randomized complete block or Latin Square. Within each main plot or cell of the orthogonal square, eight observations are obtained (one for each time) for each task force. But, of course, neither time nor targets can be randomized as required for a split-plot design.

In considering uni-variate analyses of variance for these experiments, these problems may be stated:

1. The rows of the square, designated as repetitions, do not conform to the usual pattern for rows and columns in a Latin Square. The rows within each column may be expected to have some unknown dependence or correlation.

This situation could be remedied if different tactical situations were presented to the crews. Admittedly more tactical situations could be developed for the one reservation being used, but this has not been done to date. We have even suggested using different terrain areas for each repetition, i.e., one situation might be at Ft. Bragg, another at Camp Polk, another at Camp McCoy, etc. Clearly, this would remove the memory element for the crew in remembering what happened to Task Force Alfa on the last run, and , thus, reduce the unknown correlations in each column. Such an experiment would seem to be somewhat unrealistic, however, in that a surveillance group would normally function within a limited terrain area for a period of time.

2. Degrees of freedom for assessing main plot treatments are too few. Should we combine four degrees of freedom error terms from successive experiments?

An alternative suggestion is to try to increase the error degrees of freedom within a single experiment. This increase may be accomplished by breaking out the individual degrees of freedom for the quantitative factors and using the higher order, cubic and quartic, effects to add to the four degrees of freedom for error.

First, Table 3 presents four degrees of freedom error terms from successive experiments.

Table 3

Error Mean Squares from Five Experiments
Listed by Target (Task Force)

Experiment Number	Mean Square for Error*					
	Target					
	0	1	2	3	4	5
1	2.97					
2	2.69					
3		1.5	0.82	7.2		
4		3.36	2.13	0.83	0.79	
5		2.1	4.8	2.7	0.62	1.1

These mean squares in Table 3 are obtained from an analysis of transformed data, i.e., natural logarithms of the original error distance scores expressed in meters. Selection of an appropriate transformation is a problem in itself which is not included in this paper. (2)

On the other hand, combining cubic and quartic effects with the error sum of squares has been carried out for some of the experiments and partial results are displayed in Table 4. From the available evidence both of the approaches suggested appear useful for increasing the sensitivity of the experiments.

3. The split-plot interpretation for time as a factor is not valid.

4. What interpretations can be made if an overall univariate analysis

*Variable analyzed is the natural logarithm of the observed error score.
Source: (1).

of variance is computed with both time and targets as apparent split-plot factors and there is interest in interactions with main plot treatments?

The questions (3) and (4) may be considered together. An example of an analysis of variance for one experiment appears in Table 5, below. The real problem is "What is the proper interpretation of that part of the analysis in Table 5 below the four degree of freedom error term?" The Model implied by the analysis seems inadequate for the experimental situation. (3) The particular example shown in Table 5 presents no problems; all the observed interaction mean squares are 'small' in relation to the residual mean square. The situation is quite different, however, for other experiments of the series.

5. The preceding questions raise the issue of alternative designs. Hence, what designs are practicable and desirable for these experiments?

Due to limitations on number of crews smaller squares, e.g., 4 x 4, and some Youden Squares have been used. Also, some non-orthogonal designs have been used since least squares analysis is easy with our computing facilities. (4) The latest design considered is an incomplete block design from the class of partially balanced designs with two associate classes. (5) Actually, factorial arrangements of the treatment combinations should be used so that most of the two-factor interactions could be measured. The 5 x 5 orthogonal square with treatments assigned in three languages is in fact a 1 in 125 fraction of the total design and does not permit assessment of any desired interactions. To date a feasible factorial arrangement has not been worked out. The limitation to three crews is severe. An examination of the National Bureau of Standards publication, AMS 48, for some of the smaller fractional designs indicates that four or eight crews might be used to block the experiment in an acceptable manner. (6) This blocking procedure, however, would affect the assessment of the "crew effect" since crews would be confounded with any other extraneous effects which the blocks are designed to remove in evaluating the treatments. It is believed that the resultant confounding would be no greater, perhaps, than the assignment of crews to columns of the Latin Squares. Some extraneous effects, e.g., such as a particular crew always working on the same day of the week, have been present in the already completed series of experiments. On the other hand, introduction of the trick of a pseudo-factor, i.e., dividing four crews into two groups of two crews each would permit direct introduction of crews as a factor in an experiment. (7)

Table 4

COMPARISON OF ERROR MEAN SQUARES WITH CUBIC AND QUARTIC COMPONENTS OF FACTOR EFFECTS FOR SELECTED EXPERIMENTS

Experiment Number	d.f.*	Target or Task Force 1	2	3
3				
Error M.S.	4	1.5	0.82	7.2
C & Q Components	8**	2.4	2.74	11.4
Combined	12	2.1	2.1	10.0
4				
Error M.S.	4	3.36	2.13	0.83
C & Q Components	8**	1.89	2.34	6.78
Combined	12	2.38	2.27	4.80
5				
Error M.S.	4	2.1	4.8	2.7
C & Q Components	8**	0.45	3.6	4.4
Combined	12	1.0	4.0	3.8

* D.f. = degrees of freedom

** The composition of these components is not the same for all three experiments. As an example, in Experiment Number 3, the components are obtained from:

1. Repetitions
2. Factor X
3. Factor Y
4. Factor Z

Source: (1)

Table 5

ANALYSIS OF VARIANCE OF SYSTEM PERFORMANCE BASED ON ORIGINAL SCORES, TRANSFORMED TO NATURAL LOGARITHMS USING THE SECOND MODEL FOR BLOCK I*

Sources of Variation	Degrees of Freedom	S.S.	M.S.
Total	199	153.39	-----
Crews	4	24.40	6.10
Repetitions	4	2.82	0.71
Factor X	4	3.51	0.88
Factor Y	4	9.01	2.25
Factor Z	4	4.38	1.10
Error	4	11.89	2.97
Time	7	19.26	2.75
Interactions			
XT	28	10.59	0.38
YT	28	16.18	0.58
ZT	28	6.90	0.25
CT	28	21.17	0.76
ET	28	8.47	0.30
Residual	28	15.30	0.55

* Refer Tables 1 and 2.

REFERENCES

(1) Brown, W. A., unpublished report, "Summary of Combat Surveillance Experiments", Institute of Science and Technology, The University of Michigan, 22 September 1960.

(2) Tukey, J. W., "On the Comparative Anatomy of Transformations", Annals of Mathematical Statistics, Vol. 28, (1957), p. 602.

(3)* Danford, M. B., Hughes, Harry M., and NcNee, R. C., "On the Analysis of Repeated Measurements Experiments", Biometrics, Vol. 16, (1960), p. 547.

(4) Brown, W. A., "An Analysis Technique for Evaluation of Combat-Surveillance-Game Experiments," Proceedings Third War Games Symposium, (36943-18-X). November, 1960, pp. 31-48.

(5) Bose, Clatworthy, and Shrikhande, "Tables of Partially Balanced Designs with Two Associate Classes," Technical Bulletin No. 107, North Carolina State Agricultural Experimental Station, (1954), (Reprinted by Institute of Statistics, University of North Carolina, Reprint Series No. 50).

(6) "Fractional Factorial Experiment Designs for Factors at Two Levels," U. S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series 48, April 15, 1957.

(7) Cochran, W. G. and Cox, G. M., Experimental Designs, 2nd Edition, New York: J. Wiley and Sons, 1957.

*Models of wider generality are described in this paper. Both univariate and multivariate analyses are outlined.

MULTIVARIATE ANALYSIS FOR PROJECT MICHIGAN EXPERIMENTS*

Emil H. Jebe
Operations Research Department
Institute of Science and Technology
The University of Michigan

In discussing this aspect of the IPT (Project MICHIGAN) experiments it is not necessary to repeat the general description given for the univariate analysis point of view. (1) A brief description of multivariate analysis may be useful in beginning this discussion. For example, in hybrid corn breeding work, the yield of corn per acre is usually the prime variable of interest. In some investigations, however, it is desirable to consider also the starch content, the oil content and the per cent protein of the yield. In an industrial context, one may conceive of bars of steel being made up with varying alloy contents and residual amounts of impurities. Then a metallurgist might measure the tensile strengths, hardness, and electrical conductivity of samples of the bars. The experimental unit in this steel example would appear to be a batch of bars and the sampling might be done so as to enable the study of variation between bars and within bars for the same batch. But the response variables are three--the average tensile strength, the average hardness, and the average conductivity--for each batch.

The aim of multivariate analysis of variance is to make a simultaneous analysis of the three responses for each batch of steel bars. Statistically, we become concerned with the analysis of a random vector rather than a single random variable, say yield, as is usual in the hybrid corn example described above.

The essential features of the extension to the multivariate situation are as follows:

We have the ith response in the trivariate case as a vector (Y_{1i}, Y_{2i}, Y_{3i}). The expectation of this vector is then (μ_1, μ_2, μ_3). With a sample of n such vectors we may form a sample matrix of sums of squares and sums of cross-products or of variances and co-variances. Thus,

*This work was conducted by the Information Processing Task (IPT) of Project MICHIGAN under Department of the Army Contract DA-36-039 SC-78801, administered by the U. S. Army Signal Corps.

$$S = \begin{bmatrix} y_1y_1 & y_1y_2 & y_1y_3 \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix}$$

where the summation over $i = 1, 2, \ldots, n$ is suppressed for each element in S. If we take $S/(n-1)$, we have

$$s = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix}$$

These matrices are, of course, symmetrical about the leading diagonal and in s, this same diagonal contains the sample variances for each of the responses. The off diagonal elements are the sample co-variances. In writing the expectation of the matrix s, a capital Σ is used and the s_{ij}'s are replaced by the σ_{ij}'s.

The description just given applies to simple random sampling from a homogenous universe. When the experimental and sampling procedures are more complex, the sums of squares and sums of cross-products may be subdivided in the usual manner by the analysis of variance.

Some further statistical features may be noted. In the expectation matrix, Σ, if $\sigma_{ij} = 0$ for $i \neq j$, then the elements of the observed vectors will usually be independent random variables. On the other hand, if $\sigma_{ij} \neq 0$, for some i and j, $i \neq j$, then the vector elements will be correlated. In the extreme case, $\sigma_{ij}/\sigma_i\sigma_j$ could equal +1 or -1. If this were true, the multivariate analysis would not add to the information. All the essential facts would be provided by a univariate analysis of variance for one of the elements of the response vector. The cases of interest then are zero correlation or moderate correlation.

From the information or signal point of view, we may say that in the trivariate case, each experimental unit gives us three signals. These signals seldom are independent. Multivariate analysis seeks to extract more information from the combination of signals than might be obtained from considering any one of the signals.

Details about computational procedures for carrying out a multivariate analysis of a response vector are omitted from this brief description. Several sources describe the computations for various situations and purposes (2,3, 4,7 and 8). An interesting example is treated in some detail by Smith and Gnanadesikan. (5) Other examples are described in varying degrees of detail in the references noted.

Now, it may be asked, "In what way may we apply the multivariate analysis concepts to these IPT experiments?" First, even for experiments 1 and 2, in which the tactical situation displayed only one moving Task Force, we have a vector of observations for each experimental unit. To repeat, one experimental unit was a single run of the station for one afternoon with a given crew and a particular repetition. The eight reports on Task Force ALFA at 20-minute intervals form the vector of observations.

I have tried to look at this Time aspect of the experiments in various ways. As described in (1) the structure of the experiment is an orthogonal square with an apparent split-plot feature provided by these observations spaced in Time within each cell of the square. Since Time is not, and cannot be in any sense, randomized as a factor or treatment within the cells the split-plot approach is not valid even though all calculations are carried out as for a split-plot experiment.

One type of multiple response view of these experiments is to consider an analogy with certain agronomic experiments. (6) Examples are perennial crops such as alfalfa and asparagus with several cuttings each season and harvest over several seasons before a field is replanted. In such experiments, all the yields over time may be added together for each experimental unit and these unit totals analyzed. Interest in these experiments also centers on the distribution of yields over time (just as there is interest in the fluctuation of the error distance scores over time). Therefore, complete partition of the total variation among the individual yields is undertaken to understand the experiments. Interpretation is, however, complicated by the correlations in time of the observed yields. The same problem exists in the information processing experiments. The eight reports over time for the same crew and repetition are obviously related in some unknown manner. Adequate replication solves part of the problem in some agronomic experiments but it appears that multivariate analysis techniques may be helpful.

Beginning with the third experiment the information processing experiments exhibit an added feature. Multiple targets were introduced, (i.e., the crews were asked to make reports on the locations of three or more Task Forces.) Thus, even for a single time, say 1500, a vector of responses is obtained. For these experiments it appears that multivariate analysis might be applied in two ways:

1. By summing or averaging over targets and using the time space as the vector of responses, or alternatively,

2. Averaging over time and using the error scores for the several targets as the vector of responses.

In full generality, it appears that each experimental unit for experiments 3, 4 and 5 provides a matrix of responses. This matrix which is R by C has one row for each target and one column for each time at which reports are given on target positions. To date I am not aware of any existing methods for dealing with a matrix of responses for each experimental unit. It has been pointed out that the data may be viewed as a vector of RC dimensions for each experimental unit.

It is clear from the description given that several approaches may be used for analyzing the data from the information processing experiments even though no methods are available currently for dealing with the matrix of responses. Restated these approaches are:

1. Univariate analyses of variance

   a. A separate analysis for each element of the matrix of responses. A total of RC analyses would be obtained.

   b. Analysis by summing the columns of the matrix and considering Time as a factor in the analysis,

   c. Analysis by summing the rows of the matrix and considering Targets as a factor in the analysis,

   d. A combination of (b) and (c) just mentioned with both Time and Targets considered as factors in the analysis.

2. Multivariate analyses of variance

a. Separate analysis for each row of the matrix (i.e., for each target) using the data from the Time space as the vector of responses,

b. Separate analysis for each column of the matrix (i.e., for each time) using the data from the multiple targets as the vector of responses,

c. Two analyses based on 1.(b) and 1.(c), above, where the response vectors are in the Time space and in the Target space, respectively.

Now, it will be useful to consider some aspects of the computations in making a multivariate analysis for one of the experiments, say experiment four with four targets and eight times. Among the references cited, (3) was found to be the most helpful in describing the procedures. Specifically, Chapter 7, Section 7d gives the details for the multivariate analysis of dispersion. The distribution theory for the test criterion (Likelihood Ratio) is complex but Chi Square and Variance Ratio approximations are available. The appropriate sample statistic is $V = -m \log \lambda$ where λ is the ratio of two determinants. The statistic V has an approximate Chi Square distribution with pq degrees of freedom. For p we may take the value eight or four depending on whether we choose the Time space or Target space vector. For q we have the value 4, the number of the degrees of freedom for the main effect to be tested. Thus, pq = 8(4) or 4(4). It would seem that a Chi Square with either 32 or 16 degrees of freedom would provide a fairly sensitive test. There is a catch, however. In the formula given for V there appears the factor m. This $m = n - \frac{p+q+1}{2}$ where n is the sum of the degrees of freedom for treatments plus error. In experiment four, the n value is 8 = 4 + 4, so

$$m = 8 - \frac{8+4+1}{2} = 1.5 \text{ or}$$

$$m = 8 - \frac{4+4+1}{2} = 3.5.$$

Thus, sensitivity of the multivariate test is measured not only by pq, the degrees of freedom for V, but also by the vector m which has implicitly embedded in it the usual degrees of freedom for error. Since $0 < \lambda < 1$, we see that a larger m value helps to obtain a significant Chi Square value.

Alternatively, we might use the Variance Ratio approximation instead of the statistic V. The F obtained has degrees of freedom pq and ms + 2 λ where pq is as already given. For ms + 2λ one obtains about -1.95 for the Time response vector and about +7.21 for the Target response vector. It is to be noted that ms + 2λ need not be integral for defining the degrees of freedom of the variance ratio. Since F is not defined for negative degrees of freedom the Time response vector cannot be considered. The Target response vector might be considered for an $F(16,7.21)$.

In summary, a multivariate analysis of variance may be computed for the information processing experiments using either the Time or Target response vectors. The V criterion is a little more direct to obtain in that slightly less computing is required. For both statistics, F or V, the main problem is again one of inadequate degrees of freedom for error. Either the orthogonal squares used should be replicated or a more sensitive design should be adopted.*

*The problem of other designs is discussed further in (1).

REFERENCES

(1) Jebe, E. H. and Brown, W. A., "Problems in the Analysis and Interpretation of Information Processing Experiments", Institute of Science and Technology, The University of Michigan*

(2) Anderson, T. W., An Introduction to Multivariate Statistical Analysis, New York: J. Wiley and Sons, 1958.

(3) Rao, C.R., Advanced Statistical Methods in Biometric Research, New York: J. Wiley and Sons, 1952.

(4) Geisser, S., "A Method for Testing Treatment Effects in the Presence of Learning", Biometrics, Vol. 15, (1959), p.389.

(5) Smith, H. and Gnanadesikan, R., "The Simultaneous Analyses of Multi-Response Experiments", Gordon Research Conference on Statistics in Chemistry, 1958.

(6) Steel, R.G.D., "An Analysis of Perennial Crop Data", Biometrics, Vol. 11, (1955), p.201.

(7) Tukey, J.W., "Dyadic Anova, An Analysis of Variance for Vectors", Human Biology, Vol. 21, (1949), p. 65.

(8) Tukey, J.W., "Components in Regression", Biometrics, Vol. 7, (1951), p. 33.

(9) Votaw, D.F., et al, "Compound Symmetry Tests in the Multi-Variate Analysis of Medical Experiments", Biometrics, Vol. 6, (1950), p.6.

(10) Danford, M.B., Hughes, Harry M., and McNee, R.C., "On the Analysis of Repeated Measurements Experiment", Biometrics, Vol. 16, (1960), p. 547.

---

*Refer to 6th Design of Experiments Conference Proceedings page 91.

COMPUTATION OF EXPECTED RESOLUTION IMPROVEMENT FACTOR OF AN INVERSE FILTER SYSTEM

Chandler Stewart
Mine Detection Branch
Engineering Research and Development Laboratories

Recent research, such as that of Bracewell in radio astronomy, and Marechal in photography, has demonstrated the principal of improving the time or position resolution of a detection system, by suitable processing of the detector output. The U. S. Army Engineer Research and Development Laboratories, Fort Belvoir, Virginia, is attempting with the assistance of Drexel Institute of Technology to predict the expected advantages and requirements of the resolution improvement principle under specific conditions, such as in land mine detection systems. These predictions are needed for guidance of experimental research or, in the case of a negative result, for saving the cost of an experimental program.

The objectives of the resolution improvement filter are shown qualitatively in the first slide. (Slides are placed at the end of this article.) The true intensity distribution of the detected property is represented by the left hand view, and the output of a detector of poor resolution is shown by the center view. By passing the detector output through a filter whose transmission spectrum is the reciprocal of that of the detector, one can expect an improvement in resolution, as indicated by the right hand view.

Of course, we expect to pay for this improvement by a reduction in signal-to-noise ratio and a loss of response accuracy. Our first computation objective is to obtain curves of resolution improvement factor versus noise cost. Slide 2 gives qualitative definitions of some of the terms we use in the one dimensional analysis. For example, the detector signal from scanning an infinitesimal particle is given in (5), and the corresponding narrower filter output pulse is shown at (9). The ratio of these is the resolution improvement factor, (I), of the filter, shown under (17), at the lower left hand corner. The right hand column lists the spectral form of each space function. For example, the upper frequency limit of the system is designated as K_c, in (10).

Using these concepts, (Slide 3) we obtained a curve of resolution improvement factor I versus noise cost for a one dimensional system by obtaining each of these variables as a function of the frequency limit K_c.

We did this by first obtaining the particle response function of the detector in the space domain. The curve is shown in Slide 4 on a semi-logarithmic chart.

Then we obtained the Fourier transform over a limited frequency range. The negative portion of the curve was inverted to permit display on a semilogarithmic chart. (Slide 5) The next step was to obtain the filter spectrum which is theoretically the reciprocal of this curve. Such a filter would have infinite gain at these zero crossings; therefore, to obtain a finite solution, we omitted an arbitrary region around these infinite poles.

Then, by the formulas given in the third slide, we calculated on an IBM 650 Computer the resolution improvement factor versus noise cost. (Slide 6) Since this result applies only to a hypothetical one dimensional system, it has very little value for guidance of experimental research. However, the mathematical steps employed may give some insight into the requirements for obtaining the corresponding expected performance of a real two dimensional system.

Comparable calculations were attempted for the two dimensional filter, and the spectral values were obtained for 121 points. (Slide 7) However, it was noted that, because of the high magnification of errors in taking the reciprocals of low spectral values, the integrated noise results were extremely dependent upon arbitrary choices, such as the relation of chosen values of the independent variables to the poles, and the width of the excluded polar regions. For this reason, these results are considered unreliable, and will serve only as springboard for further research, and as an interim guide pending more accurate results. It is interesting to note that the area resolution improvement factor seems to follow the square of the one dimensional resolution improvement factor. This is approximately what one would expect.

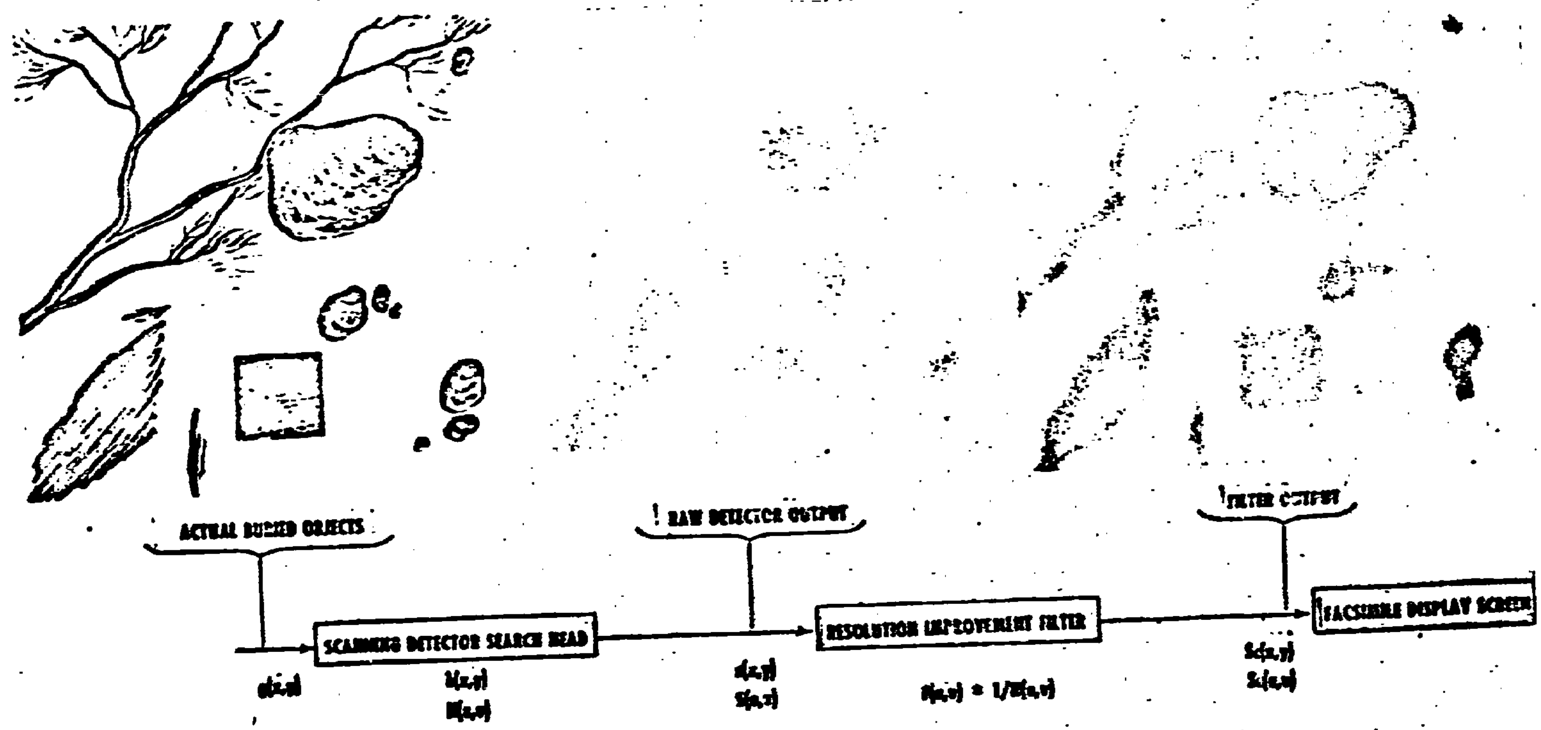
Drexel Institute has been searching unsuccessfully for computational short cuts on this problem, and has just recently turned its attention to developing a convoluation integral procedure which would eliminate the need for Fourier transform calculations. This study is still in progress.

Because of our failure, after a year of trying, to obtain adequate filter performance predictions, it seems wise to proceed with preliminary experimental research without benefit of the hoped-for theoretical guidance.

A search for techniques for carrying out the two dimensional inverse filter process physically has yielded only two proposals. One is the optical analog system shown, in which the input is used to modulate a light beam. (Slide 8) This light is analyzed, by means of lenses, into the spectral distribution of the input, at which point a spectral filter performs the necessary filter function by its light transmission properties. Another lens reconstitutes the image back into the space domain. One limitation of this system is the need for maintaining coherency of the light throughout the process, and the consequent requirement for maintaining optical dimensional tolerances on the light transmission components.

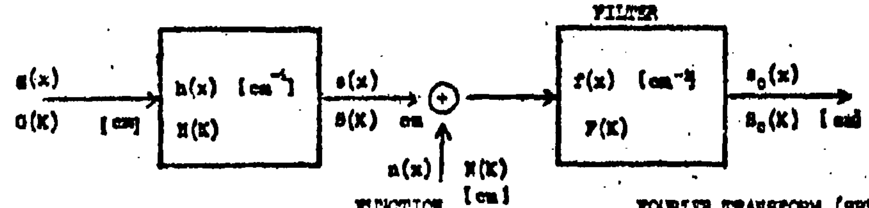
This limitation is avoided in the second proposal, in which the light modulation is maintained in the space domain throughout, and the input is cross correlated with the filter function, also distributed in the space domain. However, in both systems the filter function is bi-polar. We haven't yet found a practical way to accommodate the negative regions without sacrificing accuracy.

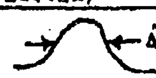
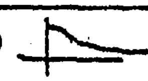
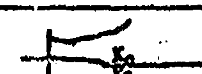
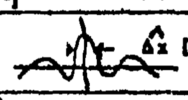
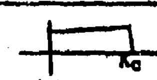
We will appreciate any suggestions which may help us to complete the two dimensional performance prediction computations, or which may lead us to the best physical design of filter. References to other groups active in this area would be especially valuable.



DETECTOR SIGNAL PROCESSING TO IMPROVE RESOLUTION

PROPOSED COMPUTATION OF RESOLUTION IMPROVEMENT FACTOR OF AN INVERSE FILTER APPLIED TO A ONE-DIMENSIONAL SCANNING DETECTOR



	FUNCTION	FOURIER TRANSFORM (SPECTRUM)
DETECTOR RESPONSE	(3) $h(x)$	(4) $H(K)$
TARGET PARTICLE PROPERTY	(1) $g(x)$  Δx [cm] $g(x) = h_o\mu(x)$ (Assume $\Delta x << \tilde{\Delta x}$)	(2) $H(K) = \frac{S(K)}{G(K)}$ h_o calibrating constant $\mu(x)$ particle property
SIGNAL AT DETECTOR OUTPUT	(5) $s(x)$  $\tilde{\Delta x}$ [cm]	(6) $S(K)$ [cm]  $K\to$
IN CASE OF NOISE, SIGNAL AT FILTER INPUT IS $s(x) + n(x)$	ASSUME NO NOISE SO THAT SIGNAL AT FILTER INPUT IS $s(x)$	
INVERSE FILTER RESPONSE	(7) $f(x)$ [cm⁻¹] $F(K) = \frac{S_c(K)}{S(K)}$	(8) $F(K)$  K_c $K\to$
OVERALL FILTERED RESPONSE OF DETECTOR TO PARTICLE	(9) $s_c(x)$  $\hat{\Delta x}$ [cm]	(10) $S_c(K)$ [cm]  K_c $K\to$

(11) IN (4) FOR $(K\Delta x)$ SMALL ENOUGH, $G(K)$ IS APPROX CONSTANT

$\therefore G(K) = k_1$

(12) IN (1) $S_c(K) = \begin{cases} k \text{ cm} & |K| \le K_c \\ 0 & |K| > K_c \end{cases}$

IN (8) $F(K) = \frac{S_c(K)}{S(K)} = \frac{k}{S(K)} \quad |K| \le K_c$

(13) NO NOISE

$S_c(K) = F(K)G(K)H(K)$

(14) MEAN SQUARE NOISE

$$E_N^2 = B\int_0^{K_c} |F(K)|^2 \, dK$$

$B = P_n R$ P_n = NOISE POWER SPECTRAL DENSITY

R = RESISTANCE

(15) MEAN SQUARE ERROR DUE TO NOISE TO FINITE OVERALL RESOLUTION

$$I = \frac{\frac{1}{K_y}\int_{-\infty}^{\infty} |s_c^2(x)| \, dx + E_N^2}{s_c^2(0)}$$

$[K_y] = \text{cm}$

(16) FILTER CUT-OFF FREQUENCY

$W_c\left[\frac{\text{radians}}{\text{sec}}\right]$ computed by minimizing I for each value of B

Solve $\frac{dI}{dK_c} = 0$ for K_c

Then $W_c = K_c V$

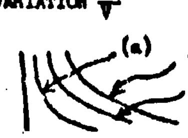
(17) RESOLUTION IMPROVEMENT FACTOR (I) VS

(a) NOISE

(b) STATIC SPEED ERROR $\frac{V_1 - V}{V}$

(c) RANDOM SPEED VARIATION $\frac{\Delta V}{V}$

$I = \frac{\tilde{\Delta x}}{\hat{\Delta x}}$



(18) $W = KV$ V = VALUE OF VELOCITY OF SCAN FOR WHICH FILTER DESIGNED

V_1 = VALUE OF VELOCITY OF ACTUAL SCAN

Δx, $\tilde{\Delta x}$, $\hat{\Delta x}$ computed as follows

Ex: $\tilde{\Delta x} = 2|x^*|$ where

$s(x^*) = \frac{s(0)}{2}$

COMPUTATION OBJECTIVES

A computation of one and two-dimensional Resolution Improvement Factor (I) versus Noise Cost, is required.

For one dimensional processing,

$$I = K_c \tilde{\Delta} X / 3.791,$$

where K_c - filter cut-off frequency, and

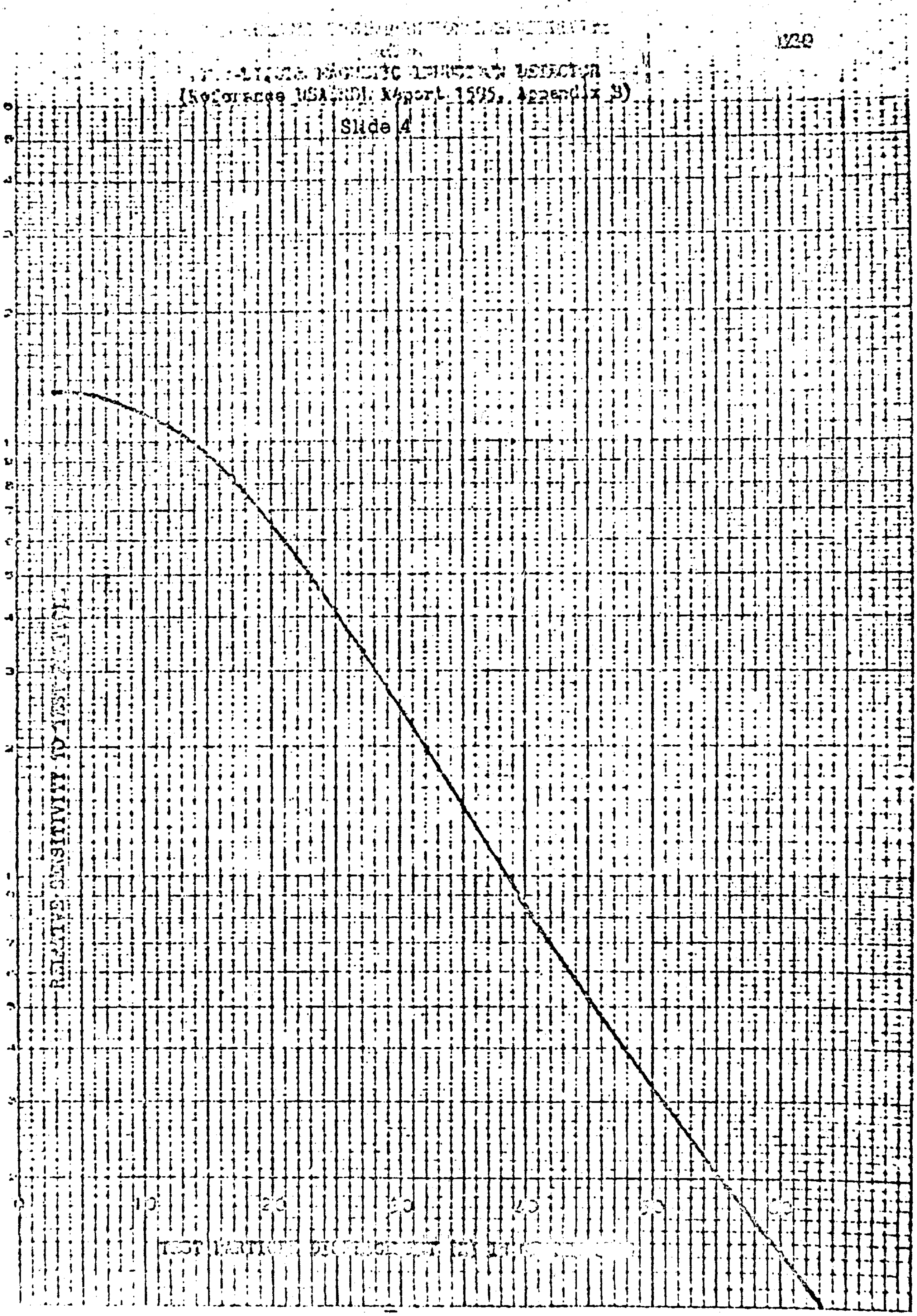
$\tilde{\Delta} X$ = width of detector response function.

$$\text{Noise Cost} = 10 \log_{10} \frac{K_{c1}^2 \int_0^{K_c} |F(K)|^2 dK}{K_c^2 \int_0^{K_{c1}} |F(K)|^2 dK} \quad \text{db}$$

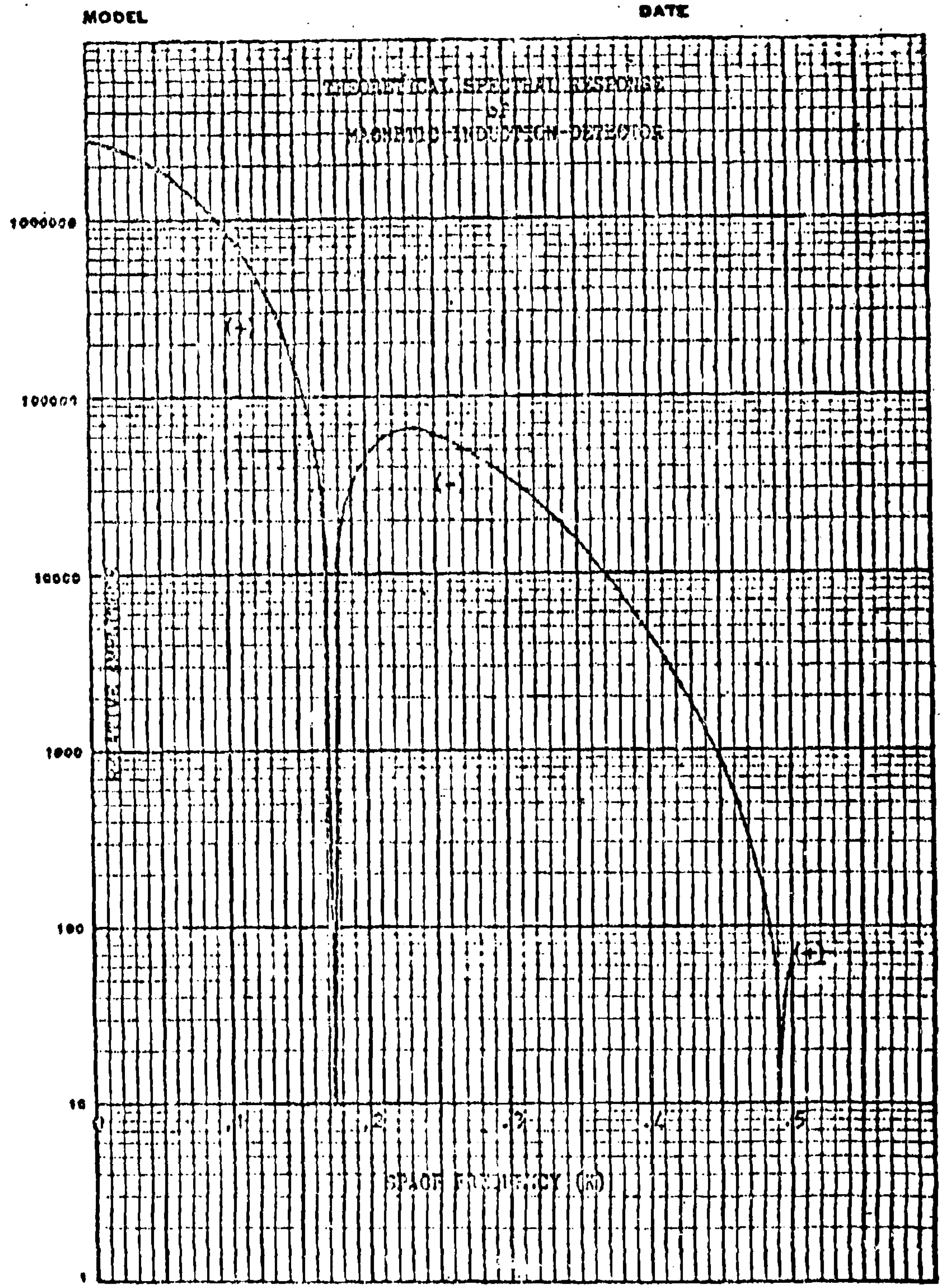
where K_{c1} = value of K_c for $I = 1$, and

$F(K)$ = spectral (Fourier) response of inverse filter.

For two-dimensional processing, similar formulas can be developed by replacing K with the wave number variables u and v.


Slide 4
0
10
20
30
40
50
60

Slide 5


MODEL
DATE
THEORETICAL SPECTRAL RESPONSE
OF
MAGNETIC INDUCTION DETECTOR
1000000
100000
10000
1000
100
10
1
(+)
(-)
(+)
0
.1
.2
.3
.4
.5
SPACE FREQUENCY (K)

$E_m = [B\int_0^{K_c} |F(K)|^2 dK]^{1/2} = 8_c(0)$ (TERMS DEFINED IN USAERDL REPORT 1595 RR, APPENDIX B, EXCEPT THAT $Z_T = Z_R = 20$ CM & $Z_T' = Z_R' = 50$ CM.)

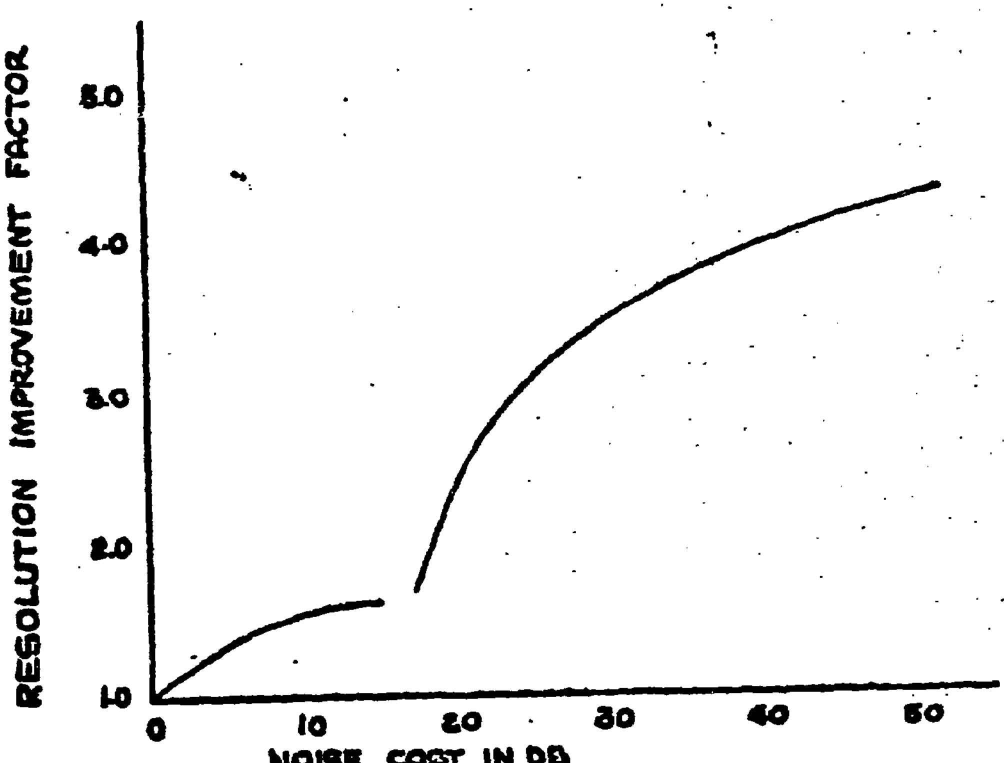
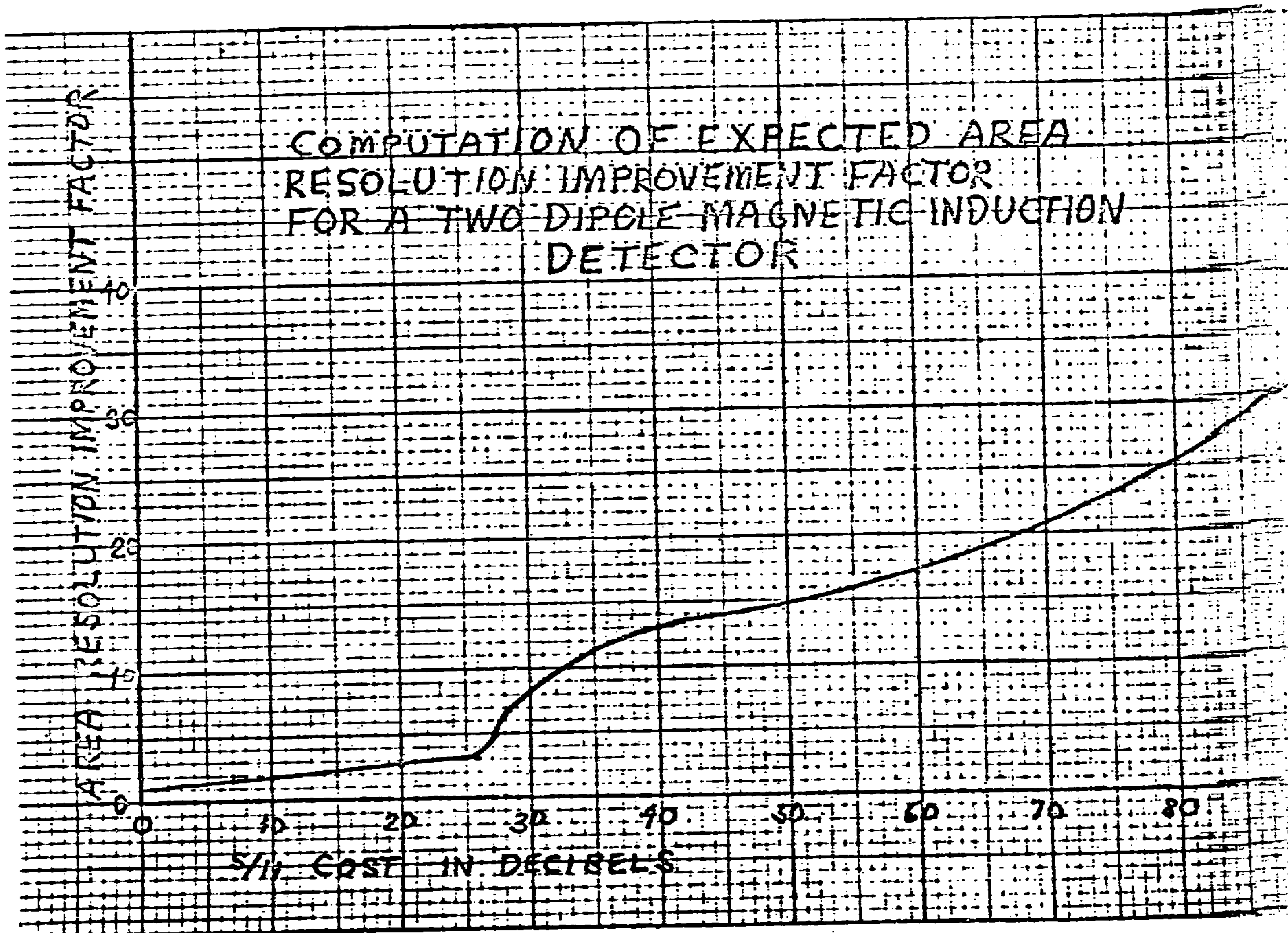
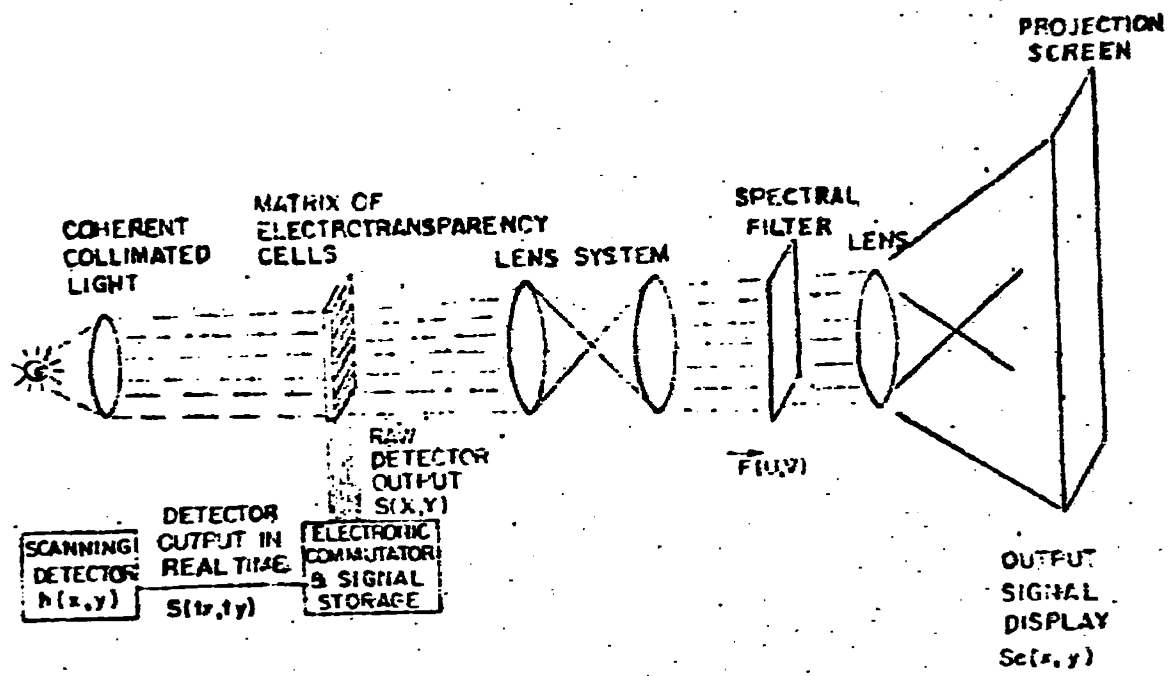


FIGURE 10.1.16 COMPUTATION OF RESOLUTION IMPROVEMENT FACTOR OF ONE DIMENSIONAL MAGNETIC INDUCTION SYSTEM.

C 13200 E 615C


COMPUTATION OF EXPECTED AREA
RESOLUTION IMPROVEMENT FACTOR
FOR A TWO DIPOLE MAGNETIC INDUCTION
DETECTOR
AREA RESOLUTION IMPROVEMENT FACTOR
0
10
20
30
40
50
60
70
80
S/N COST IN DECIBELS



OPTICAL ANALOG RESOLUTION IMPROVEMENT FILTER SYSTEM

(REFERENCE: UNIVERSITY OF MICHIGAN WILLOW RUN LABS REPORT 2900-123T FIGURE 2)

G13200E 7064

PANORAMIC VIEWING UTILIZING HYPERBOLIC ELLIPSOIDAL REFLECTING OPTICS

Donald W. Rees
U. S. Army Ordnance Tank-Automotive Command
1501 Beard
Detroit 9, Michigan

The increased need for ballistic and radiological protection in Ordnance vehicles, together with the need for increased surveillance of the area surrounding the vehicle, has dictated the development of a viewing system capable of covering a large field of view from a small aperture. In the recent past many attempts have been made at increased field viewing. In the motion picture industry (in particular) the pursuit of wider angle presentations has led to the development of many commercial optical systems. The following table lists several of these systems, together with their horizontal coverage and aspect ratio.

Name	Horizontal Coverage	Aspect Ratio
Cinerama	146°	2.06 to 1
Cinemascope	114°	2.55 to 1
Cinema 160	160°	2.26 to 1
Todd AO	128°	2.00 to 1
Circarama	360°	5.14 to 1

The optics used in these systems are of a refractive nature. In general, refractive optical arrangements are more desirable for use in imaging because of their compactness and physical sturdiness. However, they appear to be less desirable than reflective optics for wide angle imaging because of their inability to capture extremely wide angle imaging without the use of several systems operating in tandem. Refractive systems also have rather low optical efficiencies compared to those employing reflective optics. Viewing systems utilizing pure reflecting optics and reflecting optics in combination with refracting optics have achieved fields of view up to 360°. Most of these optical arrangements have been developed to imitate visual movement in connection with various types of ride and flight simulators. The University of California, Cornell Aeronautical Laboratory, Douglas Aircraft, and Curtis Wright are presently engaged in the development of simulators utilizing wide angle visual presentation.

The viewing system currently under development by U. S. Army Ordnance Tank-Automotive Command utilizes a convex hyperbolic mirror as an image collector and a concave ellipsoidal surface as a viewer.

Figure 1 illustrates a proposed application of the system in a closed pod vehicle.

The hyperbolic image collector is mounted on the vehicle in such a manner as to give an unobstructed view of the surrounding area. The vertical image of the mirror is picked up by a television camera using a wide angle lens and conveyed to a closed circuit television projection system. The image is projected into the elliptical screen from the outer focus of the ellipse. The scene is then viewed from the inner focus of the ellipse. To date, a television link has not been integrated into the arrangement. A sixteen millimeter motion picture camera utilizing both color and black and white film is being used to determine such parameters as lens focal lengths and optimum shapes for image collectors and viewers.

An illustration of the typical image collector is shown in Figure 2. Due to the geometric configuration of the real object and the virtual image, the center of focus of the pickup lens should be at the outer focus of the hyperbola. Location in any other position will tend to create distortion.

Figures 3, 4, and 5 illustrate three possible methods of image display. Projection directly into a diffuse elliptical screen (Figure 3) is the simplest of the three methods.

The image projector is located on a line between the inner and outer foci. The distance from the screen to the image projector is determined by the focal length of the lens. Thus, the shorter the focal length of the lens used, the closer the projector may be placed to the elliptical screen. The screen is then viewed from the proximity of the inner focus. The focal spot is not critical in this case since the system is diffuse. The viewer need only limit himself to a spherical area approximately 18 inches in diameter surrounding the inner focus. Despite its simplicity, this method has one serious drawback. When a scene is viewed in the lower area of the ellipse, the distance between the image and the viewer's eyes is quite small. This makes eye focus and convergence rather difficult and tends to cause eye strain.

If the diffuse elliptical screen is now replaced by a specularly reflecting ellipsoid, the foci of the system becomes much more critical. The image projector must be located exactly at the far focus of the screen. The projection lens must then have an exact focal length determined by the image required. The inner focus of the ellipsoid is, in this case, a very sharp focal point. Since the focal spot is small, viewing this system necessitates using only one eye at a time. This condition seriously restricts this system's use as a viewing device. Any movement of the viewer's eye from the focal spot would tend to introduce extreme distortion. Figure 4 illustrates the optical geometry involved in the specular ellipsoid. From this figure it may be observed that the image plane takes on a spherical configuration with the center at the inner focus of the ellipse. The spherical radius is equal to the optical distance from the image projector to the viewing focal spot.

If a diffuse screen is now inserted into the ellipse as shown in Figure 5, a combination of several of the characteristics of each of the two previous systems results. The diffuse screen is a spherical section with a radius of curvature equal to the distance between the outer focus of the ellipse and the intersection point of the minor axis and the elliptical surface.

The image projector location in this case is dependent upon the focal length of the lens as in the case of the diffuse ellipse Figure 3. The inner focus is again enlarged to a spherical configuration of about 18 inches in diameter. The main advantage of this viewer over the diffuse elliptical type lies in the position of the image plane. As may be seen in Figure 5, the image plane takes on a spherical configuration similar to that in Figure 4. The radius of the spherical image plane is, in this case, equal to the length of the optical path from the inner focus to the diffuse screen. This radius is somewhat smaller than the radius of the image plane in the purely reflective system. It is, however, large enough to eliminate the eye focus and convergence problem encountered in the use of the diffuse ellipsoid. It is felt that, of the three viewing methods previously mentioned, the method involving the use of a diffuse screen and specularly reflecting ellipsoid is more readily adaptable for use in the system.

If the diffuse screen is removed and replaced with a television monitor tube having a face with a similar radius of curvature, a geometric configuration of optical paths equal to those shown in Figure 5 will result. The monitor tube arrangement, shown in Figure 6, is considerably smaller and less cumbersome than any of the previous systems. The size and position of the scan lines however, may cause some loss in resolution.

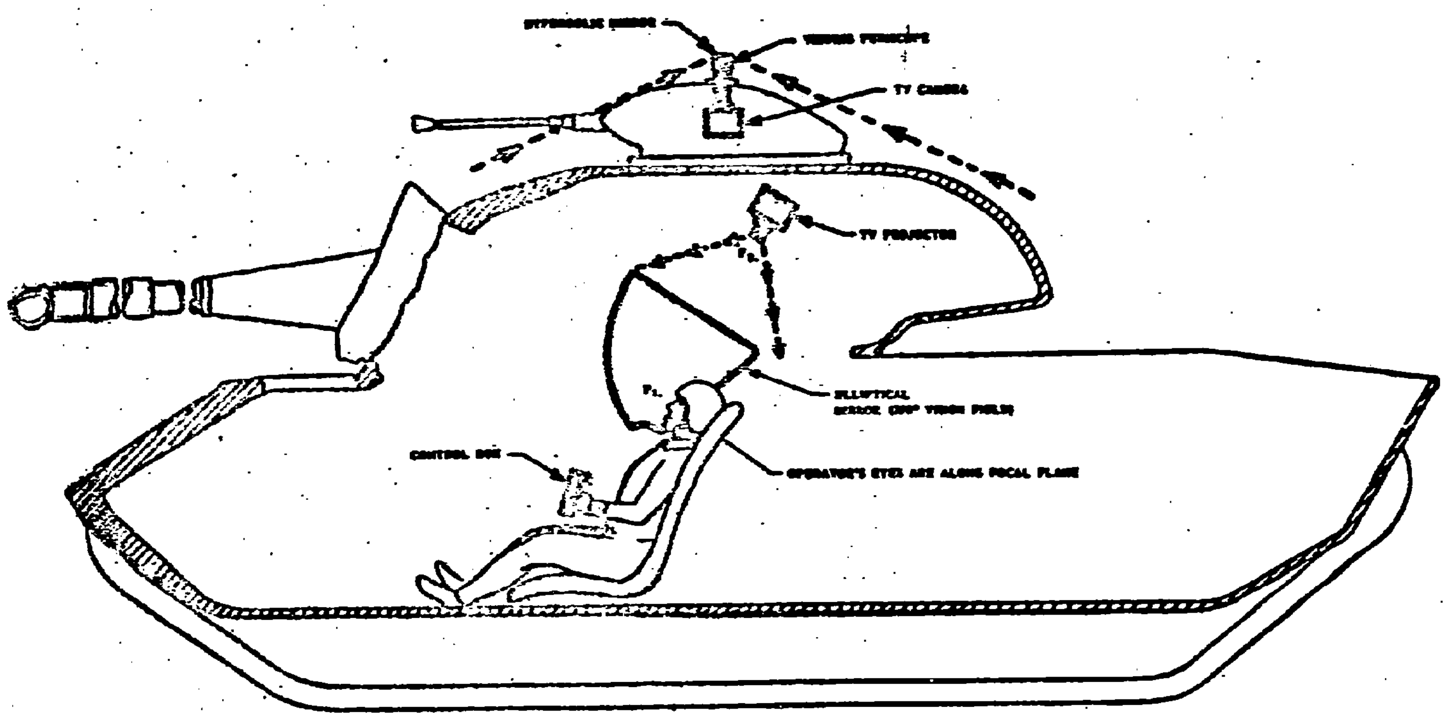
Motion pictures using the hyperbolic pickup were taken from both a stationary tripod, as shown in Figure 7, and a moving vehicle, as shown in Figure 8. The location of the pickup on a vehicle creates a problem which may cause driver discomfort. If the image former were located in the driver's compartment of the vehicle, as illustrated in Figure 8, the location of the pickup and the image former would be quite different with respect to the center of gravity of the vehicle. Any roll, pitch, or yaw encountered by the vehicle would have a magnitude at the collector different from that at the image former. The vehicle operator, sitting in the imager, will feel one magnitude of motion and see another. This sensation may cause motion sickness in some extreme cases.

The experimental set up of the diffuse ellipsoid is shown in Figure 9. It was constructed by molding glass fiber mat over a male elliptical form. The projector shown is a 16mm Kodak Analyst with a Weinberg Watson Modification which enables the film to be single framed for closer study.

Several of the basic problems involved in the development of a panoramic viewer may be states as: (a) the determination of the optimum shape and size of the hyperbola and ellipse; (b) the selection of projection and collection lenses of the proper focal length; (c) the determination of the possible problems caused by the difference in position of driver and image collector; (d) the evaluation of each of the three methods of image forming along with some of their modifications in order to determine the one best suited for this system.

It is felt that most of the problems in the system are caused by the lack of availability of adequate hardware. In the future, the construction of a larger hyperbolic pickup and a specular ellipse are planned. The purchase of a closed circuit television system is also planned. Existing motion picture equipment will be used however, until enough parameters are established to accurately define the characteristics of the television system needed.

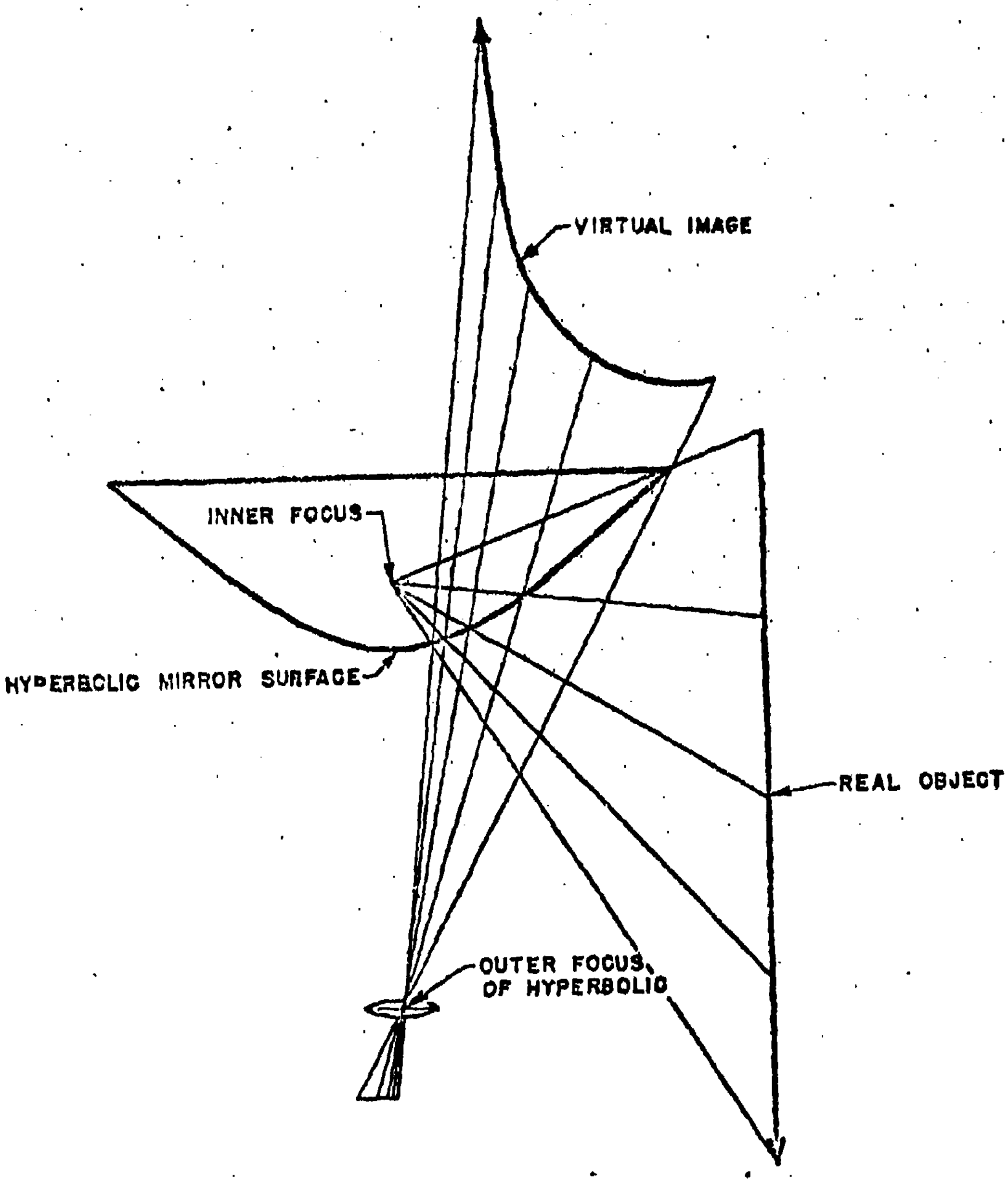
PROPOSED SYSTEM UTILIZING 360° VIEWING SYSTEM



—DETROIT ARSENAL—

FIGURE 1

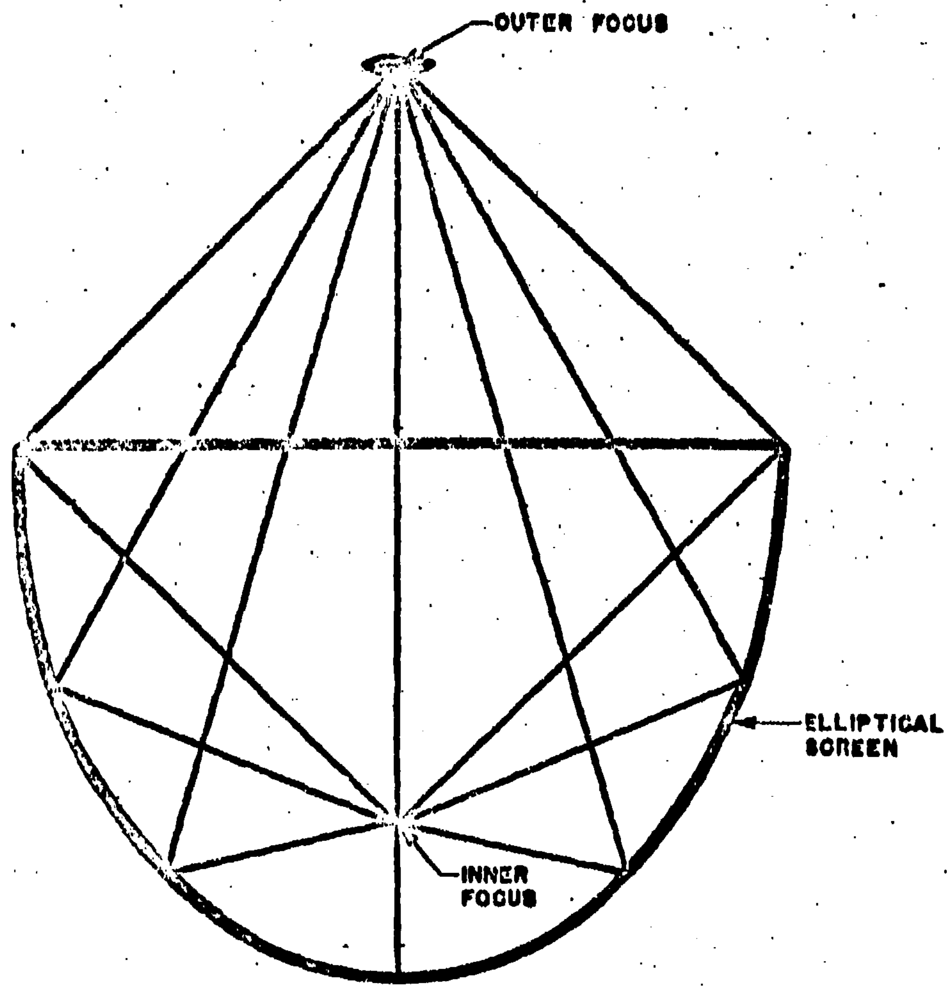
REFLECTING HYPERBOLA



—DETROIT ARSENAL—

FIGURE 2

DIFFUSE ELLIPTICAL SCREEN



—DETROIT ARSENAL—

FIGURE 3

SPECULARLY REFLECTING ELLIPSOID PROJECTOR MUST BE AT FAR FOCUS OF SCREEN

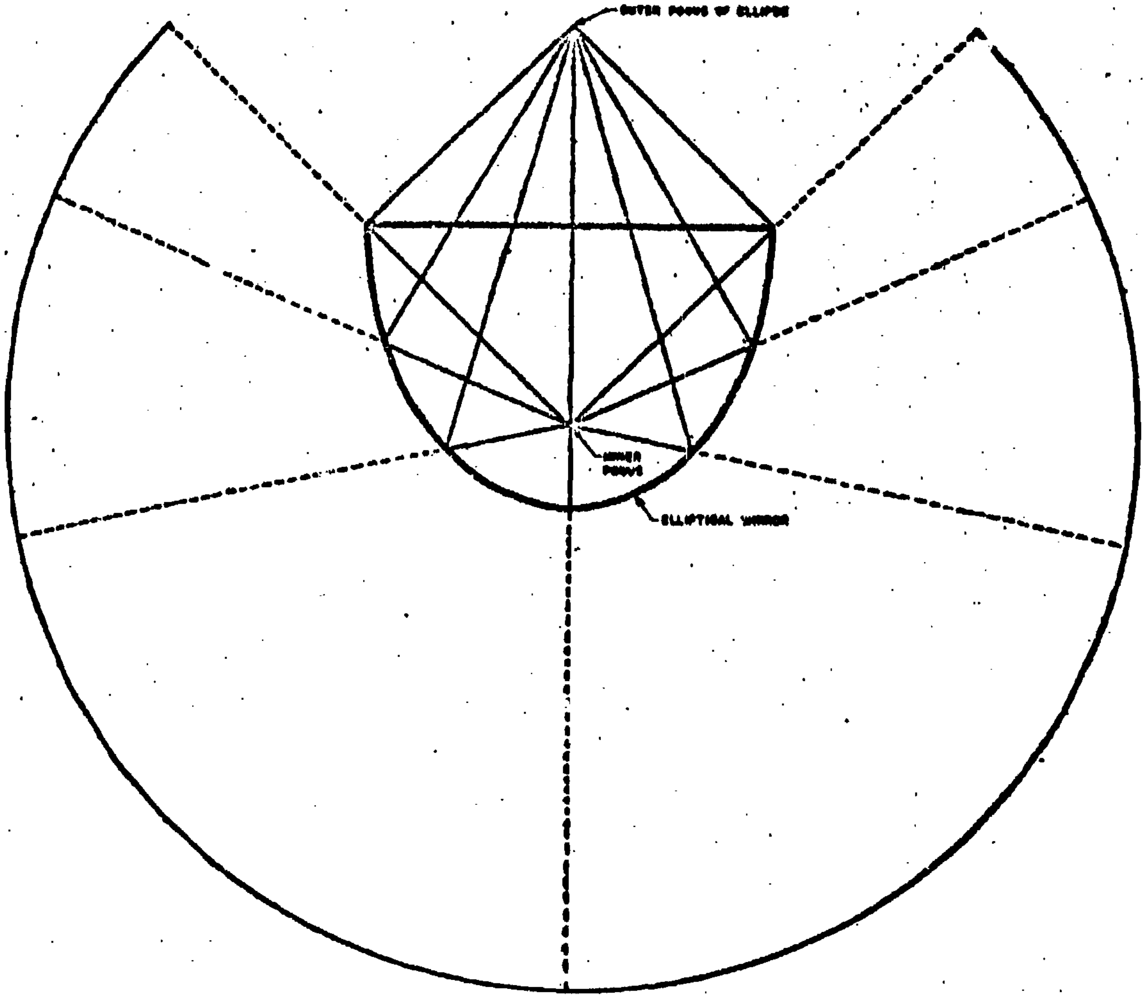
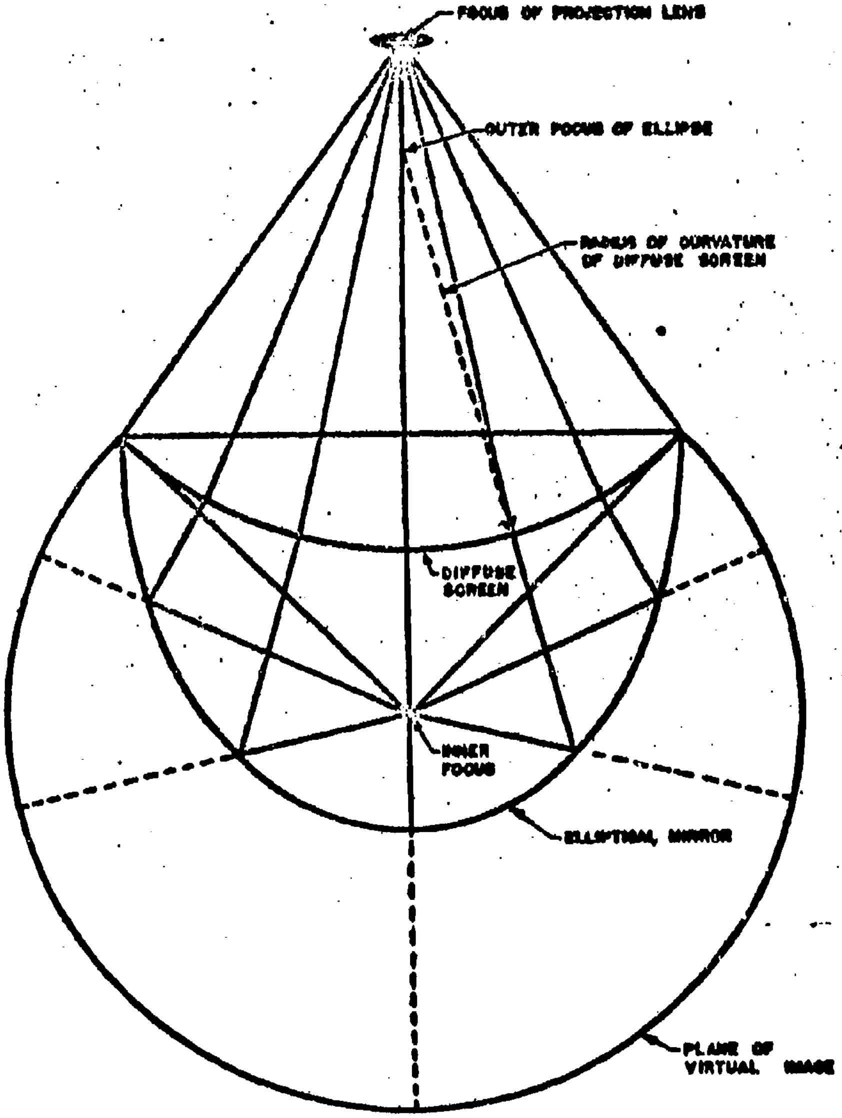


FIGURE 4

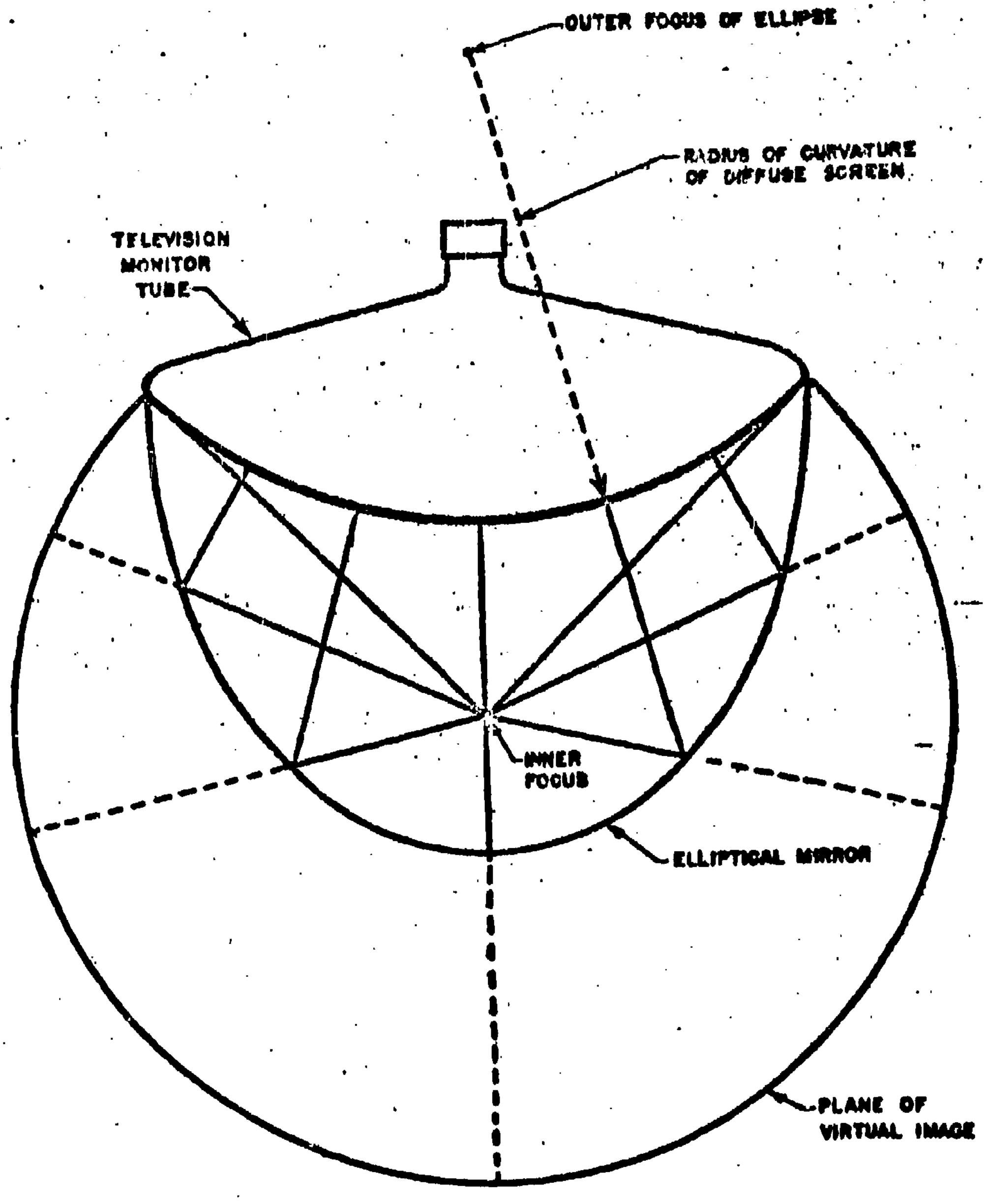
DIFFUSE TRANSMITTING SCREEN (SPHERICAL) AND SPECULARLY REFLECTING ELLIPSOID PROJECTION DISTANCE NOT CRITICAL



—DETROIT ARSENAL—

FIGURE 5

SPECULARLY REFLECTING ELLIPSOID UTILIZING KINESCOPE TUBE



—DETROIT ARSENAL—

FIGURE 6

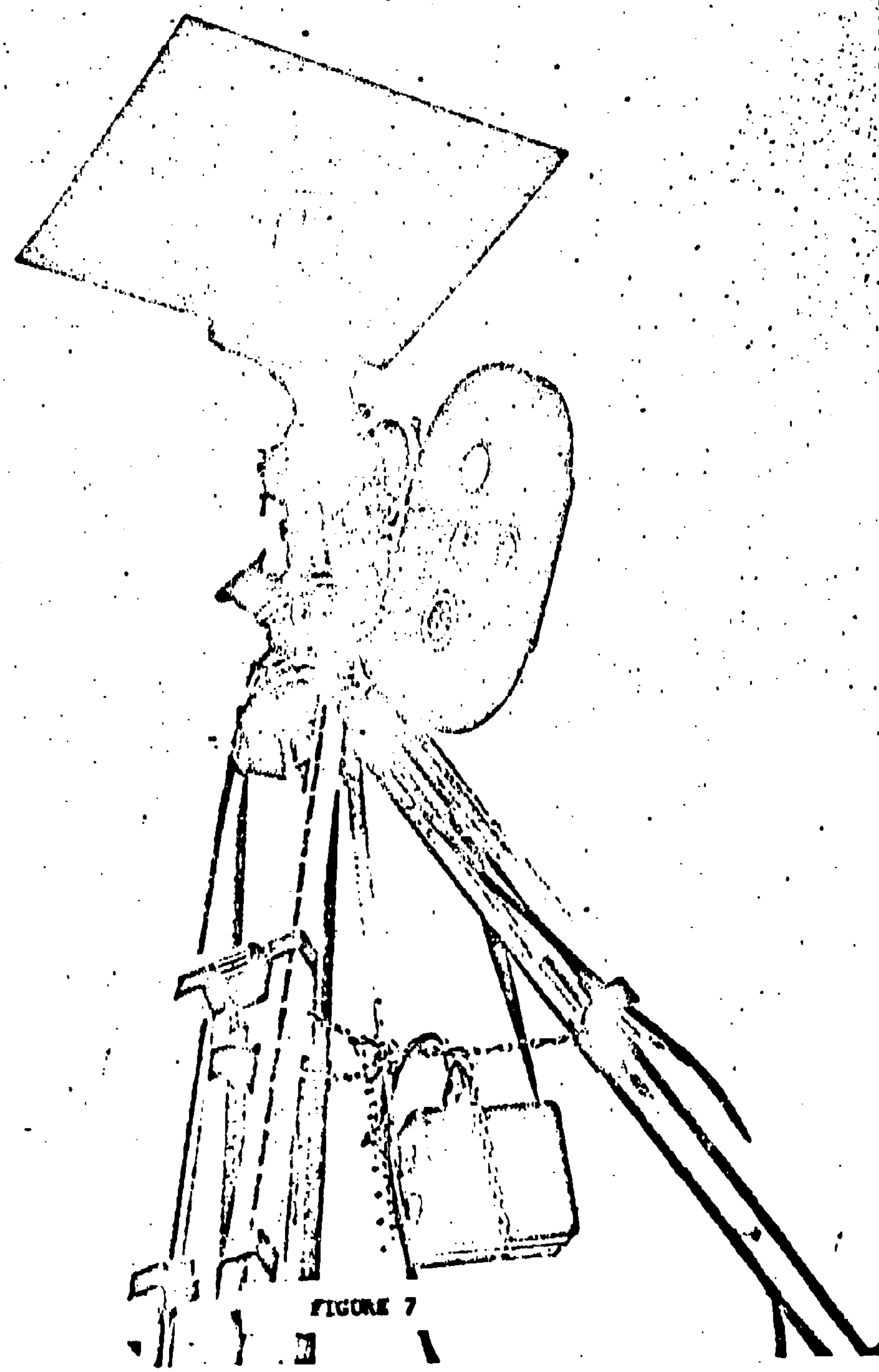


FIGURE 7

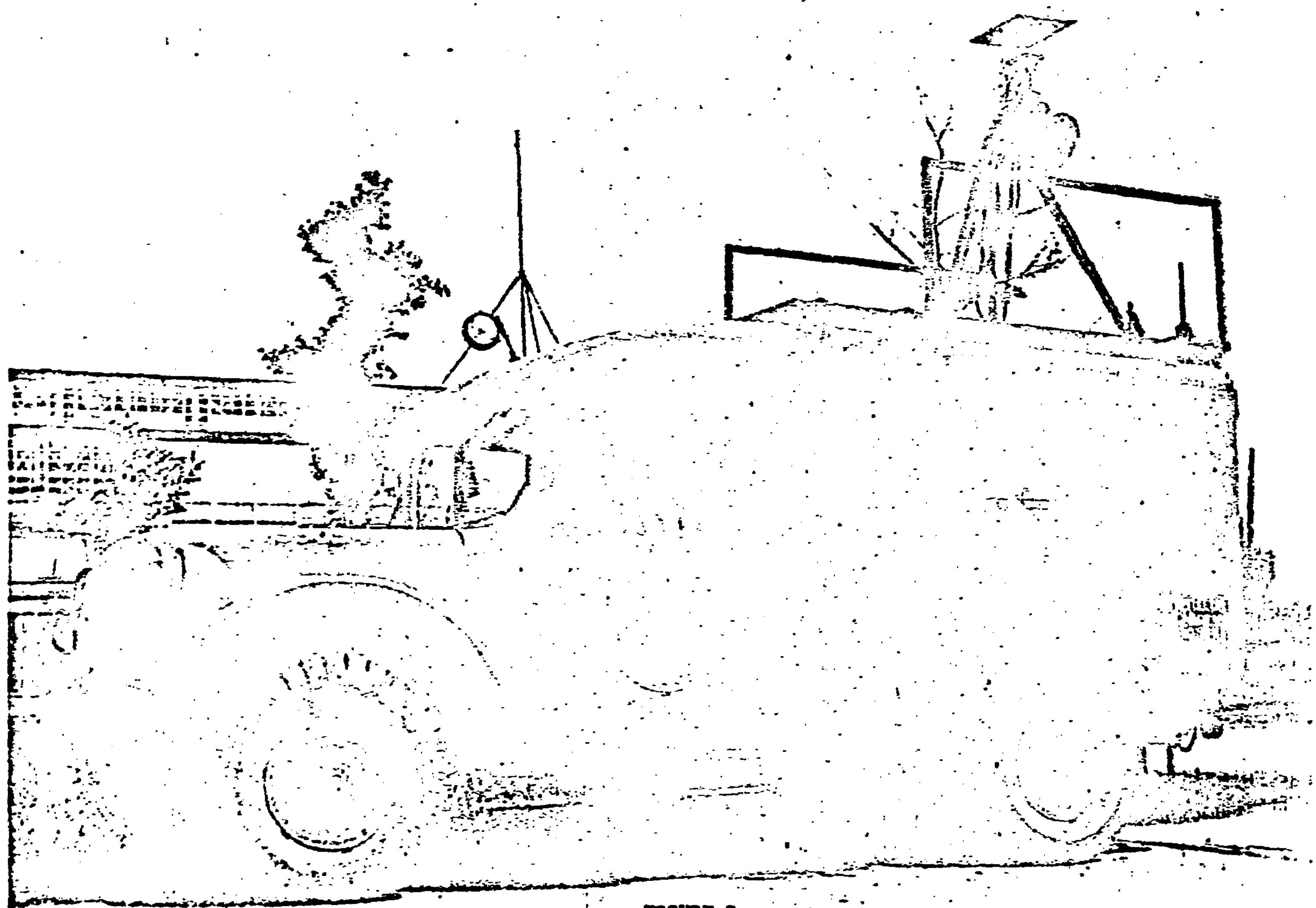


FIGURE 8

SOME STATISTICAL PROBLEMS RELATED TO MISSILE SAFETY

Paul C. Cox
Reliability and Statistics Office, Ordnance Mission
White Sands Missile Range

I. INTRODUCTION. One of the most important, yet difficult, phases of missile system evaluation is providing adequate assurance that the system will not cause serious injury or death to friendly troops as a result of misfire, pre-detonation, etc. The primary reason why this aspect of testing is so difficult is because of the necessity for demonstrating, with a high level of confidence, that the probability of serious injury to friendly troops will be very, very small; and all of this must usually be based upon the results obtained from a small sample and accomplished with a limited budget. It is the purpose of this presentation to discuss some of the possible approaches and some areas in which it appears additional study and research should be conducted. The general theme of this presentation is to obtain the desired confidence in the weapon safety and at the same time keep the sample size down to a reasonable figure.

To make the examples concrete, it will be assumed for this presentation that the safety requirement will be a 99% confidence that the probability of injury to friendly troops will be less than .0005 (one in 2000).

Before proceeding, four abbreviations will be introduced:

(1) (ECIP)--Event which might cause injury to friendly personnel.
(2) (PECIP)--Probability that an (ECIP) will occur.
(3) (ED)--Friendly personnel are actually seriously injured if an (ECIP) has occurred.
(4) (PED)--Probability of an (ED) on the condition an (ECIP) has occurred.

Inasmuch as this is a clinical paper, its purpose is to present problems for solution. These problems are listed on the last sheet (Appendix A) and will be referred to at the appropriate time during the presentation. Actually, [1] *, the first item in Appendix A is to urge the group to consider the solutions presented and think of a better approach to the overall problem of safety.

II. THE USE OF ATTRIBUTE TESTING. It can be shown that, if a random sample of N rounds has been selected and tested and if at the completion of the test, the number of (ECIP) is observed to be f, one may be 99%

*Numbers in brackets refer to questions posed in Appendix A.

confident that the (PECIP) for the entire population of rounds will be less than .0005. Values for N and f are listed in Table 1.

f	N
0	9,213
1	13,280
2	16,820
3	20,100
4	23,200

TABLE 1. Sample Size Required to Assure With a 99% Confidence That The (PECIP) $\leq$ .0005.

It is evident that the values for N, listed in Table 1, are entirely unrealistic for most weapon systems. Two other criticisms are: (1) Such a test will probably not indicate which sets of environmental conditions will assure safety and which will not; and (2) if the system is not safe, attribute testing will not, as a rule, indicate why the system is not safe.

III. LABORATORY TESTING OF COMPONENTS. The second method will be based upon laboratory testing of critical components. The approach will be:

(1) Isolate the components of the system which could result in an (ECIP).
(2) Determine those variables which can be used to verify the likelihood of the component causing and (ECIP).
(3) Determine the sets of environments under which the component is expected to operate.
(4) Design an experiment, conduct the test, analyze the data, and attempt to evaluate safety, giving full consideration to the results of (1), (2), and (3).

It is believed that this technique offers the greatest promise of any suggested within this presentation. It is quite possible that by using this method the desired probability may be verified with adequate confidence and from a relatively small sample. A second reason why this technique is desirable is because it may not only be used to indicate whether the weapon is safe or not, but it will very likely show why the system is not safe in the event it is not.

There are many interesting problems associated with this procedure, but due to limited time I will proceed to other techniques without going into further detail. [2]

IV. <u>THE APPROACH OF BREAKING DOWN THE CAUSES OF INJURY TO FRIENDLY PERSONNEL.</u> In Section II it was pointed out that if a sample of 9213 weapons are randomly selected and if none of these indicate an unsafe condition, then we may be 99% confident that the probability of an unsafe condition is less than .0005. If we find there is no reasonable alternative to attribute testing, one possible method for reducing the sample size is by breaking the problem down into two parts. The first is to test the likelihood of an event occurring which might cause injury to friendly personnel (PECIP) and then conduct a second test to estimate the probability that if such an event did occur it would actually injure friendly personnel (PED). If the likelihood of either of these events occurring is very small, one might establish the desired confidence with quite a small sample. For example, suppose a sample of n_1 systems were selected and were operated normally, then a sample of n_2 systems were selected and were induced to create a malfunction which might cause injury to friendly personnel (perhaps the motor might be induced to go "high order"), and if no unsafe malfunctions occurred in the first instance nor were any injuries noted in the second, then we may be 99% confident that $n_1 \cdot n_2 \cdot$ (PECIP) $\cdot$ (PED) ≤ 5.302 *. It is then clear that if $n_1 \cdot n_2 = 10,604$ (PECIP) x (PED) $\leq .0005$, then this value for $n_1 \cdot n_2$ can be satisfied if n_1 and n_2 are each equal 103. Thus with a total sample of 206, it may be possible to achieve as much as with a sample of 9213 <u>when the test is not broken down into two parts.</u>

It is felt that the idea presented in this section has a great deal of merit and should be explored further. Actually, one might conceivably break the safety problem down into three, four, or more causes and reduce the total sample size with each step. The procedure might easily break down, however, because: (1) The (PED) may be too large for the plan to be feasible; and (2) The cost of testing for the (PED) may be prohibitive.

The only statistical problem I am aware of in connection with this procedure is the limited supply of tables of confidence intervals for the products of binomial parameters. The table by Buehler is the only one with

---

*See confidence intervals for the product of Binomial Parameters, R. J. Buehler, P 402, Vol. 52, No. 280, Journal of the American Statistical Association.

which I am familiar, and it is rather limited. [3]

V. <u>THE USE OF SEQUENTIAL ANALYSIS.</u> When the need for economizing on sample size becomes evident, a great many people think immediately of using "Sequential Analysis", in particular, Abraham Wald's "Probability Ratio-Test". In fact, the "Probability Ratio Test" has been somewhat of a curse in the sense that so many consider it a virtual panacea when sample size becomes a problem. However, since sequential analysis is nearly always recommended as a method for reducing the sample size required, it is felt that a few questions about this approach should be presented to this group.

We will proceed by applying the methods, found in chapter 5 of Wald's test "Sequential Analysis", using the following entries:

$$\alpha = .01 \quad P_0 = .0005^*$$

$$\beta = .01 \quad P_1 = .005$$

The equations for accepting or rejecting the system are as follows:

$$a_m = -1.992 + .001956\ m$$

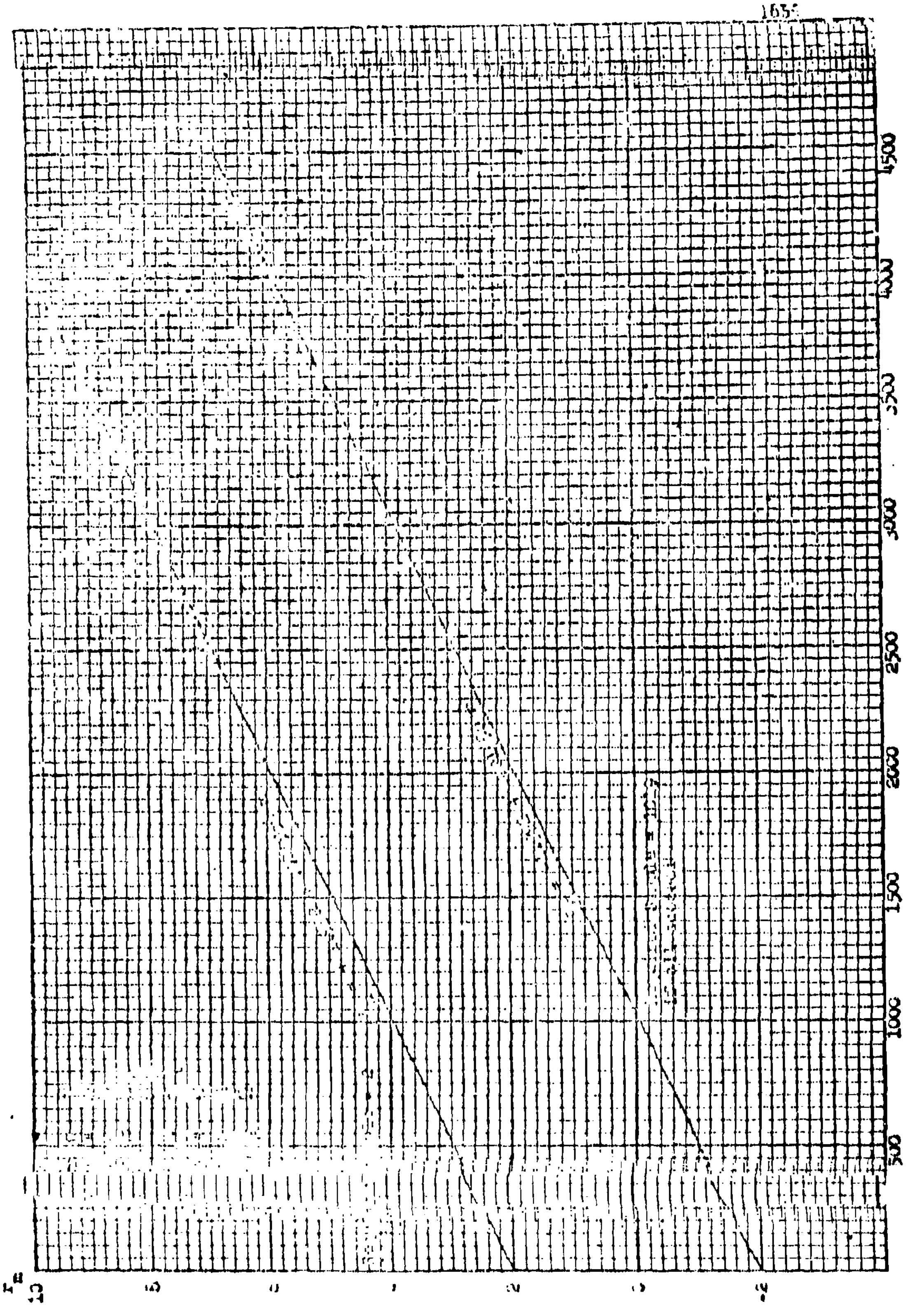
$$d_m = +1.992 + .001956\ m$$

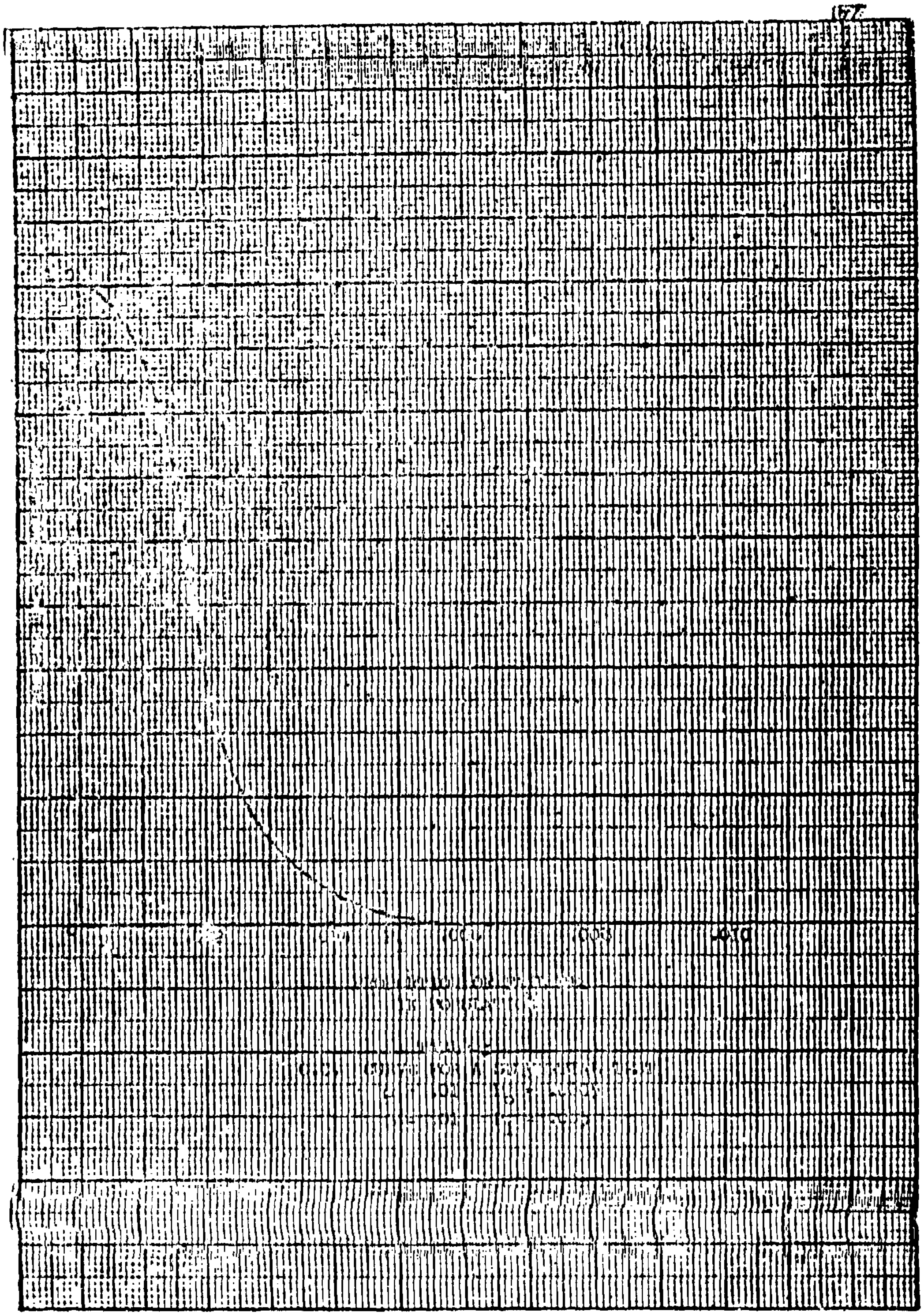
The graph of these equations is given by Table 2, and the O.C. Curve by Table 3. The following information can easily be obtained at this point: (1) If no failures occur among the first 1019 rounds tested, the system will be accepted; (2) The ASN curve has not been included, but its maximum value is approximately 2000 rounds.

From these two observations, it appears that a tremendous saving has been effected by introducing a sequential plan. However, if one investigates the O.C. Curve in Table 3, it appears that we are simply trying to answer the wrong questions. If we are answering anything at all in the area in which we are concerned we may be obtaining a 99% confidence that the (PECIP) $\geq$ .0005 <u>if the system is rejected.</u>

---

*I have chosen these values because there have been occasions when they have been offered as the appropriate values to determine, with a 99% confidence, that the (PECIP) $\leq$ .0005.


500
1000
1500
2000
2500
3000
4000
4500



It is at this point I wish to ask a few questions:

(1) Is there any existing method for computing binomial confidence limits by a sequential approach? [4] I have presented this question because quite often we are required to obtain certain confidence limits, and by the very nature of the problem the sequential approach is the appropriate one.

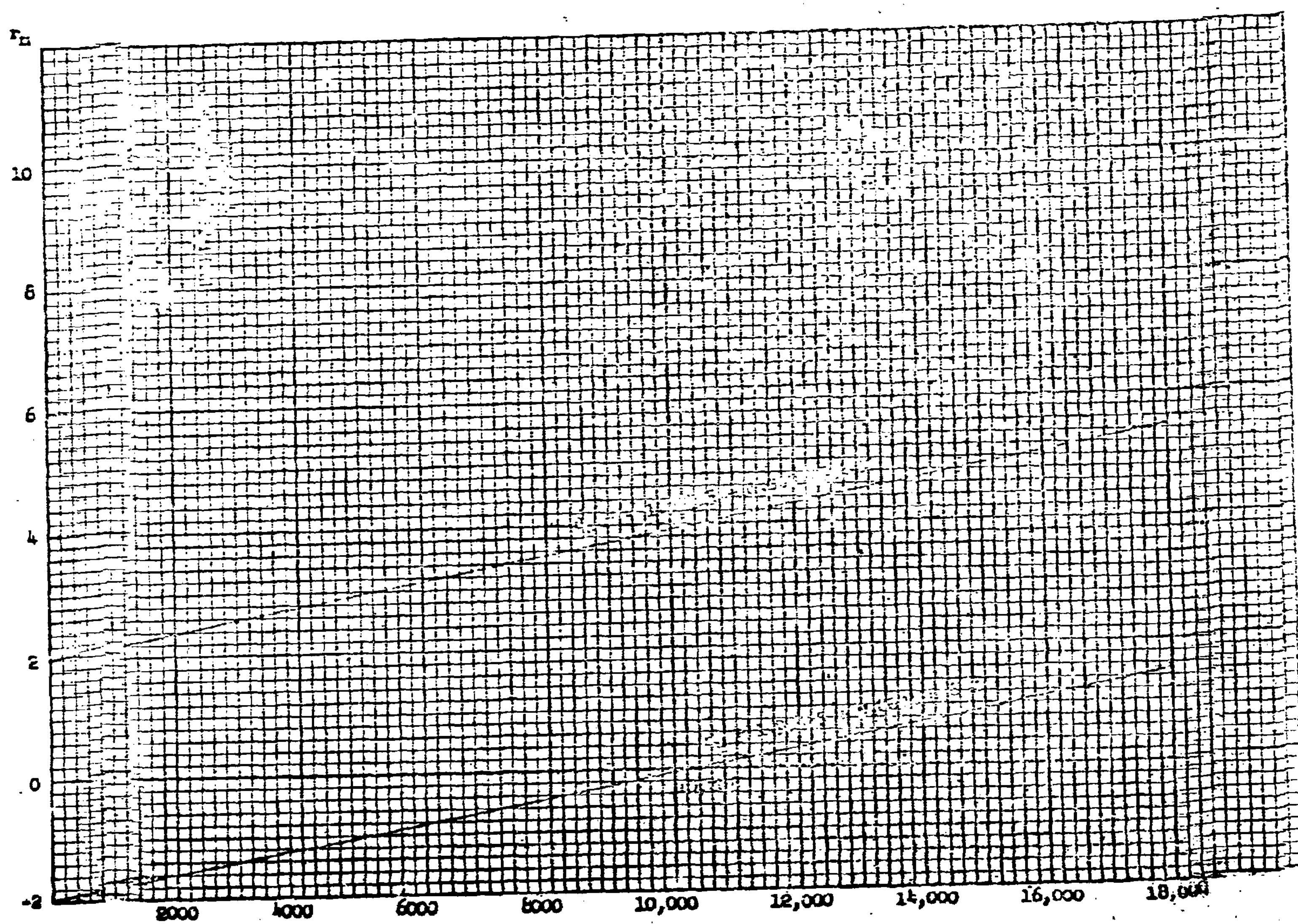
(2) Is it possible to use the well-known "Probability Ratio Test" to determine confidence limits in a sequential manner? [5]

(3) In the example just discussed, if at the termination of the test, the system is accepted, can we be 99% confident that the (PECIP) $\leq$ .005? [5] Similarly, if at the termination of the test the results indicate the system should be rejected, can we be 99% confident that the (PECIP) $\geq$ .0005?

(4) If it is not possible to obtain confidence limits from the "Probability Ratio Test", are we not obtaining something which is equally satisfying? [6] That is to say, we set up a test such that the probability of accepting the system is less than 1% if the proportion of failures in the entire population exceeds .005, then the results of the test indicated we should accept the system. Is this not as satisfying as a 99% confidence limit?

Let us assume we are obtaining confidence limits, or something equally satisfying, from the "Probability Ratio Test", then it is clear that if we wish to answer the original question of this paper (i.e., to establish with a 99% confidence that a (PECIP) $\leq$ .0005) it will be necessary to choose β = .01 and P_1 = .0005, while α and P_0 may be chosen arbitrarily or based upon some other consideration. Let us therefore choose α = .01 and P_0 = .00005. The O.C. Curve for this plan is the broken line in Table 8, and it will be discussed later. The sequential plan is given by Table 4, where it may easily be seen that if the first 10,208 rounds tested contain no (ECIP) the system will be accepted, if one and only one among the first 15,360 occurs, it will be accepted, and if only two among the first 20,000 occur it will be accepted. Obviously, this is considerably more than the sample size required by Table 1, and it is clear that any advantage gained by going to a sequential approach is not found in reduction of the sample size.

There remains another question about the use of sequential testing. It appears that both α and P_0 may be chosen arbitrarily. Actually it makes



r_H
10
8
6
4
2
0
-2
2000
4000
6000
8000
10,000
12,000
14,000
16,000
18,000

some sense to choose P_o and α small since the occurance of a single (ECIP), regardless of how many rounds have been previously tested in which no (ECIP) occurred, will undoubtedly result in a thorough investigation and perhaps suspension of production and use of the weapon.

Consequently, let us vary P_o and see what happens. The number of tests required to accept the system for 3 values of P_o (assuming no failures occur in the sample) is given by Table 5.

P_o	N
.00005	10,208
.000005	9,304
.0000005	9,197

TABLE 5. Required number of tests to accept the system when $\alpha = .01$; $\beta = .01$; $P_1 = .0005$; and no failures occur.

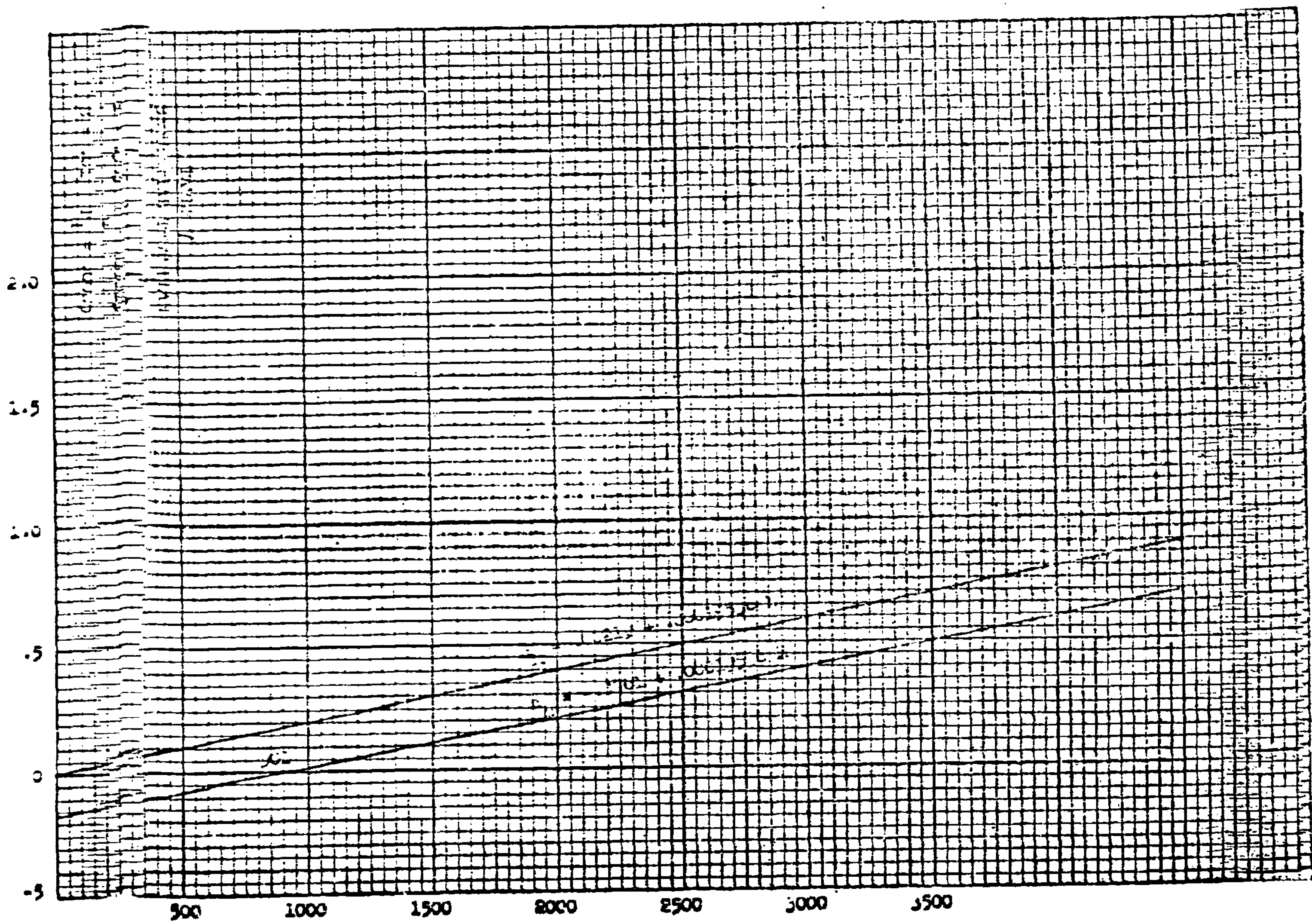
The following facts may be observed from Table 5: (1) the value 9197 is approximately equal the value 9213 listed in Table 1; and (2) regardless of how small P_o is chosen, the value for N will never become much smaller than the 9197 listed above.

Now, suppose we vary the values of α. The required sample size for various values of α (assuming no failures occur in the sample) is given by Table 6.

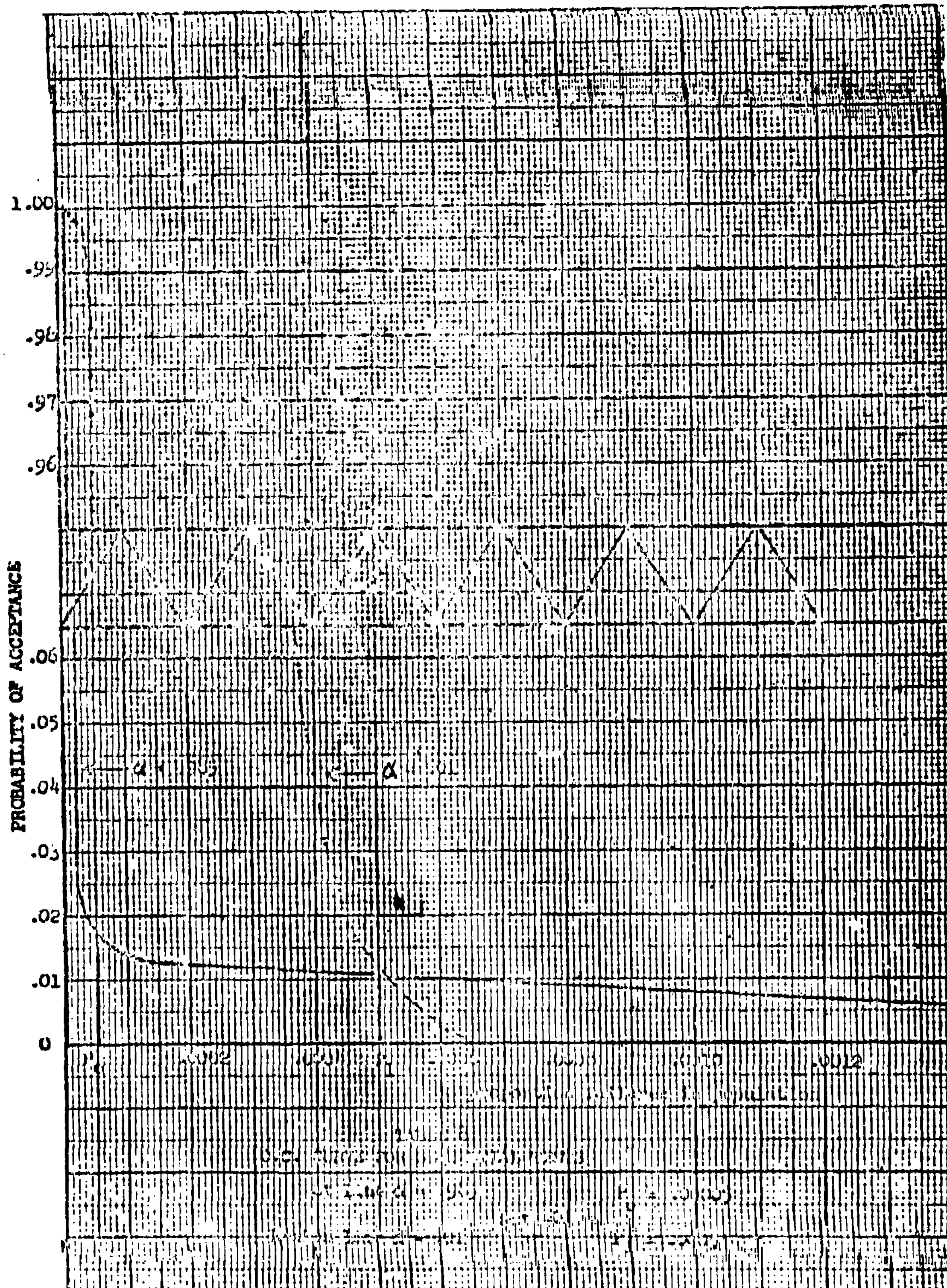
α	N	α	N
.01	10,208	.98	1541
.10	10,000	.985	901
.50	8,695	.989	212
.90	7,154	.9899	22
.95	3,577		

TABLE 6. Required number of tests to accept the system when $\beta = .01$; $P_o = .00005$; $P_1 = .0005$, if no failures occur.

Tables 7 and 8 give the sequential test plan and the O.C. Curve for the obviously absurd case in which $\alpha = .985$; $\beta = .01$; $P_o = .00005$ and $P_1 = .0005$.



2.0
1.5
1.0
.5
0
-5
500
1000
1500
2000
2500
3000
3500


PROBABILITY OF ACCEPTANCE
1.00
.99
.98
.97
.96
.06
.05
.04
.03
.02
.01
0

The results at this point appear to be somewhat ridiculous. It is certainly illogical that by taking α as close to .99 as we like, we can make N as small as we please as is evident from Table 6. Furthermore, Table 7 is peculiar in that testing must cease with the 901 round. That is, if a failure occurs in rounds one to 900 the system is immediately rejected, but if the first 901 rounds are good, the system is accepted.

The O.C. Curve in Table 8 may give some clue to the fallacy. While both curves pass through the point (.0005, .01), it may be observed that the broken line approaches zero rapidly while the smooth line approaches zero very slowly, after passing the point (.0005, .01). [7] Thus while both plans may give equal assurance at p = .0005, using α = .01 will give much better assurance of rejection at p = .0012. Nevertheless, there appears to be a serious fallacy in our reasoning, it seems we are getting something for nothing, and I would like the answer why we can't choose something like α = .985 and obtain the desired assurance with fewer rounds.

APPENDIX A

PROBLEMS SUGGESTED BY THE PRESENTATION

1. Do you have a better approach to the solution of the safety problem than any suggested in this presentation?

2. Do you have any comments concerning the use of laboratory testing of components in the determination of safety?

3. To my knowledge the article by R. J. Buehler, "Confidence Intervals for the Product of Binomial Parameters", p. 482, Vol. 52, No. 280, Journal of the American Statistical Association, is the only table of such confidence limits. There is considerable need for the preparation of additional tables in this area.

4. Is there any existing method for comp''ing binomial confidence limits by a sequential method?

5. Is it possible to adapt the well known "Probability Ratio Test" to determine confidence limits?--To be more specific, if at the end of the sequential test the product is accepted, can we be 100 $(1-\beta)$ % confident that the probability of a failure is less than P_1; and similarly, in the event the product is rejected, can we be 100 $(1-\alpha)$ % confident that the probability of failure is greater than P_0?

6. If the answer to 5 is negative, are we not obtaining something which may be equally satisfying from the "Probability Ratio Test"? That is to say, we set up a test such that the probability of accepting a system is less than β if the proportion of failures in the polulation exceeds P_1. The test indicates we should accept the system. Is it not possible this may be as satisfying as a 100 $(1-\beta)$ % confidence limit?

7. Is it conceivable that one might be concerned only with the values of β and P_1 when using the probability ratio test? If so, why can we alter the required sample size so drastically by varying α and P_0?

APPENDIX B

ABBREVIATIONS USED IN THE PRESENTATION

1. (ECIP) -- Event which might cause injury to friendly troops.
2. (PECIP) -- Probability that an (ECIP) will occur.
3. (ED) -- Friendly personnel are actually seriously injured if an (ECIP) has occurred.
4. (PED) -- Probability of an (ED) on the condition an (ECIP) has occurred.

DISCUSSION OF

"SOME STATISTICAL PROBLEMS RELATED TO MISSILE SAFETY"

The discussants for this clinical paper included: Dr. Jerzy Neyman, Dr. William Bechhofer, Dr. F. J. Anscombe, Dr. H. A. David, and Dr. J. E. Jackson. The suggestions made by these five and others in the audience were excellent, and we have the opportunity to include some of their comments.

To begin with, there are several reports which are related to this subject. Some of these are listed below:

(1) Anscombe, F. J. (1949). "Large-sample theory of sequential estimation". Biometrika 36, 455-8.

(2) Anscombe, F. J. (1953). "Sequential estimation". J. Roy. Stat. Soc. Ser. B 15, 1-29.

(3) Armitage, Peter. "Numerical Studies in the Sequential Estimation of Binomial Parameters". Vol. 45, Biometrika, 1958, p.1.

(4) DeGroot, M. H. (1959). "Unbiased sequential estimation for binomial populations". Ann. Math. Stat. 30, 80-101.

(5) Ray, W. D. "Sequential Confidence Intervals for the Mean of a Normal Population with Unknown Variances". J. Roy Stat. Soc. Ser. B, 19, 133.

In addition to the five reports listed above, Dr. F. J. Anscombe is currently preparing a paper "Testing to Establish a High Degree of Safety in Reliability", which will be offered for presentation at certain statistical meetings and for publication in one of the statistical journals, in the near future. Dr. Anscombe's report deals with many of the questions which were raised in the clinical paper and gives some valuable guide lines toward their solution.

The following comments concerning the clinical paper were made by Dr. Anscombe:

"(1) The formulas given by Abraham Wald for the probability ratio sequential test are approximations and should be used only when the boundaries are fairly far apart, i.e. when $-h_0$ and h_1 are (say) 2 or more. They are hopelessly inadequate when $h_0 = -.076$ and $h_1 = .002$ (as in Table 7).

"(2) The example illustrated in Table 7 is nothing but a fixed sample plan, having a sample of size 901 and acceptance number 0, with the understanding that if a failure occurs before all 901 rounds have been fired, the test may as well be stopped at that point. The O.C. Curve can easily be calculated, and is very different from that shown in Table 8.

"(3) Tests can and must be carried out sequentially. If the requirement of 99% confidence that the probability of failure will be less than .0005 is taken seriously, then the acceptance boundary must be identical with (or at least very close to) that given in Table 1. To obtain a fully explicit sequential procedure, we must add another boundary for "abandon the trial". It will be proper to abandon the trial either because it seems likely that $p > .0005$, or because it seems likely that to complete the trial and reach the acceptance boundary will be too expensive.

"(4) The binomial probability ratio sequential test of Wald is for the purpose of comparing two simple hypotheses, $p = p_0$ versus $p = p_1$. There are no two such simple hypotheses here. The present problem seems closer to what is generally thought of as an estimation problem than to a testing problem. Anyway, there is no reason to suppose that Wald's type of sequential test is particularly appropriate.

"(5) There is no magical economy in sequential procedures, such that you get something for nothing. In a good sequential plan, observations continue until enough information is obtained, and then they stop. The only economy is the economy of not stopping too soon before a sufficiently decisive result is obtained, nor continuing unnecessarily long. In the present case, the trial should stop as soon as one of the conditions in Table 1 is met, or as soon as the results are decisively discouraging."

The following comments were made by Dr. J. E. Jackson. Dr. Jackson's comments are direct answers to questions 4-7 in Appendix A.

"4. Since Anscombe was also on this panel, he can give you more information, probably, than anyone else in the world on sequential estimation. However, a few relevant references might be: ...." (Dr. Jackson listed several references which are included in the list above.)

"5. This is to some extent covered in No. 4. The other question is whether or not confidence limits are really the important thing. See No. 6.

"6. The point that Neyman raised is a good one, although I am not sure I am qualified to answer it. His point was that you really should be using a significance test all along since you have essentially a problem of deciding whether or not to use a particular missile system. He feels that the important thing is the decision; worry about the confidence limits later. In this case you are testing the null hypothesis:

$$H_O: \pi < .0005$$

against the alternative

$$H_1: \pi \geq .0005.$$

"7. If this is to be treated as a significance test, what risks should be used? Using $p_1 = .00005$, $p_2 = .0005$, $\alpha = .01$ and $\beta = .01$, you find you need from 10,000 to 20,000 rounds. Since p_2 and β must be kept where they are, what happens when α is increased? While it is true that increasing α to .985 will decrease the sample size required considerably, for a value of $p_1 = .00005$, it would also mean that you would hardly ever accept a missile system since if you had only one malfunction in 20,000 rounds, you would still reject the system almost all of the time. While this would guarantee your β-risk at a minimum cost, it won't likely obtain any improvements in your missile systems. However, on page 6, second paragraph from the bottom, you state that the occurance of a single [illegible] regardless of how many rounds had been fired would result in an investigation of the system. In that case, it would seem that all you would need would be a single sample plan which resulted in rejection if a single defective round was found in the sample. However, using Molina's Tables, it appears that you would need a sample size of 9,200 rounds to get a probability of .01 that you would fail to reject a system having a .0005 probability of failure. This checks out pretty well with your results on page 1. Evidently, no matter how you work it, you need tremendous sample sizes to guarantee the risks you wish to impose.

"This might suggest another possible approach but I am certainly not qualified to pass judgment on this one. This would involve a re-evaluation of the risks, particularly the choice of p_2 and β. I realize that the tactics of war have changed a great deal in the last twenty years, but in my experience as a rifleman in World War II, it seemed to me that the human element was more to be feared than the mechanical. By that I mean, it seemed to us that we would encounter more trouble from wrong firing orders on the part of the artillery and wrong identification on the part of the airforce (not that either occurred very often) than from short rounds. In other words, one short round would not be nearly as damaging as one misdirected salvo. If things are still that way, maybe p_2 is too small. Again, this sort of decision is not in my field but it is a suggestion. It doesn't appear as though the sample size can be markedly reduced otherwise".

Dr. Herbert David made the following comment:

"I certainly agree with your main conclusion that any type of attribute testing would require an inordinately large sample size. Professor Anscombe commented very adequately on the sequential procedures you discuss. In spite of Armitage's 1958 Biometrika paper, I doubt that sequential estimation procedures have a great deal to offer over and above fixed sample procedures except as a by-product of sequential tests."

APPLICATION OF FACTORIAL EXPERIMENT AND BOX TECHNIQUE TO BALLISTIC DEVICES

D. J. Katsanis and C. L. Fulton
Frankford Arsenal

ABSTRACT. The factorial experiment and Box technique have been applied to ballistic experiments with recoilless rifles, aircraft seat ejection catapults and rockets, high-low guns and Davis gun, for the purpose of reducing time and cost of ballistic experimental development projects. The result has been a reduction in the number of rounds fired with little or no reduction in the validity of the analysis of variance. A detailed discussion of application of the Box technique to the factorial data is presented. This application results in the determination of a "zone of suitable performance" which makes use of interaction effects to provide greater flexibility in the selection of design parameters.

INTRODUCTION. In the experimental development of ballistic systems at Frankford Arsenal we are faced with a wide variety of experimental problems. For example, in recent years we have been concerned with recoilless weapon systems, aircraft seat ejection catapults and rockets, thrusters, high-low guns, and reactionless launchers. Some of these systems are required to function repeatedly with performance variations of the order 0.1 percent, others are one-shot devices which must function reliably with performance over, under or within certain prescribed limits. Sample size varies a great deal, as well as the type of performance requirement. In development of items such as small arms cartridges we can fire thousands of experimental rounds while a seat ejection catapult, for example, limits us to twenty or thirty to fifty rounds.

By use of factorial experimental design techniques and analysis, combined with physical interpretation of the data in terms of response surfaces, as suggested by Dr. Box*, a tremendous flexibility of standard statistical practices is achieved. This method has been applied in one way or another to the devices mentioned previously. As examples, our studies with the reactionless launcher, an analog computer simulation of a thruster, and the "BOX" of a seat ejection catapult will be discussed. The presentation herein, illustrates in chronological order a step by step experimental evaluation of the technique. The experimental evaluation

*Box, G. E. P.; "The Exploration and Explanation of Response Surfaces: Some General Considerations and Examples", Biometrics Vol 10, No. 1, Mar 1954.

was preceded by an abstract evaluation which is not reported here. First, existing data from a seat ejection catapult development was studied to determine in a preliminary way the method's effectiveness, the required type of experiment, and some of the experimental pit-falls. Secondly, we report a theoretical study of a thruster from which we learned something about the response surfaces and methods of interpolation. We finally "wrap up the story" with a discussion of the reactionless launcher. This study was conducted from start to finish using the experimental design methods we propose.

MODIFIED M5 CATAPULT. The possibility of applying the Box technique* to existing data for the modified M5 seat ejection catapult was considered. Although a carefully controlled experiment as performed in the reactionless launcher study (to be discussed later) is required to obtain fully valid results, a preliminary analysis of existing data by the Box technique was expected to give some indication of its effectiveness. Data from 24 firings of the modified M5 catapult were analyzed using three variables: temperature (T), charge (C) and web (W), each at two levels for two propellant compositions (lot 5655.1 and lot 5656.1).

The requirements the modified M5 catapult was to meet at that time were as follows: The peak acceleration (g) and the rate of change of acceleration ($\dot{g}$) were not to exceed 25 g's and 300 g/second, respectively, the final velocity (v) to equal or exceed 80 fps.

The least square method was employed to fit plane surfaces** to the experimental data for g, $\dot{g}$ and v, yielding the following equations:

$$g = -308.3W + 0.13C + 0.097T + 46.1$$

$$\dot{g} = -1354W + 0.358C + 1.463T + 287.6$$

$$v = 152.1W + 0.3075C + 0.0908T + 60.08$$

*ibid.

**The functions are not really plane surfaces. To simplify the calculations a limited range of the parameter is chosen so that the variables can be considered a linear function of the parameters within that range. Caution must, therefore, be exercised when interpolating or extrapolating. For example, the origin (W=T=C=O) is not a valid point on those planes.

where Web, W, is in inches, Temp, T, in ^{o}F, and Charge, C, in gas.

The equations were plotted for constant values of C, W, and T, i.e., the intersections of the g, $\dot{g}$ and v responses with the six planes formed by choosing constant values of C, W, and T were graphed (See Figures 1 through 3). The lines on these graphs represent the intersection of the response surface with the constant planes. For example, Figure 3A depicts the intersection of the g, $\dot{g}$ and v response surfaces with the plane formed by taking the temperature as 70^{o}F. The arrows indicate the direction of increasing magnitude of the v response surface and decreasing magnitude of the g and $\dot{g}$ surfaces.

The next step was to form the six constant planes into a box. The response surfaces within the cube were obtained by joining the corresponding curves for g, $\dot{g}$ and v. Photo 1 (see end of this article) shows this box. The thickness of the response surfaces is a result of round to round variation in ballistic performance. This illustration is qualitative, actual thickness must be determined from analysis of variance of the data.

An operating point (W_o, C_o, T_o) which satisfies performance requirements for this model must be within the cube volume defined by the three response surfaces. It is seen that the g and $\dot{g}$ requirements* are not met by all points within this space, except points in front of these planes (in the direction of the arrows). For example, the coordinates of point W_o = 0.158 in., C_o = 121 gm and T_o = 85^{o}F give a web, charge, and temperature at which acceleration is less than 25 g's, and rate of acceleration change is less than 300 g's/sec with a velocity greater than 80 fps. We see further that there is a volume surrounding this point over which the specifications will be met. This volume we will call the zone of suitable response. It has limiting values determined by the geometry of the response surfaces.

A better operating point might be found by extending the v, g, and $\dot{g}$ response surfaces outside the limits of the Box. For example, it appears that a new constant web plane for webs greater than W = 0.16

*g = 25 ft/second and $\dot{g}$ = 300 g/second

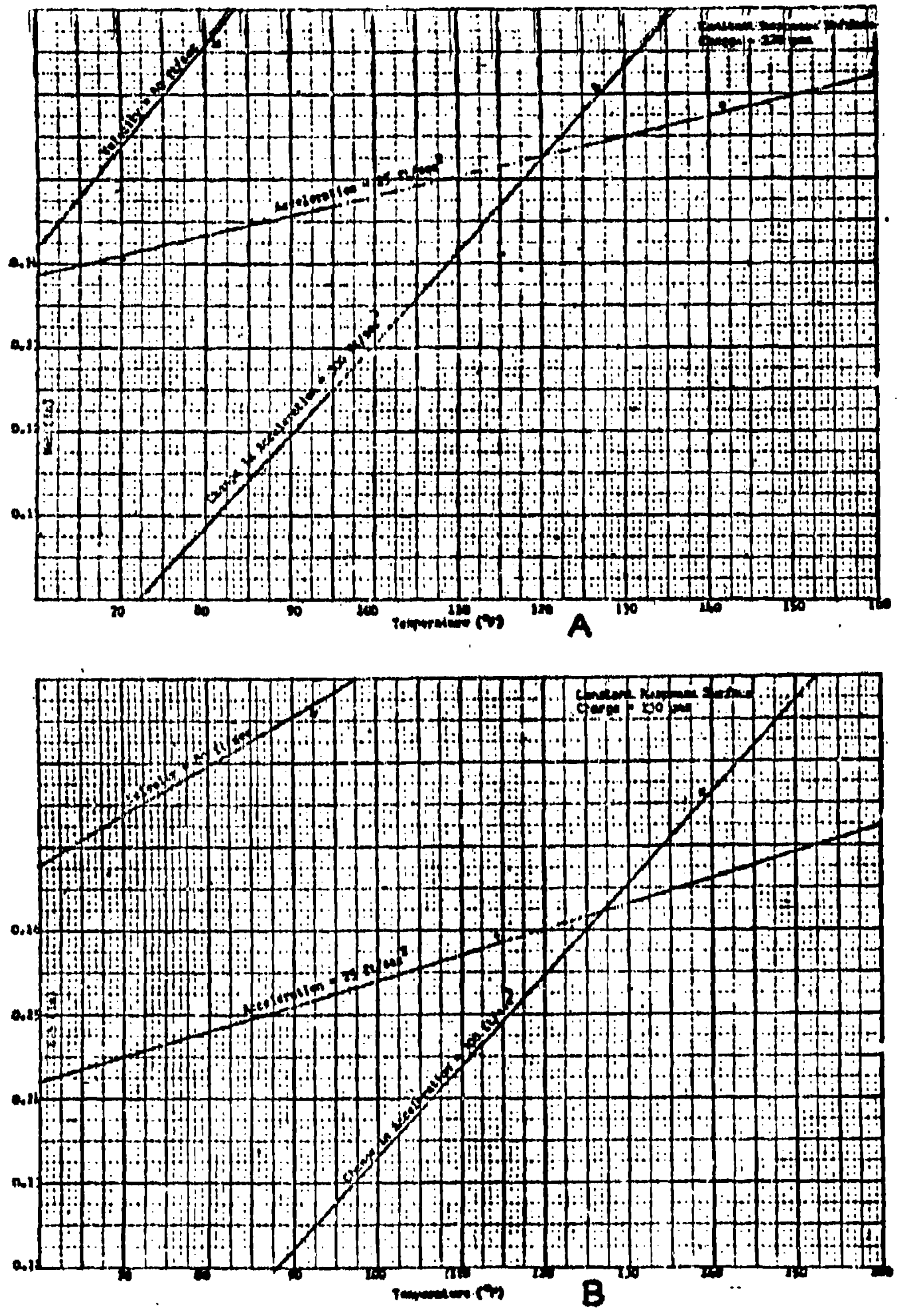


Figure 1 Constant response surface
A – Charge = 120 gm
B – Charge = 130 gm

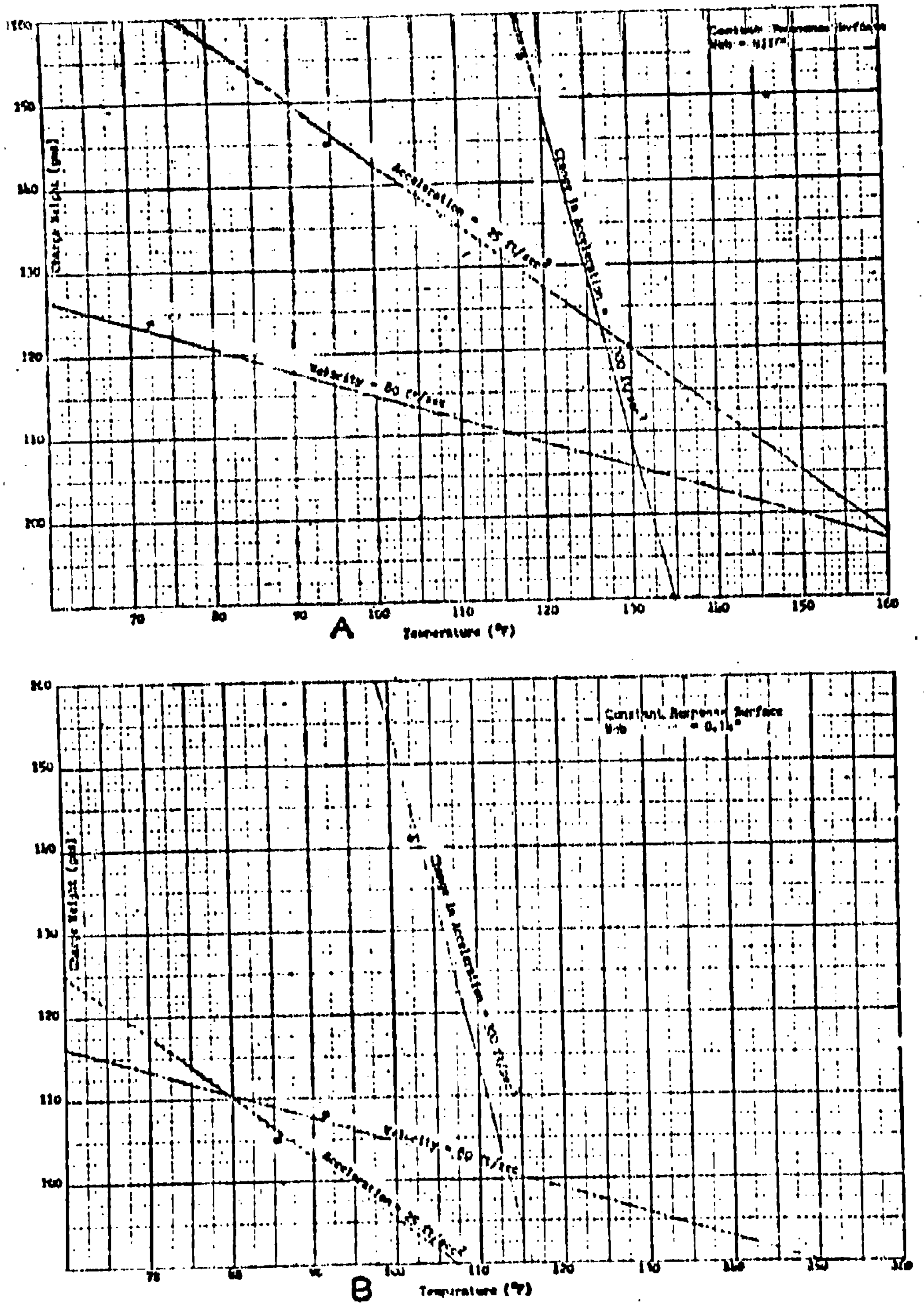


Figure 2 Constant response surface
A - Web = 0.16 in.
B - Web = 0.14 in.

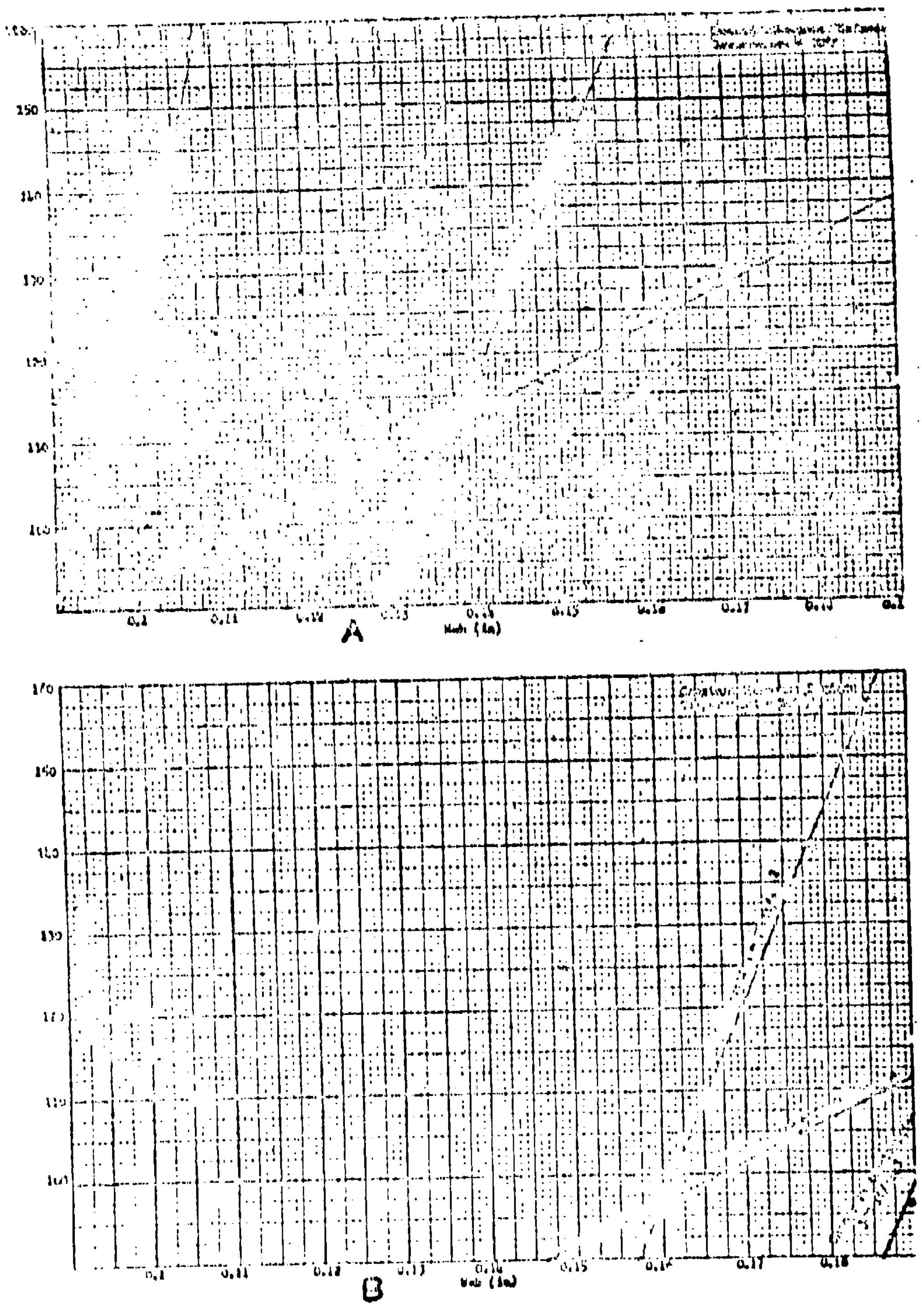


Figure 3 Constant response surface
A – Temperature = 70° F
B – Temperature = 160° F

will increase the temperature range over which desired performance is achieved.

In addition, the response surfaces may be extended in the direction of increasing or decreasing charge or temperature; thus a volume space can be obtained over any desired range of webs, charge, and temperatures (other values, such as internal volume, expansion ratio, etc., could be used instead of those chosen for this particular model) on the basis of a relatively few firings. Any extension of the response surfaces outside the cube which represents experimental values is only as valid as the assumption that the response surfaces are planes. It becomes important then to learn something about the response surface. In particular the hazards involved in interpolation and extrapolation should be studied. A start was made in this direction with a theoretical study of a thruster.

THRUSTER. An analog computer was used to develop theoretical response surfaces for a thruster which moves a 500 lb load vertically.* Two restrictions were imposed:

1. Maximum pressure to be less than 7000 psi

2. Final velocity to be greater than 7.5 fps.

About 60 computer runs were made for various design parameter combinations. The ballistic design parameters which were considered are Charge (C), propellant web (W), and chamber volume (V_C). The intersection lines of the response surfaces with the planes were obtained graphically from the results of the 60 simulations.

Figure 4 illustrates the intersection of the response surfaces with the plane: charge = 3 grams, while Figure 5 is the intersection with the plane: V_c = 1.3 in.^{3} and Figure 6 the plane: Web = 0.11 in.

* Details of computer simulation of ballistic devices can be found in the following references: Boritz Report; L. Stuart & W. A. Dittrich Report; Frankford Arsenal Report No. R-1313, "An Analog Computer Study of Interior Ballistics Equations", L. Stout & W. A. Dittrich; Frankford Arsenal Report, "Analog Computer Study of Interior Ballistics of Propellant Actuated Devices", R. Boritz & S. Narise.

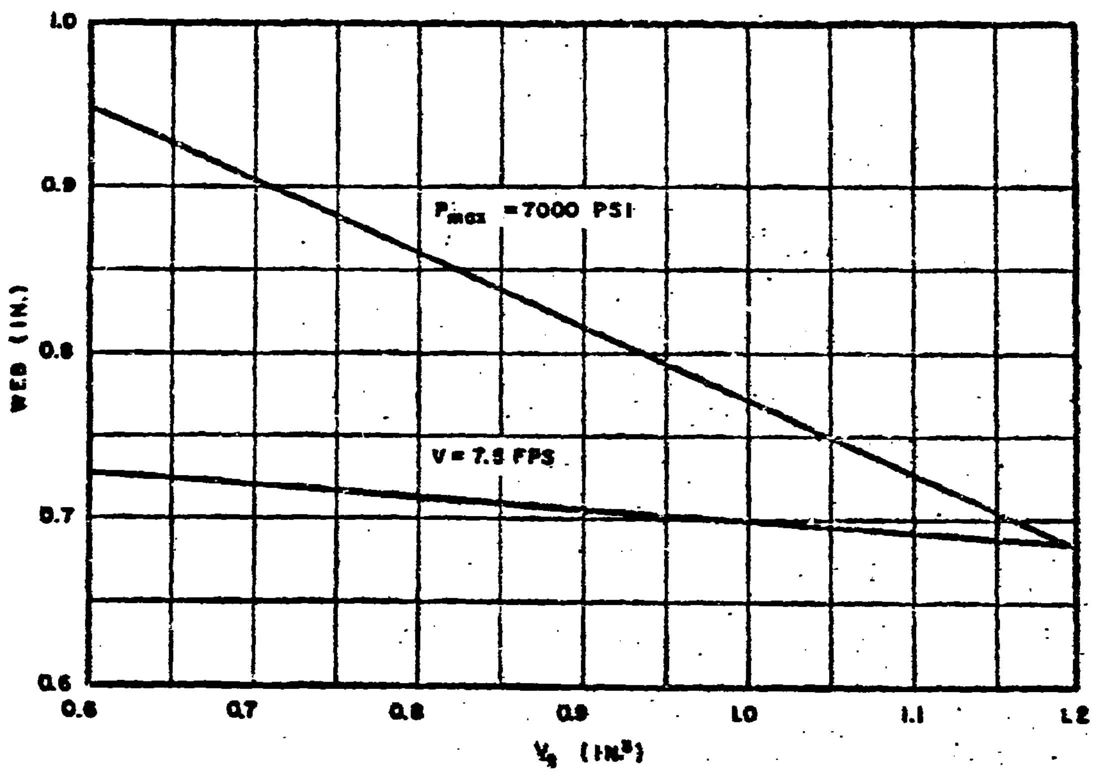


Figure 4. Intersection of the response surfaces with the plane, charge = 3 grams

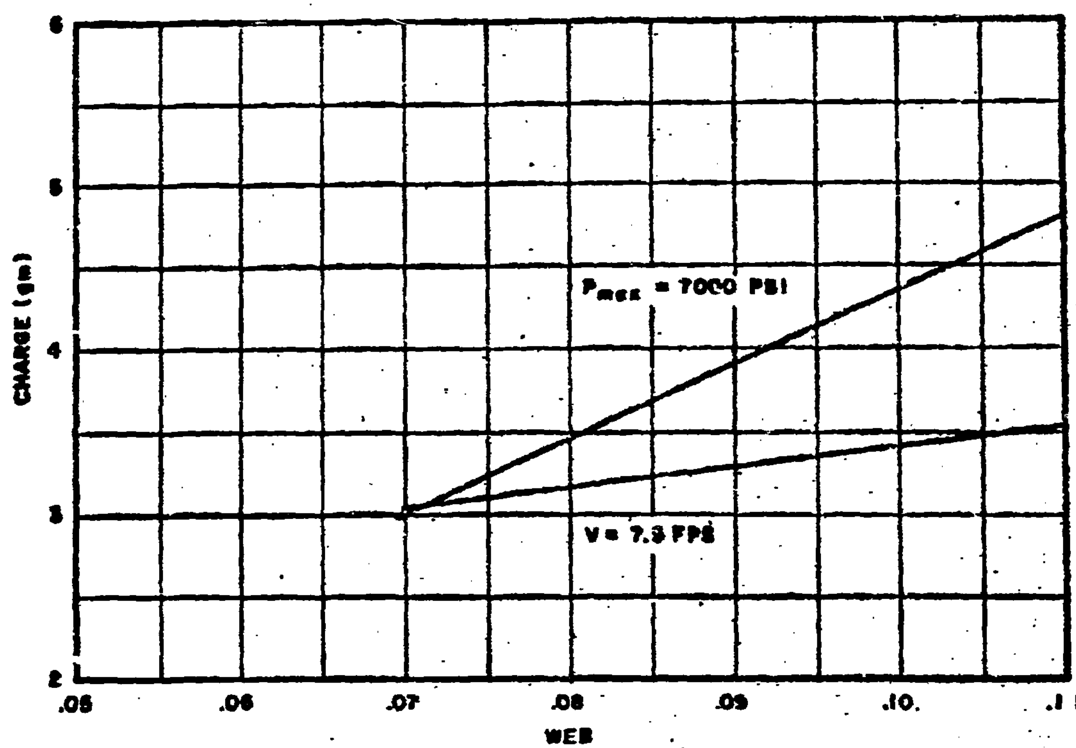


Figure 5. Intersection of the response surfaces with the plane, V_o = 1.3 in.^{3}

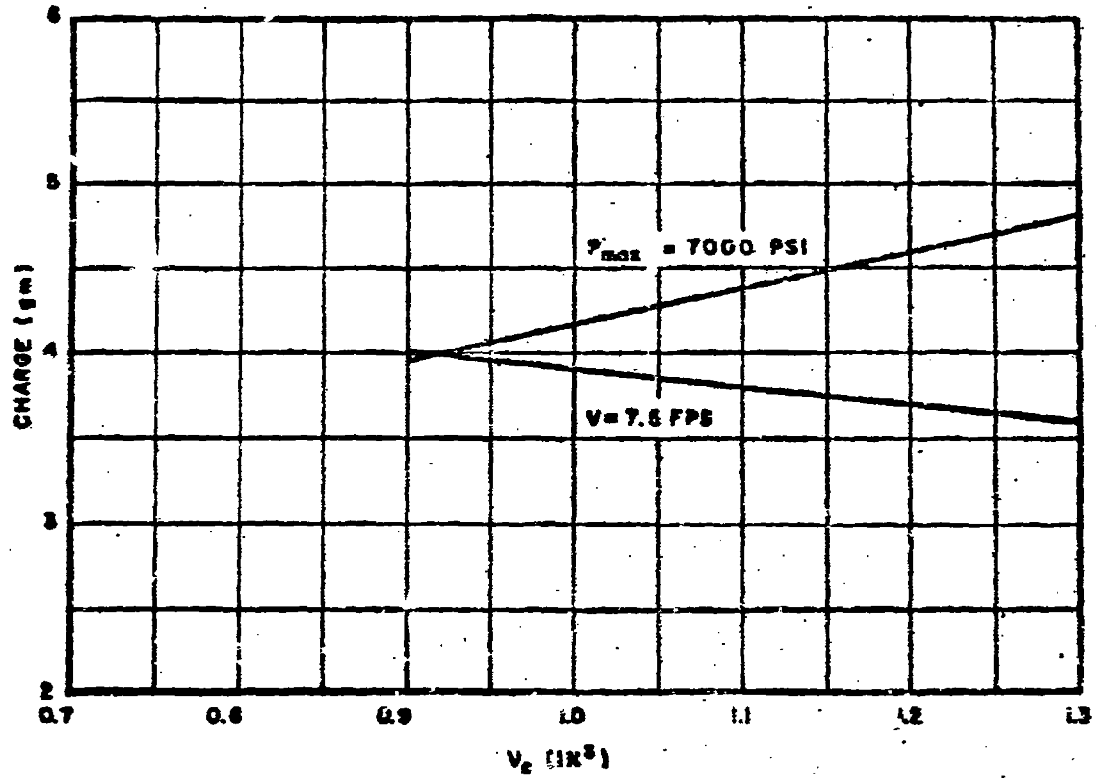


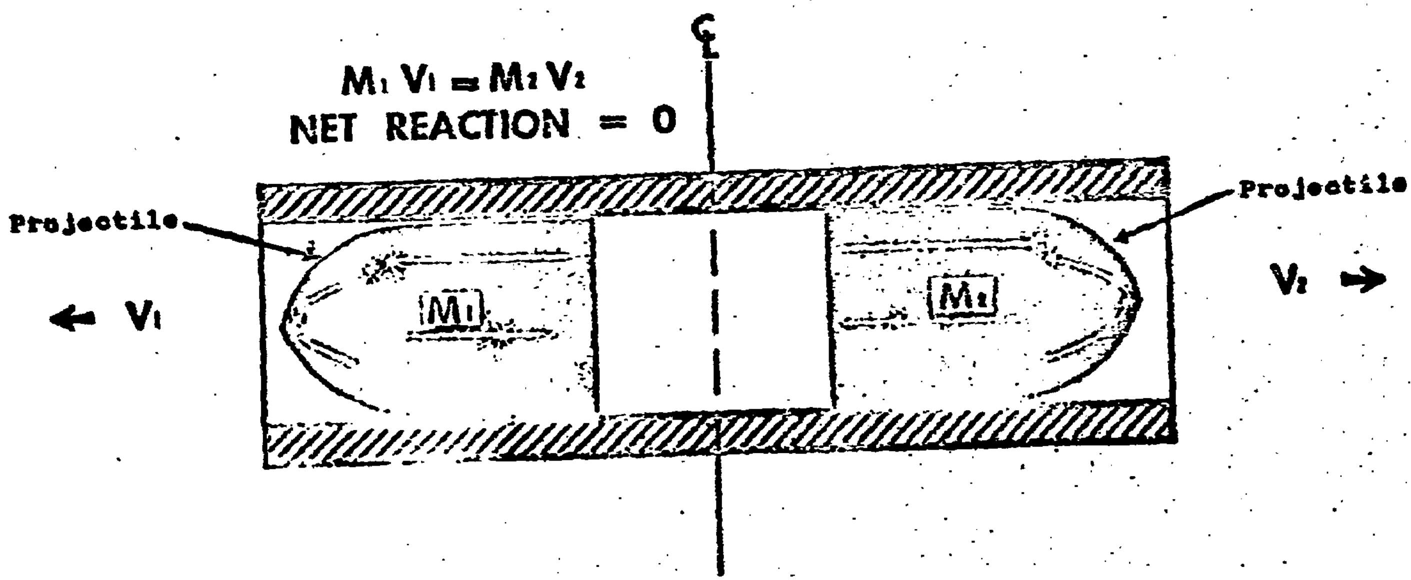
Figure 6. Intersection of the response surfaces with the plane, web = 0.11 in.

The three dimensional representation of the two response surfaces (pressure = 7000 psi and velocity = 7.5 fps) are shown in Photo 2. Some warping of the response surfaces can be seen. This illustrates a non-linear response. However, the nonlinearity is well behaved. No oscillations, peaks or humps occur. A linear interpolation should, therefore, be adequate if the box is small enough. At most, second order terms would be necessary. Size of the box should be small compared to nonlinearities but large compared to nonuniformities. Preliminary experimental work in ballistic development should be directed toward determining linearity and uniformity. This information is essential before setting up the factorial experiment so that the differences in performance levels will be significant, and so that the complexities of non-linear interpolation of the data can be avoided. In addition, this information should give some idea of the range of validity of extrapolations. However, it is a good practice always to verify extrapolation experimentally. Proper preliminary work should eliminate the need for extrapolation.

The operating volume or zone of suitable response is seen to be triangular in cross section opening up in the direction of increased chamber volume and corresponding increased web. Thus for an increased chamber volume, the range of web and charge over which the two restrictions would be met is greater. Picking a set of values for C, W, and V_c approximately in the center of the zone of suitable performance would thus minimize the chance of violating our restrictions because of manufacturing tolerances. A larger chamber volume would allow substantial reduction of these tolerances. The actual chamber volume allowable of course is subject to the physical size of the thruster and other ballistic considerations such as ignition and expansion ratio.

REACTIONLESS LAUNCHER. The reactionless launcher is a Davis type recoilless gun for ejecting masses from a ballistic missile during flight. In the particular project to be discussed here, these masses were intended to decoy anti-missile missiles. The launcher holds two projectiles as shown in Figure 7.

The decoys are of many sizes and weights and are launched at a wide range of velocities. The weight range considered was 20 to 60 pounds and the velocity ranged from 50 to 110 feet per second. The wide range of performance required two types of interior ballistic systems, direct and high-low, as shown in Figure 8. We had two types of projectiles, the bullet type (full caliber) that fits directly in the bore of the



REACTIONLESS LAUNCHER

Figure 7.

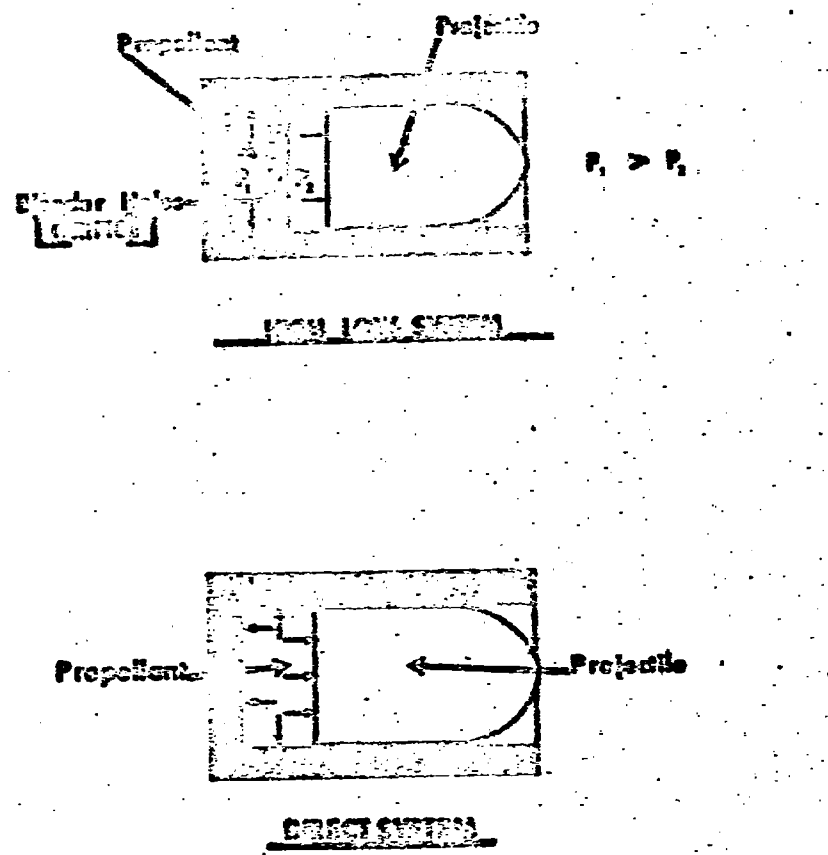

Propellant
Projectile
P1 > P2
Propellant
Projectile
DIRECT SYSTEM

Figure 8.

gun, and the spigot type which has a rod that fits in the gun barrel with the pay load outside the gun, as shown in Figure 9.

The entire study involved a total of eight variables: charge type, decoy type, charge weight, decoy weight, shot-start breaking pressure, expansion ratio, web, orifice area. To blindly set up a factorial experiment at two levels would require the firing of 2^8, or 256 rounds. Replicating three times, which is reasonable for this type of study, would lead to firing more than 750 (of the order 1000) rounds. Instead we isolated factors with no interactions, such as the type of chamber. The high-low chamber was studied separately from the direct chamber. We divorced the spigot projectile from the bullet type for the direct system but not for the high-low system since the high-low performance would not be expected to depend strongly on the type of projectile.

As a result of this disjoining process, the study was split into three programs, A, B and C. In program A, a high-low chamber was used with a bullet type projectile. The main variables were:

Charge weight.

Shot-start static breaking pressure.*

Orifice area of high pressure chamber.

In program B, a direct chamber was used with a bullet type projectile. The main variables were:

Charge weight.

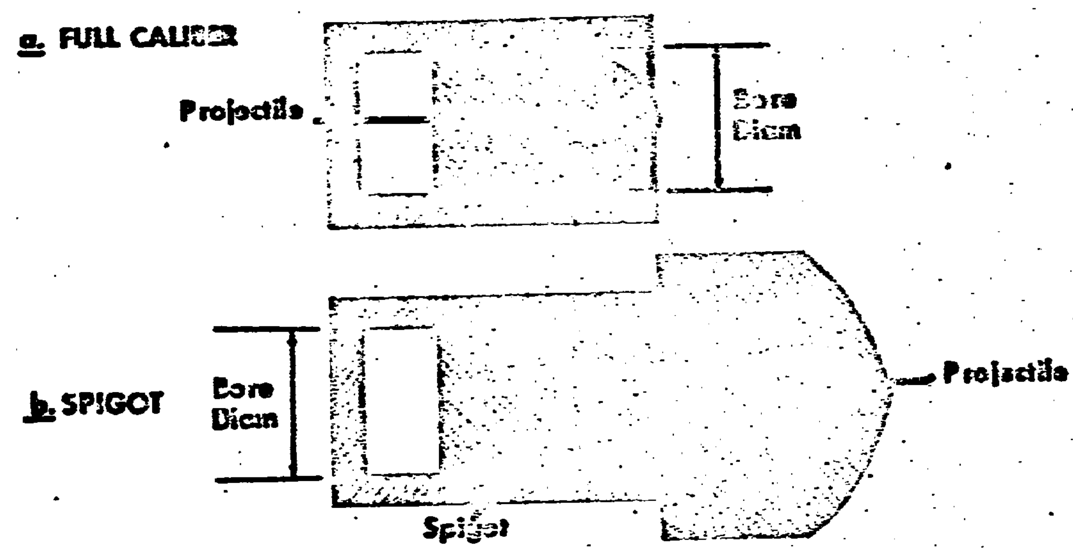
Propellant web.

Shot-start static breaking pressure.

In program C, a direct chamber and spigot projectile were used. The main variables were:

Decoy weight.

*Shot-start is a rod which restrains projectile motion until chamber pressure reaches a predetermined level.



FULL [illegible] &
SPIGOT [illegible]

Figure 9

Shot-start breaking pressure.

Spigot design; i.e., expansion ratio.

We fired factorial experiments at two levels for these variables (eight rounds for each program). For the three programs (A, B, and C) which were replicated three times, we fired a total of 8 x 3 x 3 or 72 rounds, a reduction by a factor of 10 in the number of rounds required.

The discussion is confined to the C program, as this amply illustrates the important points and the other programs are similar.

The statistical method used is found in Kempthorne*. The data taken were peak chamber pressures, peak acceleration, and the muzzle velocities of the projectiles. In addition, several other ballistic parameters, such as piezometric efficiency and ballistic efficiency, were examined.

Each result was treated separately, in the manner outlined in Kempthorne* to obtain the effects of each variable and the interactions between variables. The significance of these values was ascertained by the use of the standard error and "t" test at both the 5 and 1 percent levels. The values of the variables investigated in this program are shown in Table I and test results obtained are shown in Table II.

The results of the factorial analysis are presented in Tables III, IV and V showing the effects and interactions of the variables on peak pressure, peak acceleration, and muzzle velocity, respectively.

Each letter in the table is used to represent the average effect of the corresponding parameter. For example P = 2490 psi in Table III represents the difference between the average peak pressure of all rounds fired with a closed spigot (closed spigot indicates large expansion ratio, consequently this was considered the upper level of this parameter) and all rounds fired with an open spigot. Two capital letters written together (W_tP for example) represent the interactions of the two corresponding parameters. Using data from Table III, $W_tP = -995$ psi, that is, $\frac{500-2490}{2}$.

The interpretation of effects and interactions is as follows: The main

*Kempthorne, Design and Analysis of Experiments.

Table I. Variables Used — "C" Program

Constants: Propellant - M2SP (9201)
Web - 0.010 in.
Bore Dia - 2.25 in.

Variables	I		II		III		IV		V		VI	
Round number	1	2	3	4	5	6	13	14	15	16	17	18
Design shot-start breaking pressure, P_{ss} (psi)	none	800	800	none	none	none	800	800	none	800	800	800
Spigot	open	open	open	closed	open	closed	closed	open	closed	open	closed	open
Projectile weight, W_t (lb)	60	20	20	20	20	60	20	60	60	20	60	20
Round number	7	8	9	10	11	12	19	20	21	22	23	24
Design shot-start breaking pressure, P_{ss} (psi)	800	none	800	none	800	800	none	none	800	none	none	none
Spigot	closed	closed	closed	open	open	closed	closed	open	closed	open	closed	open
Projectile weight, W_t (lb)	60	20	60	60	60	20	60	20	20	20	20	60

Table II. "C" Program Test Results

Results	I		II		III		IV		V		VI	
Round number	1	2	3	4	5	6	13	14	15	16	17	18
Peak pressure (psi)	1630	1810	1600	3450	1580	4350	3600	1710	4350	1740	4500	1800
Peak acceleration (g)	108	316	314	686	269	328	753	113	283	115	301	314
Final velocity (fps)	72.8	130.4	131.2	103.3	133.5	93.8	164.8	70.9	93.8	71.8	93.6	133.4
Recoil impulse (lb-sec)	0.28	0.22	0.24	0.06	0.20	0.31	0.24	0.57	0.15	1.45	0.10	0.21
Round number	7	8	9	10	11	12	19	20	21	22	23	24
Peak pressure (psi)	4600	3500	4400	1630	1740	3800	4350	1400	3750	1440	3500	1470
Peak acceleration (g)	305	608	297	108	115	755	266	393	748	301	654	98
Final velocity (fps)	93.0	164.8	94.5	72.3	70.9	166.8	82.3	131.2	165.8	131.9	164.5	65.8
Recoil impulse (lb-sec)	0.07	0.12	0.15	0.24	0.81	0.10	0.13	0.15	0.46	0.10	0.35	0.12

Velocity for C-24 was obtained from P-T curve, using a planimeter.

Best Available Copy

Table III Effects and Interactions of the Variables on the Mean Peak Pressure, "C" Program*

	Average Effects	Spigot: Open	Spigot: Closed	Projectile Wt: [illegible] lb	Projectile Wt: [illegible] lb	Shot-Start: [illegible] psi	Shot-Start: [illegible]
Spigot (P)	2450			$P + PW_t$ -2700	$P - PW_t$ -2100	$P + P_{ss}P$ -2250	$P - P_{ss}P$ -2425
Projectile weight (W_t)	500	$W_t + W_tP$ 200	$W_t - W_tP$ 810			$W_t + W_tP_{ss}$ 480	$W_t - W_tP_{ss}$ 520
Shot-start (P_{ss})	190	$P_{ss} + P_{ss}P$ 125	$P_{ss} - P_{ss}P$ 250	$P_{ss} + P_{ss}W_t$ 170	$P_{ss} - P_{ss}W_t$ 200		
Standard error	18						

Mean peak pressure = 2000 psi
5% significant level = 39 psi
1% significant level = 50 psi

*All values expressed in psi

Table IV. Effects and Interactions of the Variables on the Mean Peak Acceleration, "C" Program*

	Average Effects	Spigot: Open	Spigot: Closed	Projectile Wt: [illegible] lb	Projectile Wt: [illegible] lb	Shot-Start: [illegible] psi	Shot-Start: [illegible]
Spigot (P)	-300			$P + PW_t$ -185	$P - PW_t$ -420	$P + P_{ss}P$ -310	$P - P_{ss}P$ -[illegible]
Projectile weight (W_t)	-310	$W_t + W_tP$ -195	$W_t - W_tP$ -430			$W_t + W_tP_{ss}$ -325	$W_t - W_tP_{ss}$ -[illegible]
Shot-start (P_{ss})	-25	$P_{ss} + W_tP$ 16.8	$P_{ss} - P_{ss}P$ 33.5	$P_{ss} + P_{ss}W_t$ 11.2	$P_{ss} - P_{ss}W_t$ 38.7		
Standard error	2.8						

Mean peak acceleration = 300 g
5% significant level = 5.4 g
1% significant level = 7.8 g

*All values expressed in g's

Table V. Effects and Interactions of the Variables on the Mean Muzzle Velocity, "C" Program*

	Average Effects	Spigot: Open	Spigot: Closed	Projectile Wt: [illegible] lb	Projectile Wt: [illegible] lb	Shot-Start: [illegible] psi	Shot-Start: [illegible]
Spigot (P)	-28			$P + PW_t$ -22	$P - PW_t$ -30	$P + P_{ss}P$ -27	$P - P_{ss}P$ [illegible]
Projectile weight (W_t)	-62	$W_t + W_tP$ -58	$W_t + W_tP$ -65			$W_t + W_tP_{ss}$ -62.5	$W_t - W_tP_{ss}$ [illegible]
Shot-start (P_{ss})	-0.06	$P_{ss} + P_{ss}P$ -0.83	$P_{ss} - P_{ss}P$ 0.95	$P_{ss} + P_{ss}W_t$ -0.08	$P_{ss} - P_{ss}W_t$ 0.20		
Standard error	1.1						

Mean muzzle velocity = 108 fps
5% significant level = 2.4 fps
1% significant level = 3.4 fps

*All values expressed in fps

effect P, for example, is the effect on the variable (Pressure in Table III, Acceleration in IV, Velocity in V) of increasing expansion ratio (changing from closed spigot to open spigot) averaged over all possible combinations of projectile weight and shot-start values. It is desired now to determine the effect of expansion ratio averaged over all shot-start values but at the low projectile weight. This is denoted symbolically $P - PW_t$.

In Table III, for example, $P - PW_t = -2180$ psi indicates that using data for 20 pound projectile weight only and averaging over all shot-start values the peak pressure is reduced 2180 psi in changing from large expansion ratio (closed spigot) to small expansion ratio (open spigot). For data from the 60 pound projectile weight and all shot-start values (symbolically $P + PW_t$) we have -2790 psi. The fact that $P - PW_t$ differs from $P + PW_t$ indicates an interaction between projectile weight and expansion ratio.

The results in Table III show that $W_t + W_tP = 200$ psi and $W_t - W_tP = 810$ psi. Therefore, the projectile weight effect when the open spigot is used is 200 psi. When used with the closed spigot, the projectile weight effect is 810 psi. The difference value of 610 psi (2790 psi - 2180 psi and 810 psi - 200 psi) is the interaction effect between expansion ratio and projectile weight.

For a pictorial representation of the results, the variables are laid out as the axis of a transparent cube. The corners of the cube represent the eight combinations of variables fired. The yields (velocity, acceleration, and peak chamber pressure) are assumed to vary along the edges of the cube according to the predictions of ballistic theory. Thus, the yields at the corners are interpolated to obtain planes of constant response. (Ideally an analog computer analysis to calculate the planes exactly is desirable, as was done for the thruster previously discussed.) The planes indicated in Photo 3 represent peak pressure: 2800 psi; velocity: 108 fps; and peak acceleration: 360 g's. Points within the transparent cube above the red surface (designated P) represent variables which result in pressures below 2800 psi. Similarly, points in front of the V surface (green) are below 108 fps, and behind the G surface, are less than 360 g's. Thus, the three surfaces enclose a polygon of triangular cross section which is the zone of suitable response.

Combinations of variables near the surface of the zone may result in unsuitable performance as a result of round-to-round variations. Analysis of variance from the results of the factorial analysis and interpolation of the variance along the cube edge, using the same technique as in interpolating the yields, allows us to ascribe a thickness to the response surfaces. To illustrate this the zone of suitable performance has been removed from the cube in Photo 4. The zone of suitable performance now appears as three boards nailed together. The hollow space is known as the zone of acceptable variables.

Performance confidence requirements, reliability requirements, and the experimental data determine the thickness of the surfaces. Only one way of applying this method is illustrated. The response surface of finite thickness would be used to construct the zones in different ways for different performance requirements. Suppose the velocity were required to be 108 ± 5 fps instead of simply greater than 108 fps, still keeping the pressure and acceleration requirements as before. Then the zone of suitable response would be represented by the green board marked V in Photo 4. The zone of acceptable variables would be represented by a surface running along the board bisecting the thickness. There is an extremely wide variety of requirements that can be treated with this technique. No unusual or exotic statistical mathematics is required.

<u>CONCLUSIONS AND RECOMMENDATIONS</u>. Our general conclusion is that the use of factorial type experimental design programs represents a definite advantage to the ballistic designer. These advantages are measured in terms of a larger number of variables investigated for fewer rounds (time and money economy). In addition, interaction effects among the variables are determined. Adding the Box technique and pictorial representation to the use of factorial experiments in ballistic research gives the experimenter a more economical and vivid picture of how the variables operate. To this picture may be added the variances of each response. Thus a zone of suitable performance may be determined in which the greatest reliability of operation is obtained.

It is recommended that in the design of ballistic devices factorial experiments be conducted and combined with a "Box" representation of the results.

M-5 AIRCRAFT SEAT EJECTION CATAPULT

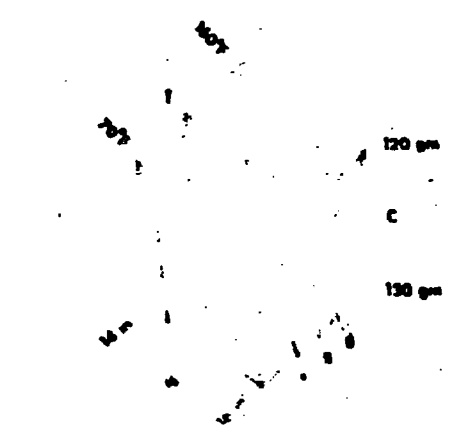


Photo I

THRUSTER

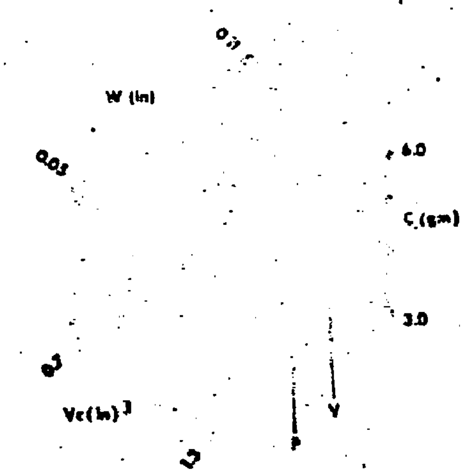


Photo II

REACTIONLESS LAUNCHER

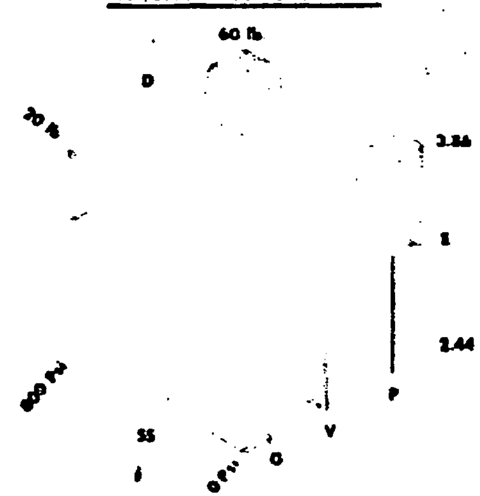


Photo III

REACTIONLESS LAUNCHER

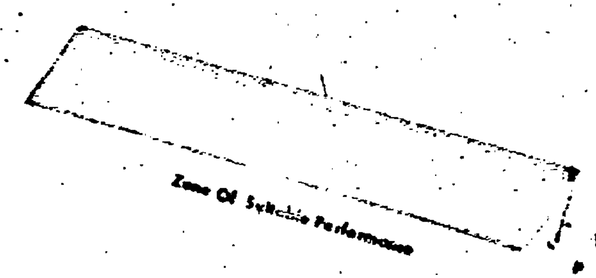


Photo IV

"BUILD-UP" OF SINGLE POINT SOURCE DATA

R. F. White
General Analysis Corporation*
Dugway Proving Ground Office

I. DIRECT AND INDIRECT BUILD-UP METHODS. Consider two types of vapor dissemination trials. In one type - the multiple round - a rocket containing a large number (about 300) agent-container bomblets is fired at a horizontal target. The bomblets are released at some point in the rocket's trajectory and, upon impact, release their contained vapor agent. The bomblet impact points are determined later and the dosage over the target area (and downwind of it) is determined by suitable samplers.

In the second type of trial - the single round - an amount of agent equal to the amount contained in one bomblet is released instantaneously from a point and the dosage is determined by suitable samplers over an area around the point and downwind of it. The "build-up" problem is to use dosage data obtained from a trial of the second type to estimate the dosage distribution to be obtained from a trial of the first type under identical meteorological and terrain conditions.

The major difficulty in this problem is that it is essentially impossible to test, with a high degree of rigor, any proposed solution. This is because identical meteorological conditions can never be obtained, even if all the relevant meteorological factors were known. Thus, if a given method of solution does fail to give build-up values reasonably similar to those actually obtained on a multiple round trial, the failure can be ascribed either to the method or to the non-identicality of meteorological conditions and it is not easy to say which is at fault. On the other hand, it is supposed, in the conduct of CW trials generally, that the relevant meteorological factors are being observed and that these do have a close determining effect on the results of a trial, for otherwise such trials would have no practical value, being impossible to extend to other situations. Hence, if a build-up method does not give similar results to a given multiple round trial, then the method can be said, at least within the framework of present knowledge, to have no practical value.

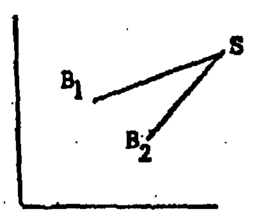
Furthermore, if similar results should be obtained, even in the face of these difficulties, then it is logical to suppose that such results are not simply due to chance, but are due to an inherent feasibility of the build-up

*This corporation is now called CEIR Inc.

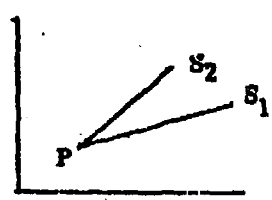
method. Therefore with some reservations, it may be assumed that a method of solution can be tested.

With this as a background, consider an ideal situation. There are two trials - a single round and a multiple round trial - with "identical" meteorological conditions and with dosage data for each. The impact point locations of functioning bomblets in the multiple round trial are given. ("Dosage" in this discussion refers to ground level total dosage.) Two general methods of building-up the single round data to the multiple round are suggested:

(1) Direct build-up. Consider a particular sampler in the multiple round trial. The position of this sampler relative to each bomblet may be approximately equated to the position of some sampler relative to the point source of the single round trial. The dosage received at the multiple round sampler from each bomblet is then estimated to be identical to the dosage received in the single round trial at the appropriate sampler. Thus in the following figures let B_1 and B_2 be bomblet locations and S a multiple round sampler position and let P be the single round point source and S_1 and S_2 sampler positions. The vectors B_1S and PS_1 are equal as are the vectors B_2S and PS_2. The dosage values observed at S_1



multiple round trial



single round trial

and S_2 are then estimated to be, respectively, the dosage contributions to S of B_1 and B_2. This process is extended to $B_1, B_2, \ldots, B_n$ where n is the number of functioning bomblets and the total of these estimated dosage contributions is then taken as the estimated dosage at S.

(2) Indirect build-up. A functional form is fitted to the single round data and this function is used to estimate the dosage at any point in the multiple round trial. For example suppose, from the single round data, a function $D(x,y)$ is fitted which gives the value of the dosage

at downwind distance x and crosswind distance y from the point source. Then if the bomblets on the multiple round trial are located at (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), the built-up dosage at (x, y) is simply

$$\mathcal{D}(x,y) = \sum_{i=1}^{n} D(x-x_i, y-y_i) \tag{1}$$

Further, with a large number of bomblets it is feasible to assume a bomblet distribution density $f(x, y)$ where

$$\iint_R f(x,y)\, dx\, dy = 1$$

and R is the bomblet impact region. Then instead of (1) we may take

$$\dot{\mathcal{D}}(x,y) = n\iint_R f(u,v)\, D(x-u, y-v)\, du\, dv \tag{2}$$

The advantage of equation (2) over equation (1) is that the use of equation (2) will usually not require specifying n bomblet coordinates. For example, if $f(u,v)$ is a bivariate normal density (perhaps truncated) then equation (2) is specified by the one, two, or at most three, parameters of $f(u,v)$. Since these parameters are relatively constant for a given ballistic situation, equation (2) will have much wider predictive ability than will equation (1). Specifically, it can be used to predict, prior to the trial, the results of a multiple round trial provided ballistic information on bomblet impact pattern distribution is obtained. As a matter of fact, build-up methods may turn out to have, as their primary function, usefulness in guiding the conduct of multiple round trials, rather than in replacing them.

II. <u>COMPARISON OF THE METHODS.</u> There are several reasons which suggest that the direct method of build-up or some modification of it will fail to give useful results:

(1) <u>Immeasurable single-round dosages</u>. An obvious difference between single and multiple round dosage results is that the apparent area of cloud travel is considerably smaller for the single than it is for the multiple round situation. This is caused by the fact that at the single round cloud edge dosages are so small as to be immeasurable by the analytic procedures employed. Such small individual bomblet contributions become

measurable and important in the multiple round situation, however. Thus, in terms of the above figures, the dosages observed at S_1 or S_2 could be zero and this would lead to a definite under-estimation of the dosage at S. To a certain extent, this disadvantage is shared by the indirect method but not to as great an extent, since the estimate of $D(x,y)$ uses data from "measurable" areas.

(2) Single round samplers unavailable at desired locations. Since the locations of $B_1, B_2, \ldots, B_n$ are random, it is necessary to take $S_1, S_2, \ldots, S_n$ to make the vectors $B_1S, \ldots, B_nS$ as only approximately equal, respectively, to $PS_1, \ldots, PS_n$. Unless the single round sampling grid is very dense, these approximations will be quite rough. Further, the area of primary interest in the multiple round situation has been taken to be at considerable downwind distance from the impact pattern. In this area the dosage contributions of individual bomblets is small and it is in this corresponding area in the single round trials that the sampling density is low. In particular the downwind distances between successive single round sampling arcs is large. The dense sampling array found close to the release point of single round trials does not contribute much to the total accuracy of build-up for large downwind distances. Conceivably, the direct method could be greatly improved by having dense sampling at large downwind distances in the single round trials. This would not help to answer reason (1), unless the analytic procedure were made more sensitive. In any case, the indirect method does not face this disadvantage at all since $D(x,y)$ is defined at every (x,y).

(3) Sampler variability and cloud heterogeneity. The direct method develops estimated dosages by local build-ups. Heterogeneity is a local phenomenon and can be ameliorated only by a statistical "smoothing" process. This means, in effect, that a process such as the indirect method, which uses all the data to estimate each point is better than a process which estimates each point by an individual observation.

(4) Non-predictability of results. An advantage of the indirect method is that equation (2) can be used to replace equation (1) and, as has been discussed, lead to predictions independent of knowledge of bomblet locations. This is not feasible with the direct method, unless some complex analog (such as assuming a bomblet pattern) were used.

In summary, it can be said that the direct method of build-up is inherently incapable of giving good estimates of multiple round dosages and leads to less useful (in being non-predictable) results.

III. APPLICATION OF THE INDIRECT METHOD. The basis of the indirect method is the Calder-Sutton instantaneous point source model:

$$D(x,y) = \exp\left[a - by^2/x^{\alpha} - c \ln(x)\right] \qquad (3)$$

where

$D(x,y)$ = ground-level total dosage at downwind distance x and crosswind distance y and a, b, c, α are parameters.

The overall procedure of the indirect method is as follows.

(1) "Fit" the model to the data of the single round trial. This amounts to an estimation (say by least squares) of the parameters a, b, c, α.

(2) Apply equation (1) or alternatively equation (2) (if a bomblet density function $f(u,v)$ can be assumed).

The problem of fitting the model is lengthy and will be discussed after the second problem - application of equation (1) or (2) - is considered. Therefore to start this discussion, let us assume that the function $D(x,y)$ has been estimated. Note that equation (3) is defined only for $x > 0$. (This means that the model assumes no upwind dosage, an assumption which is not strictly true, but which can be considered as having a compensating error due to the fact that upwind dosages will be ignored in both the single round and the multiple round situation.) Hence in applying equation (1) we must take:

$$\mathcal{D}(x,y) = \sum_{x_i < x} D(x - x_i,\ y - y_i) \qquad (4)$$

as the estimated built-up dosage at (x,y). Nothing more can be said about this. We do not recommend use of this procedure, inasmuch as it is tedious (although quite suitable for solution on a high-speed computer) and does not lead to "predictability". We therefore proceed to the more interesting question of application of equation (2).

A. Normally Distributed Bomblet Pattern.

Suppose we assume that f(u,v) is a bivariate normal density (distributed around the impact pattern center):

$$f(u,v) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{u^2}{\sigma_x^2} + \frac{v^2}{\sigma_y^2} - \frac{2\rho uv}{\sigma_x\sigma_y}\right]\right\} \quad (5)$$

Then we take, for equation (2),

$$\mathcal{D}(x,y) = n\int_{-\infty}^{x-\delta}\int_{-\infty}^{\infty} f(u,v)\ D(x-u,y-v)dv\ . \quad (6)$$

Taking the crosswind limits of integration as $-\infty$ and $+\infty$ is a simplifying approximation and should not make much difference in the results since the tails of the normal distribution rapidly become small. Note that the downwind upper limit of integration is $x-\delta$. This is because the integrand is discontinuous at $u=x$ in that

$$\lim_{u\to x-} D(x-u,y-v) = \infty;\quad v=y$$
$$= 0;\quad v\neq y$$

The problem is to choose a reasonably small δ.

If we consider the bomblets upwind of x, it is clear that the integration should be limited to the most downwind of these bomblets. In fact it is reasonable, in making practical use of equation (6), to integrate up to the expected downwind coordinate of this most downwind bomblet. This concept is difficult to explain briefly without the following mathematical development.

Let x be any downwind axis coordinate and let n(x) be the number of bomblets upwind of x (i.e., the number of bomblets whose downwind axis coordinates are to the left of x.) Let u(x) be the largest of these coordinates (i.e., u(x) is the largest downwind coordinate of those bomblets which are upwind of x). Now u(x) is a random variable and it seems reasonable to take equation (6) as

$$\mathcal{D}(x,y) = n \int_{-\infty}^{\bar{u}(x)} du \int_{-\infty}^{\infty} f(u,v) D(x-u, y-v) dv \tag{7}$$

where $\bar{u}(x) = Eu(x)$ is the expected value of $u(x)$, for a given x. Consider now the evaluation of $\bar{u}(x)$.

First, the marginal probability distribution function of bomblet downwind axis coordinates is

$$F(x) = \int_{-\infty}^{x} du \int_{-\infty}^{\infty} f(u,v) dv$$

and so the probability of any given value of $n(x)$ is given by the binomial distribution as

$$p(n(x)) = \binom{n}{n(x)} \left[F(x)\right]^{n(x)} \left[1-F(x)\right]^{n-n(x)}; \quad n(x) = 0, 1, 2, \ldots, n.$$

Now $u(x)$ is defined only when $n(x) \geq 1$ the probability of which is

$$1 - \left[1-F(x)\right]^{n}.$$

Hence the conditional probability of any given value of $n(x)$ under the condition that $n(x) \geq 1$ is

$$p^{*}(n(x)) = \frac{\binom{n}{n(x)} \left[F(x)\right]^{n(x)} \left[1-F(x)\right]^{n-n(x)}}{1 - \left[1-F(x)\right]^{n}} \tag{8}$$

$$n(x) = 1, 2, \ldots, n.$$

The conditional probability distribution function of $u(x)$, for a given value of $n(x)$, is

$$G(u(x) \mid n(x)) = \left[\frac{F(u(x))}{F(x)}\right]^{n(x)}; \quad u(x) \leq x \tag{9}$$

Hence the marginal probability distribution of $u(x)$ is

$$G(u(x)) = \sum_{n(x)=1}^{n} p^*(n(x))G\left[u(x)\,\middle|\,n(x)\right]$$

$$= 1 - \frac{1-\left\{1+F\left[u(x)\right]-F(x)\right\}^n}{1-\left[1-F(x)\right]^n} \quad ; \quad u(x) \leq x. \tag{10}$$

Therefore

$$\bar{u}(x) = Eu(x) = \int_{-\infty}^{x} u\, dG(u). \tag{11}$$

Unfortunately, this integral is not simple and can be evaluated only by tedious numerical methods. It seems reasonable therfore to approximate $\bar{u}(x)$ by $\hat{u}(x)$ where

$$F\left[\hat{u}(x)\right] = E\,F\left[u(x)\right]$$

$$= \int_{-\infty}^{x} F(u)dG(u) = \frac{n}{1-\left[1-F(x)\right]^n}\int_{-\infty}^{x} F(u)\left[1+F(u)-F(x)\right]^{n-1} dF(u)$$

$$= \frac{F(x) - \dfrac{1-\left[1-F(x)\right]^{n+1}}{n+1}}{1-\left[1-F(x)\right]^n}$$

$$= F(x) - \frac{1-\left[1-F(x)\right]^{n+1}}{n+1}$$

$$+ \frac{\left[1-F(x)\right]^n}{1-\left[1-F(x)\right]^n}\left\{F(x) - \frac{1-\left[1-F(x)\right]^{n+1}}{n+1}\right\} \tag{12}$$

Thus we take $\hat{u}(x)$ as the solution to

$$F\left[\hat{u}(x)\right] = \frac{F(x) - \dfrac{1-\left[1-F(x)\right]^{n+1}}{n+1}}{1-\left[1-F(x)\right]^n} \tag{13}$$

which, since the integral $F(x)$ is tabled, is not difficult.

For cases where $F(x) > 1/2$ (i.e., x is downwind of the bomblet downwind axis mean) and reasonably large n, equation (12) is very well approximated by

$$E\,F\left[u(x)\right] = F(x) - \frac{1}{n+1}$$

and instead of (13), take

$$F\left[\hat{u}(x)\right] = F(x) - \frac{1}{n+1} \qquad (13a)$$

As previously defined,

$$F(x) = \int_{-\infty}^{x} du \int_{-\infty}^{\infty} f(u,v)\,dv$$

which, from equation (5),

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\sigma_x} \exp\left\{-u^2/2\right\} du = N\left(\frac{x}{\sigma_x}\right), \text{ say.} \qquad (14)$$

The integral $N(\frac{x}{\sigma_x})$, the normal probability integral, is well tabled. Thus, if equation (13a) is used, the value $\hat{u}(x)$ such that

$$N\left[\frac{\hat{u}(x)}{\sigma_x}\right] = N\left(\frac{x}{\sigma_x}\right) - \frac{1}{n+1}$$

is obtained for each x. In any case, values of $\hat{u}(x)$ are computed for corresponding values of x and, instead of equation (7), we take

$$\mathcal{D}(x,y) = n\int_{-\infty}^{\hat{u}(x)} du \int_{-\infty}^{\infty} f(\hat{u},v)\,D(x-u,y-v)\,dv \qquad (15)$$

Now, from (3) and (5),

$$\int_{-\infty}^{\infty} f(u,v)D(x-u,y-v)dv = \frac{\exp\left\{a - c\,\ln(x-u) - \dfrac{u^2}{2(1-\rho^2)\,\sigma_x^2}\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_x^2}}$$

$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{v^2}{\sigma_y^2} - \frac{2\rho uv}{\sigma_x\sigma_y}\right] - \frac{b(y-v)^2}{(x-u)^{\alpha}}\right\}dv$$

$$= \frac{k\exp\left\{-\dfrac{u^2}{2(1-\rho^2)\sigma_x^2} - \dfrac{by^2}{(x-u)^{\alpha}} + \dfrac{\left[\dfrac{\rho u}{\text{[illegible]}} + \dfrac{2yb}{\text{[illegible]}}\right]^2}{2h}\right\}}{(x-u)^c\sqrt{h}}$$

$$= g(u,x,y), \text{ say} \tag{16}$$

where

$$k = \frac{e^a}{\sigma_x\sigma_y\sqrt{2\pi(1-\rho^2)}}$$

and

$$h = \frac{1}{(1-\rho^2)\sigma_y^2} + \frac{2b}{(x-u)^{\alpha}}$$

so that

$$\mathcal{D}(x,y) = n\int_{-\infty}^{\hat{u}(x)} g(u,x,y)du \tag{17}$$

The problem then reduces to evaluating equation (17) for a "grid" of points (x,y). The integral is not simple and requires tedious numerical methods which will not be explored here. However, a reasonable approximation can be given briefly. The function $g(u,x,y)$ can be factored

$$g(u,x,y) = g_1(u,x,y)g_2(u,x,y)$$

where

$$g_1(u,x,y) = \exp\left\{-\frac{u^2}{2(1-\rho^2)\sigma_x^2}\right\}$$

$$g_2(u,x,y) = \frac{k \exp\left\{-\frac{by^2}{(x-u)\alpha}\right\} + \frac{\left[\frac{\rho u}{\sigma_x^2(1-\rho^2)} + \frac{2vb}{(x-u)\alpha}\right]^2}{2h}}{(x-u)^c \cdot \sqrt{h}}$$

then, by the mean value theorem, there exists u^* $(-\infty < u^* < u(x))$ such that

$$\mathcal{P}(x,y) = n g_2(u^*,x,y) \int_{-\infty}^{\hat{u}(x)} g_1(u,x,y)\,du \tag{18}$$

where, approximately,

$$u^* = \frac{\int_{-\infty}^{\hat{u}(x)} u g_1(u,x,y)\,du}{\int_{-\infty}^{\hat{u}(x)} g_1(u,x,y)\,du} \tag{19}$$

Little further error is introduced by replacing $\hat{u}(x)$ by x in equation (19), and this makes the calculation of $\hat{u}(x)$ unnecessary. Thus,

$$\int_{-\infty}^{x} g_1(u,x,y)\,du = \sigma_x\sqrt{2\pi(1-\rho^2)}\; N\left[\frac{x}{\sigma_x\sqrt{1-\rho^2}}\right] \tag{20}$$

where, as before, N is the normal probability integral,

$$N\left[\frac{x}{\sigma_x\sqrt{1-\rho^2}}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x/\sigma_x\sqrt{1-\rho^2}} e^{-t^2/2}\,dt$$

Also,

$$\int_{-\infty}^{x} u g_1(u,x,y)\,du = -\sigma_x^2(1-\rho^2)\exp\left\{-\frac{x^2}{2\sigma_x^2(1-\rho^2)}\right\} \tag{21}$$

Hence, taking

$$\mathcal{D}(x,y) = n g_2(u^*,x,y) \int_{-\infty}^{x} g_1(u;x,y)\,du$$

gives

$$\mathcal{D}(x,y) = \frac{n \exp\left\{a - \frac{by^2}{(x-u^*)^{\alpha}} + \frac{\left[\frac{\rho u^*}{\sigma_x \sigma_y} + \frac{2yb}{(x-u^*)^{\alpha}}\right]^2}{2h^*}\right\}}{\sigma_y (x-u^*)^c \sqrt{h^*}} \times$$

$$N\left[\frac{x}{\sigma_x \sqrt{1-\rho^2}}\right] \tag{22}$$

where

$$u^* = -\frac{\sigma_x \sqrt{1-\rho^2}}{\sqrt{2\pi}\; N\left[\frac{x}{\sigma_x \sqrt{1-\rho^2}}\right]} \exp\left\{-\frac{x^2}{2\sigma_x^2 (1-\rho^2)}\right\}$$

and

$$h^* = \frac{1}{(1-\rho^2)\sigma_y^2} + \frac{2b}{(x-u^*)^{\alpha}}$$

If a circular impact pattern can be assumed then

$\rho = 0$

and $\sigma_x = \sigma_y$ (= σ, say);

so that

$$\mathcal{D}(x,y) = \frac{n \exp\left\{a - \frac{by^2}{(x-u^*)^{\alpha} + 2b\sigma^2}\right\} N(\frac{x}{\sigma})}{(x-u^*)^{c-\alpha/2} \sqrt{(x-u^*)^{\alpha} + 2b\sigma^2}} \tag{23}$$

where

$$u^* = \frac{-\sigma}{\sqrt{2\pi}\; N(\frac{x}{\sigma})} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \tag{24}$$

If a circular impact pattern is accepted as the most practical case, equation (23) is the most useful build-up equation so far developed in this paper. It is interesting in that it may also be used as a model for multiple round data. The values of x and y are respectively the downwind and crosswind distances from the impact pattern center.

B. Use of Equation (23) as a Multiple Round Model.

It is not entirely out of place at this time to discuss equation (23) as a multiple round model, having five parameters: a, b, c, α, σ. If the bomblet positions $(x_1, y_1), \ldots, (x_n, y_n)$ are measured from the pattern center then a simple estimate of the parameter σ is

$$\hat{\sigma} = \sqrt{\frac{\sum_i (x_i^2 + y_i^2)}{2n}} \tag{25}$$

Now consider certain functions derived from equation (23). First, for a given downwind sampling row, we have the "crosswind maximum dosage":

$$\mathrm{CWMD}(x) = \mathcal{D}(x,0) = \frac{n\, e^a\, N\left(\frac{x}{\sigma}\right)}{(x-u^*)^{c-\alpha/2}\sqrt{(x-u^*)^{\alpha} + 2b\sigma^2}} \tag{26}$$

and the "crosswind integrated dosage":

$$\mathrm{CWID}(x) = \int_{-\infty}^{\infty} \mathcal{D}(x,y)\,dy = \frac{n\, e^a\, N\left(\frac{x}{\sigma}\right)\sqrt{\frac{\pi}{b}}}{(x-u^*)^{c-\alpha/2}} . \tag{27}$$

(Each of these quantities is obviously observable experimentally, for each downwind sampling row.) The estimate of makes $N(x/\sigma)$ observable as is (see equation (24)) u^*. Then various functions of (26) and (27) may be plotted against $(x-u^*)$ to obtain estimates of $b, c,$ and α. For example, consider

$$z(x) = \left(1/\pi\right)\left(\frac{\mathrm{CWID}(x)}{\mathrm{CWMD}(x)}\right)^2 - 2\sigma^2 = \left(1/b\right)(x-u^*)^{\alpha} \tag{28}$$

so that if $z(x)$ is plotted against $x-u^*$ on log log paper a straight line should result with intercept = log (1/b) and slope = α. The success of such techniques will be very much dependent on the reliability of the dosage data. Nothing further will be said on this subject in this paper, inasmuch as the present problem is not the estimation of parameters from multiple round data, but is rather the "building-up" of single round data.

C. General Impact Pattern Distributions.

Suppose that at the time of release of bomblets, the rocket is at position (x_o, y_o, z_o) and suppose a given bomblet has initial velocity vector (v_x, v_y, v_z). Then with the effects of wind ignored, the velocity vector of the bomblet at time t after release is

$$(v_x, v_y, v_z - gt)$$

where g is the gravity acceleration constant. The position of the bomblet at time t is

$$(x_o, y_o, z_o) + \int_o^t (v_x, v_y, v_z - gt)dt =$$

$$(x_o + v_x t, y_o + v_y t, z_o + v_z t - gt^2/2) \tag{29}$$

The time of ground impact is such that the vertical coordinate is zero. That is, the impact time is such that

$$z_o + v_z t - gt^2/2 = 0$$

the positive root of which is

$$t_1 = \frac{v_z + \sqrt{v_z^2 + 2gz_o}}{g} . \tag{30}$$

The ground position of the bomblet is therefore

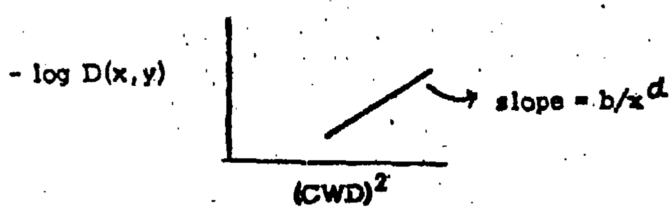
$$\left.\begin{aligned} x_1 &= x_o + v_x t_1 \\ y_1 &= y_o + v_y t_1 \end{aligned}\right\} . \tag{31}$$

Conceivably, a bomblet pattern distribution can be deduced from simple considerations of equations (30) and (31) and of the distribution of x_o, y_o, z_o, v_x, v_y, and v_z. For example, if x_o, y_o, and z_o are considered non-random, and the effect of v_z is small relative to the effect of gravity, and if v_x and v_y are bivariate normally distributed then x_1 and y_1 are bivariate normally distributed. We are continuing our exploration of this problem.

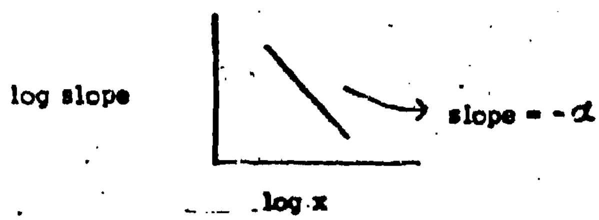
IV. "FITTING" THE SINGLE ROUND DATA. Consider now the problem where total dosage data has been obtained over a sampling grid in a trial of the second type - the single round. If a wind direction can be assumed and if the sampling grid is such that it contains rows perpendicular to the wind direction, the model fitting is greatly simplified. Consider the dosage model

$$D(x,y) = \exp(a - by^2/x^{\alpha} - c \ \ln x)$$

shows that if, for a constant x, the negative of the log dosage is plotted against y^2, the square of the crosswind distance, a straight line with slope = b/x^{α} should result:



If this slope is estimated for each crosswind row and if the set of slopes is plotted against x on log paper, a straight line with slope = $-\alpha$ should result:



By this means an estimate of α can be obtained.

After α is estimated the estimation of a, b, and c is quite simple inasmuch as ordinary least squares techniques can be applied to

$$z(x,y) = a - b\ell - cm$$

where

$$z(x,y) = \ell n\, D(x,y)$$

$$\ell = y^2/x^{\hat{\alpha}}$$

$$m = \ell n\, x$$

Thus the least squares estimates are

$$\hat{a} = \bar{z} + \hat{b}\bar{\ell} + \hat{c}\bar{m}$$

$$\hat{b} = -(s_{\ell z} + \hat{c}s_{\ell m})/s_{\ell\ell}$$

$$\hat{c} = (s_{\ell z}s_{\ell m} - s_{\ell\ell}s_{mz})/s_{\ell\ell}s_{mm} - s^2_{\ell m} \qquad (32)$$

where $\bar{z}$, $\bar{\ell}$, $\bar{m}$ are the means, respectively of the z's, ℓ's, and m's and

$$s_{\ell z} = \sum(\ell - \bar{\ell})(z - \bar{z})$$

$$s_{\ell m} = \sum(\ell - \bar{\ell})(m - \bar{m})$$

$$s_{\ell\ell} = \sum(\ell - \bar{\ell})^2\text{, etc.}$$

There is one difficulty with application of the model to such data. The model describes diffusion due to wind currents. "Close" to the release point another mechanism - the munition blast - becomes more important in cloud travel. Therefore, not all the data is suitable for use in equations (32). Some judgment about this may possibly be obtained by observing the plot of log slope against log x indicated above. For small values of x, the plot may show erratic departures from linearity and this is an indication that data from the corresponding crosswind rows is unsuitable for fitting.

Other observable functions can be used for aiding in the fit. For example the crosswind maximum dosage:

$$\mathrm{CWMD}(x) = D(x,0) = e^{a}/x^{c}$$

and the crosswind integrated dosage:

$$\mathrm{CWID}(x) = \int_{-\infty}^{\infty} D(x,y)\,dy = \frac{e^{a}\sqrt{\pi/2}}{b x^{c-\alpha}}$$

so that

$$\frac{\mathrm{CWID}(x)}{\mathrm{CWMD}(x)} = \frac{x^{\alpha}\sqrt{\pi/2}}{b} \qquad (33)$$

A plot of this latter quantity against x on log paper at once gives an estimate of α.

The above estimation procedures depended on having crosswind sampling rows perpendicular to the wind direction. In fact, this will seldom be the case. The problem is partially avoided by having circular sampling arcs, but this still leaves some difficulty. An approximation suggests that a circular arc with large radius will exhibit similar dosage results to a straight line perpendicular to the wind.

Finally, establishment of a wind direction is not as simple as it sounds. For the purpose of fitting, it seems better to fit a line through the maximum dosages on the crosswind sampling arcs and call this the "virtual wind line", rather than to rely on a wind track obtained by meteorological observations.

The subject of fitting is lengthy and cannot be effectively discussed without an actual example. It is hoped that what has been said will serve as an adequate introduction to the problem.

Panel Discussion

COMMON PITFALLS IN THE DESIGN AND ANALYSIS OF EXPERIMENTS

Chairman: G. E. P. Box, The University of Wisconsin
Panel Members: Cuthbert Daniel,* Private Consultant
J. S. Hunter, Mathematics Research Center, The University of Wisconsin
W. J. Youden, National Bureau of Standards
Marvin Zelen, The University of Maryland

Prior to the start of the conference, each Panel Member sent to Dr. Box a brief on the Common Pitfalls in the Design and Analysis of Experiments which he planned to discuss at the Aberdeen meeting. We publish here in outline form these briefs.

Cuthbert Daniel

1. In design, the mistake I make most often, is to start planning experiments with insufficient understanding of the substantive problem.

2. In analysis, the commonest mistake I make is to assume prematurely that I know what went on.

3. After these two are out of the way -- or forgotten -- the commonest defect in data is the presence of a very small number of very bad values. The pitfall is to fail to notice these bad values. By a bad value I mean one whose observed or recorded magnitude controls or dominates the interpretation of the whole set of data.

4. In balanced, and especially in factorial, experimentation, defective randomization, usually in the direction of plot-splitting is the commonest error of experimenters and its presence undetected is then the common pitfall in trying to interpret the data.

J. S. Hunter

1. Considerable arithmetical dexterity is required to perform the analysis of variance associated with many experimental designs and their associated

*Mr. Daniel was unable to attend the meeting. He telephoned his comments to Dr. F. E. Grubbs. These were read by Professor G. E. P. Box.

mathematical models. Unfortunately, the experimenter frequently considers his data analyzed once this arithmetic is completed. Methods of data analysis used with great profit long before the invention of the ANOVA are thus neglected (graphs, histograms, effects of changes in scale, the search for abberant observations, trends, etc.). After completing the regular ANOVA experimenters also frequently fail to review the residuals for additional signals or to think in terms of alternative mathematical models that might be expressive of the data.

2. There exists an unjustifiable high regard for the ability of regression equations to unfold and identify information within a large accumulation of data. This faith in multiple regression techniques is particularly strong when the data have been haphazardly collected with little regard for either randomization or good experimental design.

W. J. Youden

My favorite "pitfall" is elementary, obvious and yet pretty treacherous. Let us suppose some ammunition stored at three temperatures and two humidities (6 combinations). Every six months samples are taken and fired and, let us say, shell velocity determined. We have a lovely trap for anyone who knows how to do an analysis of variance. Suppose duplicate firings.

Rel. Hum.	Temp. °C	Period stored - months				
		6	12	18	26	....
	0	*	*	*		
50	20		etc.			
	40					
	0					
90	20					
	40					

The error of the "duplicates" may not be applicable to comparing shots fired 6 months apart. The error for any one of the time curves may not apply for comparisons among the six storage conditions. After all, presumably only one storage chamber was available for each storage condition. I need not elaborate further. Overlooking the fact that the error of split plot comparisons is usually less than that for between plots is all too common.

Marvin Zelen

Often non-statistical pitfalls may invalidate an entire experiment or even cause incorrect conclusions to be made. No amount of good statistics will be able to rectify a non-statistical blunder. Three kinds of non-statistical pitfalls discussed are:

1. Analysis of data without really understanding how the experiment was executed may create an incorrect analysis.

2. When cooperative experiments are being carried out with groups not in complete contact with one another, the groups will often differ in their administration of treatments and evaluation of responses.

3. Extrapolation of data over a different range of experimental conditions.

SOME TESTS FOR OUTLIERS[1]

C. P. Quesenberry[2] and H. A. David
Virginia Polytechnic Institute

1. INTRODUCTION. This paper is concerned with the problem of detecting outlying observations when in addition to the normal sample $x_1, x_2, \ldots, x_n$ at hand an independent mean-square estimate s_ν^2 of the common variance σ^2 is available. The same situation has been considered by Nair (1948). To test for one outlier at a specified end of the sample he proposes using the ratio of the extreme deviate from the sample mean to s_ν. For two-sided testing the extreme absolute deviate from the sample mean divided by s_ν has been proposed by Halperin, et al. (1955).

The two statistics mentioned above do not make use of the variance estimate s^2 from the sample and for this reason do not possess certain desirable optimal properties. We shall propose statistics with the same numerators as those above but with s_ν in the denominators replaced by the pooled estimate

$$s^* = \left\{ \left[(n-1)s^2 + \nu s_\nu^2 \right] / (n+\nu-1) \right\}^{1/2}.$$

Kudo (1956) has shown that among a suitably restricted class of tests these statistics maximize the probability of rejecting the null hypothesis of homogeneity of the sample in the presence of a single outlier.

We shall develop a method for computing percentage points of these statistics and present the computed tables. These tables are also immediately applicable to the problem of slippage of means in normal samples. Two examples illustrate the procedures.

2. NOTATION AND DEFINITIONS. Let $x_1, x_2, \ldots, x_n$ denote the sample in the order drawn. Then define

---

[1]This research supported in part by a National Science Foundation Fellowship and in part by the Office of Ordnance Research, U. S. Army.

[2]Now at Montana State College.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \, , \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \, ,$$

s_ν^2 is an independent mean-square estimate of variance, σ^2, with ν degrees of freedom,

$$S^2 = (n-1)s^2 + \nu s_\nu^2 \, ,$$

$$b_i = (x_i - \bar{x}) / S \, , \qquad (2.1)$$

$$b = \max_i b_i = \frac{x_{max} - \bar{x}}{S} \, , \qquad (2.2)$$

$$b^* = \max_i |b_i| \qquad (2.3)$$

Then b and b* are essentially the one-sided and two-sided statistics, respectively, discussed in section 1. It should be noted that S^2 has not been divided by its degrees of freedom.

The special case $\nu = 0$ has been treated by Pearson and Chandra Sekar (1936), Grubbs (1950) and Borenius (1958).

3. DISTRIBUTION THEORY. For the work in later sections the distribution of b_i and the joint distribution of b_i and b_j are needed. We shall now obtain these distributions, taking for definiteness $i = 1$ and $j = 2$. An extremely complicated derivation of essentially the same distributions has been given by Doornbos, Kesten and Prins (1956) in an article concerned with slippage tests.

As is well known, $(n-1)s^2$ may be decomposed into two independent components

$$\frac{n}{n-1}(x_1 - \bar{x})^2 + \chi^2_{(n-2)}\,\sigma^2 \, ,$$

which are distributed respectively as $\chi^2 \sigma^2$ with 1 and n − 2 degrees of freedom. With the same notation we have, therefore

$$s^2 = \frac{n}{n-1}(x_1 - \bar{x})^2 + \chi^2_{(n+\nu-2)}\sigma^2 \qquad (3.1)$$

Then b_1 may be written as

$$b_1 = \left(\frac{n}{n-1} + \frac{\chi^2_{(n+\nu-2)}\sigma^2}{(x_1 - \bar{x})^2} \right)^{-1/2}$$

$$= \left(\frac{n-1}{n} \right)^{1/2} \left(1 + \frac{n+\nu-2}{t^2_{(n+\nu-2)}} \right)^{-1/2}$$

where $t_{(n+\nu-2)}$ denotes a t-variate with $n+\nu-2$ degrees of freedom. It follows that the density function of b_1 is

$$f(b_1) = \left(\frac{n}{n-1} \right)^{1/2} \cdot \frac{\Gamma[(n+\nu-1)/2]}{\sqrt{\pi}\,\Gamma[(n+\nu-2)/2]}$$

$$\times \left[1 - nb_1^2/(n-1) \right]^{1/2(n+\nu-4)}, \qquad (3.2)$$

$$-\left(\frac{n-1}{n} \right)^{1/2} \leq b_1 \leq \left(\frac{n-1}{n} \right)^{1/2}.$$

This generalization of a result due to Thompson (1935) has also recently been pointed out by Anscombe (1960).

Continuing the decomposition of (3.1) one step further we have

$$S^2 = \frac{n}{n-1}(x_1 - \bar{x})^2 + \frac{n-1}{n-2}(x_2 - \bar{x}')^2 + \chi^2_{(n+\nu-3)}\sigma^2$$

with

$$\bar{x}' = \frac{1}{n-1}\sum_{i=2}^{n} x_i \quad .$$

Let $b'_2 = (x_2 - \bar{x}')/S'$,

where $S'^2 = S^2 - \frac{n}{n-1}(x_1 - \bar{x})^2$.

Clearly, the distribution of $\underline{b'_2}$ is of the form (3.2) with $\underline{n}$ replaced by n - 1. Moreover, $\underline{b'_2}$ is independent of $\underline{b_1}$. To see this suppose $\underline{s^2_\nu}$ is based on a random sample of size $\nu + 1$, with mean $\underline{\bar{x}_1}$, taken from a $N(\mu_1, \sigma^2)$ parent. This assumption is unnecessarily restrictive but does not essentially affect the argument. Then $\underline{\bar{x}}$, $\underline{\bar{x}_1}$ and $\underline{S^2}$ are complete sufficient statistics for μ, μ_1 and σ^2. Since the distribution of $\underline{b_2'}$ does not involve these parameters, b_2' is independent of the joint distribution of $\underline{\bar{x}}$, $\underline{\bar{x}_1}$ and $\underline{S}$ (Basu, 1955). Also $\underline{b'_2}$ does not involve $\underline{x_1}$ so that it must be independent of $\underline{b_1}$.

The joint density function of $\underline{b_1}$ and $\underline{b'_2}$ is therefore

$$f(b_1, b_2') = \left(\frac{n}{n-2}\right)^{1/2}\left(\frac{n+\nu-3}{2\pi}\right)\left(1 - \frac{n}{n-1}b_1^2\right)^{(1/2)(n+\nu-4)}$$

$$\times\left(1 - \frac{n-1}{n-2}\,b_2'^{\,2}\right)^{(n+\nu-5)/2} ,$$

$$-\left(\frac{n-1}{n}\right)^{1/2} \leq b_1 \leq \left(\frac{n-1}{n}\right)^{1/2},$$

$$-\left(\frac{n-2}{n-1}\right)^{1/2} \leq b_2' \leq \left(\frac{n-2}{n-1}\right)^{1/2}.$$

Since $b_2' = \dfrac{b_2 + b_1/(n-1)}{\left[1 - n b_1^2/(n-1)\right]^{1/2}}$

we obtain

$$f(b_1, b_2) = \left(\frac{n}{n-2}\right)^{1/2} \frac{(n+\nu-3)}{2\pi}$$

$$\times \left(1 - \frac{n-1}{n-2} b_1^2 - \frac{2b_1 b_2}{n-2} - \frac{n-1}{n-2} b_2^2\right)^{(1/2)(n+\nu-5)} \tag{3.3}$$

over the ellipse

$$\frac{n-1}{n-2} b_1^2 - \frac{2b_1 b_2}{n-2} + \frac{n-1}{n-2} b_2^2 \leq 1,$$

and

$f(b_1, b_2) = 0$, elsewhere.

Moments of b and b*

Since the distributions of b and b* do not involve μ, μ_1, and σ^2 it follows as above that b and b* are distributed independently of S. This result is well known for the special case $\nu = 0$ and may indeed be proved in a similar fashion, as was pointed out to us by Dr. G. E. P. Box. As a consequence of this independence the moments (about zero) of b and b* are the ratios of the moments of their respective numerators and denominators. Thus we have, for example,

$$\mathcal{E}(b^r) = \frac{\mathcal{E}(x_{max} - \bar{x})^r}{\mathcal{E}\, S^r}$$

In this case (but not so readily for b*) the right hand side can be evaluated numerically since the cumulants of $x_{max} - \bar{x}$ are related to the tabulated cumulants of the extreme (Ruben, 1954) by equations which for $\mu = 0, \sigma = 1$ become (McKay, 1935)

$$K_r(x_{max} - \bar{x}) = K_r(x_{max}) \qquad r = 1, 3, 4, 5, \ldots$$

$$K_2(x_{max} - \bar{x}) = K_2(x_{max}) - 1/n.$$

The distribution of b may therefore be approximated by a Pearson Type curve. However, for the purpose of obtaining upper percentage points the approach of the following section, applicable to both b and b*, is preferable.

4. THE COMPUTATIONAL PROCEDURES.

(a) One-Sided Case, b

We now consider a procedure for computing significance points of b defined by (2.2). For a given value of n and ν let D_α be the required α significance point of b. By Bonferroni's inequalities (cf. David, 1956) we have for any D

$$nPr(b_i > D) - \binom{n}{2} Pr(b_i > D,\ b_j > D) \leq Pr(b > D) \leq n\ Pr(b_i > D). \tag{4.1}$$

For sufficiently large D the right side serves as a first approximation to $Pr(b > D)$ and the left side as a second approximation. If the first approximation is set equal to α, then the resulting equation, i.e.

$$Pr\ (b_i > D) = \alpha/n$$

can be solved for a value D_1, which is an upper bound of the value D_α sought. From (3.2) this equation can be written as

$$\alpha/n = C_1 \int_{D_1}^{\left(\frac{n-1}{n}\right)^{1/2}} \left[1 - \frac{n}{n-1}\, b_1^2\right]^{(1/2)(n+\nu-4)} db_1 ,$$

where $C_1 = \left[\frac{n}{\pi(n-1)}\right]^{1/2} \frac{\Gamma\left[(n+\nu-1)/2\right]}{\Gamma\left[(n+\nu-2)/2\right]}$.

The following equivalent equation is more convenient to work with

$$B_1 = \int_0^{D_1} \left[1 - \frac{n}{n-1}\, x^2\right]^{(n+\nu-4)/2} dx$$

$$= D_1 + \sum_{r=1}^{\infty} \frac{(-1)^r \left[n+\nu-4\right]\left[n+\nu-6\right]\ldots\left[n+\nu-2(r+1)\right] n^r D_1^{2r+1}}{r!\cdot 2^r \cdot (n-1)^r \cdot (2r+1)}, \tag{4.2}$$

where

$$B_1 = \frac{1/2 - \alpha/n}{C_1} .$$

By transposing the second term on the right to the left of (4.2) the equation can be identified with

$$D_{1,i-1} = h\left(D_{1,i-1}\right) ,$$

so that Newton's iterative formula

$$D_{1,i} = D_{1,i-1} - \frac{h\left(D_{1,i-1}\right)}{h'\left(D_{1,i-1}\right)}$$

can be used to solve for D_1. Note also

$$h'(D_1) = \left(1 - \frac{n}{n-1} D_1^{2}\right)^{(1/2)(n+\nu-4)} .$$

The initial value of D_1 used to start the iteration procedure was $D_{1,0} = B_1$, for B_1 as given above.

While D_1 is an upper bound for D_α, a lower bound by (4.1) satisfies

$$n\Pr(b_i > D_2) - \binom{n}{2} P_r(b_i > D_2, b_j > D_2) = \alpha .$$

A first approximation $D_{2,0}$ to D_2 is, therefore, given by

$$n\Pr(b_i > D_{2,0}) = \alpha + \binom{n}{2} \Pr(b_i > D_1, b_j > D_1). \qquad (4.3)$$

On replacing D_1 in (4.3) by $D_{2,0}$ a second approximation $D_{2,1}$ is obtained. The process can be continued until $D_{2,t+1}$ and $D_{2,t}$ agree to three decimal places. In the present case $D_{2,0}$ was found to be sufficiently accurate in all but a few cases.

The second term on the right side of equation (4.3) is evaluated by numerical integration. The joint density $f(b_1, b_2)$ given by (3.3) is integrated over the region for which $b_1 > D_1$ and $b_2 > D_1$. This region is shaded in Fig. 1. This numerical integration was performed on an I.B.M. 650 computer. The numerical method used is equivalent to fitting an increasing number of planes to the density surface until the desired accuracy is achieved.

An examination of Fig. 1 shows that if $D_1 > \left[(n-2)/2n\right]^{1/2}$ then the second term on the right side of (4.3) is zero. Then $d_1 = D_2 = D_\alpha$ is the exact percentage point of b. This is important in that it allows the exact calculation of a number of percentage points for lower values of n and γ.

The lower and upper bounds for the percentage points were found to agree so well for the values of α considered here (.01, .05) that only one value had to be tabulated. Tables 1 and 2 give the 1 and 5 per cent points, respectively, for selected values of γ and n.

(b) The Two-Sided Case, b*

Essentially the same procedure is used to obtain the significance points of b* as for b. The Bonferroni inequalities in this case give

$$n\Pr(|b_1|>D) - \binom{n}{2} \Pr(|b_1|>D, |b_j|>D) \leq \Pr(b^*>D) \leq n\Pr(|b_1|>D). \qquad (4.4)$$

From the symmetry of $f(b_1)$ we have

$$\Pr(|b_1|>D) = 2\Pr(b_1 > D).$$

Let D^*_α be the desired significance point of b*. Then an upper bound

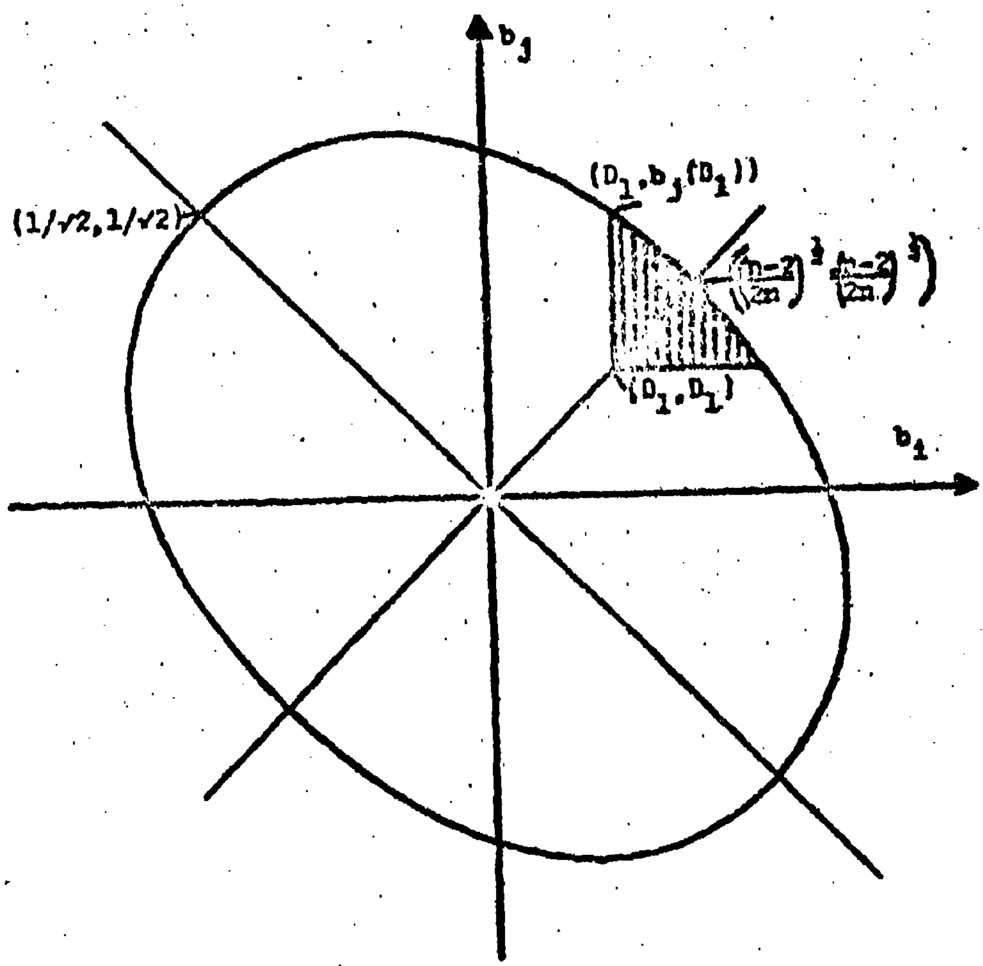


Figure 1. Region over which joint density is integrated for the one-sided case

D_1^* can be obtained from (4.2) by replacing α by $\alpha/2$ in B_1 and solving. A first approximation $D^*_{2,0}$ to a lower bound D^*_2 on D^*_α is given by

$$n\Pr(b_1 > D^*_{2,0}) = \frac{1}{2}\left[\alpha + \binom{n}{2} \Pr(|b_i| > D^*_1, |b_j| > D^*_1)\right]. \quad (4.5)$$

A second approximation $D^*_{3,0}$ can be obtained by replacing D^*_1 by $D^*_{2,0}$ in (4.5), etc. The second term on the right of (4.5) is evaluated this time by integrating $f(b_1, b_2)$ over the area in each quadrant where $|b_i| > D^*_1$ and $|b_j| > D^*_1$. The bounds on D^*_α do not agree as well as for the one-sided case. Tables 3 and 4 give bounds for D^*_α for $\alpha = .01$ and $\alpha = .05$, respectively. When the bounds agree to three places only one value is tabulated.

5. THE SLIPPAGE PROBLEM. The statistics b and b^* are useful in treating the slippage problem for normal populations. Here we have the sample

$$\left.\begin{array}{lll} x_{11} & x_{12} \cdots x_{1n_1} \\ x_{22} & x_{22} \cdots x_{2n_2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_{k1} & x_{k2} \cdots x_{kn_k} \end{array}\right\} \quad (5.1)$$

We wish to test the hypothesis that the entire array is from a common normal parent against the alternative that the i th sample $(x_{i1}, x_{i2}, \ldots, x_{in_i})$ is from a normal parent with a different mean, where i is unspecified.

$$\text{Let } \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

and s_t^2 be an independent mean-square estimate of error with t degrees of freedom.

An important special case is that of all equal subsample sizes, i.e. $n_1 = n_2 = \ldots = n_2 = m$. For this special case the statistics

$$\max_i \frac{\sqrt{m}\,(\bar{x}_i - \bar{x})}{S_1}, \tag{5.2}$$

and

$$\max_i \frac{\sqrt{m}\,|\bar{x}_i - \bar{x}|}{S_1}, \tag{5.3}$$

where

$$S_1^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} (x_{ij} - \bar{x})^2 + t\, s_t^2 ,$$

are distributed as $\underline{b}$ and $\underline{b}^*$, respectively. The significance points of these statistics are obtained from the tables of $\underline{b}$ and $\underline{b}^*$ with the parameters $\underline{n}$ and $\underline{\nu}$ of the tables replaced by $n = k$ and $\nu = k(m-1) + t$. These slippage tests possess the same desirable properties as do outlier tests based on $\underline{b}$ and $\underline{b}^*$ (see Paulson, 1952).

If the sample sizes n_i are not all equal but are approximately so the tables of b and b* can be used to obtain approximate tests.
Put

$$\left.\begin{aligned} \bar{x}_w &= \frac{1}{k}\sum_{i=1}^{k} \sqrt{n_i}\,\bar{x}_i \,. \\ &\text{and} \\ N &= \sum_{i=1}^{k} n_i \,. \end{aligned}\right\} \tag{5.4}$$

Then the statistics

$$\max_i \frac{\sqrt{n_i}\;\bar{x}_i - \bar{x}_w}{s_i} \tag{5.5}$$

and

$$\max_i \frac{\left|\sqrt{n_i}\;\bar{x}_i - \bar{x}_w\right|}{s_i} \tag{5.6}$$

give approximate tests based on the tables of b and b*. The parameters n and ν of the tables are here $n = k$ and $\nu = N + t - k$.

6. EXAMPLES. We now give two examples to illustrate the use of the tables.

Example 1.

Squibs are small devices for igniting the rocket motors of missiles. Watertightness and shock resistance are important characteristics of squibs. In order to study these characteristics of a large batch a random sample of size 48 was drawn. The sample was randomly subdivided into 3 equal groups. The first group was used as a control unit and received no treatment, the second group was submerged in water and the third group was dropped from a fixed height. Each squib in the entire sample was tested by having a current of 5 amperes passed through it and its time to failure recorded.

From previous experience it is felt that these delay times are approximately normally distributed. It is also known from previous experience that occasionally extremely large delay times occur. Because of this an outliers test was used on each subgroup to guard against such spurious observations. The variance was assumed to be constant throughout the experiment. The data are given in Table 6.1.

First we test each of the subgroups for outlying observations. For each test the variance estimates from the two remaining subgroups are used as an independent estimate of error. For the control group we have from (2.1)

$$b = \frac{.76 - .4438}{\sqrt{.8698}} = .3390 .$$

Table 2 gives the 5 per cent point for $n = 16$ and $\nu = 30$ as approximately $b(.05; 16, 30) = .384$, so the above value does not attain significance.

For the watertightness group

$$b = \frac{1.09 - .5188}{.93274} = .6124$$

and this is compared with the same value, $b(.05; 16, 30) = .384$, as before. This is highly significant and the observation 1.09 is rejected.

Table 6.1

Control (x_{11})	Watertightness (x_{21})	Shock (x_{31})
.38	.53	.51
.26	.35	.63
.41	.38	.46
.33	.45	.47
.33	1.09	.42
.37	.46	.45
.54	.57	.41
.76	.47	.39
.51	.39	.35
.55	.74	.41
.53	.32	.49
.41	.74	.40
.47	.48	.58
.49	.37	.46
.42	.52	.38
.34	.44	.48

(Data furnished by Ordnance Missile Laboratories, ARGMA, AOMC, Redstone Arsenal, Alabama.)

$\sum x_{11} = 7.10$	$\sum x_{21} = 8.30$	$\sum x_{31} = 7.29$
$\bar{x}_1 = .4438$	$\bar{x}_2 = .5188$	$\bar{x}_3 = .4556$
$\sum x_{11}^2 = 3.3686$	$\sum x_{21}^2 = 4.8768$	$\sum x_{31}^2 = 3.4021$
$\frac{\left(\sum x_{11}\right)^2}{16} = 3.1506$	$\frac{\left(\sum x_{21}\right)^2}{16} = 4.3056$	$\frac{\left(\sum x_{31}\right)^2}{16} = 3.3215$
$SS(1) = .2180$	$SS(2) = .5712$	$SS(3) = .0806$

Now, since the 1.09 observation in the 2nd group was unduly inflating the variance estimate for the first test, we shall recompute that test with this observation omitted from the variance estimate. The test statistic becomes

$$b = \frac{.76 - .4438}{\sqrt{.5217}} = .4378 ,$$

and is to be compared with b (.05; 16, 29). Table 2 gives b (.05; 15, 24) = .410 and increasing either n or ν tends to decrease the percentage point, so the above statistic is significant. The observation .76 is omitted from the control group. The sum of squares based on the remaining 15 observations is .1113.

For testing the 3rd (shock) group b = .2707. This is not significant at the .05 level. Further tests in the subsamples lead to no more discarded observations.

The purpose of the experiment is to test the significance of the water and shock treatments. We are interested in testing the hypothesis that either one or both treatments increased the mean delay time. A two-sided test is appropriate here. For the two-sided test will have a probability of rejection higher than for the null situation under any alternative except when one treatment effect is exactly twice the other (both non-zero).

Since the subsample sizes are large and nearly equal, the approximate test discussed in section 5 should give accurate results. Here $n_1 = n_2 = 15$ and $n_3 = 16$. The weighted means are $\sqrt{n_1}\,\bar{x}_1 = 1.637$, $\sqrt{n_2}\,\bar{x}_2 = 1.862$, $\sqrt{n_3}\,\bar{x}_3 = 1.822$ and $\bar{x}_w = 1.774$. Then (5.6) gives

$$\frac{1.774 - 1.637}{.6442} = 0.213 .$$

This is to be compared with b (.05; 3, 43). Table 4 gives b (.05; 3, 40) $\cong$ 0.283 so the value 0.213 is not significant at the .05 level. We conclude that the treatments had no effect.

Example 2.

A sample of six observations was drawn from a table of random normal numbers and a randomly selected observation was increased by two standard deviations. The observations obtained were 265, 223, 291, 105, 43 and 477. A sample of six observations was drawn from a table with the same variance but with a different mean to give an independent estimate of variance. These observations were 171, 111, 185, 271, 68 and 217. The mean and sum of squares about the mean for the first set of observations are 234 and 116,504 respectively. The sum of squares about the mean for the second set is 26,519. So the one sided test statistic is from (2.1)

$$b = \frac{477 - 234}{\sqrt{143,023}} = .643.$$

Table 2 gives $b(.05;\ 6,\ 5) = .638$, so that the observation 477 is rejected at the .05 level. Table 4 gives $b^*(.05;\ 6,\ 5) = .681$, so that the observation 477 is not rejected at the .05 level by the two-sided test.

REFERENCES

(1) Anscombe, F. J. (1960). Rejection of outliers. Technometrics, 2, 123-147.

(2) Basu, D. (1955). On statistics independent of a complete sufficient statistic. Sankhya, 15, 377-380.

(3) Borenius, G. (1958). On the distribution of the extreme values in a sample from a normal population. Skandinavisk Aktuarietidskrift, 3, 131-166.

(4) David, H. A. (1956). On the applications to statistics of an elementary theorem in probability. Biometrika, 43, 85 - 91.

(5) Doornbos, R., Kesten, H. and Prins, H. J. (1956). A class of slippage tests. Report S 206 (VP8), Mathematisch Centrum, Amsterdam.

(6) Grubbs, F. E. (1950). Sample criteria for testing outlying observations. Ann. Math. Statistics, 21, 27 - 58.

(7) Halperin, M., Greenhouse, S., Cornfield, J. and Zalokar, Julia (1955). Tables of percentage points for the Studentized maximum absolute deviate in normal samples. J. Amer. Statist. Assn. 50, 185 - 195.

(8) Kudo, A. (1956). On the testing of outlying observations. Sankhya, 17, 67 - 76.

(9) McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. Biometrika, 27, 466 - 472.

(10) Nair, K. R. (1948). The distribution of the extreme deviate from the sample mean and its Studentized form. Biometrika, 35, 118 - 144.

(11) Paulson, E. (1952). An optimum solution to the k-sample slippage problem for the normal distribution. Ann. Math. Stat., 23, 610 - 616.

(12) Pearson, E. S. and Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika, 28, 308 - 320.

(13) Ruben, H. (1954). On the moments of order statistics in samples from normal populations. Biometrika, 41, 200 - 227.

(14) Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. Ann. Math. Stat., 6, 214 - 219.

APPENDIX

THE TABLES

Tables 1 and 2 give the 1 and 5 per cent points, respectively, of b [see (2.1)]. The 1 per cent points are correct to 1 unit in the fourth place. Only a few values for n and ν large are questionable at all in the last digit. The five per cent points of b are correct to three places except for a few large values of n and ν. For n = 20 and $\nu = 50$ the value given may be as much as 2 units large in the third place. No other values in the 5 per cent table are incorrect by more than one unit in the third place.

Tables 3 and 4 give lower and upper bounds for the percentage points of b^* [see (2.2)]. For each combination of parameter values lower and upper bounds are given except when these bounds agree to three decimal places and then only one value is given.

The upper 1% points of $(x_{max} - \bar{x})/S$ for

$$S^2 = \sum_{1}^{n} (x_i - \bar{x})^2 + \sum_{1}^{\nu+1} (y_i - \bar{y})^2 \text{, and } y_i \text{ independent of } x_j.$$

Table 1

The upper 5% points of $(x_{max} - \bar{x})/S$ for

$$S^2 = \sum_{1}^{n} (x_i - \bar{x})^2 + \sum_{1}^{\nu+1} (y_i - \bar{y})^2 \text{, and } y_i \text{ independent of } x_j.$$

Table 2

The upper 1% points of $\max|x_i - \bar{x}|/S$ for

$$S^2 = \sum_{1}^{n} (x_i - \bar{x})^2 + \sum_{1}^{\nu+1} (y_i - \bar{y})^2 \text{, and } y_i \text{ independent of } x_j.$$

Table 3

The upper 5 % points of $\max |x_i - \bar{x}|/S$ for

$$S^2 = \sum_{1}^{n} (x_i - \bar{x})^2 + \sum_{1}^{\nu+1} (y_i - \bar{y})^2,$$ and y_i independent of x_j.

Table 4

Table 1

Table of 1% points of the distribution of b

ν \ n	3	4	5	6	7	8
0	0.8165	0.8617	0.8739	0.8695	0.8566	0.8394
1	0.8111	0.8431	0.8478	0.8400	0.8263	0.8104
2	0.7904	0.8155	0.8176	0.8094	0.7971	0.7833
3	0.7614	0.7844	0.7865	0.7800	0.7698	0.7579
4	0.7299	0.7532	0.7570	0.7527	0.7444	0.7341
5	0.6990	0.7238	0.7297	0.7274	0.7207	0.7120
6	0.6703	0.6968	0.7045	0.7037	0.6987	0.6918
7	0.6442	0.6720	0.6812	0.6819	0.6786	0.6729
8	0.6204	0.6491	0.6597	0.6620	0.6599	0.6554
9	0.5986	0.6282	0.6401	0.6436	0.6424	0.6389
10	0.5788	0.6091	0.6219	0.6263	0.6263	0.6237
12	0.5441	0.5723	0.5895	0.5956	0.5971	0.5962
15	0.5017	0.5333	0.5489	0.5566	0.5600	0.5607
20	0.4482	0.4795	0.4962	0.5055	0.5106	0.5132
24	0.4158	0.4463	0.4633	0.4732	0.4792	0.4826
30	0.3779	0.4073	0.4240	0.4346	0.4412	0.4455
40	0.3328	0.3600	0.3763	0.3869	0.3940	0.3990
50	0.3006	0.3261	0.3416	0.3519	0.3591	0.3642

Table 1

(continued)

ν \ n	9	10	12	15	20
0	0.8211	0.8032	0.7687	0.7228	0.6614
1	0.7942	0.7780	0.7465	0.7048	0.6483
2	0.7688	0.7541	0.7260	0.6879	0.6356
3	0.7450	0.7320	0.7070	0.6723	0.6239
4	0.7229	0.7116	0.6890	0.6576	0.6127
5	0.7026	0.6926	0.6724	0.6438	0.6020
6	0.6837	0.6748	0.6548	0.6306	0.5917
7	0.6659	0.6581	0.6422	0.6182	0.5822
8	0.6495	0.6428	0.6286	0.6066	0.5727
9	0.6341	0.6284	0.6156	0.5956	0.5638
10	0.6198	0.6148	0.6031	0.5851	0.5556
12	0.5935	0.5899	0.5808	0.5654	0.5398
15	0.5597	0.5576	0.5513	0.5393	0.5183
20	0.5140	0.5136	0.5104	0.5031	0.4880
24	0.4844	0.4850	0.4837	0.4785	0.4668
30	0.4480	0.4496	0.4501	0.4479	0.4393
40	0.4023	0.4047	0.4071	0.4074	0.4038
50	0.3681	0.3711	0.3744	0.3766	0.3751

Table 2

Table of 5% points of the distribution of b

ν \ n	3	4	5	6	7	8
0	0.8154	0.844	0.836	0.815	0.791	0.768
1	0.789	0.800	0.789	0.771	0.752	0.733
2	0.741	0.752	0.745	0.732	0.717	0.701
3	0.692	0.707	0.705	0.697	0.686	0.673
4	0.648	0.668	0.671	0.666	0.658	0.648
5	0.610	0.634	0.640	0.638	0.633	0.625
6	0.577	0.604	0.613	0.614	0.610	0.604
7	0.549	0.578	0.589	0.591	0.590	0.586
8	0.524	0.554	0.567	0.571	0.571	0.569
9	0.502	0.533	0.547	0.553	0.554	0.553
10	0.483	0.515	0.530	0.536	0.539	0.538
12	0.450	0.482	0.499	0.507	0.511	0.512
15	0.411	0.443	0.461	0.471	0.476	0.479
20	0.363	0.395	0.413	0.425	0.431	0.436
24	0.335	0.366	0.384	0.396	0.403	0.408
30	0.303	0.332	0.350	0.362	0.370	0.375
40	0.266	0.292	0.309	0.321	0.329	0.335
50	0.239	0.264	0.280	0.291	0.299	0.305

Table 2

(continued)

v \ n	9	10	12	15	20
0	0.746	0.725	0.689	0.644	0.586
1	0.714	0.697	0.666	0.626	0.574
2	0.686	0.672	0.644	0.609	0.562
3	0.661	0.648	0.625	0.594	0.550
4	0.638	0.627	0.607	0.579	0.540
5	0.617	0.608	0.591	0.566	0.530
6	0.598	0.591	0.575	0.553	0.520
7	0.580	0.574	0.561	0.541	0.511
8	0.564	0.559	0.548	0.530	0.502
9	0.550	0.546	0.536	0.520	0.494
10	0.536	0.533	0.524	0.510	0.486
12	0.511	0.509	0.503	0.492	0.472
15	0.480	0.479	0.476	0.468	0.452
20	0.438	0.439	0.439	0.435	0.424
24	0.411	0.413	0.415	0.413	0.405
30	0.379	0.382	0.385	0.385	0.381
40	0.339	0.342	0.347	0.349	0.349
50	0.310	0.313	0.318	0.322	0.323

Table 3

Table of 1% points of the distribution of b*

ν \ n	3	4	5	6	7	8
0	0.8165 *	0.864 *	0.881 *	0.882 *	0.874 *	0.860 *
1	0.814 *	0.851 *	0.862 *	0.858 *	0.847 *	0.833 *
2	0.800 *	0.830 *	0.837 *	0.831 *	0.821 *	0.808 *
3	0.778 *	0.805 *	0.809 *	0.805 *	0.796 *	0.785 *
4	0.751 *	0.777 *	0.782 *	0.779 *	0.772 *	0.762 *
5	0.724 *	0.750 *	0.757 *	0.755 *	0.749 *	0.741 *
6	0.698	0.725 *	0.733 *	0.733 *	0.728 *	0.721 *
7	0.673	0.702	0.711	0.712	0.708 *	0.703
8	0.651	0.680	0.690	0.692	0.690	0.685
9	0.629	0.659	0.671	0.674	0.673	0.669
10	0.610 0.609	0.640	0.653	0.657	0.657	0.654
12	0.576 0.573	0.607	0.621	0.626	0.628	0.626
15	0.533 0.529	0.564	0.580	0.587	0.590	0.590
20	0.478 0.473	0.509 0.508	0.526 0.525	0.534	0.539	0.541
24	0.444 0.439	0.475 0.473	0.492 0.491	0.501	0.507	0.510
30	0.405 0.399	0.434 0.432	0.451 0.450	0.461 0.460	0.468 0.467	0.471 0.470
40	0.357 0.351	0.385 0.382	0.401 0.399	0.411 0.410	0.418 0.417	0.423 0.422
50	0.323 0.317	0.349 0.346	0.365 0.363	0.375 0.373	0.382 0.381	0.387 0.386

Table 3

(continued)

ν \ n	9	10	12	15	20
0	0.843 *	0.827 *	0.794 *	0.750 *	0.688
1	0.819 *	0.804 *	0.773 *	0.732 *	0.674
2	0.795 *	0.781 *	0.753 *	0.715 *	0.662
3	0.772 *	0.759 *	0.735 *	0.700	0.651
4	0.751 *	0.740 *	0.717 *	0.686	0.639
5	0.731 *	0.721 *	0.701	0.671	0.628
6	0.713 *	0.704	0.685	0.658	0.617
7	0.695	0.687	0.671	0.646	0.608
8	0.679	0.672	0.657	0.634	0.599
9	0.664	0.657	0.644	0.623	0.591
10	0.649	0.644	0.632	0.612	0.581
12	0.623	0.619	0.609	0.593	0.565
15	0.588	0.586	0.579	0.566	0.544
20	0.542	0.541	0.537	0.529	0.512
24	0.511	0.511	0.509	0.503	0.490
30	0.474 0.473	0.475 0.474	0.475 0.474	0.471	0.462
40	0.426 0.426	0.428 0.427	0.430 0.429	0.429	0.424
50	0.390 0.389	0.393 0.392	0.396 0.395	0.397 0.397	0.395 0.394

Table 4

Table of 5% points of the distribution of b*

ν \ n	3	4	5	6	7	8
0	0.8162 *	0.855 *	0.857 *	0.844 *	0.825 *	0.804 *
1	0.803 *	0.824 *	0.820 *	0.807 *	0.789 *	0.771 *
2	0.769 *	0.786 *	0.782 *	0.771 *	0.757 *	0.741 *
3	0.729 *	0.747 *	0.746 *	0.738 *	0.727 *	0.714 *
4	0.690	0.711 *	0.713 *	0.708 *	0.670	0.689
5	0.655 0.654	0.678	0.684	0.681	0.675	0.667
6	0.623 0.620	0.649	0.657	0.657	0.652	0.646
7	0.596 0.590	0.623 0.622	0.633	0.635	0.632	0.627
8	0.571 0.563	0.600 0.598	0.611 0.610	0.614	0.613	0.609
9	0.549 0.540	0.579 0.576	0.592 0.590	0.596 0.595	0.596 0.595	0.593
10	0.529 0.519	0.560 0.556	0.573 0.572	0.579 0.578	0.580 0.579	0.579 0.578
12	0.495 0.483	0.526 0.522	0.542 0.539	0.549 0.547	0.551 0.550	0.551 0.550
15	0.454 0.441	0.486 0.480	0.502 0.499	0.511 0.509	0.515 0.514	0.517 0.516
20	0.403 0.389	0.434 0.427	0.452 0.447	0.462 0.459	0.468 0.466	0.471 0.469
24	0.373 0.359	0.403 0.395	0.421 0.416	0.432 0.428	0.438 0.436	0.442 0.440
30	0.338 0.324	0.367 0.359	0.384 0.379	0.395 0.391	0.403 0.399	0.408 0.405
40	0.297 0.283	0.323 0.315	0.340 0.334	0.351 0.346	0.359 0.355	0.364 0.361
50	0.267 0.254	0.292 0.284	0.308 0.302	0.319 0.314	0.326 0.323	0.332 0.328

Table 4

(continued)

v \ n	9	10	12	15	20
0	0.783*	0.763*	0.727*	0.681	0.621
1	0.753*	0.736*	0.704	0.663	0.608
2	0.726*	0.711*	0.683	0.646	0.596
3	0.701	0.688	0.664	0.630	0.584
4	0.678	0.667	0.646	0.616	0.573
5	0.658	0.648	0.629	0.602	0.563
6	0.638	0.630	0.613	0.589	0.553
7	0.621	0.614	0.599	0.577	0.544
8	0.604	0.598	0.586	0.566	0.535
9	0.589	0.584	0.573	0.555	0.526
10	0.575	0.571	0.561	0.545	0.518
12	0.550 0.549	0.547 0.546	0.539 0.539	0.526 0.525	0.503 0.502
15	0.517 0.516	0.515 0.514	0.511 0.510	0.501 0.500	0.483 0.482
20	0.473 0.471	0.473 0.472	0.472 0.471	0.466 0.465	0.453 0.452
24	0.445 0.443	0.446 0.444	0.446 0.445	0.443 0.442	0.433 0.432
30	0.411 0.408	0.413 0.411	0.415 0.413	0.414 0.413	0.408 0.407
40	0.368 0.365	0.371 0.368	0.374 0.372	0.376 0.375	0.374 0.372
50	0.336 0.333	0.339 0.336	0.344 0.341	0.347 0.345	0.347 0.345

NOTE ON PRECISION OF GRADED VS. ALL-OR-NONE RESPONSE IN BIOASSAY

Francis Marion Wadley
U. S. Army Chemical Corps Biological Laboratories

For purposes of biological assay, two types of response are generally available. The graded response furnishes for each subject a measure of effect, while the "quantal" or all-or-none response provides for each subject merely the fact that it did or did not respond. In the latter case, in order to obtain an arithmetic figure for analysis, it is necessary to use a number of subjects and to record the proportion responding to a given stimulus. This is the procedure of dosage-mortality studies; a single test involves determination of proportions responding to several concentrations of the test material.

The graded response is attractive because it provides a definite measure of response for each subject, and because it has a simple relationship with basic regression analysis. Since the measure of extent of an effect is more precise than the mere statement that it passed or did not pass a certain point, a successful graded response obviously gives more information per subject than an all-or-none response. This added precision, with more efficiency in expensive experimentation, is of considerable importance to biologists.

It frequently is true that a good graded response is not available, or that the all-or-none response is the only one that will answer the question at issue. However, it is important to keep both possibilities in mind in the choice of an experimental program. To be of maximum usefulness, a graded response must show a consistent and strong relationship with the concentration of the material tested, and all subjects must respond to some extent.

A well-established method of treating all-or-none data is the log-probit method, which has considerable evidence of validity. It assumes a normal distribution of the logarithms of tolerance or susceptibility among individual subjects. Finney (1952a) states that the mean of this normal distribution is estimated by the log of the LD_{50} or $\underline{m}$; the standard deviation by the reciprocal of the probit regression coefficient, $\underline{1/b}$.

If it were possible, for the individual subjects in a probit test, to read individual log tolerances directly, the variance of log tolerances would be given by $1/b^2$; the variance of m, the mean, by $1/b^2n$.

(1) $$Vm = 1/b^2n$$

A successful graded response must obviously have a high correlation with log tolerance; and if the log tolerance would be read directly, it would constitute an excellent graded response with mean and variance as stated. For the probit solution, where x is log concentration and n.w is the probit weight, the variance of m is estimated as:

$$(2) \qquad Vm = 1/b^2 \left[1/\sum nw + (m - \bar{x})^2 / \sum nw\,(x - \bar{x})^2 \right]$$

With good choice of concentrations, the second term in brackets is often negligible when data are well balanced around 50 per cent.

The two expressions for variance may serve to give a preliminary comparison of precision in graded and all-or-none response. For the probit, (2) may be simplified to $1/\sum nwb^2$ to compare with $1/nb^2$. Since Finney's "Sn" or $\sum n$ is equivalent to the "n" of (1), the differing factor is w, the average weighting coefficient used in (2). This w may average 0.5 in a probit solution with well-spaced concentrations, but is more often a little lower. Thus the comparison of $1/nb^2$ with $1/nwb^2$ shows the graded response with a variance a little less than half the variance for the all-or-none response. It indicates that a good graded response may make about twice as efficient a use of subjects as a good quantal response. This relation was brought out by Gaddum (1933).

To attempt tests of this result with actual data, it is necessary to relate variances of estimates by the two methods. Hewlett and Plackett (1956) compare 1/b (all-or-none) with s/b_g (graded response), where b_g is regression of graded response on concentration and s is the standard deviation from regression. They tabulate about 50 values of each from the literature and show that 1/b and s/b_g have similar means and ranges. The quantity s/b_g is the basis of error calculations with graded response (Finney 1952b). Hewlett and Plackett cite these sets of data from vertebrate subjects; the sets, of course, were unpaired. It may be assumed that the responses were well adapted or else they would not have been published.

In this study, for a more direct comparison, several sets of graded response data were adapted to all-or-none study. A particular level of response was defined as "critical", and for each dose level the percentage of subjects reaching or failing to reach this "critical" level was determined. The critical value was defined so as to be near the mean and to provide a usable series of percentages. The variance of log LD_{50}, from probit analysis of these percentages, is compared with the variance of log concentration needed for the critical graded response, as defined by regression.

The first set of data was taken from an article by the writer (Wadley 1949, Table 4). Guinea pigs were the subjects, and diameter of the irritated area after tuberculin injection was the response. Three lots of tuberculin were tested at each of 3 concentrations at 10-fold intervals. Each concentration of each lot was injected into 4 guinea pigs. The 3 lots were quite similar in potency, and only the various concentrations produced significant differences in the level of response. Hence the 36 observations were grouped into 12 tuberculin reactions in each of 3 concentrations. The "critical" response level was defined as an irritated area with a diameter of 12 millimeters. In the 3 groups of animals this level was reached by 8.3 per cent, 66.7 per cent and 100 per cent for the low, medium and high concentrations, respectively, giving a basis for probit analysis.

The value found for s/b_g was 0.41, while $1/b$ was estimated at 0.36, a close correspondence. The variance of $\underline{m}$ was estimated as 0.008 for graded, 0.013 for all-or-none.

A second set of data is from Finney (1952b, Table 9.1) on weight gains of rats following vitamin doses, with 10 responses at each of 3 concentrations. The critical response indicated was 36 units, which gave a percentage series of 10, 70 and 80 per cent. Two more sets from Fort Detrick data involved the exposure of guinea pigs to toxic aerosols. Log of survival time in hours was the response. The critical response was taken as the mean log (about 2.00); one percentage series was 25, 38, 94; the other was 7, 12, 31, 69, 100. These were used in analysis as with the tuberculin data. Results are brought together in a table below.

Comparison of All-or-None with Graded Response

Response	Probit Results		Graded Response		Vm Ratio
	$1/b$	Vm	s/b_g	Vm	Probit/Graded
Tuberculin reaction	0.36	0.013	0.41	0.008	1.62
Weight gain	0.22	0.0031	0.18	0.0012	2.58
Log survival time	0.53	0.0297	0.60	0.0120	2.48
Log survival time	0.79	0.0188	0.69	0.0060	3.13

These somewhat artificial but valid comparisons are compatible with the idea that precision of graded response may be a little more than double that of a quantal response, and that $1/b$ and s/b_g tend to be close. Such values as 0.53 and 0.60, for instance are of the same order. The comparisons are undertaken only to give a rough check to the theory of relation of responses, and are not advised as a procedure for experimenters.

The agreement of closely corresponding responses, such as those in the table, does not indicate agreement of all responses. In the problem represented in the third row of the table, when percentage of deaths was used as an all-or-none response, $1/b$ was much above s/b_g. In the problem of the fourth row, $1/b$ for death as a response, was lower than s/b_g. The approximate agreement seems to occur when both are about equally well adapted.

At Fort Detrick some study has been given to use of graded response in the hope of a gain in precision over the often difficult all-or-none tests. Responses studied have included time to death, time to onset of symptoms and weight loss. In some cases the graded responses have shown some gain over quantal responses; in others they have given difficulty and have failed to compete with the all-or-none.

Since for equal precision $1/b^2\bar{w} = s^2/b^2_g$, and since $\bar{w}$ may be about 0.5 or a little less, $2/b^2$ may be equated to s^2/b^2_g. Solution of this equation should give the b required for approximate equivalence to a given s/b_g or vice versa. For example, in a recent test s/b_g was estimated at 0.76. Writing $2/b^2 = s^2/b^2_g$ (or 0.58), and solving, a value of $b = 1.86$ is indicated as competitive in precision. Perhaps 2.5 would be a better factor than 2 for this comparison since 0.4 is probably nearer the usual average weight than 0.5. This procedure has been helpful in the writer's work, and should prove of value to experimenters.

In making a choice between responses, the first criterion will be the adaptation of available responses. The equation just above may be of help. If graded and all-or-none response are equally well adapted, the graded response may be expected at least to double the precision of the all-or-none.

REFERENCES CITED

1. Finney, D. J.

1952. a. Probit Analysis, 2d ed., pp 318. Cambridge
1952. b. Statistical Method in Biological Assay, pp 661. New York

2. Gaddum, J. H.

1933. Methods of Biological Assay Depending on a Quantal Response. Spec. Rept. Ser. 183, Med. Res. Council (England)

3. Hewlett, P. S. and R. L. Plackett

1956. The Relation Between Quantal and Graded Response in Drugs. Biometrics 12:72 - 78

4. Wadley, F. M.

1949. The Use of Biometric Methods in Comparison of Acid-Fast Allergens. Amer. Review Tuberc., 60:131 - 139

A COMPARISON OF LABORATORY EVALUATION AND FIELD WEAR OF MILITARY FABRICS

William S. Cowie
Quartermaster Research and Engineering Command
Natick, Massachusetts

The Textile, Clothing & Footwear Division of the Quartermaster Research & Engineering Command is charged with the responsibility for development of new and improved textile fabrics for military garments and other textile end items such as tentage and personal equipage for members of the armed forces. This paper deals with certain problems which have arisen in the evaluation of new textiles for clothing items.

The normal pattern of development for a new textile fabric for military use is as follows:

1. ENGINEERING OF THE FABRIC. Engineering design of new textile fabrics is accomplished by textile technologists located in the Textile, Clothing & Footwear Division at Natick, Mass. Dependent upon the functional characteristics desired in the end item for which the fabric is intended, appropriate weights, weaves, yarn sizes, textures and finishes are decided upon and tentative specifications are prepared. In many instances the feasibility of fabrication of an experimental fabric is discussed with the representatives of the textile industry before these tentative specifications are finalized.

2. PROCUREMENT OF SAMPLE YARDAGE. The textile industry is invited to bid on contracts for the fabrication of sample yardages of experimental fabrics. Such contracts usually call for the production of from 500 to 1000 yards of the new fabric. It is customary for one of the textile technologists involved in each new textile design to visit the contractor's plant in order to observe the fabrication process and to discuss with the contractor any difficulties which may have arisen during manufacturing operations.

3. LABORATORY EVALUATION OF THE FABRIC. When fabrication of the experimental item has been completed, the cloth is shipped to the Textile Engineering Laboratory at Natick in order that a complete physical evaluation may be accomplished. The laboratory is a completely equipped textile testing facility containing all of the test instruments prescribed by the American Society for Testing Materials as well as research equipment as required.

New textile fabrics are checked for both constructional and physical requirements contained in the tentative specification. Amongst the former are weave, weight, yarn size, yarn count, texture (the number of warp and filling yarns per inch), thickness, type of fiber or fibers etc. Some of the important physical requirements are breaking strength, tearing strength, bursting strength, resistance to abrasion, porosity, elongation, water repellency and others. Of course the relative importance of these physical characteristics is dependent upon the end item use for which the new fabrics are candidates. In every instance new items are compared to the existing fabric which they may replace. For those characteristics which may be expressed in terms of numerical data (such as breaking strength for example), differences between the standard and experimental item are tested statistically by means of such standard techniques as the "t" and F tests, and the analysis of variance.

4. FIELD EVALUATION. Those experimental fabrics which show promise on the basis of laboratory evaluation are fabricated into garments (specifically trousers) at the clothing factory of Philadelphia Quartermaster Depot. These trousers are forwarded to the Field Evaluation Agency of the Quartermaster Research & Engineering Command which is located at Ft. Lee, Va. Here they are subjected to accelerated field wear on a specially designed fabric wear course (Figure 1 is placed at the end of this article). Standard trousers for comparison are worn over the course simultaneously. All garments are worn by military test subjects.

The fabric course is a quarter of a mile long and consists of thirty obstacles. The test subjects climb a stone embankment, crawl across a section of railroad track, and slide down a steep cobblestone incline. They crawl across a single log bridge, through concrete culverts, across terrain consisting of cinders, sand, gravel and boulders. They also crawl through trenches and across rough terrain.

Two traversals of the course constitute one cycle. After each cycle garments are laundered and a wear score is obtained based on visual examinations by trained military personnel who chart the scores. Frays, holes, tears and wear areas are all considered in computation of the total wear score. Depending upon the severity of each of these types of wear, a point value is assessed. At the completion of ten cycles, wear scores for both the experimental and standard garments are totalled. The results are compared statistically by the analysis of variance technique and a formal written report is prepared and submitted to Headquarters Quartermaster Research & Engineering Command.

5. USER TEST. If the results of field evaluation of an experimental fabric indicate significant improvement over the standard fabric, then consideration is given to fabrication of a large number of garments containing the new fabric. If the decision is made to do so, these garments are placed in the hands of troops located at various military installations. The choice of location depends upon the specific garment for which the new fabric is intended. For example, fabrics for cold weather clothing might be tested in Alaskan bases, and garments containing lightweight tropical fabrics might be worn in Panama. A report is prepared of the user's reaction to the new fabric.

6. STANDARDIZATION. When the user's reaction to new fabrics is favorable, action is taken to finalize the temporary requirements for the fabric which have been in effect up to this time. A final formalized specification is prepared over which large quantities of the fabric may be procured and the new fabric replaces the present standard. In the event that this new fabric is intended for an end item which is of interest to services other then the Army then the new specification is coordinated with these other services prior to promulgation.

It is realized that the above background material is very much of an oversimplification, however, the information is provided only to introduce the specific problem with which this paper is concerned.

As was noted earlier, only those new fabrics which show promise in the laboratory are evaluated in the field. Historically, good correlation has been obtained between the results of laboratory flex abrasion testing and wear scores obtained on the fabric wear course. Reasonably good correlation has also been obtained between tear resistance as determined in the laboratory and field evaluation results (Ref. 1). Finally, good correlation has been found between accelerated field wear on the fabric course and actual field wear on infantry training troops. Over the period of years between 1945 to 1958 the presence of good laboratory - field correlation was verified on numerous occasions. However, all of the studies conducted during this period dealt with all-cotton garments.

In recent years, due to the increased demands for durability and protection imposed by modern warfare concepts, interest in blends of cotton with synthetic fibers has increased. It is considered from the known physical properties of nylon for example, that a more durable utility garment could be developed from proper blending of nylon with cotton. As a result, a fabric was recently engineered from a cotton/nylon blend (approximately 70% cotton and 30% nylon). When tested in the laboratory samples of the new fabric showed from twice to six times as much resistance to flex abrasion on conventional laboratory equipment as did the all-cotton fabric which is currently used in the standard utility garment. Such differences based on past experiences would indicate that markedly superior resistance to wear would be demonstrated on the wear course. However, when the two fabrics were manufactured into garments and worn on the wear course in a two phase evaluation no significant difference was found between the wear scores of the two items. In fact in one phase of the test the all-cotton fabric appeared to be slightly more resistant to wear than did the cotton/nylon blend.

Both laboratory and fabric course test results were carefully re-evaluated. No testing afterfacts were uncovered which could in any way account for the findings. As a result the following actions were taken.

A senior Textile Technologist from the Natick laboratories visited Ft. Lee and personally "ran" the Combat Course. Following this experience the technologist returned to Natick and designed a new type of laboratory abrading instrument which in his opinion more nearly reproduced the type of wear encountered on the fabric course then did existing laboratory test equipment. The Sand Abrader is composed of a block of iron measuring at the bearing surface 2" × 3" to which a 1/2 inch thick wool felt is cemented. The fabric to be tested is clamped or sewed over the felt covered surface. By means of a pivoted arm the block of iron with fabric attached is pushed back and forth over a cement block at the rate of 88 strokes per minute. Sand that has passed through a #16 screen is constantly being dropped onto the cement block. The sand is sifted through a #30 screen before being used again. The pressure on the fabric used by the weight of the arm and iron block is 0.5 pound per square inch which is the pressure of a man's thigh when lying prone. (Most of the wear on the trousers in the Fabric Evaluation Course is on the front of the trousers between the knees and crotch).

Fifteen samples of the two fabrics included in the above fabric course wear studies have been abraded for 3000 cycles on this instrument. Visual examination of the abraded samples by a panel of three textile technologists revealed no less wear on the cotton/nylon items than on the all-cotton standard. Additionally, tear strength values were obtained on new samples of both materials and on the abraded items. Losses in tear strength following abrasion were almost identical in the filling direction of both fabrics and only slightly greater in the warp direction of the all-cotton fabric. Thus, these preliminary results show much better agreement with the fabric course than did the results of conventional laboratory testing.

These results are considered encouraging insofar as providing a possible means of improved correlation between laboratory and accelerated field wear. Samples of other new blends which are scheduled for fabric course evaluation in the near future are being similarly tested.

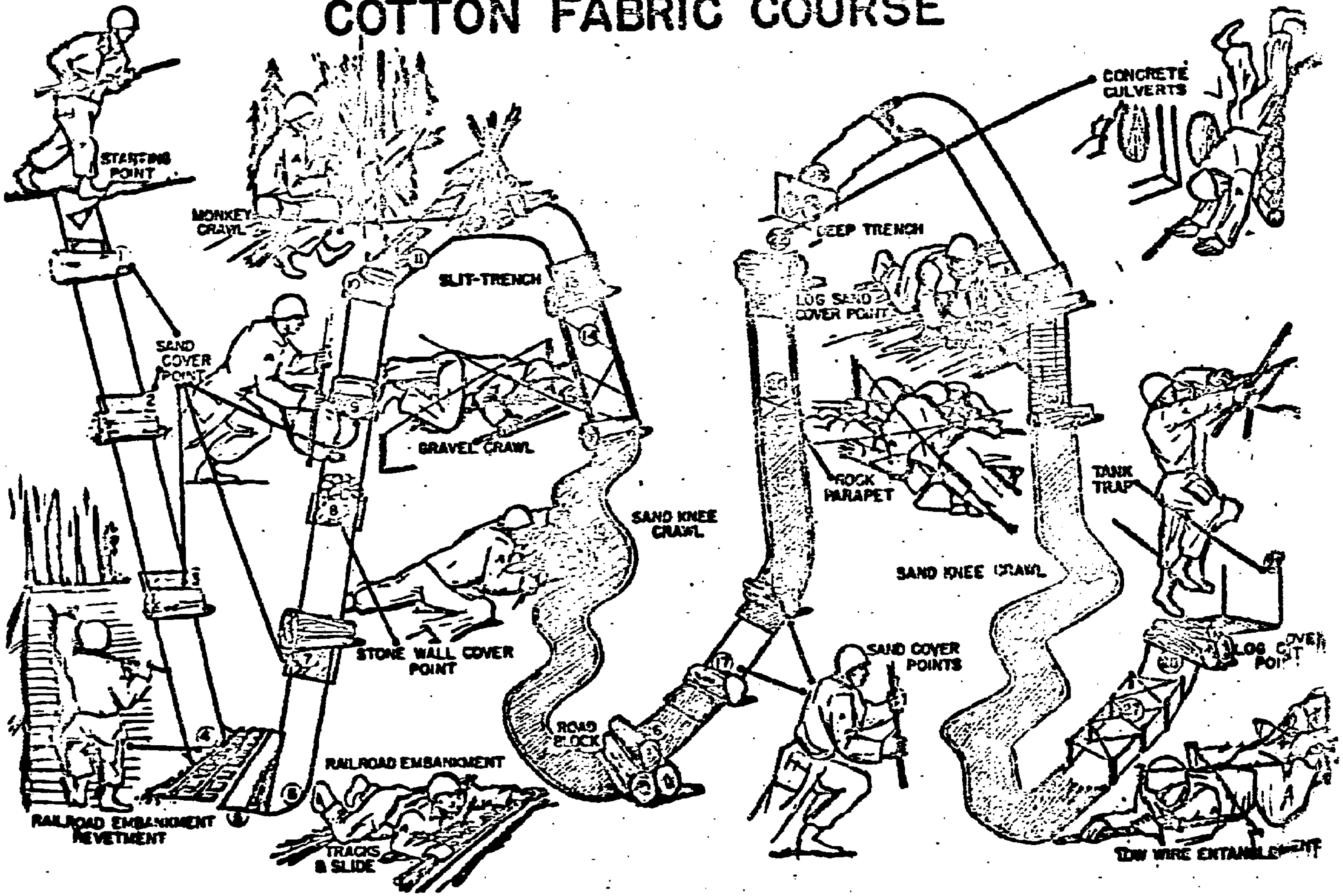
However, it is considered by the textile engineering staff of the QM R and E that the solution to the problem may not lie in the area of fabric course - laboratory correlation. No knowledge exists of the behavior of similar blended fabrics in actual field wear. It may be that the previous correlation between accelerated wear on the fabric course and actual field wear which existed on all-cotton fabrics may not be found when part synthetic garments are tested. Therefore, in addition to the development, a small scale pilot study is presently underway on combat troops. Twelve members of an artillery battalion are wearing utility ensembles fabricated from a cotton/nylon blend in maneuvers and during training. These artillerymen have also been issued new all-cotton uniforms for comparison. Admittedly, this is not a controlled or designed experiment. Neither fabric or resources are available for a more ambitious effort at this time. It is anticipated, however, that valuable information may be obtained which will assist in the design of a formal field trial which is planned for the spring of 1961. At that time a large scale field wear evaluation will be conducted at Ft. Jackson, S. C. on two full platoons of soldiers wearing experimental vs. standard utility uniforms. Although this design is not complete at this time, it is hoped that appropriate statistical techniques will help enable the Quartermaster Corps to glean the greatest amount of possible information from this study and will also provide means for assessing new and promising fabric developments for military garments with the least expense and shortest lead time.

REFERENCE

1. Quartermaster Research and Development Laboratories, Textile Materials Engineering Laboratory Report No. 110A, "A Survey of Quartermaster Studies of the Wear Resistance of Cotton Fabrics", by Oscar Mandel dated March, 1953.

FIGURE I

COTTON FABRIC COURSE



GROUP SCREENING DESIGNS

W. S. Connor
Research Triangle Institute

1. INTRODUCTION. Recently, G. S. Watson [1] considered a particular approach to the problem of screening a large number of factors of which only a few affect the response variable. The general approach is similar to that of Dorfman [2] to the biological problem of the detection of a rare defect among the members of a large population. Dorfman suggests that pooled blood samples be tested, and that the individual samples which form a pooled sample be tested whenever the latter gives a positive result. A 100 per cent screening may be achieved with substantial saving in the number of blood tests. The present paper modifies the development in [1] so that orthogonal designs may be used.

2. THE GROUP SCREENING DESIGN WHEN THERE IS NO EXPERIMENTAL ERROR. Beginning with the case when the experimental error is negligible, Watson makes the following assumptions:

(i) all factors have, independently, the same prior probability, p ($q = 1 - p$), of being effective,

(ii) a factor is effective if it produces a non-zero change in the response,

(iii) none of the factors interact,

(iv) the directions of possible effects are known, and

(v) the number of factors $f = gk$, where g = the number of groups and k = the number per group.

A typical group screening design is illustrated by $f = 9$. Before discussing it, we note that for a single stage design, in the absence of experimental error, ten runs are sufficient to determine which factors are effective.

This paper was initially issued as Technical Report No. 3, S - 10 of the Research Triangle Institute.

This work was done under OOR Project No. 2579, Contract No. DA-01-009-ORD-816.

Of course, this design will not permit uncorrelated estimates of the main effects. In the group screening design, the nine factors denoted by A, B, ..., I, are divided into $g = 3$ groups of $k = 3$ factors each, to form group-factors (A, B, C), (D, E, F), and (G, H, I), which may be denoted respectively by X, Y, and Z. Then, adopting the convention that the upper and lower levels of the factors A, B, ..., I are defined so that their effects, if any, are to give greater and lesser values to the response, so that the main effects, if any, are positive, the upper and lower levels of the group-factors are defined as follows:

(2.1) Definitions of Levels of Group-Factors

	Group-Factors		
Levels	(A, B, C):X	(D, E, F):Y	(G, H, I):Z
Lower level	(0, 0, 0):1	(0, 0, 0):1	(0, 0, 0):1
Upper level	(1, 1, 1):x	(1, 1, 1):y	(1, 1, 1):z

The first stage design is for the group-factors. One may use a 1/2 replicate of a 2^3 design, as

(2.2) $x, y, z, xyz.$

In terms of the factors, these treatment combinations are

(2.3)

$$\begin{aligned} x&: (1, 1, 1, 0, 0, 0, 0, 0, 0) \\ y&: (0, 0, 0, 1, 1, 1, 0, 0, 0) \\ z&: (0, 0, 0, 0, 0, 0, 1, 1, 1) \\ xyz&: (1, 1, 1, 1, 1, 1, 1, 1, 1) \end{aligned}$$

The usual functions of the responses to the treatment combinations are indicated below:

(2.4)

Treatment Combination (Group-factors)	Mean	Main Effect X	Y	Z
x	1	1	-1	-1
y	1	-1	1	-1
z	1	-1	-1	1
xyz	1	1	1	1

For our purposes, the divisor will be taken as the number of responses in the design, which in this instance is four. In view of (iii), this design will estimate the main effects of the group-factors, and in view of (iii) and (iv), there will be no cancelling out of effects within the group-factors.

Every group-factor which contains at least one effective factor will itself be effective and, because of the absence of experimental error, will be detected. If a first stage experiment reveals that one or more group-factors are effective, a second stage experiment will be carried out on the factors which comprise them to find out which factors are effective. For example, if group-factor X is effective, but group-factors Y and Z are not effective, then the second stage experiment will involve factors A, B, and C. If group-factors X and Y are effective, but not group-factor Z, then the second stage experiment will involve factors A, B, C, D, E, and F. Finally, if group-factors X, Y, and Z are effective, then all nine factors will be included in the second stage experiment. Accordingly, depending on the outcome of the first stage experiment, there may be no second stage at all, or there may be a second stage involving 3, 6, or 9 factors.

If h factors are to be studied at the second stage, then only h runs are needed at the second stage.* This is because one run from the first stage can be used in the analysis of the responses from the second stage. To illustrate, suppose that only group-factor X is effective from the first stage. Then factors A, B, and C must be studied in the second stage. One way that this can be done is by running treatment combinations

*Actually, it can be demonstrated that if n group-factors are effective, then only n(k-1) runs are needed at the second stage, not nk = h as stated here.

(2.5) $$bc = (0, 1, 1, 0, 0, 0, 0, 0, 0)$$
$$ac = (1, 0, 1, 0, 0, 0, 0, 0, 0)$$
$$ab = (1, 1, 0, 0, 0, 0, 0, 0, 0)$$

at the second stage. Then, remembering that $x = abc$, the effects of A, B, and C are determined from

(2.6) $$x - bc, \; x - ac, \text{ and } x - ab,$$

respectively.

By assuming some value for p, it is possible to calculate the probabilities of various numbers of runs at the second stage, and thereby to compare the group screening design with the single stage design. For $p = .15$, $fp = 1.35$, so that it is expected that one or two of the nine factors will be effective. The probabilities that a group-factor will contain 0, 1, 2, or 3 effective factors are given below:

(2.7) Probabilities that a Group-Factor Will Contain 0, 1, 2, or 3 Effective Factors, for $p = .15$

Number of Effective Factors	Formula	Numerical Value
0	q^3	.614
1	$3pq^2$	.325
2	$3p^2q$	.057
3	p^3	.003

If there are 1, 2, or 3 effective factors, the three factors will be run in a second stage. The probability of this event is

(2.8) $$1 - q^3 = 1 - .614 = .386 = r, \text{ say.}$$

There will be no second stage if there are no effective factors. The probability of this event is q^9. The probabilities that the second stage will require 0, 3, 6, or 9 runs, and therefore that the two stages together will require 4, 7, 10, or 13 runs are given below:

(2.9) Probabilities of Various Numbers of Runs

Number of Runs: Second Stage	Number of Runs: Total	Probability: Formula	Probability: Numerical Value
0	4	$(1-r)^3$	.23
3	7	$3r(1-r)^2$	.44
6	10	$3r^2(1-r)$	.27
9	13	r^3	.06

From this table, the expected total number of runs is calculated to be

(2.10) $$4 \times 0.23 + 7 \times 0.44 + 10 \times 0.27 + 13 \times 0.06 = 7.48,$$

which is an average saving of 2.52 runs from the 10 runs which are required by a single stage experiment.

The saving is greater for smaller p. For $p = .10$, the expected number of runs is 6.43 and for $p = .05$, it is 5.29. Of course, the number of runs cannot drop below 4. For p greater than .15, the saving is less. For $p = .20$, the expected number of runs is 8.39 and for $p = .29$, the expected number of runs is 10, so that for still larger p, the single stage design is preferable. These observations illustrate the general principle that, for fixed f, the expected saving varies inversely with p.

3. THE GROUP SCREENING DESIGN WHEN THERE IS EXPERIMENTAL ERROR. For the case when the experimental error is appreciable, $\sigma \neq 0$, Watson modified assumption (ii) to

(ii) effective factors have the same effect, $\Delta > 0$.

This assumption implies that the effect of a group-factor is one of the values $0, \Delta, 2\Delta, \ldots, k\Delta$ and that if the effect is $s\Delta$, $(s = 0, \ldots, k)$, then the group-factor contains s effective factors and $(k - s)$ ineffective factors. But, of course, in the real problem, the effect of a group-factor may be some value other than Δ, and effects $s\Delta$ may be achieved by adding effects from $s' \neq s$ factors. Although this assumption is somewhat arbitrary and unrealistic, it perhaps results in shedding some light on the characteristics of group screening designs. It is akin to the real problem in that the levels of the factors may be chosen in such a way that there is a common least change in response, say Δ, which is worth detecting.

Another assumption made by Watson is that

(vi) the errors of all observations are independently normal with a constant known variance σ^2.

The procedure is further specified by assuming that

(vii) estimated main effects of group-factors are tested at significance level α, and if one or more of them is significantly different from zero, a second-stage experiment is carried out. Tests of whether the main effects of the factors are zero are made at significance level β.

Because of the nice properties of orthogonal designs, it will be assumed that such designs are used at both stages. Orthogonal fractional factorial designs exist having 2^m treatment combinations, which can be used to estimate the main effects of $2^m - 1$ factors. This is a rather thin series. However, Plackett and Burman [4] give orthogonal designs having 4t treatment combinations, which will accommodate $4t - 1$ factors.

For the example under discussion, suppose that the first stage design is the one already described above. If there is only one group-factor which is found to be significant, then three factors are varied in the second experiment, and the same design can be used. If two group-factors are found to be significant, then eight treatment combinations are needed at the second stage. Such a design is given below:

(3.1) Treatment Combinations for Six Factors

A	B	C	D	E	F	
0	0	0	0	0	0	0
0	0	0	1	1	1	1
0	1	1	0	0	1	1
0	1	1	1	1	0	0
1	0	1	0	1	0	1
1	0	1	1	0	1	0
1	1	0	0	1	1	0
1	1	0	1	0	0	1

A seventh factor could be added at the levels shown in the last column, or the three remaining factors all could be held constant throughout.

If all three group-factors turn out to be significant, then twelve treatment combinations are required at the second stage. A suitable design is the following:

(3.2) Twelve Treatment Combinations for Nine Factors (Plackett and Burman)

A	B	C	D	E	F	G	H	I		
1	0	1	0	0	0	1	1	1	0	1
1	1	0	1	0	0	0	1	1	1	0
0	1	1	0	1	0	0	0	1	1	1
1	0	1	1	0	1	0	0	0	1	1
1	1	0	1	1	0	1	0	0	0	1
1	1	1	0	1	1	0	1	0	0	0
0	1	1	1	0	1	1	0	1	0	0
0	0	1	1	1	0	1	1	0	1	0
0	0	0	1	1	1	0	1	1	0	1
1	0	0	0	1	1	1	0	1	1	0
0	1	0	0	0	1	1	1	0	1	1
0	0	0	0	0	0	0	0	0	0	0

A design accommodating two more factors can be obtained by assigning levels as indicated in the last two columns.

The expected number of runs, the expected number of effective factors detected, and the expected number of ineffective factors wrongly declared to be effective are quantities which describe the operating characteristics of the method.

The power of the test of a group-factor depends on its mean. If the mean is $s\Delta$, the power of the t-test of it will be

$$\pi_1 = \pi_1(s\,\phi_1,\ \alpha), \tag{3.3}$$

where

$$\phi_1 = \sqrt{2\left[\frac{q}{4}\right]}\ \frac{\Delta}{\sigma}$$

is the parameter used in Table 10 of Pearson and Hartley [3]. Then the probability that a group-factor will be declared significant is

$$\pi_1^{*} = \sum_{s=0}^{k} \binom{k}{s} p^s q^{k-s}\ \pi_1(s\,\phi_1,\ \alpha), \tag{3.4}$$

and that an effective group-factor will be declared significant is

$$\pi_1' = \left[\sum_{1}^{k} \binom{k}{s} p^s q^{k-s}\ \pi_1(s\,\phi_1,\ \alpha)\right] / (1 - q^k),$$

and, of course, the probability that an ineffective group-factor will be declared significant is α.

The power of the t-test of a factor will be

(3.5) $$\pi_2 = \pi_2(\phi_2, \beta),$$

where

$$\phi_2 = \sqrt{2\left[\frac{nk}{4}\right]}\,\gamma\,\frac{\Delta}{\sigma},$$

$\gamma = 0$ for an ineffective factor and 1 for an effective factor, and $\left[\frac{nk}{4}\right]$ is the least integer greater than $\frac{nk}{4}$, except that $\left[\frac{nk}{4}\right] = 0$ when $n = 0$. Of course, $\pi_2(0, \beta) = \beta$.

It can be shown that the expected number of effective factors to be declared effective (significant) is

(3.6) $$E = kp' \sum_{n=0}^{g} n\,\pi_2\left(\sqrt{2\left[\frac{nk}{4}\right]}\,\frac{\Delta}{\sigma}; \beta\right)\binom{g}{n}\pi_1^{*\,n}(1-\pi_1^*)^{g-n},$$

where $p' = p\,\pi_1'/\pi_1^*$, and that the expected number of ineffective factors to be declared effective is

(3.7) $$\bar{E} = fq\beta\left[\pi_1^* - q^{k-1}(\pi_1^* - \alpha p)\right] / (1 - q^k).$$

Also, the expected number of runs is

(3.8) $$R = 4\left[\frac{g}{4}\right] + \sum_{n=1}^{g} 4\left[\frac{nk}{4}\right]\binom{g}{n}\pi_1^{*\,n}(1-\pi_1^*)^{g-n}.$$

These formulas may be calculated for the example under consideration. They are as follows:

$$\phi_1 = \sqrt{2}\,\Delta/\sigma$$

$$\pi_1^* = 0.614\,\alpha + 0.325\,\pi_1(\phi_1, \alpha) + 0.057\,\pi_1(2\,\phi_1, \alpha) + 0.003\,\pi_1(3\,\phi_1, \alpha)$$

$$\pi_1' = 0.844\,\pi_1(\phi_1, \alpha) + 0.148\,\pi_1(2\phi_1, \alpha) + 0.008\,\pi_1(3\,\phi_1, \alpha)$$

$$E = \frac{3\,\pi_1'}{\pi_1^*}\,(0.15)\sum_{n=0}^{3} n\,\pi_2\left(\sqrt{2\left[\frac{nk}{4}\right]}\,\frac{\Delta}{\sigma}, \beta\right)\binom{3}{n}\pi_1^{*n}(1-\pi_1^*)^{3-n}$$

$$\overline{E} = 19.1\,\beta\left\{\pi_1^* - 0.723\,(\pi_1^* - \alpha \times 0.15)\right\} = 19.1\beta\left\{0.277\,\pi_1^* + 0.108\alpha\right\}$$

$$R = 4 + \sum_{n=1}^{3} 4\left[\frac{n3}{4}\right]\binom{3}{n}\pi_1^{*n}(1-\pi_1^*)^{3-n}.$$

From (2.7) it is seen that for $p = .15$, the probability of two or more effective factors occurring together in the same group-factor is equal to .06. Accordingly, in practice, one would not have to know the directions of possible effects, and two-sided tests would be used. Some calculations for E, $\overline{E}$, and R have been made for two-sided tests. For $\Delta/\sigma = \sqrt{2}$, 2, 3; $\alpha =$ 0.01, 0.05; and $\beta =$ 0.01, 0.05, the results are shown in the accompanying table.

TABLE OF EXPECTED VALUES

$\frac{\Delta}{\sigma}$	β		α = 0.01	α = 0.05
$\sqrt{2}$	0.01	E	.67	.92
		$\bar{E}$	.01	.02
	0.05	E	.80	1.03
		$\bar{E}$	.07	.10
		R	7.14	8.24
2	0.01	E	1.23	1.30
		$\bar{E}$	.02	.02
	0.05	E	1.26	1.32
		$\bar{E}$	.10	.11
		R	8.40	8.88
3	0.01	E	1.35	1.35
		$\bar{E}$	.02	.02
	0.05	E	1.35	1.35
		$\bar{E}$	.10	.11
		R	8.69	8.99

As might have been anticipated, increasing α or Δ/σ results in more group-factors being declared significant and, hence, results in more factors being tested in the second stage. It therefore increases the average number of runs, the average number of effective factors which are identified, and less markedly, the number of ineffective factors which are declared to be effective. Increasing β has no effect on the average number of runs, but does increase the average number of effective factors which are identified and the average number of ineffective factors which are declared to be effective.

REFERENCES

[1] Watson, G. S., "A study of group screening designs," Technical Report No. 2, S-10, Statistics Research Division, Research Triangle Institute; to be published in Technometrics.

[2] Dorfman, R., "The detection of defective members of large populations," Annals of Mathematical Statistics, Vol. 14 (1943), pp. 436 - 440.

[3] Pearson, E. S. and Hartley, H. O., Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press, 1954.

[4] Plackett, R. L. and Burman, J. P., "Design of optimum multi-factorial experiments," Biometrika, Vol. 33, 1946.

MULTIVARIATE ANALYSIS ILLUSTRATED BY NIKE-HERCULES:

I. Separation of product and measurement variability.

II. Acceptance sampling.

J. Edward Jackson
Eastman Kodak Company

Abstract

This expository paper is concerned with the application of some multivariate techniques to some of the problems involved in evaluation of missile testing data. Examples illustrating these techniques are drawn from actual Nike booster test data.

The first part of the paper is concerned with methods separating the total variability of test results into components representing product and measurement variability. Specific techniques discussed are (1) regression analysis, (2) principal components, and (3) the methods of Grubbs, Kruskal and David for analyzing related pairs of observations.

The problems of acceptance sampling when more than one variable is involved are discussed in the second part of the paper. Because of the cost of testing involved in missile evaluation, a sequential multivariate type of inspection plan is developed. Considerable emphasis is placed on the problems involved in incorporating the product specifications into these sampling plans. A second procedure is developed to sequentially test for excess dispersion within a lot.

MULTIVARIATE ANALYSIS ILLUSTRATED BY NIKE-HERCULES

J. Edward Jackson
Eastman Kodak Company, Rochester, New York

I. SEPARATION OF PRODUCT FROM TESTING AND MEASUREMENT VARIABILITY

1. Introduction. In the past few years, the techniques associated with multivariate analysis have begun to come into their own in the field of industrial statistics. Prior to that time, say World War II, most of the available literature was devoted to the use of factor analysis in education and psychology or the use of discriminant analysis in genetics and archeology. However, by the close of World War II, we in the industrial world began to discover that control charts and the analysis of variance would not solve all of our problems and that one of the reasons for this was that several factors in a production problem seemed to vary all at once, but not in a completely random fashion. Starting with Hotelling's bombsight paper,[4] the number of references in the literature has grown slowly but steadily and eventually the generalized T^2-statistic, multivariate analysis of variance and the method of principal components should become regular members of the kit of tools of the industrial statistician. There still is a great need for some fairly non-technical literature on multivariate methods and more work in methodology to "translate" the great amount of theoretical work now available into a form readily usable by the practicing statistician.

Most published works so far have been related either to control chart procedures or component and factor analysis. To show that there are other industrial applications of multivariate techniques, the first part of this paper will deal with some methods used to break the total variability of a system into components representing product variability and testing and measurement variability. Some of the techniques used are not classical multivariate techniques as such but all of them are nevertheless multivariate in the sense that they all take into account the relationships among two or more variables. These techniques will be illustrated with numerical examples dealing with static tests of Nike boosters carried on by the Hercules Powder Company at Radford Arsenal, Virginia. It should be pointed out that the purpose of these examples is to illustrate the various techniques and does not reflect in any way on the quality assurance policies exercised by the Radford Arsenal.

It will be seen that no one of these techniques can completely separate product variability from testing and measurement variability and it would seem that an integrated system employing perhaps several of these techniques would be necessary to obtain optimum results.

2. Regression Analysis. While not everyone may construe the term "regression analysis" to be a part of multivariate analysis, it nevertheless should be thought of as such since this technique takes into account the relationship of each of the independent variables with each other as well as with the dependent variable. One of the steps in regression analysis is to obtain a sum of squares "due to regression" and from this, break the total variability of a system into components representing explained and unexplained variability. In experimental design problems, the factors which one generally studies are ones associated with the product and hence the residual variability of an analysis of this type of data generally represents experimental error. In the present case, the procedure will be reversed. The variables to be considered will be ones associated with the testing and measurement phase of the program so that the major portion of the residual could be assumed to represent product variability. Hotelling used this method to determine the residual variability of bombsights after removing the effect of guidance systems, crews and flight and bombing patterns. In the present case, we shall be concerned with the action times of Nike boosters as determined from static tests.

The procedure in static testing consists of conditioning a round to a specific temperature, then placing it in a firing bay with strain gages fastened to its nose in such a way that pressure, thrust and action time (essentially a measure of the time it takes to burn all of the powder in a round) can be measured. The measurements are recorded either on an oscilloscope or an electronic integrator using a fairly complex system of components. Since this is a complicated procedure, a number of production records are kept on anything which might affect the results and these become the independent variables in our regression analysis.

The present example will deal with the variability of the action time measurements for a production lot with the tested rounds being conditioned to a temperature of -10°F. The independent variables are:

x_1 = Propellant temperature. (Although the propellant temperature is supposed to be -10° F, it may actually vary a degree or two from this.)

x_2 = Length of time propellant was conditioned.

x_3 = Outside temperature. (This can be an important factor for -10° F rounds if it is 90° F in the shade.)

x_4 = Length of time elapsed from the time the round is removed from its conditioning box until it is fired.

x_5 = Length of elapsed time from the time the ignitor is removed from its box until the round is fired.

x_6 = Ignitor resistance.

x_7 = Ignitor delay.

x_8 = 0 if the measurements were obtained on instrument table #1.

= 1 if the measurements were obtained on instrument table #2.

x_1', x_2', ... x_k' represent the different strain gages used to measure pressure. (Action time is defined as the length of time during burning that the chamber pressure is above a certain amount.)

The cross-products of x_1, x_3, and x_4 were also used.

For a particular production lot studied, these variables explained nearly 60 per cent of the total variability of the action time measurements. When the residuals of this analysis were further related to production variables, about half of the remaining variability was accounted for.

There is nothing new about using regression analysis but the main reason for discussing it here is to show that the measurement and testing variability can sometimes appear in the explained sums of squares as well as the residual. Another reason for mentioning this technique is to re-echo some of the warnings that have been sounded in the past regarding the misuse of regression analysis. When regression analysis first appeared on the scene, it was so overworked that many examples of "nonsense" correlations began to appear and this necessitated the development of partial regression and correlation methods. After that, regression analysis became a little more respectable and more reputable results began to appear. I feel that with the advent of high speed computers (the above problem takes about a minute on a 700-Series IBM Computer) we are approaching another problem era. It is relatively easy to include a large number of independent variables and this may tempt people to throw in everything but the kitchen sink and in every inconceivable combination. Eventually, the Type I errors can get so large that erroneous results can almost be guaranteed. Also, this "kitchen sink" technique is apt to decrease the number of degrees of freedom associated with the residual rather rapidly so that high correlations can result solely because the number of parameters fitted is nearly as large as the number of observations. There is no easy way out of this but I feel that it is well that these problems be mentioned occasionally to prevent fingers from being pointed at us collectively again.

An obvious criticism of this example is that it is essentially PARC analysis and we are being continually advised that this is not the way to do things. There are a number of valid reasons why PARC analysis should not be used, not the least of which is that a particularly important factor may not vary much over the period of time represented by the data and consequently may not appear to be very important. However, in ballistic missile testing, one cannot conduct designed experiments - the present testing procedure is expensive enough - and hence we must make do with what we have. Fortunately, in this case, the primary goal was to obtain measures of product and testing and measurement variability rather than to obtain a functional relationship among the measurement variables and the test data.

3. Principal Components. The method of principal components has been around a long time. Karl Pearson[9] suggested this technique around the turn of the century. Hotelling[3] developed methods for its use in the mid-thirties and this technique coupled with factor analysis has kept many psychometricians occupied for some time. Only in the last ten years has much use been made of principal components in industry where it has been used as a control tool[7] and a method of prediction as well as its primary use, the determination of the structure of a system. As somewhat of a by-product of this technique, we may also use it to obtain an approximate method for separating product from measurement variability. This can be illustrated by a specific example:

In the static testing of ballistic missiles such as the Nike booster, for a characteristic such as total impulse, four actual measurements are made on each round during a test firing. Each round has two thrust gages attached to it and each of these gages is in turn recorded on both an oscilloscope and an electronic integrator. We can then treat this as a four variable problem. Since these variables are highly correlated, one would expect that the first characteristic vector associated with a covariance matrix of these variables would represent product variability and the remainder might give some insight into the measurement errors. Since the sum of the characteristic roots associated with these vectors equals the trace of the covariance matrix, the trace may be assumed to be a rough measure of the total variability of the system. The ratio of each root to the total may then be considered a measure of the variability explained by that particular principal component. It should be emphasized that this method can be best employed if the original variables are all in the same units and have the same variances although for prediction purposes, this requirement is not necessary.

A particular example using Nike booster data has been discussed in Technometrics[5]. The first vector, as might be expected, represented, essentially, the average of the four readings and explained 78 per cent of the trace. The other three vectors, accounting for 22 per cent of the trace represented gage differences and integrator vs. oscilloscope differences. One should not infer that 78 per cent of the total variability was product variability since this component actually represents variability common to all four measurements which would consist not only of product variability but some of the factors mentioned in the regression example. One could infer however, that approximately 22 per cent of the variability of individual measurements could be attributable to instrumentation variability. The first transformed variate could then be used as a starting point for other studies.

4. The Methods of Grubbs, Kruskal and David*. Quite often in industrial work, one obtains two sets of data which are functionally related in some way. Quite often this will involve duplicate measurements on a series of items. If both measurements are made with the same equipment and personnel, the data are commonly analyzed by a one-way analysis of variance, the "between" sum of squares representing the variability among the items themselves (and changes in level of the equipment if any such exist) and the "within" sum of squares representing measurement variability. On the other hand, one of each pair of observations can be made with one piece of equipment and the second observation on another. If these same two pieces of equipment are employed for a series of items, then the duplicate variability may not necessarily be entirely random since a bias may exist between the two pieces of equipment. Data of this sort are commonly examined by means of a randomized block analysis with the "treatments" being the items, the "blocks" being the pieces of equipment and the residual representing the inherent variability. (Actually, the residual is an item x equipment interaction but in most cases this can be considered inherent variability.) These two situations represent the most widely used approaches

* These methods were first proposed for ballistic missile static tests by D. E. Thompson, Allegany Ballistics Laboratory[10].

in industry to the separation of product and measurement variability. The second of these situations deserves a bit more consideration, however.

When the duplicate measurements are made on different pieces of equipment or by different persons, there is no guarantee that the inherent variability will be the same for both sets of data. This is one of the basic assumptions of the analysis of variance although we are now consoled with the conclusion that the violation of this assumption is not too important. Nevertheless, if the inherent variabilities of the two sets of data are different, the industrial statistician might want to know what they are. Two possibilities arise:

Case I. Mean values of the two sets of data are the same or differ by a constant. (Grubbs' method.)[2]

Let y_{ij} represent the j-th measurement on the i-th item. This can be expressed in the following manner:

$$y_{ij} = \mu + x_i + \delta_j + e_{ij}$$

where μ is the overall mean, x_i is the deviation of the i-th item from the mean, δ_j is the bias of the j-th piece of equipment and e_{ij} the inherent variability associated with the ij-th measurement. x_i and e_{ij} are random variables and suggest the following variance component models for the two pieces of equipment:

$$\sigma_{y_1}^2 = \sigma_x^2 + \sigma_{e_1}^2$$

$$\sigma_{y_2}^2 = \sigma_x^2 + \sigma_{e_2}^2$$

Under the assumption that e_{i1} and e_{i2} are uncorrelated, the estimates of these components are obtained as follows:

$$s_x^2 = s_{y_1 y_2}$$

$$s_{e_1}^2 = s_{y_1}^2 - s_{y_1 y_2}$$

$$s_{e_2}^2 = s_{y_2}^2 - s_{y_1 y_2}$$

Example:

In the static testing of ballistic missiles, two strain gages are employed to measure each of the attributes of pressure and thrust. The information from each of these gages is then relayed back through an electronic system with the final result appearing either on an electronic integrator or an oscilloscope as we have already mentioned. The present example consists of the results on thirty-seven pressure measurements of Nike boosters all from one production lot and conditioned at 130° F. Two observations (P.S.I. units) have been obtained for each round, one from the integrator related to each pressure gage so the present method will be employed to estimate the inherent variabilities of each measurement system. The variance of each set of data and the covariance were found to be:

$$s_{y_1}^2 = 606, \quad s_{y_2}^2 = 605 \text{ and } s_{y_1 y_2} = 580.$$

The "product" variability is then $s_x^2 = s_{y_1 y_2} = 580$. The measurement variabilities are given by

$$s_{e_1}^2 = s_{y_1}^2 - s_{y_1 y_2} = 26 \text{ and } s_{e_2}^2 = s_{y_2}^2 - s_{y_1 y_2} = 25.$$

From this it can be seen that the inherent variabilities of the two systems are about the same. (Had a difference existed, a number of reasons could be suggested such as a difference in the variability of the gages used, number of tubes replaced in the system, etc.) Furthermore, it can be concluded that the inherent variability is, on the average, about one-sixth of the total variability. The remainder cannot be considered solely product variability since it has already been suggested in the earlier examples that such things as temperature affect the overall results of both gages and hence s_x^2 might be considered a measure of product and testing variability. As stated in the introduction, no one of these methods will singlehandedly resolve the problem of completely separating these variabilities. A comprehensive study would probably involve the use of several of these techniques.

It should be pointed out that Grubbs has also derived methods to handle triple and quadruple measurements although the mechanics are not as simple as the ones shown here.

Case II. Mean values of the two sets of data have a fixed ratio. (The method of Kruskal and David.)

This problem must be stated in a slightly different manner since it can cover a wider range of cases. Again, we have two sets of data but these need not be the same type of measurements on each item. They may be duplicate measurements or they may be related measurements such as pressure and thrust. Since they may be different types of measurements, the model now changes to:

$$y_{ij} = \mu_j + x_{ij} + e_{ij}$$

where μ is the mean of the j-th type of measurement, x_{ij} is the deviation of the j-th measurement of the i-th item from its mean and e_{ij} is the inherent variability associated with the j-th measurement on the i-th item. In relating this case to the model in Case I, $\mu_j = \mu + \delta_j$ except that δ_j is not necessarily a bias and x_{i1} is not necessarily equal to x_{i2} as was the case previously. The x_{ij} are random but no longer independent being restricted by the relationship:

$$\frac{\mu_1 + x_{i1}}{\mu_2 + x_{i2}} = \xi$$

where ξ is a fixed and known constant (known in the sense of any parameter - we may still have to estimate it).

For any individual item, any variation of

$$\frac{y_{i1}}{y_{i2}} = \frac{\mu_1 + x_{i1} + e_{i1}}{\mu_2 + x_{i2} + e_{i2}}$$

from ξ then is due to e_{i1} and e_{i2} and from this it is possible to obtain estimates of $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ as well as $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$. These estimates can be obtained from the following relationships:

$$s^2_{y_1} = s^2_{x_1} + s^2_{e_1}$$

$$s^2_{y_2} = s^2_{x_2} + s^2_{e_2}$$

$$s_{y_1 y_2} = s_{x_1 x_2} \text{ (i.e. errors are uncorrelated)}$$

$$= s^2_{x_1}/\xi = \xi s^2_{x_2} .$$

Therefore, all of the estimates can be obtained from $s^2_{y_1}$, $s^2_{y_2}$, $s_{y_1 y_2}$ (as in Case I) and a knowledge of ξ. If $\xi = 1$, we have Case I.

Example:

This technique has been proposed as a method to obtain measures of the variability for thrust and pressure measurements using the assumption that the ratio of the time-integrals for thrust and pressure is constant for a given production lot. From the same lot mentioned in the previous example the time integrals (averaged over the two gages) were obtained for thrust and pressure. Since the true ratio of these quantities varies from lot to lot, it is necessary to estimate ξ from the data itself. If y_1 designates $\int Fdt$ and y_2 designates $\int Pdt$, then $\bar{y}_1 = 148162$, $\bar{y}_2 = 3276$, $\hat{\xi} = 45.23$, $s^2_{y_1} = 385752.25$, $s^2_{y_2} = 543.26$ and $s_{y_1 y_2} = 4975.66$. From this, $s^2_{x_1} = 22504.91$, $s^2_{x_2} = 110.01$, $s^2_{e_1} = 160703.15$ and $s^2_{e_2} = 433.25$.

From these results, it would appear that roughly 40 per cent of the total variability of thrust measurements and 80 per cent of the pressure variability could be attributable to measurement variability.

Two points should be mentioned in connection with this example:

1. Although the proportion of measurement variability seems fairly high, the coefficients of variation for the total variability of these measurements are of the order of four-tenths to seven-tenths of one per cent.

2. There has been considerable evidence to indicate that the errors in thrust and pressure measurements may be correlated which of course would invalidate the use of this technique to solve problems of this type until it has been modified to allow for correlated errors. The method itself is, however, perfectly valid as long as the model holds.

II. ACCEPTANCE SAMPLING[5]

1. Introduction. The motivation for this part stems from the acceptance sampling programs used in the evaluation of production lots of ballistic missiles such as the Honest John and the Nike. These particular missiles are operational and have been on a production basis for some time. They are currently produced in lots and subjected to the ordinary acceptance sampling schemes used in quality control. Since the testing of these missiles is very expensive, judgments on lots should be made with as little inspection as possible consistent with the prescribed risks of accepting poor lots and rejecting good ones. This suggests sequential sampling which has come into widespread use in the past few years and, in fact, some types of missiles are now inspected in that manner.

There are several important parameters to be inspected on each round in missile production; a few such characteristics are action time, thrust or impulse, and some measure of chamber pressure. These variables are interrelated and hence the problem is a multivariate one. In present day operations, separate sequential plans must be set up for each parameter. It is, therefore, possible to get conflicting answers about the quality of a lot, sampling may terminate for one characteristic before another and there is no appreciation of the true sampling risks involved in the overall program. It is obvious that a sequential multivariate technique should be used.

In this article we will give some multivariate sequential inspection schemes for the characteristic averages both for the case where the population covariance matrix Σ is known or assumed to be known (a typical quality control situation) and where it must be estimated from the sample. When the covariance matrix is known, we use a sequential χ^2-test; when the covariance matrix is estimated from the sample, we use a sequential T^2-test.

2. Sequential Univariate and Multivariate Procedures for Testing Means. In univariate situations, test procedures have been constructed to test the null hypothesis

$$H_0 : \mu = \mu_0 \text{ (or } \mu - \mu_0 = \delta)$$

against the alternative

$$H_1 : \mu \neq \mu_0 \text{ (or } \mu - \mu_0 \neq \delta).$$

When these procedures are extended to the sequential case, it is customary to replace these with more specific hypotheses, viz:

$$H_0 : \mu = \mu_0 \text{ (or } \mu - \mu_0 = \delta_0)$$

$$H_1 : \mu = \mu_1 \text{ (or } \mu - \mu_0 = \delta_1).$$

For the case where the population variance is known, the sequential procedures have been worked out by Wald and for the case where the variance is not known, by Wald, Rushton and others.

In the multivariate case, these expressions could be replaced by p-variate vectors. The null hypothesis could be given as

$$H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$$

but it becomes very difficult to specify a meaningful single alternative since there are, presumably, infinitely many points in p-space that are of equal importance as alternatives and even a hypothesis of the type

$$H_0 : \underset{\sim}{\mu} - \underset{\sim}{\mu}_0 = \underset{\sim}{\delta}_0$$

would be difficult to specify. It is easier to operate with the surfaces of p-dimensional ellipsoids. For instance, the statements $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ and $(\underset{\sim}{\mu} - \underset{\sim}{\mu}_0) \times \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)' = 0$ are identical but the quadratic form of the latter expression can also be set equal to some scalar quantity viz:

$$(\underset{\sim}{\mu} - \underset{\sim}{\mu}_0) \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)' = \lambda_0^2$$

which represents the entire surface of a p-dimensional ellipsoid while the expression $\underset{\sim}{\mu} - \underset{\sim}{\mu}_0 = \underset{\sim}{\delta}_0$ would represent only one point. Similarly, the alternative hypothesis would be of the same form but equal to a larger scalar value. Our hypotheses become

$$H_0: \quad (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0) \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)' = \lambda_0^2 \text{ (quite often zero)}$$

$$H_1: \quad (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0) \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)' = \lambda_1^2 \; (\lambda_1^2 > \lambda_0^2).$$

3. Test Procedures. Although the form of the sequential procedures differ for the case $\sum$ known or $\sum$ unknown, they are quite similar in administration. All sequential procedures, in the Wald sense, employ a sequential probability ratio p_{1n}/p_{0n} which is evaluated after each new observation is taken. Let α and β denote the usual Type I and Type II errors. If, after n observations have been made, $p_{1n}/p_{0n} \le \beta/(1-\alpha)$ we accept H_0; if $p_{1n}/p_{0n} \ge (1-\beta)/\alpha$ we accept H_1; if $\beta/(1-\alpha) < p_{1n}/p_{0n} < (1-\beta)/\alpha$ we infer that we do not have enough information and proceed to take another observation and repeat the entire procedure.

If the population covariance matrix is known, we have the sequential χ^2 -test with

$$p_{1n}/p_{0n} = e^{-n(\lambda_1^2 - \lambda_0^2)/2} \, {}_0F_1(p/2;\, n\lambda_1^2 \chi_n^2/4) / {}_0F_1(p/2;\, n\lambda_0^2 \chi_n^2/4)$$

where $\bar{x}$ is a p-element vector of sample means, based on n observations, $\chi_n^2 = n(\bar{x} - \mu_0) \sum^{-1} (\bar{x} - \mu_0)'$ and $F(c;x)$ represents the generalized hypergeometric function:

$$F(c;x) = 1 + \frac{x}{c} + \frac{x^2}{c(c+1)\,2!} + \frac{x^3}{c(c+1)(c+2)\,3!} + \cdots\cdots$$

If $\lambda_0^2 = 0$, the probability ratio reduces to:

$$p_{1n}/p_{0n} = e^{-n\lambda_1^2/2} \, {}_0F_1(p/2;\, n\lambda_1^2 \chi^2/4).$$

If the population covariance matrix is not known and must be estimated from the same sample as the sample means, we have the sequential T^2-test with

$$p_{1n}/p_{0n} = e^{-n(\lambda_1^2 - \lambda_0^2)/2} \; {}_1F_1\left[n/2,\ p/2;\ n\lambda_1^2 T_n^2/2\ (n-1+T_n^2)\right]$$

$$\div\ {}_1F_1\left[n/2,\ p/2;\ n\lambda_0^2\ T_n^2/2\ (n-1+T_n^2)\right]$$

where $\underset{\sim}{S}$ represents the sample covariance matrix based on n observations, $T_n^2 = n\,(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'\ \underset{\sim}{S}^{-1}\ (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)$ and ${}_1F_1\ (a,\ c;\ x)$ represents another generalized hypergeometric function

$${}_1F_1\ (a,\ c;\ x) = 1 + \frac{ax}{c} + \frac{a(a+1)x^2}{c(c+1)2!} + \frac{a(a+1)\ (a+2)x^3}{c(c+1)\ (c+2)3!} + \ldots$$

which is more familiarly known as the confluent hypergeometric function. If $\lambda_0^2 = 0$, the probability ratio reduces to:

$$p_{1n}/p_{0n} = e^{-n\lambda_1^2/2}\ {}_1F_1\left[n/2,\ p/2;\ n\lambda_1^2 T_n^2/2\ (n-1+T_n^2)\right].$$

Both of these procedures terminate with probability unity and the risks of incorrectly accepting H_1 and H_0 are approximately equal to α and β respectively.

rejected when the true means of all three characteristics are on standard. It will further be assumed that only a meager amount of information is available concerning the variability of these characteristics. From this information, it is inferred that the individual tolerances constitute limits of $\pm 3\sigma$ for individual observations about their standards and that there is no evidence that the variables are correlated. (The assumption of independence regarding tolerances is not always valid. In some types of missiles, total impulse and action time must be negatively correlated to insure a fixed range for their flight.)

Considering each characteristic separately, the requirement that 97.5 per cent of the lot must be within tolerances implies that the true lot mean for that characteristic cannot be closer than 2.24σ to either tolerance limit; conversely, the true mean must be within $.76\sigma$ of the standard since the tolerances were assumed to be $\pm 3\sigma$ limits. Several possibilities exist. One possibility would involve inscribing an ellipsoid inside the rectangular solid bounded by $\mu_i \pm .76\sigma_i$; $i = 1, 2, 3$. This would be rather restrictive and would be employed in the case where the specifications were to be strictly employed. The non-centrality parameter under τ_1 for a given characteristic would be $(.76)^2$ and since the occurance of an out-of-specification condition for any one of the variables is sufficient reason to reject the lot, $\lambda_1^2 = .5776$. Considering the crudeness of the determination of λ^2 in the first place, a value of $\lambda_1^2 = .5$ is probably quite adequate. Our hypotheses have been restated as:

$$H_0: \lambda^2 = 0$$

$$H_1: \lambda^2 = .5$$

Since the covariance matrix is unknown, we now employ the sequential T^2-test. On the average, we may expect to test 27 rounds if H_0 is true or 24 rounds if H_1 is true whereas the corresponding fixed-sample-size test would require 37 rounds. The fact that so many rounds are required in the example indicates that this procedure is considerably more exacting than the current methods now employed or conversely that the value of λ_1^2 implied by current procedures is considerably larger than .5.

A second method would be to circumscribe an ellipsoid around the rectangular solid representing the specifications. This is more conservative than the other method and would result in more material being accepted. This method can be used when the lot would be acceptable even if all of the characteristics were borderline. This would yield a value of $\lambda_1^2 = 3(.76) = 1.73$. Rounding to a value of $\lambda_1^2 = 2.0$, this procedure would require, on the average, 6 rounds to reach a decision under H_0 and 10 rounds under H_1 as compared with a fixed-sample size of 13. This should demonstrate quite adequately the problem associated with specifying the hypotheses in multivariate analysis. Other possibilities for setting up these hypotheses could involve acceptance sampling for variables techniques or sequential estimation procedures but such techniques have not yet been developed.

Case II: $\underset{\sim}{\Sigma}$ known

We will now assume that in the time which has elapsed since the operations carried on in the preceding section, sufficient information has been gathered so that the population covariance matrix can be assumed to be known. We may now use a sequential χ^2-test. Suppose that it turned out that the variances of these three variables were smaller than originally supposed so that the original tolerances were larger than $\pm 3\sigma$. This suggests several possibilities again. One possibility would be to use the natural tolerances of the process and allow λ_1^2 to remain at .5. This procedure is often employed when the acceptance sampling program is also used to control the process but in the case of Nike boosters, too much material of acceptable quality would be rejected and, considering the cost factor, this would not be a recommended procedure.

Other possibilities would involve inscribing or circumscribing an ellipsoid based on the now known covariance matrix about the tolerances. When variables are highly correlated, circumscribing can lead to acceptance of a fair amount of unsatisfactory material. Suppose that by inscribing an ellipsoid, we arrive at a value of $\lambda_1^2 = 1.0$. A sequential χ^2-test of this type would require, on the average, 13 rounds under H_0 and 9 rounds under H_1 compared with a fixed-sample size of 18.

5. Computations. It would be only fair to state that the computational requirements for the sequential T^2-test are not modest since the sample covariance matrix must be inverted for each observation. However, in ballistic missile testing, this cost is still negligible when compared to the actual cost of the round itself. The computations for the sequential χ^2-test are quite straight-forward requiring only vector by matrix multiplication.

6. Generalized χ^2-statistics. An additional technique which may be employed for Case II, the situation where the covariance matrix is known, involves the use of the generalized χ^2-statistics developed by Hotelling.[4] This allows us to compare not only the sample mean of a lot with the standard but also the covariance matrix of the sample with the previously established covariance matrix. It is, of course, possible for the mean of the lot to be close to standard but for the variability of individual rounds to be excessive enough to impair the overall quality of the lot anyhow. Techniques are now available to test sequentially the mean, covariance matrix, and if appropriate, the overall variability of the lot, although this last test is not as discriminating as the others and by the time this test had rejected a lot one of the other two would have probably already rejected it.

7. Acknowledgements. The research for the Part II of this paper was sponsored by the Office of Naval Research, Department of the Navy: Contract Number: NONR-2352(01), Task Order NR 042-019 with the Virginia Polytechnic Institute, Ralph A. Bradley, Principal Investigator. An article illustrating these techniques with numerical examples based on Radford Arsenal data is in preparation and should appear shortly.

Bibliography

1. Freund, R. J. and Jackson, J. E. "Tables to facilitate multivariate sequential testing for means." Technical Report #12. The Development of Statistical Methods for Experimental Designs in Quality Control and Surveillance Testing. Virginia Polytechnic Institute, Blacksburg, Virginia, September, 1960.

2. Grubbs, F. E. "On estimating precision of measuring instruments and product variability." J. Amer. Stat. Assoc., Vol. 43, 1948. pp. 243-264.

3. Hotelling, H. "Analysis of a complex of statistical variables into principal components." J. Educ. Psychol., Vol. 24, 1933. pp. 417-441, 498-520.

4. Hotelling, H. "Multivariate quality control." Techniques of Statistical Analysis, Ed. by Eisenhart, Hastay and Wallis, McGraw-Hill, New York, 1947, pp. 111-184.

5. Jackson, J. E. "Quality control methods for several related variables." Technometrics, Vol. 1, 1959. pp. 359-377.

6. Jackson, J. E. and Bradley, R. A. "Multivariate sequential procedures for testing means." Technical Report #10. The Development of Statistical Methods for Experimental Designs in Quality Control and Surveillance Testing, Virginia Polytechnic Institute, Blacksburg, Virginia; August 1959.

7. Jackson, J. E. and Morris, R. H. "An application of multivariate quality control to photographic processing." J. Amer. Stat. Assoc., Vol. 52, 1957, pp. 186-199.

8. Kruskal, W. H. and David, H. T. "Estimating of variances in bivariate sampling with partial information about parameters and both inherent variation and measurement errors present." Univ. of Chicago Report, SRC-50331 DK 22, March 31, 1955.

9. Pearson, K. "On lines and planes of closest fit to systems of points in space." Phil. Mag., Vol. 2 (6th Series), 1901, pp. 559-572.

10. Thompson, D. E. "Estimation of inherent propellant variation and errors of measurement (Confidential)." Allegany Ballistics Laboratory Report ABL/B-14, November 1956.

A TRIAL COMPARING CERTAIN SIDE EFFECTS OF TWO NERVE GAS ANTIDOTES, USING HUMAN SUBJECTS

C. A. deCandole, MD*
Defence Research Medical Laboratories
Downsview, Ontario, Canada

and

B. A. Richardson
Canadian Army Operational Research Establishment
Ottawa, Ontario, Canada

INTRODUCTION. This paper describes the experimental design of a trial performed at Camp Borden, Ontario, in February of 1960, with the cooperation of the Canadian Forces Medical Service, to compare certain side effects of two nerve gas antidotes, using human subjects. It must be made clear at the outset that the human subjects used were not exposed to nerve gas, since interest lay only in the side effects of the drugs under test.

Certain particulars of the drugs and dosages used in the trial and the numerical results obtained are security classified. In the present paper, therefore, reference to these topics will be made in coded or qualitative terms.

The accepted common nerve gas antidote, atropine, tends to induce undesirable side effects when used in the rather high dosage levels recommended. These effects include blurring of vision, nausea, disturbance of the pulse, dizziness, and a tendency to faint on sudden rising to the feet; all of which are clearly serious defects from the military point of view.

In Canada, C. A. deCandole has investigated a treatment that showed promise of both enhanced protective action against nerve gas as well as reduced side effects, when tested in the laboratory on animals and on a small number of human volunteers. By 1959 research had reached a point where a test on humans (for side effects) in a trial on a fairly large scale seemed worthwhile. In that trial, a modified form of the conventional atropine treatment (here called "Treatment A") was compared with de Candole's treatment (here called "Treatment B").

---

* Now at Suffield Experimental Station, Ralston, Alberta, Canada.

AIM. As relatively little was known about Treatment B, it was thought prudent first to learn something about the basic physiological consequences of its use before attempting to assess the effects of the two treatments upon military performance directly. A limited aim was therefore established, namely:

"To compare the physiological effects of Treatments A and B".

RESPONSE METAMETERS. The physiological effects to be recorded were as follows. First, obviously, any visible reactions; not only instances of fainting but also the less dramatic effects, if any appeared--tremor, restlessness, pallor, flushing, and so on. Next, subjective effects--dizziness, nausea, headache, thirst, for example. If these were not revealed by the test subject spontaneously, they were to be disclosed by direct questioning.

But qualitative and subjective data would not be enough: quantitative, objective data were needed as well. Now, it happens that some of the side effects of military importance are associated with disturbance of the cardiovascular system--the heart and blood vessels. Fainting, for example, can occur if the pulse pressure falls too low. It was therefore relevant to record the blood pressures and pulse rates, defined as follows:

a. The Systolic Blood Pressure is the peak pressure reached during the contraction of the heart.

b. The Diastolic Blood Pressure is the low during relaxation of the heart while it is refilling.

c. The Pulse Pressure is the difference between the first two.

d. The Pulse Rate is the number of systolic peaks per minute.

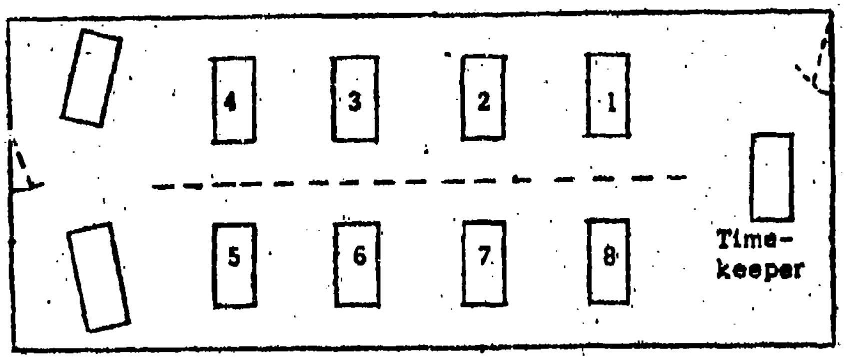
Three aspects of these four physiological parameters were of interest, namely:

a. their absolute values;

b. their response to the drugs; that is, their departure from normal after treatment;

c. the differential response; that is, the difference between the response to A and the response to B.

EQUIPMENT. The physical layout of the facilities provided for the trial is shown in Figure 1: a hospital ward with sufficient equipment for testing 8 subjects at a time, with one M.D. and one assistant at each testing station to act as observers and recorders. Three American medical doctors participated. We are pleased at their interest in the trial and grateful for their help.



LAYOUT OF WARD

Figure 1

The eight testing stations were arranged four along each wall, with a screen running down the center aisle so that the test subjects would not directly face one another. At each end of the room were tables for the use of the Project Officer and his assistants one of whom acted as timekeeper.

Each testing station was outfitted with a tilt table, for the test subject to lie upon. Also (not shown in Figure 1), at the head (outer) end of each tilt table there was a small table to hold instruments and other equipment; and at the foot, a drawing board with blank data-recording forms, and an over-bed table for other necessary equipment. The tilt table could be laid flat or tilted quickly upright to an angle of 85 degrees. The subject was strapped lightly to the table to keep him from pitching forward on his face on tilting. The purpose of this tilting was to simulate, in a standard fashion, the act of suddenly rising to the feet.

The pulse rates were recorded by means of the electrocardiograph. The electrocardiograph is more accurate than digital palpation, and it gives a permanent record that can be examined at leisure. Blood pressures were measured by the ordinary sphygmomanometer, since no satisfactory instrument of the recording type was available at Camp Borden.

CONTROL OF SOURCES OF VARIATION IN RESPONSE. There are many possible sources of variation in response, as biological systems can be extremely sensitive to small changes in the conditions to which they are subjected. The more important disturbing factors in the present instance fall into four categories: those dependent on:

a. the drugs used;

b. the test subject;

c. the environment;

d. the observational technique.

First, the drug factors: there are at least four of these:

a. the route of administration;

b. the size (volume) of the dose administered;

c. other ingredients in the formulation;

d. the concentration of the active ingredients.

The effects of these factors were eliminated by restricting the scope of the trial. Both treatments, A and B, were administered by the same one route: intramuscular injection. The dose for each treatment, that is, the quant. y of active ingredients injected, was kept constant, not varied in proportion to the subject's body weight. Only one formulation of each mixture was tested. Concentrations were chosen so that the volume injected would be the same for both treatments.

The test subjects introduce personal factors of two types: fixed and variable. The fixed factors include:

a. body weight;

b. height;

c. age;

d. normal blood pressures and pulse rate;

e. medical history;

f. idiosyncrasies.

The variable factors include:

a. time;

b. posture;

c. rate of tilt;

d. physical state;

e. mental state.

Of the fixed factors, the effects of body weight, height, age, and normal levels of blood pressure and pulse rate can be reduced by using matched pairs of subjects, giving A to one and B to the other. Alternatively, one can use the same subject twice; but this changes his medical history and so is an advantage only if his response characteristics (his idiosyncrasies) remain unaffected. A "used" man might conceivably give a worse comparison of the two treatments than would be obtained using a second, "fresh" one. So to get the advantage either way, matched pairs were used and each man was exposed twice, alternating the treatments.

The total number of test subjects was 56. The personal characteristics of this group were relatively homogeneous, except with regard to body weight. Body weight was thus the major difference among the resulting 28 pairs.

Time, posture and rate of tilt are the factors whose effects were under examination, and so these were varied deliberately.

The effects of hunger, fatigue and other physical and physiological states were minimized by restricting the free-time activities of the subjects: special meal schedules, no heavy exercise, no alcohol, early to bed. To control mental and psychological factors, the subjects were given a thorough briefing in advance to reduce possible apprehension and fear. Excitement was reduced by maintaining an atmosphere of calm and relaxation in the testing ward.

Three environmental factors can be identified:

a. the climate of the ward;

b. the location of the testing station;

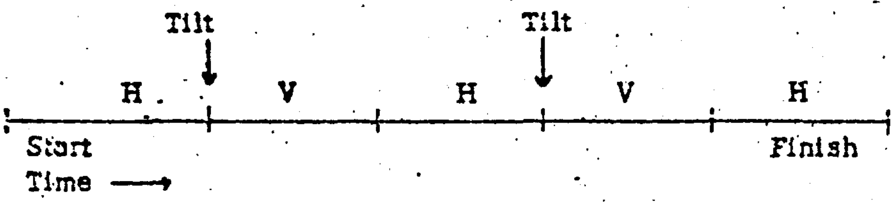
c. incidents occurring during the course of a test that might affect neighboring subjects.

These factors were minimized by testing the subjects in groups of 8 (that is, 4 pairs). The effects of any change in the general climate of the ward, temperature, humidity and noise level and so on, would thus show up in the variation between groups. To control differences between stations the subjects were assigned at random; but they were tested at the same station on both test occasions. Sporadic events occurring during the trial might affect response but they would surely be distributed at random.

The observing teams were assigned to stations at random, at the beginning of the trial, but rotated two positions for the second round of tests so that they would not handle the same subject twice. To eliminate personal bias in the physicians--reading blood pressure is still more an art than a science--they were drilled in a standard technique and required to use it regardless of their own inclinations. In addition, neither they nor anyone but the Project Officer knew the identity of the treatment given in any instance.

The instruments were calibrated before the trial. Their residual variation is regarded as negligible.

PROCEDURE. During the course of the trial the subjects underwent a series of posture changes as illustrated in Figure 2: horizontal, vertical, horizontal, vertical, horizontal. Such a series of changes will be called a sequence. Blood pressures and pulse rate were recorded at a set of intervals within each sequence. The readings and posture changes were made simultaneously at all 8 stations, on signal from the timekeeper. The same schedule was used for all sequences.



POSTURE SEQUENCE

Figure 2

Each subject underwent two sequences in immediate succession. The drugs were injected at the end of the first or control sequence, thus marking the start of the second, or drug sequence. The pair of sequences together constitute a session.

The 56 test subjects required 7 sessions in all. Every subject was tested on two occasions separated by an interval of 2 days. The order of testing was the same on both occasions. The two members of each pair or mates, were always tested together in the same session. One received Treatment A and the other Treatment B on each occasion, with the treatments reversed for the second. The mates, therefore, fall into two classes according to the order of treatment: the A-firsts, or AB's, and the B-firsts or BA's.

TRIAL DESIGN. The shape of the trial design is illustrated in Figure 3. The basic structure consists of 28 Latin Squares of order 2 aranged in seven blocks of 4. One such block is shown here. Each Latin Square represents one pair of subjects. Rows represent occasions, columns are mates, and the treatments occupy the diagonals. In the third dimension, the four quadrants represent the sessions. The quadrants could be shown as split in this dimension, representing the control and drug sequences, but if we confine our attention to Response, -- that is, Drug minus Control -- the split aspect disappears and the picture is as shown here.

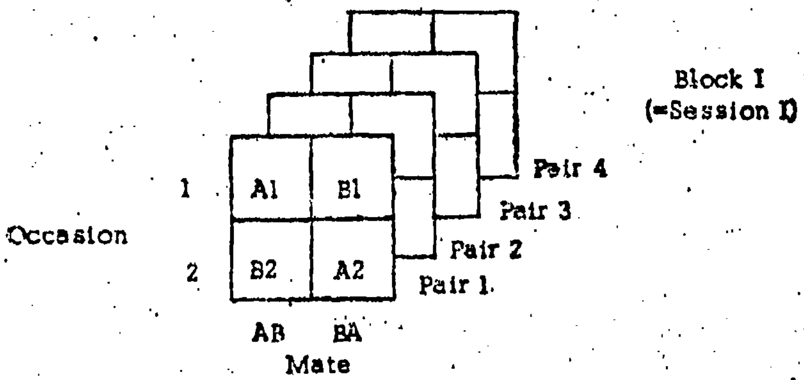


Figure 3

To visualize the entire trial, the other six blocks can be imagined as receding into the background behind Block I. The points in time might be represented by rows of blocks in succession left to right, and the four parameters thought of as stacked one on top of another.

RESULTS. During the trial two test subjects fainted, and among the others a number suffered reactions to Treatment B severe enough to make their withdrawal from the trial advisable. As a result, many readings were lost--so many that use of a missing-values routine was out of the question. Analysis was therefore confined to pairs of subjects for whom complete sets of readings were available, each reading time being considered independently.

As an example of the kind of results obtained, the responses for each of the four physiological parameters at the time of maximum effect--just before and just after the second tilt to vertical--are shown in Tables 1 and 2. These represent the period of greatest interest; a presentation of the complete results would be out of place here. The tabular entries are the results of F tests of the mean and of each of the four main factor effects, using the residual variance as error. The symbols have the conventional meanings. For significant means, the direction of the response is also given, plus and minus signifying increase and decline, respectively. Similarly, for each significant factor effect, the level showing the greatest response in the same direction as the mean is indicated. N is the number of (complete) pairs available in each case.

TABLE 1

Effects at Time of Maximum Response: Subjects Horizontal

Contrast	Physiological parameter			
	Systolic BP (N=23)	Diastolic BP (N=23)	Pulse Press. (N=22)	Pulse Rate (N=21)
Mean (vs zero)	**(+)	**(+)	*(+)	**(+)
Treatment (A vs B)	**(B)	**(B)	**(B)	NS
Occasion (1st vs 2nd)	NS	NS	NS	NS
Order (AB vs BA)	NS	NS	NS	NS
Between pairs	**	**	NS	**

TABLE 2

Effects at Time of Maximum Response: Subjects Vertical

Contrast	Physiological parameter			
	Systolic BP (N=19)	Diastolic BP (N=14)	Pulse Press. (N=14)	Pulse Rate (N=22)
Mean (vs zero)	**(-)	**(+)	**(-)	**(+)
Treatment (A vs B)	**(A)	**(B)	*(A)	**(A)
Occasion (1st vs 2nd)	NS	NS	NS	**(1)
Order (AB vs BA)	NS	NS	NS	NS
Between pairs	NS	NS	NS	**

CONCLUSIONS. We are not here concerned with a specific physiological interpretation of the results; it is sufficient to note that the results were unequivocal. The main conclusions were as follows:

a. Both treatments produce definite responses in subjects in either posture, horizontal or vertical.

b. The two treatments distinctly differ in magnitude of response, except in that of pulse rate in horizontal subjects.

c. Treatment B effectively maintains the pulse pressure on change of posture from horizontal to vertical, as intended, but at an unacceptable price in new and unforeseen side effects.

d. Day-to-day differences and order of administration of the treatments can probably be safely disregarded in any future trials of Treatment B or modifications of it.

e. There was significant variation among pairs of subjects. Now, this difference represents the combined effects of all personal factors plus any variation between sessions. Body weight is probably the dominant factor; if so, the recorded body weights should account for most of the difference observed. This portion of the data would no doubt repay further analysis.

SUMMARY. This paper has described a trial performed on human subjects to compare the physiological side effects produced by two nerve gas antidotes, "Treatment A" and "Treatment B". Treatment B was designed to avoid certain side effects that tend to accompany Treatment A, these side effects being undesirable from the military point of view.

The trial clearly showed Treatment B to be superior to Treatment A in the one respect of special interest, but revealed that it introduced new and unforeseen side effects that were themselves undesirable. The trial thus illustrates the need to proceed with caution in complex circumstances, and to examine fundamentals before attacking even more complex problems such as the evaluation of military performance.

This trial is presented as an example of the usefulness of formal experimental design and analysis of variance in an area where it is not yet regularly applied. We hope its success in producing some clear-cut answers in the presence of many complicating factors will help spread awareness of these valuable techniques.

DESIGN OF AN EXPERIMENT FOR THE MOST EFFICIENT CONDUCT OF SAFETY, RELIABILITY AND PERFORMANCE TESTS OF FUZES IN THE DESIGN AND DEVELOPMENT STAGES

Gertrude Weintraub
Missile Warhead and Special Projects Laboratory, Picatinny Arsenal

PROBLEM. To design an experiment for the efficient conduct of tests to determine the operational and safety characteristics of fuzes being developed for a missile warhead.

STATEMENT OF PROBLEM. Fuzes are designed to accomplish a particular mission in the successful operation of ordnance ammunition such as mines, warheads and missiles.

Before they are used in their ultimate mission, they are subjected to various environmental treatments such as vibration and waterproofness tests to insure their proper functioning and safety for use. During the conduct of these tests, yes-no responses as well as quantitative measurements are obtainable. Moreover, tests are generally conducted using rather small size samples to determine the functioning and safety characteristics of the fuze at the design level. Based upon the results of these tests, performance and reliability estimates are made. In addition, engineering judgment and previous experience are also usually involved in making reliability estimates. The fuze is then judged to function properly and to be safe for use.

a. Fuze Design Characteristics

Fuzes are generally composed of several major components whose co-functioning affects the overall fuze performance. Also, incorporated in the fuze design are various safety characteristics.

b. Types of Environmental Treatments

The various types of environmental treatments to which fuzes are likely to be exposed include the following:

(1) Transportation Vibration

(2) Rough Handling

(3) Aircraft Vibration

(4) Temperature and Humidity

(5) Vacuum-Steam Pressure

(6) Salt Spray

(7) Waterproofness

(8) Weathering (Exposed)

(9) Jolt and Jumble

(10) Low Drop

(11) Detonator Safety

c. Evaluation Tests

These measure the functioning and safety of the fuzes after they have been exposed to the environmental tests. They include the following:

(1) Inspection

(2) Self-Destruction

(3) 40 Ft. Drop

(4) Functioning

PROPOSED STATISTICAL PROGRAM TO BE IMPLEMENTED. The following types of statistical programs are currently being used at Picatinny Arsenal in the development of fuzes. The first is a factorial experiment designed to detect design and material differences in various components. The second plan makes use of increased severity testing to reduce the required sample sizes to allowable limits.

a. For Component Testing

Before testing the overall fuze performance for functioning and safety, it is mandatory that each of the major components of the fuze be quality controlled to insure against defective material. Also, in the event of the possible application of alternate materials for particular components, each of the alternate materials should be subjected to test in order to insure that the best material (the one which yields the highest degree of functioning reliability or accuracy) is selected. A factorially designed experiment incorporating the various alternate materials together with the various environmental treatments can be set up. Test results would indicate the particular environmental treatment or treatments which show significant failure rate due to the effect of such treatments. Also obtainable therefrom will be the failure rate for each of the alternate types of materials for the component. After these are determined, engineering judgment can be invoked to determine causes for failure and modifications can be made to remedy such causes. Also, the material yielding the lowest failure rate can be isolated and further design effort concentrated on that type of material which yields the highest degree of functioning reliability. After the major components have been pre-tested and "bugs" withdrawn, they can be incorporated into the overall fuze design.

b. For Complete Fuzes

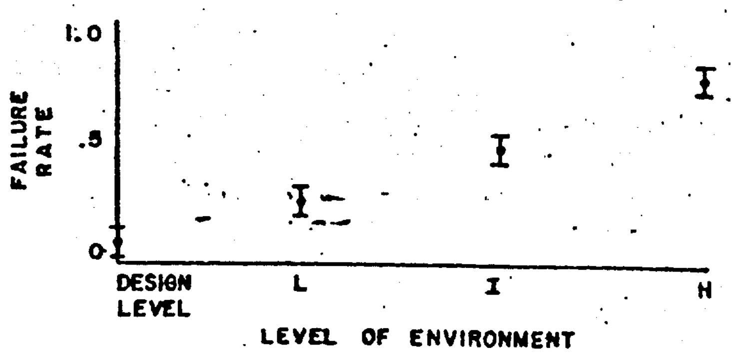
Fuzes will be subjected to a series of environmental treatments in sequence in a manner simulating that in which they are normally expected to occur. Increased severity levels of these environmental treatments will be selected on the basis of engineering judgment as being those which appear to be most likely to cause failures during use. The statistical test plan encompasses fractional factorial designed experiments which would subject a minimum of fuze samples to environmental treatments in sequence. Each environmental treatment would consist of two levels, the absence of the particular treatment and the presence of the particular treatment. Moreover, two types of response data could be elicited therefrom. One, namely, attribute data and the other, variable data which measures a continuous function like arming time, self-destruction time, sustainer switch functioning time, etc. The statistical test plan philosophy provides for the deliberate inducing of increased severity levels* of each of the environmental treatments. These levels will include the following:

High Level - - - - - - - - - - Where a high proportion of failures is expected.
Intermediate Level - - - - - - Where a moderate proportion of failures is expected.
Low Level - - - - - - - - - - Where a small proportion of failures is expected.

* The selection of levels will be based on engineering judgment.

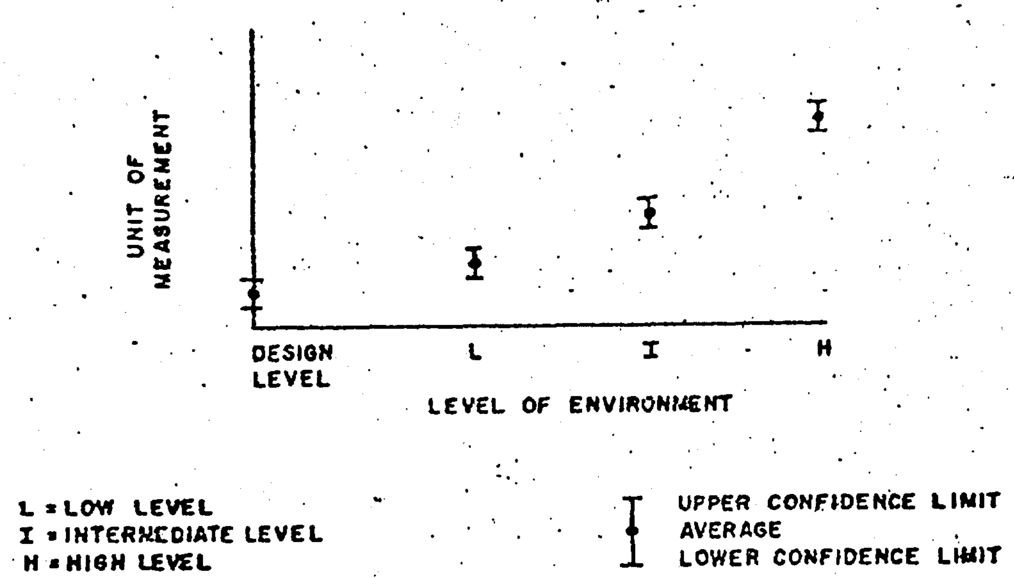
The test plan will generate information of the failure rate in the absence and in the presence of a particular environmental treatment. Also, it will produce variable type response data in the absence and in the presence of a particular environmental treatment which later can be translated to probability of successful functioning. Failure rate distribution curves will be obtainable for individual environmental treatments. These curves will show failure rate as a function of level of severity of an individual environmental treatment and of multiple environmental treatments. The following types of curves will be obtainable:

FAILURE RATE AS A FUNCTION OF LEVEL OF ENVIRONMENTAL TREATMENT
(ATTRIBUTE DATA)



VARIABLE MEASUREMENT AS A FUNCTION OF ENVIRONMENTAL TREATMENT

II

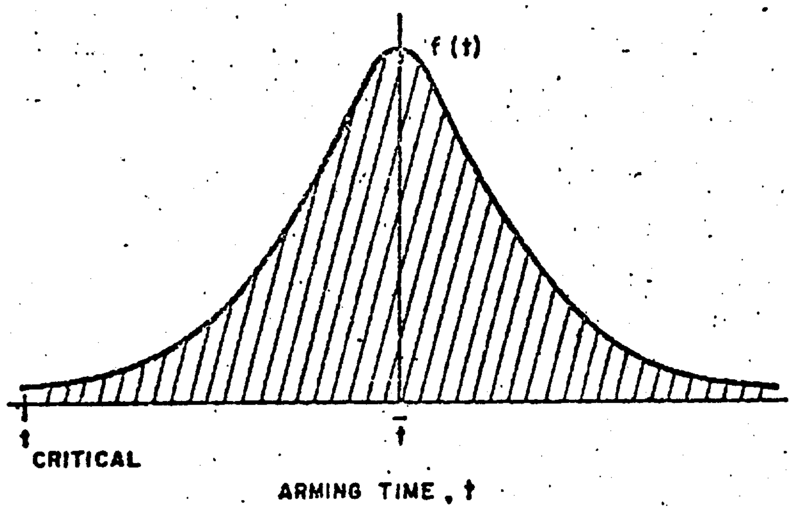


Curve I will be obtainable for each of the environmental treatments and 2-factor interactions for the attribute response.

Curve II will be obtainable for each of the environmental treatments and 2-factor interactions for each of variable responses. Also, it is expected that Curve II can be translated to a curve of probability of successful functioning as a function of level of environmental treatment.

The aforementioned plan is presented as an alternate approach to testing a prohibitively large number of samples at the design level in order to insure reliable functioning. This approach proposes to accomplish the same objective with a small number of samples by obtaining failure rate distributions over the range of increased severity levels for each of the environmental treatments imposed upon the fuzes. Also, actual measurement data of specific fuze functions will be obtainable over the range of increased severity levels of each of the environmental treatments. These data can then be translated to probability of functioning. For example, if we define the probability of a successful function as the probability of the continuous variable, e.g., arming time, being greater than a given critical arming time or lying within a given acceptable region, the probability of success or the reliability can be computed from Curve II data. Thus the probability of arming time being greater than a minimum arming time, t critical, is as follows:

$$P(t \geq t_{critical}) = \int_{t_{critical}}^{\infty} f(t)\, dt$$



It should be noted that the aforementioned proposed program represents the approach currently being implemented since it appears to be the best approach to date to our reliability problem. Although the program does not encompass the correlation aspects for the case of multiple responses, it is contemplated that such aspects are entirely possible. However, since co-relationship may exist among several attribute responses, among several variable responses, and also among attribute and variable responses, the manner in which these multiple responses can be handled still remains to be investigated.