AD NUMBER

AD476507

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; Oct 1965. Other requests shall be referred to Army Research Office, Research Triangle Park, NC.

AUTHORITY

USARO ltr, 21 Nov 1966

ARO-D Report 65-3

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

Sponsored by

The Army Mathematics Steering Committee

on Behalf of

THE OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 65-3
October 1965

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

Sponsored by the Army Mathematics Steering Committee

Host

The Army Research Office, Office Chief of Research and Development
Department of the Army
Washington, D. C.
4-6 November 1964

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

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd) Page

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

TABLE OF CONTENTS (cont'd) Page

---

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

TABLE OF CONTENTS (cont'd) Page

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

TABLE OF CONTENTS (cont'd)

* We anticipate that this paper will appear in the Proceedings of the Eleventh Conference on the Design of Experiments

FOREWORD

The Army Research Office, Office of the Chief of Research and Development, Department of the Army, served as host for the Tenth Conference on Design of Experiments in Army Research, Development and Testing. The Conference was held in Washington, D. C. during 4-6 November 1964.

The continued success of these conferences is a tribute to the foresight of Professor Samuel S. Wilks who conceived the idea of holding such conferences and chaired the Program Committee for the first nine conferences. Unfortunately, due to his untimely death, Professor Wilks could not participate in this Tenth Anniversary Conference. His effort in connection with these Conferences was only one of Professor Wilks' many contributions to the Army. His wise counsel and advice will be missed. As a small recognition for his services to the Army, this Tenth Anniversary Conference was dedicated to the memory of Professor Wilks.

Almost 300 statisticians, engineers and physicists from the Army, other government agencies, Army contractors, and universities attended the conference. This number far exceeds the attendance at any of the previous conferences and reflects, in part, the esteem for Professor Wilks in the statistical community.

One surprising feature was the announcement that Mr. Philip G. Rust of Thomasville, Georgia, had contributed funds for a Samuel S. Wilks Award to be presented annually at the Design of Experiments Conference. It is especially gratifying that a long-time civilian employee of the U. S. Army, Dr. Frank Grubbs, Associate Technical Director of the Ballistic Research Laboratories, was the recipient of the initial award. We are appreciative that the American Statistical Association has accepted the responsibility for determining future Award winners.

Because of the particular significance of this Tenth Conference, the Program Committee invited several distinguished statisticians to deliver papers: Professor H. O. Hartley, Professor Oscar Kempthorne, Dr. M G. Kendall and Professor John W. Tukey. Professor Gerald J. Liebeiman served as chairman of the Panel Discussion on Regression Analysis and arranged for Professor G.E. P. Box, Professor Jack C. Kiefer, and Professor Ingram Olkin to give pertinent papers and for

Professor Robert Bechhofer to serve as the invited discussant. In addition to these invited addresses, 13 papers were given in the Clinical Sessions and 18 papers in the Technical Sessions. Additional highlights of the meetings were the after dinner presentations by Dr. Churchill Eisenhart and Dr. W. J. Youden.

It is fitting to give recognition for the particular activities of two groups with regard to these Conferences. The Army Mathematics Steering Committee (AMSC), currently chaired by Dr. I. R. Hershner, Jr. is commended for its strong support of these Conferences because of the actual and potential gains obtained by Army facilities. The members of the Tenth Conference Program Committee are commended for their work in obtaining speakers, selecting a location and planning the overall program. The members of this Committee were: Dr. F. G. Dressel (Secretary), Mr. Fred Frishman, Dr. Walter D. Foster, Dr. Frank E. Grubbs (Chairman), Professor Boyd Harshbarger, Professor H. L. Lucas, Dr. Clifford J. Maloney, Professor Henry B. Mann and Professor Geoffrey S. Watson. Special credit is given to Dr. F. G. Dressel for performing all of the necessary details regarding the program, invitations and the publication of these Proceedings.

It is planned to have an Eleventh Conference at Picatinny Arsenal in 1965. As is well known, these Conferences have been held to assist Army statisticians and their parent organizations. It is hoped that Army statisticians will continue to support these conferences both by the presentation of scientific papers and by their attendance.

WALTER E. LOTZ, JR.
Director of Army Research

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

4-6 November 1964

Wednesday, 4 November

0800-0900 REGISTRATION -- Mezzanine Floor in Foyer No. 3 of the Statler-Hilton Hotel

0900-0920 CALLING OF CONFERENCE TO ORDER -- South American Room, Fred Frishman, Chairman on Local Arrangements

0920-1200 GENERAL SESSION 1

Chairman: Major General Austin W. Betts, Deputy Chief of Research and Development

THE STIMULUS OF S. S. WILKS TO ARMY STATISTICS
Major General Leslie E. Simon (Ret'd), Winter Park, Florida

THE SAMUEL S. WILKS AWARD

Announcement: Don Riley, American Statistical Association

Presentation: Philip G. Rust, Thomasville, Georgia

BREAK

DEVELOPMENT OF THE DESIGN OF EXPERIMENTS OVER THE PAST TEN YEARS
Professor Oscar Kempthorne, Iowa State University, Ames, Iowa

1200-1320 LUNCH

Technical Sessions I and II and Clinical Session A will start at 1320 and run to 1500. After a break Technical Sessions III and IV and Clinical Session B will convene at 1540 and run to 1710.

1320 - 1500 TECHNICAL SESSION I -- New York Room

Chairman: W. H. Ewart, Research and Development Division, Army Missile Command, Redstone Arsenal, Alabama

APPLICATION OF DIMENSION THEORY TO MULTIPLE REGRESSION ANALYSIS
David R. Howes, U. S. Army Strategy and Tactics Analysis Group, Bethesda, Maryland

THE USE OF REGRESSION ANALYSIS FOR CORRECTING OF MATRIX EFFECTS IN THE X-RAY FLUORESCENCE ANALYSES OF PYROTECHNIC COMPOSITIONS
R. H. Myers and B. J. Alley, Virginia Polytechnic Institute, Blacksburg, Virginia, Rep. Redstone Arsenal

1320 - 1500 TECHNICAL SESSION II - South American Room

Chairman: Henry Ellner, Directorate for Quality Assurance, Edgewood Arsenal, Maryland

SAMPLING FOR DESTRUCTIVE OR EXPENSIVE TESTING
Joseph Mandelson, Directorate of Quality Assurance, U. S. Army Edgewood Arsenal, Edgewood Arsenal, Md.

TOTAL SAMPLE STATISTICS FROM SUBSAMPLE STATISTICS
Paul C. Cox, Reliability and Statistics Office, Army Missile Test and Evaluation Directorate, White Sands Missile Range, New Mexico

1320 - 1500 CLINICAL SESSION A -- California Room

Chairman: Ira A. DeArmon, Jr., Operations Research Group, Army Chemical Corps, Edgewood Arsenal, Md.

Panelists:

Dr. Frank E. Grubbs, Army Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

Professor H. C. Hartley, Institute of Statistics, Agricultural and Mechancial College, College Station, Texas

Panelists (cont'd):

Dr. Emil H. Jebe, Institute of Science and Technology, The University of Michigan, Ann Arbor, Michigan

Professor Gerald J. Lieberman, Stanford University, Stanford, California

Professor H. L. Lucas, Institute of Statistics, North Carolina State of the U.N.C., Raleigh, North Carolina

SYSTEM CONFIGURATION PROBLEMS AND ERROR SEPARATION PROBLEMS
Fred S. Hanson, Plan and Operations Directorate, White Sands Missile Range, New Mexico

AN EXPERIMENT IN MAKING TECHNICAL DECISIONS USING OPERATIONS RESEARCH AND STATISTICAL METHODS
Andrew H. Jenkins, U. S. Army Missile Command, Huntsville, Alabama, and Edwin M. Bartee, School of Engineering, University of Alabama

1500 - 1540 BREAK

1540 - 1710 TECHNICAL SESSION III -- New York Room

Chairman: Morris A. Rhian, Operations Research Group, Army Chemical Corps, Edgewood Arsenal, Md.

IMPROVEMENT CURVES: PRINCIPLES AND PRACTICES
Jerome H. N. Selman, Stevens Institute of Technology, Rep. the U. S. Army Munitions Command, Dover, N. J.

THE EFFECT OF VALIDITY, LENGTH, AND SCORE CONVERSION ON A MEASURE OF PERSONNEL ALLOCATION EFFICIENCY
Richard C. Sorenson and Cecil D. Johnson, U. S. Army Personnel Research Office, Washington, D. C.

1540 - 1710 TECHNICAL SESSION IV -- South American Room

Chairman: Joseph R. Lane, Technical Evaluation Office, Army Research Office-Durham, Durham, N. C.

A QUANTITATIVE ASSAY FOR CRUDE ANTHRAX TOXINS
Bertram W. Haines, U. S. Army Biological Labs., Fort Detrick, Frederick Maryland

AN INVESTIGATION OF THE DISTRIBUTION OF DIRECT HITS ON PERSONNEL BY SELF-DISPERSING BOMBLETS
David M. Moss and Theodore W. Horner, Booz-Allen Applied Research, Inc., Bethesda 14, Maryland
Rep. Biomathematics Division of Fort Detrick, Maryland

1540 - 1710 CLINICAL SESSION B -- California Room

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

Panelists:

Dr. O. P. Bruno, Surveillance Group, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.

Dr. Donald S. Burdick, Duke University, Durham, N. C.

Professor Clyde Y. Kramer, Virginia Polytechnic Institute, Blacksburg, Virginia

Dr. R. L. Stearman, C-E-I-R, Inc., Los Angeles, Calif.

Dr. William Wolman, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland

EXPLOSIVE SAFETY AND RELIABILITY ESTIMATES FROM A LIMITED SIZE SAMPLE
J. N. Ayres, L. D. Hampton and I. Kabik, U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland

COMPARING THE VARIABILITIES OF TWO TEST METHODS USING DATA FOR SEVERAL POPULATIONS
Manfred W. Krimmer, U. S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois

Thursday, 5 November

Technical Session V and Clinical Session C and D will run from 0830-1010. After the break General Session 2 will convene at 1050. After lunch Technical Sessions VI and VII and Clinical Session E will start at 1300 and end at 1420. The Panel Discussion is scheduled to be conducted from 1450 to 1710. Following the banquet, which starts at 1900, there will be two short talks.

0830-1010 TECHNICAL SESSION V -- South American Room

Chairman: R. H. Myers, Statistical Laboratory, Virginia Polytechnic Institute, Blacksburg, Virginia

CYCLIC DESIGNS
H. A. David and F. W. Wolock, University of North Carolina and Virginia Polytechnic Institute, Rep. Army Research Office-Durham

SOME RESULTS ON THE FOUNDATIONS OF STATISTICAL DECISION THEORY
Bernard Harris, J. D. Church, F. V. Atkinson, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

0830-1010 CLINICAL SESSION C -- California Room

Chairman: Dr. Erwin L. LeClerg, Biometrical Services Division, U. S. Department of Agriculture, Plant Industry, Beltsville, Maryland

Panelists:

Dr. Walter D. Foster, Biometrics Division, Army Biological Warfare Laboratories, Fort Detrick, Md.

Dr. Samuel W. Greenhouse, Biometrics Branch, National Institute of Mental Health, Bethesda, Maryland

Panelists (cont'd):

Professor Clyde Y. Kramer, Virginia Polytechnic Institute, Blacksburg, Virginia

Professor H. L. Lucas, North Carolina State of the UNC, Raleigh, North Carolina

Dr. Clifford J. Maloney, Division of Biologics Standards, National Institutes of Health, Bethesda, Maryland

DISINFECTION OF AEROSOLIZED PATHOGENIC FUNGI ON LABORATORY SURFACES
Richard H. Kruse, Theron D. Green, Richard C. Chambers and Marian W. Jones, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

THE EFFECT OF SNAKE VENOM AND ENDOTOXIN ON CORTICAL ELECTRICAL ACTIVITY
James A. Vick, Henry P. Ciuchta, Edward H. Polley, and James Manthei, Directorate of Medical Research, Chemical Research and Development Laboratories, Edgewood Arsenal, Maryland

COMPUTER ANALYSIS OF RHESUS MONKEY IN VISUAL DISCRIMINATION TESTING
John C. Atkinson, Directorate of Medical Research, Chemical Research and Development Laboratories, Edgewood Arsenal, Maryland

0830-1010 CLINICAL SESSION D -- New York Room

Chairman: Lee W. Green, Jr., Florida Research and Development Center, Pratt and Whitney Aircraft, West Palm Beach, Florida

Panelists:

Professor R. E. Bechhofer, Cornell University, Ithaca, New York

Professor G. E. P. Box, the University of Wisconsin, Madison, Wisconsin

Panelists (cont'd):

Dr. T. W. Horner, Booz-Allen Applied Research, Inc., Bethesda, Maryland

Professor G. J. Lieberman, Stanford University, Stanford, California

Dr. H. B. Mann, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

FATIGUE - LIMIT ANALYSES AND DESIGN OF FATIGUE EXPERIMENTS
A. H. Soni and R. E. Little, Oklahoma State University, Stillwater, Oklahoma. Representing Army Research Office-Durham

GETTING REGRESSION ANALYSIS IMPLEMENTED
W. H. Ammann, U. S. Army Aviation Materiel Command, St. Louis, Missouri

1010 - 1050 BREAK

1050 - 1150 GENERAL SESSION 2 -- South American Room

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

ASSESSMENT AND CORRECTION OF DEFICIENCES IN PERT
Drs. H. O. Hartley and A. W. Wortham, Institute of Statistics, Texas A and M University, College Station, Texas

1150 - 1300 LUNCH

1300 - 1420 TECHNICAL SESSION VI -- South American Room

Chairman: Leonard Pepper, Concrete Division, U. S. Army Engineer Waterways Experiments Station, Vicksburg, Miss.

TEQUILAP: TEN QUANTITATIVE ILLUSIONS OF ADMINISTRATIVE PRACTICE
Clifford J. Maloney

COMBAT VEHICLE FLEET MANAGEMENT
C. J. Christianson and Mr. G. F. Cooper, Research Analysis Corporation, McLean, Virginia

1300 - 1420 TECHNICAL SESSION VII -- New York Room

Chairman: Eugene F. Smith, Concrete Division, U. S. Army Waterways Experiment Station, Vicksburg, Miss.

APPLICATION OF STATISTICS TO EVALUATE SWIVEL HOOK TYPE CROSS CHAIN FASTENERS FOR MILITARY APPLICATIONS OF TIRE CHAINS
Otto H. Pfeiffer, Components Research and Development Labs., Army Tank-Automotive Center, Warren, Michigan

SOME FACTORS AFFECTING THE PRECISION OF CO-ORDINATE MEASUREMENTS ON PHOTOGENIC PLATES
Desmond O'Connor, Research and Analysis Division, U. S. Army Engineer Geodesy, Intelligence and Mapping Research and Development Agency, Fort Belvoir, Virginia

1300 - 1420 CLINICAL SESSION E -- California Room

Chairman: Joseph Mandelson, Directorate of Quality Assurance, Edgewood Arsenal, Maryland

Panelists:

Professor Donald S. Burdick, Duke University, Durham, North Carolina

Dr. Bernard Harris, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wis.

Professor Ingram Olkin, Stanford University, Stanford, California

Dr. H. M. Rosenblatt, Statistical Research Division, Bureau of the Census, Washington, D. C.

Professor G. S. Watson, The Johns Hopkins University, Baltimore, Maryland

ERROR ANALYSIS PROBLEMS IN THE ESTIMATION OF SPECTRA
Virginia Tipton, Plans and Operations Directorate, White Sands Missile Range, New Mexico

VALIDATION PROBLEMS OF AN INTERFERENCE PREDICTION MODEL
William B. McIntosh, Army Electronics Proving Ground, Fort Huachuca, Arizona

1420 - 1450 BREAK

1450 - 1710 GENERAL SESSION 3 -- South American Room

PANEL DISCUSSION ON REGRESSION ANALYSIS
Chairman: Professor Gerald J. Lieberman, Stanford University

Panelists and the Titles of their Addresses:

USE AND ABUSE OF REGRESSION
Professor G. E. P. Box, The University of Wisconsin

OPTIMUM EXTRAPOLATION AND INTERPOLATION DESIGNS
Professor Jack C. Kiefer, Cornell University

ESTIMATION FOR A REGRESSION MODEL WITH CONVARIANCE
Professor Ingram Olkin, Stanford University

Discussant: Professor Robert Bechhofer, Cornell University

1900 BANQUET

Evening Session Chairman: Dr. I. R. Hershner, Jr., ARO

SAM WILKS AS I REMEMBER HIM
Dr. Churchill Eisenhart, National Bureau of Standards, Washington, D. C.

AN OPERATIONS RESEARCH YARN AND OTHER COMMENTS
Dr. W. J. Youden, National Bureau of Standards, Washington, D. C.

Friday, 6 November

Technical Sessions VIII and IX as well as Clinical Session F run from 0830 to 0950. General Session 4 will start at 1020 and end at 1220.

0830-0950 TECHNICAL SESSION VIII -- South American Room

Chairman: Donald S. Burdick, Duke University, Durham, N. C.

THE DESIGN OF COMPLEX SENSITIVITY EXPERIMENTS
D. Rothman and J. M. Zimmerman, Mathematic and Statistics Group, Rocketdyne, A Division of N. American Aviation, Canoga Park, Calif. Rep. George C. Marshall Space Flight Center, NASA, Huntsville, Alabama

FACTORS AFFECTING SENSITIVITY EXPERIMENTS
J. R. Kniss and W. Wenger, U. S. Army Ballistic Research Labs., Aberdeen Proving Ground, Maryland

0830-0950 TECHNICAL SESSION IX -- New York Room

Chairman: Ralph E. Brown, U. S. Army Munitions Command, Philadelphia, Pennsylvania

A COMPARISON OF RECONNAISSANCE TECHNIQUES FOR LIGHT OBSERVATION HELICOPTERS AND A GROUND SCOUT PLATOON
Harrison N. Hoppes, Barry M. Kibel, Arthur R. Woods, Research Analysis Corporation, McLean, Virginia

A STUDY OF PROBABILITY ASPECTS OF A SIMULATANEOUS SHOCK WAVE PROBLEM
Edward C. Hecht, Nuclear Engineering Directorate, Picatinny Arsenal, Dover, New Jersey

0830-0950 CLINICAL SESSION F -- California Room

Chairman: Dr. B. W. Haines, U. S. Army Biological Laboratories, Fort Detrick, Maryland

Panelists:

Professor R. E. Bechhofer, Cornell University, Ithaca, New York

Mr. David R. Howes, U. S. Army Strategy and Tactics Analysis Group, Bethesda, Maryland

Dr. R. J. Lundegard, Logistics and Mathematical Statistics Branch, Office of Naval Research, Washington, D. C.

Professor Ingram Olkin, Stanford University, Stanford, California

Professor G. S. Watson, The Johns Hopkins University, Baltimore, Maryland

A DATA COLLECTION PROCEDURE FOR ASSESSING NEURO-MOTOR PERFORMANCE IN THE PRESENCE OF MISSILE WOUNDS
William H. Kirby, Jr., William Kokinakis, Larry M. Sturdivan and William P. Johnson, Ballistic Research Aberdeen Proving Ground, Maryland

PROBLEMS IN THE DESIGN OF STATISTICS-GENERATING WAR GAMES
William H. Sutherland, Research Analysis Corporation, McLean, Virginia

0950 - 1020 BREAK

1020 - 1220 GENERAL SESSION 4 -- South American Room

Chairman: Dr. Frank E. Grubbs, Chairman of the Conference, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

THE FUTURE OF PROCESSES OF DATA ANALYSIS
Professor John W. Tukey, Princeton University, Princeton, New Jersey

STATISTICS AND MANAGEMENT
Dr. M. G. Kendall, C-E-I-R, London, England

THE STIMULUS OF S. S. WILKS TO ARMY STATISTICS

Leslie E. Simon
Major General, USA (Ret.)

ABSTRACT. The stimulus of S. S. Wilks to the scientific community is discussed briefly, followed by a more detailed account of his originating the idea of a series of Army-wide conferences on design of experiments in Army research, development and testing. The Army's rather satisfactory progress in statistical methodology prior to the conference series is discussed, with comments on its limitations and less than ideal direction of procedure. Wilks' apparent perception of the situation, his courage in undertaking a large and difficult task, and his surprisingly large measure of success is discussed. The importance of carrying on the spirit of Wilks is emphasized, and the creation of The Wilks Award, as a measure to that end is mentioned.

ORIGIN OF THE CONFERENCE SERIES. Mr. Chairman, Fellow Conferees, Ladies and Gentlemen, Samuel Stanley Wilks was my very good friend most of his professional life. Whereas I am aware of many of Wilks' dedicated and outstanding services at a national, if not a world level, I prefer to concentrate my remarks on an area of Wilks' career that is close to home to me: the very valuable services that he did voluntarily for the Army. I am sure that others more able than I will cover his broader services as a teacher, both academic and extra curricular; as a research worker, as an organizer, and as a competent and inspiring leader. Frederick Mosteller has presented an excellent outline of Wilks' worldwide work in the April, 1964 issue of "The American Statistician", under the title, "Samuel S. Wilks; Statesman of Statistics". Mosteller's paper should serve as a guide for other papers on Wilks. However, I cannot help observing that although Mosteller's title is justified, I hope that he will forgive me if I observe that Wilks was by his own choice somewhat lacking in the formality associated with statesmanship. Contrary to one's concept of dignity, Sam was "just folks", whether he was talking with a first-rate scientist, a neophyte in Applied Statistics or a man primarily a soldier. He knew and understood people; and, by nature was ever-ready to give any help within his competence to anyone who genuinely needed it. It was in the latter two capacities, that I had my entree to Wilks.

It was over fifteen years after our initial meeting that Wilks made a proposal that has helped much in improving Army organization, doctrine, tactics and weapons; and, at the same time contributed to improving the morale of Army personnel, and to saving time and expense in military research and development.

In late 1954 or early 1955, when I was Assistant Chief of Ordnance for Research and Development, U. S. Army, Wilks proposed that the Army establish a series of Army-wide conferences on design of experiments in Army research, development and testing. Dr. Frank E. Grubbs, who, under the authority of my office, had chaired an Ordnance symposium on Statistical Methods in 1953 [1], strongly indorsed Wilks' proposal for Army-wide conferences, devoted primarily to design of experiments; and, of course, I concurred. The Army Mathematics Advisory Panel* (later, designated as the Army Mathematics Steering Committee) operated under the Office of Ordnance Research (now Army Research Office-Durham); and consequently the responsibility for the conferences was assigned to that office. Wilks' proposal was made pursuant to a survey made by the Army Mathematics Steering Committee in which they investigated over 30 Army facilities. They found that one of the most frequently mentioned needs expressed by the scientific personnel was for greater knowledge of modern statistical theory of the design and analysis of experiments. The First Conference on Design of Experiments, in Army Research, Development and Testing was held on October 19-21, 1955 at the Diamond Ordnance Fuze Laboratories and The National Bureau of Standards. Wilks chaired all the conferences up to the present Tenth Conference.

I believe that observing as best we can the time-rate-of-change of the character of these conferences and the concurrent increase of basic understanding of the interrelationships of men, weapons, organization, doctrine, tactics, and research and development, will throw light on the beneficial influence of Wilks on National Defense. I do not mean to infer that all Statistical progress is due to Wilks; but I am sure that much of the progress is due to the spirit of cooperation that he infused, to his influence and to his

---

*The Army Mathematics Advisory Panel, of which Wilks was a member was operated by the Ordnance Corps for the Office of the Chief of Research and Development, U. S. Army. I am indebted to Colonel P. N. Gillon (Ret.), who was both the Commanding Officer of the Office of Ordnance Research (Durham) and the very able Chairman of the Army Mathematics Advisory Panel for the clear, curt minutes and records that he left, and especially for reference [2].

personal contributions. Similarly, I believe that the history of Wilks in this relatively small sub-field of his very active life is a close parallel to the fruitfulness of his activity in other fields to which he devoted far more time. Let us, then, observe the status of Army statistics up to 1953; trace, at least approximately, the conferences on Design of Experiments in Army Research, Development and Testing; and observe the present-day status of Army statistics.

Incidentally, the Army was neither without statistical sophistication in 1953, nor is its knowledge optimum today.

SUMMARY OF ARMY STATISTICAL PROGRESS, BETWEEN WORLD WAR I AND II. Historically, the application of probability theory to the dispersion of shots on a target appears to be about the only Army use of Statistics, prior to World War I. There was a jump in mathematical sophistication during World War I, due to A. A. Bennett [3], Fowler [4], Moulton [5], and others in connection with progress in applying statistics to Ballistic problems. Between World Wars I and II, Kent, Dederick, McShane and others developed further applications of Statistics in connection with Ballistics. The staff of the Bell Telephone Laboratories, especially Dr. Walter A. Shewhart and Harold F. Dodge, was most fruitful in the discovery of Statistical techniques, and the Army was a shameless plagiarist in adapting them to its problems. Shewhart's work [6] led to the Army's first full-scale industrial use of Statistical Quality Control in manufacture at Picatinny Arsenal, Dover, New Jersey, which also was certainly one of the first few of such uses in the world. The Army Ammunition Surveillance [7] (Stockpile Reliability) System (circa 1939) was based largely on what was very recent work at that time. The Dodge-Romig Sampling Tables, not yet in book form [8], appeared just in time for use for ammunition inspection and acceptance tests in World War II. During the period shortly before World War II, the Army felt a bit smug about its statistical competence.

ARMY STATISTICAL PROGRESS DURING WORLD WAR II. World War II saw great progress in the military use of Statistics, due primarily to the availability to the war effort of men of competence. The National Defense Research Council (later, Office of Scientific Research and Development), the staff of the BRL, and, to a lesser extent, the staffs of Ordnance Arsenals acquired many Mathematicians and Statisticians of competence. Procedures for specifications of materiel, sampling, testing and interpretation of data (both planned data and the salvaging of unplanned data) were greatly improved. Indeed, Operations Research was being born even then. The Army* was not unmindful of the possible adaptation of any new Statistical "tool" to its work.

*References to the Army do not imply that the Navy and Air Force did not also make progress.

In addition to the above uses of Statistical Methods substantial progress was made by the Army during World War II of which there is little or no record. Many new techniques such as Sequential Sampling and Reliability were actually used in the Army, at least in an empirical way, before they were later designated by appropriate specific names. Of course, needed theory was not worked out in a formal way at that time. For example, the formal presentation of sequential sampling had to await the work of Dr. Abraham Wald, which was not published in book form until 1947 [9].

ARMY STATISTICAL PROGRESS, WORLD WAR II - 1953. After World War II, progress continued, although its rate was diminished due both to decrease in staff and to loss of some of the more competent people. Apparently, experiments that involved Factorial Designs were the first instances of full use of Experimental Designs in the Army. Factorial designs were used at the Ballistic Research Laboratories in the study of armor plate (1946-47)*, in the mammoth experiment on Aircraft Vulnerability (1946-50)*, and even on Project Stalk (a tank-fire control study under field conditions)* circa 1953. In 1953-1954 Reliability [10], in its present day sense, was used by Ordnance Research and Development, in a full-scale organizational and technical way, as a means of rescuing the Country's first operational guided missile, the NIKE, from a serious threat of failure.

With this rather glowing account of Army progress and status, one might well question wherein was the Army laggard, and where was the failure or potential threat of failure? What great work was there left to be done by the series of conferences on design of experiments under Wilks? I shall show that a very great deal was wrong with the Army's use (or lack of use) of statistical methods; that the task of righting the wrong was formidable, both in magnitude and in potential obstacles; and that astonishing progress has been made on the task during the nine years of the conferences.

From the survey of the 30 Army facilities, Wilks must have understood rather well what the Army needed, and have understood also the need for newly organized and sustained effort to supply the need. His skill as a teacher must have fortified him from fear of failure in undertaking to change the mode of operation of a large segment of the Army.

---

*Ballistic Research Laboratories Publications.

WHAT WAS WRONG. Let us observe that the origin, growth, and use of Statistical Methods in the Army was not only unplanned, but actually tended to progress in the least advantageous direction; i.e., from end-point to origin, rather than from origin to end. Roughly speaking, we can regard the military regime as consisting of the following steps or stages: doctrine, tactics, organization, selection of equipment, fabrication of equipment, test of equipment, and use of equipment. Logically, a powerful medium for the improvement of a stage should be first applied to the preceding stage or stages to which it is applicable. For example, a big improvement in use of equipment, (e.g., accuracy of ammunition) loses much of its potentially beneficial effect if either the tactics, organization, or weapons system is poor.

Contrary to the above observation, the earliest use of probability theory by the Army was for use of equipment, viz, the adjustment of artillery fire. The use of techniques based on the Gaussian Distribution, or Normal Probability Law, in connection with artillery fire probably is exceeded in antiquity only by the use of elementary probability theory in connection with games of change [11].

Decades elapsed before the next major step. In 1936, the Army began to use Statistical Quality Control in the manufacture of equipment, viz, the production of ammunition at Picatinny Arsenal, Dover, New Jersey. Kindred techniques such as sampling theory and statistical methods for analyzing data soon spread to improve specifications, inspections and acceptance tests.

During World War II almost all fabrication of military equipment was better, cheaper, and quicker, due largely to these techniques. During World War II, one strange reversal occurred in the inverse order of progress. Operations Research was born out of military sponsorship and was actually used to a limited degree by the staffs of high military planners in connection with the planning of the operations of large combat forces.

After World War II, it began to be more and more realized that since Statistical Methods improved the quality of equipment and reduced costs it would be a good idea to use similar techniques with the research, developing and testing in connection with new designs of equipment, thereby making better and more useful equipment designs at the out-set. Except for the invention of Reliability, which was a distinct child of necessity, this is just about where Wilks came in.

WILKS' TASK. When Wilks toured the 30 Army installations with the Army Mathematics Advisory Panel, it was he who articulated, "the most frequently mentioned needs expressed by the scientific personnel were for greater knowledge of modern statistical theory of the design and analyses of experiments." Thus, it is clear that Wilks recognized at least a major part of what was wrong with the Army; i.e., insufficient use of Design of Experiments in Research, Development and Testing.*

Certainly Wilks was not the first person to recognize the fact that an improvement in the early stages of the Army regime, i.e., doctrine, tactics, organization, etc., has greater leverage power than an improvement in later stages such as selection of equipment, fabrication, and use. The trend toward "up-stream" improvement began long before he appeared on the scene; and ranged from such measures as advocacy of industrial preparedness, as an important measure towards preserving the peace, to various stratagems for introducing sophistication in the upper stages of the Army's evolutionary process. Many persons deplored the fact that traditionally we had been forced to begin wars with the weapons left over from the previous war. Army Ordnance began to take measures against this ill shortly after World War I, and the then infant Army Ordnance Association (now the American Ordnance Association) lent a patriotic and helping hand, pursuant to its slogan advocating industrial preparedness as an insurance against war; i.e., a large production capacity should exist to meet a war demand for munitions of the latest designs. Army Ordnance realized that it must have an eye to the future and an ear to the ground regarding the plans and needs of the combat soldier, and therefore sent selected Ordnance Officers to the Army Schools ranging from the Command and General Staff College to the National War College to give them a close understanding of the combat soldier. Liaison officers from the combat arms were assigned to Aberdeen Proving Ground, Maryland, to assist in the realization of combat viewpoints, and in the development tests of materiel. Shortly after World War II, a number of persons, including some Ordnance, advocated the establishment of a scientific staff at Headquarters, Army Field Forces, Fort Monroe, Virginia, to assist in analyzing Army needs and in stating needs for new materiel in valid form. Such a group was partially formed and existed for a year

*The Army was not new to Wilks. In 1948 he was awarded a Joint Army-Navy Certificate of Merit for his war-time contributions to anti-submarine warfare and the solution of convoy problems.

or two*. However, it was Wilks who undertook systematically the task of greatly accelerating the spread of powerful and useful statistical techniques to the upper echelons of the Army regime, where the improvements that they enhanced would have the greatest leverage power.

Even if Wilks recognized the full nature of the job that he was doing, certainly, he did not have opportunity to finish the job. Much remains to be done. The real point in this discourse is the breadth and extent of the progress made in the nine years of Wilks' kindly and sympathetic leadership, effective persuasion, and his engendering of mutual cooperation and helpfulness between men of competence with whom he dealt. Let us try to note the progress, before any attempt to assess the remaining task.

ASSESSING THE PROGRESS. I hope that by the foregoing discussion I have led no one to believe that I have an objective method of measuring the progress of use of statistical methods in the Army during the 1955-63 period. I might say that the measuring of progress in a field of science or engineering is perhaps one degree more difficult than measuring the quantity and quality of output of research by laboratory; and whereas many have tried to do this, I know of no one who has really succeeded. The cold statistical facts are briefly these:

All the design of experiments conferences were for three days each, held in October or November, and conducted at a number of Army R&D establishments.

The number of registrants or conferees was always of the order of 200. Attendance was by invitation and the number of invitations was undoubtedly conditioned by the available accommodations.

The number of papers presented at each conference was of the order of 30. This appears to be about the number of papers that can be presented in a three-day conference.

All conferences were of a three-part character: Invited papers by distinguished Statisticians, technical sessions in which there were discussions of recent accomplished work, and clinical sessions in which work in progress was discussed from the viewpoint of inviting advice and criticism.

---

*Later, a permanent group was formed.

It thus appears that based on documental evidence the progress of the conferences can be judged only by the kinds of scientific and technical fields covered by the papers and by the inherent quality of the papers.

CHARACTER OF PAPERS PRESENTED. By and large, the place at which the conference was held had a strong influence on the character of the papers presented. This is undoubtedly due to the fact that the program committee gave some degree of precedence to the host institution, e.g., more papers bearing on the field of medicine were presented at the Eighth conference held at Walter Reed Medical Center than at other conferences. However, in the statistical fields there was a constantly increasing emphasis over the the nine years on the more sophisticated phases of design of experiments, screening theory, simulation stratagems, reliability, and techniques for evaluation of experiments. It is thus apparent that expertise on the part of the participants increased and also evident that the use of statistical experts in various fields of Army activities was increased both in number of experts and in variety of fields of activity.

Whereas, at the beginning of the conferences papers centered largely around items of Ordnance materiel, as the conferences proceeded the subject matter of the conferences expanded to include more emphasis on systems analysis. Similarly, with the penetration of statistical methods into new fields of activity, more papers were devoted to other than Ordnance equipment. With the broader use of statistical designs, papers appeared on the relation of equipment to organization, and to new theoretical developments having immediate application in Army use.

A further change in the character of the papers is the noticeable effect of learning to do by doing. It is apparent that whereas designed experiments gave greatly improved results, the same experiments also showed deficiencies in understanding what one's work was really about. For example, biases in results could be detected that were readily attributable to repeated use of the same personnel over the same terrain. Command exercises had to be altered and new stratagems employed (such as randomization techniques) to screen out the biases which passed unnoticed when experiments were of less sophisticated character. In fact it was precisely the acquirement of such evidence that convinced even non-statisticians that there was need for more movement "up-stream". This was a very fortunate circumstance because it drew military commanders into participation in the planning of the experiments and resulted in a constant movement of the sphere of activity of statisticians into the domain of persons who were concerned with policy, tactics, and doctrine. Thus, non-statisticians saw the gains made through experiments in which they, themselves participated.

It is quite one thing to make a presentation on the efficacy of a technique, and quite another thing to convince the hearer that the use of the technique is important to his job. Successful experiments in which one himself has participated (although a step-wise process) are an effective method convincing one of the value of the methods used. By way of contrast, I believe that it would be quite impossible to suddenly inject into the military service (or into any other organizational sphere, for that matter) the concept and attitude which is expressed by the following quotation taken from a Combat Developments Experimentation Center (CDCEC) pamphlet:

> "The ability of the Army to carry out its goals in the future depends upon the success it has in achieving its combat developments goals today . . . of developing future concepts, doctrine, tactics, and techniques, and providing requirements for weapons, equipment, and appropriate organizations."

It is indeed heartening to read such a quotation. This Experimentation Center has an area of over a quarter of a million acres, a brigade of troops, a contract with Stanford Research Institute for Statistical Support, a variety of sophisticated equipment, including facilities for computer simulation of field experiments. Nevertheless, we know well that the tasks expressed in the quotation are only beginning and that only the first fruits have yet been achieved. From the foregoing example of CDCEC we can infer (a) that the advance of Statistical Methods in the Army, during the past nine years have been great, and (b) that the remaining part of the task, i.e., achieving the full nature of the job that Wilks undertook is still a large one.

WILKS' METHODOLOGY. If we hope to carry on in substantial measure the task that lies ahead we should take a good look at Wilks' methods. Wilks was a scientist for the sake of science, but he was also a realist and wished to see the practical results of applied science come to full fruition.

This is a rare combination of qualities.* Despite his many high scientific achievements and the respect in which he was held by his colleagues, he never assumed an authoritative position. On no occasion did he attempt to do a whole job himself to the exclusion of others. On the contrary, he always invited the cooperation of every person who could contribute substantially to getting the job done. He could organize and delegate without being obvious about it. In this way he secured the enthusiastic support of the men around him. If anything, he was more the servant of others than one demanding services. He had confidence in himself, but he also inspired confidence in others that led them to venture to cooperate, to work with him and to work together; and the work became an interesting enterprise to the point of preoccupation. In closing, I would like to give a brief example of how the spirit of Sam Wilks worked towards getting things done whether they were large or small.

AN EXAMPLE OF WILKS' WORK. About a year and a half ago, a gentleman in Georgia, a former member of the war-time team at The Franklin Institute, who is intensely interested in small arms fire asked several statisticians including Wilks some questions about the inter-relations of various measurements of central tendency and dispersion of shots on small arms targets, although he did not express it in these terms. In order to answer his questions, one needed to know the probability density distributions of several statistical measures whose distributions were unknown. These questions set off a kind of chain reaction. It was possible that answers to the small arms problem could well be answers to other, and probably more important, problems. Scientific men of good will, infused by the spirit of cooperation and scientific inquiry contributed what they knew to the general problem; but it became evident that a complete answer could be achieved only by some research that would add a modicum of knowledge to our existing store. Perhaps the most important contributions came (later) from Wilks, Grubbs, and one or two other colleagues in connection with their work on the analysis of tracking data on firings of long range missiles at the

*In writing for the Journal of the Royal Statistical Society, July, 1964, the noted British Statistician, E. S. Pearson says, "... it is hard to think of any mathematical statistician of the past 30 years who combined to a greater extent an excellence in the field of theory with a power of inspiring confidence in government agencies, national research institutions, and educational authorities, as a wise counseller in practical affairs."

Atlantic Missile Range, The work turned out to be so important that it has been carefully written up by Grubbs in a forthcoming monograph. This illustrates the humbleness, the spirit and the methods of Wilks. First, he was willing to lend his powers to anything that appeared to be a valid scientific enterprise; second, he had a keen perception of what is fundamentally important even though the context in which it was presented made it appear somewhat of casual interest if not unimportant; third, he could engender the spirit of true scientific inquiry into his colleagues; fourth, he could bring a matter to a crux so as to make it a permanent addition to the useful knowledge of mankind.

THE WILKS' AWARD. It is important that the spirit of Sam Wilks be carried on, both for an unselfish reason and a selfish reason. Our first reason is that of honoring his memory in gratitude for what he had done for us. The second and selfish reason is that carrying on the spirit of his work will contribute much to advancing the solutions for the great task that he loved and to which he devoted himself. We shall never achieve the task in full; but each solution or partial solution will contribute to the improvement of the military posture and safety of our Country. I am sure that Sam would approve this second motive. Through the generosity of Mr. Philip G. Rust of Thomasville, Georgia, and the good offices of the American Statistical Association, it appears that a means has been found of achieving, at least in part, both of the above purposes. An award will be created which by its character will help to carry on the stimulus of Wilks to Army Statistics.

REFERENCES

1. BRL Report No. 897, Proceedings of the First Symposium on Statistical Methods; Sampling Techniques.

2. Summary Report on the Establishment of the Army Mathematics Center. 14 October 1955. Office of Ordnance Research (now Army Research Office - Durham).

3. A. A. Bennett, On the Accuracy of Sampling. 6 January 1918. Unpublished paper in the files of the Ballistic Research Laboratories.

4. Fowler, Gallip, Lock, and Richmond, Aerodynamics of a Spinning Shell, Philosophical Transactions of the Royal Society of London, 1920.

5. F. R. Moulton, New Methods in Exterior Ballistics. University of Chicago Press. 1920.

6. W. A. Shewhart, Economic Control of Quality of Manufactured Product. Van Nostrand, 1931.

7. The Proposed System of Surveillance of War Reserve Ammunition, Zornig and Simon, BRL Report No. 115, 1938.

8. Dodge, Harold F. and Harry G. Romig, "Sampling Inspection Tables -- Single and Double Sampling", John Wiley and Sons, New York 1944; 2nd edition, 1959.

9. Abraham Wald, Sequential Analysis. John Wiley and Sons, 1947.

10. L. E. Simon, The Relation of Engineering to Very High Reliability. Proceedings of the Tenth National Symposium on Reliability and Quality Control. January 1964. IEEE.

11. Games, Gods and Gambling, F. N. David, 1962, Hafner Publishing Company, New York.

THE WILKS AWARD

Introduction of Mr. Donald C. Riley by Major General Leslie E. Simon

Mr. Chairman, Fellow Conferees, Ladies and Gentlemen, what the next two speakers have to say is so closely associated with my discourse on Wilks that I have been designated to introduce them.

As I implied at the end of my talk, the establishment of the Wilks Award was a tri-partite undertaking: And involved the Army as principal beneficiary, The American Statistical Association as the bearer of the burden of administration, and Mr. Philip G. Rust who endowed the award. Secretary Hawkins has personally expressed to the ASA his gratitude for its competent and patriotic services.

Mr. Donald C. Riley, Secretary-Treasurer and Executive Director of the American Statistical Association has rendered invaluable assistance in getting swift solutions to procedural problems he has been so kind as to agree to indicate to you the duties and obligations of the ASA in carrying out the Wilks Award; and he will also announce the recipient of the initial Wilks award. Don Riley!

INITIAL WILKS AWARD PRESENTED TO DR. FRANK E. GRUBBS

Donald C. Riley, Executive Director,
American Statistical Association

Many members of the American Statistical Association, as well as I, are glad to be present at this, the Tenth Annual Design of Experiments Conference. This is a very special occasion and the American Statistical Association is glad to participate during a uniquely auspicious time in its long history. This year is the 125th Anniversary of the establishment of the American Statistical Association which recent research at Stanford has found to be the second oldest national professional society in the United States.

The American Statistical Association has always worked closely, although usually quite informally, with agencies of the Federal Government. For example, during the year it was founded, 1839, it began to press for the improvement of decennial censuses and its representatives played a major part in the design of four of the six census schedules for the 1850 Census. As statistics and statistical methodology proliferated vastly since that time, almost all areas of research have felt their impact. Certainly the whole area of design of experiments has had the closest association with statistics. The annual Design of Experiments Conference has become an institution. General Simon has reminded you of the close association of Professor Samuel S. Wilks with this Conference. Most of you know that relationship by heart. Sam lent his aid readily, unstintingly and effectively in many areas. This was part of the genius of the man.

I should note also that Wilks was the President of the American Statistical Association in 1950 and that he had always done much for the Association. He also helped to carry on in another area the close relation between the Association and the Federal Government. Just the day before he died he participated as a member of the Advisory Committee on Statistical Policy to the Office of Statistical Standards in the Bureau of the Budget. The Office of Statistical Standards requires consultation from time to time at a high level in its work as the central statistical coordinating body of the Federal Government. This Advisory Committee consists largely of former ASA Presidents and Wilks was one of its "founding fathers."

As mentioned in General Simon's address, the ASA has recently had the opportunity to be of further service. By joint agreement between representatives of the Army, Mr. Philip G. Rust and the ASA, the Samuel S. Wilks Award has been established. The Award will consist of a medal and an honorarium. The ASA has accepted the obligation of administering the Award in accordance with guidance and criteria which are consonant with law and with the wishes of Army representatives, Mr. Rust and the ASA.

Annually, ASA has agreed that an appropriate committee be selected (or appointed) to select the awardee, based on the criterion that he is a person whom the committee regards as deserving of the award, based primarily on his contribution (either recent or past) to the advancement of scientific or technical knowledge, ingenious application of existing knowledge, or successful activity in the fostering of cooperative scientific efforts which have only coincidentally benefited the Army. The award shall be made with the intent of recognizing the personal and intellectual accomplishments of the individual and shall not be given with the intent of supplementing the individual's salary, providing him with compensation, or advancing the interests of the donor or trustee of the endowment.

The American Statistical Association has been asked to invest the funds so generously turned over to it for this purpose and I am sure that its Board of Directors, which has given its wholehearted approval, feels honored in being asked to join in honoring Sam Wilks. ASA will need to consult very closely with those of you who have helped to develop the annual Design of Experiments Conferences, in the selection of an Annual Sam Wilks Award Committee. I believe that Dr. Albert H. Bowker, the President of the American Statistical Association this year, will be able to announce this Committee shortly.

As executive Director of the ASA, I have the honor to announce that Dr. Frank E. Grubbs of the Army's Ballistic Research Laboratories has been selected to receive the "initial," not the first, Samuel S. Wilks Award. As is not unusual in the initial award of an honor, Dr. Grubbs was selected not by the process governing the first and subsequent recipients, but rather by unanimous agreement of those concerned with the establishment of the Award. He is so selected because of his close working relationship with Wilks, and especially because of his contributions along with Wilks to solutions and clarification of simple measures of dispersion, which are deemed useful to riflemen, ballisticians, and statisticians in general.

I have no medal to present to Dr. Grubbs, because the medal has not yet been struck; but it will be presented at the earliest appropriate opportunity, after it is available.

Incidentally, I will not be able to attend the banquet here tomorrow evening because I agreed long ago to attend the inauguration ceremonies in New York of Dr. Bowker as Chancellor of the combined Universities of the City of New York which was organized a few years ago.

The American Statistical Association will want to continue to advise closely with the Conference and will be glad to ask its auditor to render a brief auditing report each year if this seems satisfactory to those who have been so close to Sam Wilks, General Simon and especially Mr. Philip G. Rust, who has been so generous and public spirited in making the award possible. I should like to join in thanking Mr. Rust most profoundly.

INTRODUCTION OF MR. PHILIP G. RUST
BY MAJOR GENERAL LESLIE E. SIMON

Mr. Chairman, Fellow Conferees and Ladies and Gentlemen.

We now come to the third and last speaker in this phase of our honoring Sam Wilks, Mr. Philip G. Rust of Winnstead Plantation, Thomasville, Georgia.

Mr. Rust is a very modest man, and more adept at understatement than a typical Britisher. It was only under pressure personally exerted by Secretary Hawkins that we succeeded, first, in overcoming his insistence that he remain anonymous, second, in getting him to attend this conference, and third, in persuading him to present the honorarium to the initial recipient of the Wilks Award, Dr. Grubbs.

Mr. Rust purports to be practically innocent of theoretical and applied statistics; but if under pressure, he can cite statistical literature by page and paragraph showing each historical advance in statistical measures of dispersion; he professes no close association with science and engineering, but I find that he was not only a research chemist for over ten years, but also returned to science and engineering during World War II; he lays claim only to being a Georgia farmer, but he has contributed to ASA the funds necessary to establish the award commemorating his old friend, Sam Wilks, contributing to the welfare of the military services, and fostering science in general.

With these cautionary remarks, I deem it a privilege and an honor to introduce Mr. Philip G. Rust.

THE CONCEPTION OF THE WILKS AWARD

Philip G. Rust
Winnstead Plantation, Thomasville, Georgia

Mr. Chairman and members of the audience you have heard a most informative talk by General Simon on "The Stimulus of S. S. Wilks to Army Statistics." Then, on Thursday we may look forward to Dr. Eisenhart's "Sam Wilks as I Remember Him."

In view of the newly established Association's Wilks Award, concisely described to you by Mr. Donald C. Riley, the Executive Director of the American Statistical Association; it is appropriate that I briefly discuss the conception of this award.

Back in the dark days of 1944, Dr. Wilks and I were headed north from Washington, by train; he to Princeton, and I to my home in Wilmington. At the time, I was at The Franklin Institute, working on .50 calibre barrel erosion, and also as the un-official translator of pertinent technical works. In passing, I would state that the Institute work was less statistical than of the ear drum rupturing variety.

On this train trip, I happened to mention, that for years, my spare time had been devoted to certain statistical measures of shots on a target. After telling Dr. Wilks about the firing of hundreds of .22 calibre targets, from rest; to get an empirical measure of the distribution of "extreme spread", he asked if I had started any theoretical work on the subject. (Incidentially, "extreme spread" is defined as the separation distance of the two widest apart shots.) His interest increased when it was mentioned that I had made a start by generating a few hundred artificial targets by using pairs of random numbers in the well-known bi-variate circular distribution. Equal likelihood of angular distribution was assumed, with no systematic errors.

The shots were laboriously plotted on cross section paper, and the extreme spread and other parameters examined. It is of interest to note that the fired targets and the plotted ones are extremely close.

About this time, my travelling companion suggested that he disembark at Wilmington, also. I had the feeling that he wanted to explore the application of these data to other, more vital matters. He stated that he had an exceptional graduate student who might be given the job of finding the true distribution of "extreme spread".

Eight or ten years went by, and our contacts were largely by phone. He assured me that he was still interested, and working on target problems; but that as yet, this distribution had not been discovered. The possibility of Monte Carlo methods on a to-be-aquired computer were discussed. Then on 10 August 1963, I received a long-hand letter saying that a 7090 computer was at hand, busily working on related matters.

While waiting for promised data from Dr. Wilks, I approached General Simon about the subject. He later discussed it with Dr. Frank Grubbs of Aberdeen, who subsequently brought forth an extremely useful manuscript, soon to be published.

Finally, on Dr. Wilk's 1963 Christmas card, he stated that the target problem was tied in with tracking work on the Atlantic Missile Range.

General Simon; with his very orderly mind, and sense of the fitting, then suggested the idea of the annual A. S. A. Wilks Award. This idea was greeted enthusiastically by all concerned.

What, then could be more fitting, than that Dr. Frank E. Grubbs should be the recipient of the initial award.

And now, it gives me great pleasure to hand Dr. Grubbs the initial honorarium and the assurance of its accompanying medal on its completion.

DEVELOPMENT OF THE DESIGN OF EXPERIMENTS OVER THE PAST TEN YEARS*

Oscar Kempthorne
Iowa State University, Ames, Iowa

INTRODUCTION. The main aspects of experimentation on which progress has been made in the past 10 years appear to be the following:

(a) the analysis of experiments
(b) the development of incomplete block designs
and (c) the investigation of multifactorial systems.

I shall have just a few words to say about the *first* two items and shall spend practically all my time on the third item.

THE ANALYSIS OF EXPERIMENTS. In the last 15 years or so, statisticians have become concerned about the assumptions that are commonly made in the analysis of comparative experiments. The common analysis is to use the matrix model

$$y = X\beta + e$$

in which y is the vector of observations, X is a matrix of known elements, β is a vector of unknown parameters, and the vector e of errors is assumed to consist of components which are normally and independently distributed around zero with constant variance. The obvious questions about such a model are:

(1) why use y, and why not a defined function of y, such as $\log y$ or $1/y$, or any of a host of other possibilities?

(2) is the model linear in the parameters, that is, is the expectation $X\beta$, correct?

(3) is the assumption about the errors correct?

In recent years there has been considerable attention to these questions, primarily by Anscombe (1961), Tukey (1962) and Anscombe and Tukey (1963), the work dating back to Tukey's one degree of freedom for non-additivity. This has led to the topic - residual analysis - which is now an every day phrase.

---

*Prepared in connection with work on Contract AF 33(615)-1737, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio.

Matters unrelated to residual analysis but part of data analysis, are topics such as the question of multiple comparisons, the effects of preliminary test on conclusions, random, mixed and fixed models, and randomization theory of experimental inference. I shall not discuss these.

THE DEVELOPMENT ON INCOMPLETE BLOCK DESIGNS. Incomplete block designs were developed to control variability among the experimental units. The original incomplete block designs were given by Yates in the 30's, and in 1939 Bose and Nair developed a fairly general class of such design. Since that time there has been a development of blocking theory with regard to

(a) The structure and existence of incomplete block plans
(b) the arrangement of factorial designs in incomplete block designs.

Such development is very desirable, but it is agreed by most, I imagine, that the impact of this work on the conduct of experiments is not great. Roughly speaking we have had for many years an array of incomplete block designs which provides an adequate basis for choice for most experimental situations.

THE INVESTIGATION OF QUALITATIVE FACTORIAL SITUATIONS. It is essential to differentiate between multifactorial situations in which the factors are qualitative and in which the factors are continuous or quantitative. In the former case the structure of the totality of possible information consists of the true yields and variability for each of the possible factor combinations. In the latter case the totality of possible information is a functional relationship of yield to the levels of the factors or variables. So in the qualitative case, if one has factors say a, b, c, . . . , with levels denoted by $a_i, b_j, c_k, \ldots,$ the underlying formula for yield will be of the form $y(a_i, b_j, c_k, \ldots,) =$ $f(i, j, k, \ldots,)$ + error where the function is defined only for the factor levels $i, j, k, \ldots,$ in the situation. In other words, the model has to be a classificatory model. Classificatory models can be linear as exemplified by

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \text{etc} + \text{error},$$

or can be non-linear, as for example

$$y_{ijk} = \frac{\alpha_i}{\beta_j + \gamma_k} + \text{error.}$$

Essentially no theory exists for non-linear classificatory models, and I am of the opinion that this is a real gap in our knowledge.

In the case of study of the full set of factorial combinations, one of the basic problems is error control and systems of confounding were developed for symmetrical systems in the 30's. There have been a few developments in recent years with regard to confounding for the asymmetrical case, and also some clarification of the mathematical structure of factorial experiments [e. g. Kurkjian and Zelen (1962)] . I imagine, however, that examination of the full set of factorial combinations is rarely appropriate except possibly

(a) when most of the factors have 2-levels, with perhaps two three-level factors,

and (b) in the case of experiments, like in agronomy, for which there is a long essentially unalterable interval of time from executing the design to obtaining the experimental results, on the basis of which to plan another experiment.

There has been one development of analysis which seems to be very informative, when the totality of treatment degrees of freedom can be partitioned into meaningful orthogonal single degrees of freedom, the half-normal plot of Daniel (1959). The idea of half-normal plotting is the very elementary one of looking at the distribution of the totality of single degree of freedom contrasts, and to observe which ones are outliers. The half-normal plot is a convenient way of doing this. In general tight rules of significance for examining the realized half-normal plot do not exist. The procedures of half-normal plotting have been generalized to the case of a multivariate response by Wilk and Gnanadesikan (1963, 1964).

In the case of the linear classificatory models it is obvious that the simplest design problem is to estimate the effects under the assumption of no interactions. Effective designs for this case have now been available for several years. The earliest example of such a design was given by Tippett and is described in Fisher's "Design of Experiments" for the testing 5 factors in 25 trials. In the 1940's the following sets of main effect plans were developed:

the Fisher series:

$$2^n - 1 \text{ factors at } 2 \text{ levels with } 2^n \text{ trials}$$

$$\frac{p^n - 1}{p - 1} \text{ factors at } p \text{ levels with } p^n \text{ trials}$$

the Plackett-Burman series (related to Mood's weighing designs)

$$4N - 1 \text{ factors at } 2 \text{ levels in } 4N \text{ trials.}$$

An additional series was developed by Addelman and Kempthorne (1961) for

$$2\frac{(p^n - 1)}{(p - 1)} - 1 \quad \text{factors at } p \text{ levels in } 2p^n \text{ trials.}$$

In all these cases p is a prime number or a power of a prime number. Tukey (1959) and Addelman (1962) showed that these symmetrical main effect plans can be used to develop very reasonable main effect plans for asymmetrical factorial situations.

In the 1940's Finney (1945) formulated the general idea of fractional replication, which is closely related to the idea of confounding. It is interesting to note, in passing that Fisher was primarily interested in systems of confounding, and it was not adequately realized for some years that he had in fact developed incidentally the series of main effect plans mentioned above. The idea of fractional replication is to use a subset of the totality of treatment combinations chosen on the basis of the definition of effects and interactions. Obvious candidates as useful designs in this class are the main effect plans, and the designs which permit estimation of all main effects and two-factor interactions.

Also in 1946, Rao (1947) formulated the idea of orthogonal arrays. An array (N, k, s, t) is a collection of N treatment combinations out of the totality s^k of treatment combinations possible with k factors each at s levels, such that every combination of every subset of t factors occurs equally frequently. The value t is called the strength of the array. An array of strength 2 is an orthogonal main effect plan. With an array of strength 3, no main effect is confounded with two-factor interactions, but some two-factor interactions are mutually confounded. An array of strength 4 enables the orthogonal estimation of all main effects and two-factor interactions, and so on. Clearly the enumeration of main effect plans, two-factor interaction plans etc. is related to the enumeration of orthogonal arrays. Box and Hunter (1961a, b) have given a rather detailed account of the possibilities of fractional replication with 2-level factors, using the term degree of resolution instead of the strength of array of Rao. A design of resolution III gives main effects estimates, which will be biassed by two-factor interactions. A design of resolution IV gives main effects unconfounded with two-factor interactions, but with the two-factor interactions somewhat interconfounded and a design of resolution V is a two-factor interaction - clear design. They show that a design of resolution III repeated with reversed signs gives a design of resolution IV. They discuss extensively the arrangement of fractionally replicated plans in blocks. They also examine the possibility of plans which estimate interactions among all of a subset of the factors with the effects of another subset of factors, the former being regarded as major variables and the latter as minor variables. For example, they give a 2^{16-11} plan which enables the estimation of all effects and interactions among 4 major variables and the main effects of 16 minor variables. Box and Hunter (1961b) give the possible two-factor interaction clear fractions in blocks for up to 11 factors. The possibilities are as follows:

No. of factors	5	6	7	8	9	10	11
No. of observations	16	32	64	64	128	128	128
No. of blocks	1	2	8	4	8	8	8 .

Addelman (private communication) has found a 2^{17-9} resolution V plan in 8 blocks of 32. These plans enable orthogonal estimation of all the effects and two-factor interactions and appear to be the minimal designs which allow orthogonal estimates.

If one is prepared to relax the orthogonality requirement, one can obtain reasonably precise estimates with irregular fractions (Addelman, 1961 and Whitwell and Morbey, 1961). For instance Addelman gives a fraction $\frac{3}{8}$ of a 2^7 factorial, $\frac{3}{16}$ of a 2^8, and $\frac{3}{16}$ of a 2^9 to estimate all main effects and 2-factor interactions. Whitwell and Morbey give a design using 96 observations which allows the estimation of the main effects and all but 3 of the two-factor interactions of 11 factors.

Fractional replication of the 3^n factorial system is much more difficult, as soon as one wishes to estimate two-factor interactions. In the case of 5 factors, for instance, the smallest plan which allows estimation of two-factor interactions is a 1/3 replicate requiring 81 observations. The problems of enumerating two-factor interaction clear plans for the 3^n factorial system appear to be rather difficult. Bose, Bush, Seiden and others have worked on the enumeration of orthogonal arrays and on the maximum number of factors which can be accommodated with a given number of observations, but the situation is still quite unclear. Obviously, the main experimental interest is in arrays of strength 4.

One possible way of examining a multifactor situation is by some use of random sampling of the totality of treatment combinations. This idea was first put forward, it appears, by Satterthwaite (1959) and attempts have been made to develop a theory of inference from such sampling, e.g. by Dempster (1960, 1961). It appears that the situation is very difficult. Ehrenfeld and Zacks (1961), Zacks (1963) and Ehrenfeld and Zacks (1963) have examined two procedures of random sampling the totality of treatment combinations which are based on fractional replication. It would appear that considerable further development is needed of ways of sampling the totality of treatment combinations and of analyzing the resultant sample.

The general moral to be drawn, then, with regard to multifactor (qualitative) experiments, is that it is easy to examine for main effects, more or less regardless of the number of levels, but that examination for interactions can in general be done at all easily only with two levels for each factor. It is likely that if the requirement of orthogonality is waived, plans requiring reasonable numbers of observations can be developed.

<u>THE INVESTIGATION OF DEPENDENCE OF A YIELD VARIABLE (y) ON k CONTINUOUS CONTROL VARIABLES $(x_1, x_2, \ldots, x_k)$.</u> It would seem that while there are many aspects of the dependence of a yield variable on k control variables which can be varied continuously, one can "spin off" one problem which is quite different in nature from all the others, and that is the optimization problem, namely to determine the values of $x_1, x_2, \ldots, x_k$, such that the yield is a maximum (or minimum). Of course there are situations in which there are several yield variables, say, $y_1, y_2, \ldots, y_m$ and the problem may be more complex, such as to determine the combination $(x_1, x_2, \ldots, x_k)$ for which y_1 is a maximum, subject to restraints of the type $y_2 < k_2$, $y_3 > k_3$ and so on.

OPTIMUM SEEKING. The work in this area was utterly naive, one factor at a time experimentation, until the work of Box and Wilson (1951) to whom great credit is due for tackling the problem with some degree of sophistication. I shall enumerate briefly the steps of the Box-Wilson procedure. They are:

(1) local exploration around a guessed optimum by means of a design which enables the fitting of the relationship

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \, ;$$

(2) proceeding along a line in the direction of steepest ascent in the units chosen to an optimum on that line;

(3) local exploration around this newly obtained optimum as in (1) ;

(4) proceeding along a new steepest ascent direction as in (2);

(5) repetition of steps (3) and (4);

(6) when there ceases to be a pay-off from this process, perform local experimentation around the achieved sub-optimum to enable the fitting of a second degree dependence of y on the x's;

(7) do a mathematical analysis of the achieved second degree relationship. That is, if one has found the relationship

$$y = \beta_0 + \Sigma \beta_i x_i + \sum_{i \leq j} \beta_{ij} x_i x_j \, ,$$

then one can make a linear transformation of $x_1, x_2, \ldots, x_k$ to say $z_1, z_2, \ldots, z_k$ so that

$$y = \gamma_0 + \lambda_1 z_1^2 + \lambda_2 z_2^2 + \ldots + \lambda_k z_k^2 \, ;$$

(8) this representation enables one to see the form of the relationship of y to the z's in the neighborhood of the sub-optimum achieved earlier. If all the λ_i are negative, the optimum is at the point where all the z's are zero. If some are zero there is a subspace of optima. If for example λ_1 is zero and the others are negative

> the optimum (maximum) is achieved wherever z_1 which is a linear function of the x's is zero. If of course any λ_i is positive the maximum is not at all defined by the fit.

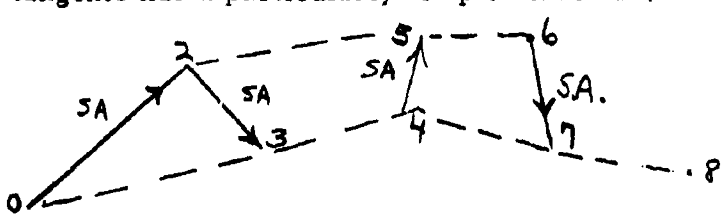
Apart from steps (6), (7) and (8) this is the standard iterated steepest ascent. Obviously the procedure was developed for the optimization of a production process in which only local experimentation is possible so as not to disrupt production.

The procedure suffers from the well-known disadvantage of steepest ascent in that progress may be excellent for the first few steps but then becomes very slow. Of course steps (6), (7) and (8) were inserted by Box and Wilson to take care of this.

A line of attack on this problem, which is closely related to the Box-Wilson approach, consists of trying to develop algorithms which will give rapid convergence to the optimum if the variable to be optimized y, say, is known without error and is of the form

$$y = b_0 + b'x + x'Cx$$

in which C is negative definite, so that a unique optimum exists. One then attempts to determine the properties of the algorithm if the relationship of the y to the x's is not of the postulated form, and if y is known only with error. The methods I know of which have this structure are the following, the method of parallel tangents due to Shah, Buehler and myself (1964), and the method of Fletcher and Powell (1963). The method of Fletcher and Powell is based on a guess of the matrix C, which would ordinarily be taken as the unit matrix, and on successive line searches, the directions of which change on the basis of previously determined gradients and on the steps to the optima on the lines. The method of parallel tangents is really just an acceleration of the initial steps of the Box-Wilson procedure which removes the necessity of fitting a second order relationship. One variant of the method of parallel tangents has a particularly simple structure:



in which the lines labelled S. A. are steepest ascent lines and the dashed lines are acceleration lines. In the absence of error and with k dimensional ellipsoidal yield contours the maximum as reached at the point labeled 2n.

There are other intuitive methods such as pattern search of Hooke and Jeeves (1962), and methods using sectioning of the factor space on the basis of tangent planes to the yield contours (Wilde, 1964).

These methods appear to use with some degree of effectiveness, the information that is accumulated by the separate local experiments. A real difficulty from a theoretical viewpoint is to evaluate the properties of all these methods, including the Box-Wilson method, in the presence of error.

Just how important it is from a practical viewpoint to establish tight clean mathematical results about the performance of these strategies in the presence of error is, I believe, a moot point. It would of course be valuable from an aesthetic viewpoint to have such information, but the difficulties of obtaining information of practical value seem to be tremendous. It is clear that the strategies described above are so loosely defined that they cannot be subjected to precise mathematical evaluation. Answers to such questions as (a) how does one explore locally? (b) what is the "spread" of the local design? (c) how does one search for the optimum on a line? (d) how does one decide when to terminate?, are not given by th procedures. They are, however, questions which the user will be able to make choices which must, of course, be somewhat arbitrary but which will be modified as information accumulates. If the local experimentation does not indicate clearly that there is a direction in which improvement can be made, more local experimentation will be done, presumably by either repeating what was done before or by "pulling in" the local design and repeating. Also, it is obvious that the experimenter will survey the totality of information obtained up to any particular point in the process and will modify the algorithms if he can spot a pattern in the response relationship.

A direct attack on the optimization problem with error was made by Kiefer and Wolfowitz (1952) with work related to that of Robbins and Monro (1951) who developed a stochastic approximation scheme for finding the value x, at which the expected value M(x) of a random variable y(x) takes a particular value. The Kiefer-Wolfowitz procedure is as follows: for the case of optimization in one dimension choose two sequences of positive numbers, c_n, a_n, such that $\lim c_n = 0$, $\Sigma a_n = \infty$, $\Sigma a_n c_n < \infty$ and $\Sigma a_n^2 c_n^{-2} < \infty$, as, for example $a_n = \frac{1}{n}$, $c_n = \frac{1}{n^{1/3}}$; take an arbitrary z_1 and then use

$$z_{n+1} = z_n + a_n \cdot \left\{ \frac{y(z_n + c_n) - y(z_n - c_n)}{c_n} \right\}.$$

Then z_n converges stochastically to the point z at which $E\, y(z)$ is a maximum. Kiefer and Wolfowitz (1952) state that there remain the problems of choices of sequences a_n and c_n which will be optimal in some sense, and the specification of a stopping rule. This line of work has been developed considerably by Blum (1954), Dvoretsky (1956), Kesten (1958) and by Sacks (1958), and others to the multidimensional case.

It is not clear at all what the attitude of the practical statistician should be to these very different approaches. Kiefer (1959) states that methods such as the Box-Wilson one or the others of the same flavor, "cannot in their present state have any role in satisfactorily solving these problems, since they have no guaranteed probability properties and are not even well-defined rules of operation." Barnard, however, in discussion of Kiefer's paper, disagreed and took the view that rules of operation which are not well-defined may be preferable to the rules which are. It would seem that the guaranteed property of convergence with probability one with an infinite number of observations is small comfort to the practical man, even though it was obviously not easy to develop procedures for which one can prove the property.

What we really lack are accounts of actual experiences with the various methods. Perhaps a good practical strategy is to use the "deterministic" schemes at first, and then turn to the stochastic schemes when the former cease to give advances.

RESPONSE SURFACE EXPLORATION. I now turn to the problem of studying the dependence of a yield variable y on continuous control variables $(x_1, x_2, \ldots, x_k)$ which has been termed a response surface exploration by Box and his co-workers.

The great bulk of the work on this problem has been by Box and his associates, stemming back to the famous Box-Wilson paper (1951). The background for the work is the paper by Box (1952) on first order multifactorial designs, which I have to review even though it was done more than

10 years ago. Here Box specified the amount of variation of each variable or factor by defining the scale unit S_i for the i-th variable as

$$S_i = \left\{ \sum_u \frac{(X_{iu} - \bar{X}_i)^2}{N} \right\}^{1/2} ,$$

where X_{iu} is the level of the i-th factor in the u-th observation. He defined the standardized variable x_{iu} as

$$x_{iu} = (X_{iu} - \bar{X}_i) / S_i .$$

He then took the design problem to be as follows:

(a) the experimenter is to specify $\bar{X}_i$, the "center" of the design and scale multiplier S_i for each variable,

(b) the designer of the experiment is to choose an array of standardized levels, x_{iu}, at which the observations are to be taken, the actual levels being

$$x_{iu} = \bar{X}_i + x_{iu} S_i .$$

In other words, the "center" of the design and the "spread" are specified by the experimenter and the only problem of the designer is to choose the x_{iu} which, of course, satisfy

$$\sum_{u=1}^{N} x_{iu} = 0, \qquad \sum_{u=1}^{N} x_{iu}^2 = N .$$

I shall comment on this basis later, but, for the present, will indicate the subsequent developments. In the case of the first order designs, the criterion was optimum estimation of the coefficients in the equation

$$y_u = \beta_0 + \beta_1 x_{1u} + \beta_2 x_{2u} + \dots + \beta_k x_{ku}$$

and the optimum design is one in which the x_{iu} are given by the columns, after the first, of a matrix $N^{1/2}O$, where O is an orthogonal matrix whose first column consists of unit elements. Box then noted that if the number of observations is k+1, the experimental points are the vertices of a regular dimensional simplex. He also noted that any rotation of this regular figure would satisfy the conditions. Box and Hunter (1957) developed in considerable detail the concept of rotatability. A design is said to be rotatable if, when the levels of the variables are standardized as stated above to be $(x_1, x_2, \ldots, x_k)$, the variance of the predicted y at a point $(x_1, x_2, \ldots, x_k)$ is a function of these x's only through $\sum_1 x_i^2$. In other words if one were to construct contours of variance of the predicted y they would be spherical with center at the 'center' of the design, when plotted in standardized levels. They stated their aim to be "to develop arrangements which generate information (equal to the reciprocal of the variance of prediction of y) symmetrically in those coordinates regarded as most relevant to the experimenter." Box and Hunter developed second order designs in 2 dimensions by taking two or more concentric rings of points, with each ring being a regular figure, for example a pentagonal design with extra center points. For 3 dimensions, they took points equally spaced on a sphere, for instance, by combining a regular tetrahedron, a octahedron, and a cube with additional center points. For more than three dimensions they suggested the combination of the points of a 2^k factorial, and 2^k points of an axial set and additional center points. Throughout attention was paid to the problem of blocking, that is, of arranging the totality of points in subsets to enable the elimination of heterogeneity between the units. Box and Behnken (1960a) developed designs by operating in a simple way on first order simplex designs. If the points of the simplex design are regarded as vectors, one can develop additional points by forming sums of the original vectors two at a time, sums of the original vectors three at a time, and so on. The configurations so developed are then scaled to satisfy the scaling and rotatability conditions. In this way they obtained, for instance, designs to examine 4 variables in two blocks of 22 observations, 5 variables in two blocks of 26 observations, 6 variables in two blocks of 34 observations, 7 variables in two blocks of 33 observations. The last one in this list is quite impressive in that it uses only 3 levels of each factor and enables all 36 coefficients of a second degree fitting to be evaluated reasonably. It is curious that all the points except the center points be on a hypersphere of radius $\sqrt{3}$ (in the standardized units). Box and Behnken (1960b) developed another series of 3-level rotatable designs by utilizing incomplete block configurations. The simplest example was the following. We have the balanced incomplete block configuration

		x_1	x_2	x_3	x_4
"Block"	1	x	x		
	2			x	x
	3	x			x
	4		x	x	
	5	x		x	
	6		x		x

If a "block" contains x_i and x_j, it is replaced by the 4 treatment combinations on x_i and x_j, (-1, -1), (-1, 1), (1, -1) and (1, 1), the other variables being taken at the zero level. Bose and Draper (1959), Draper (1960a) and others have constructed classes of second order rotatable designs. Gardner, Grandage and Hader (1959) and Draper (1960b, 1961, 1962) have developed third order rotatable designs. Throughout it appears that the designs are based on the combination of symmetrically placed points on spheres in the standardized factor space. The ideas of Box have led to the development of a considerable array of designs, all based on the concept of rotatability. Many of the designs are remarkable in that they allow the fitting of functions of the second or third degree with relatively low redundancy of experimental points. Also by choosing odd moments up to particular order equal to zero, one can prevent bias in the regression coefficients from third order coefficients in the polynomial representation.

The motivation for the development of the array of rotatable designs seems to be summarized by Box and Behnken (1960a, page 840) in the following quotation,

> "At a particular stage we are interested in the behavior of the response function 'in the neighborhood' R of some particular point P. We have in mind that the operability region O, that is the region in the space of the variables in which experiments could be conducted, is fairly extensive and that P is not close

to the boundary of O. We suppose that the neighborhood of interest about P is a region R which nowhere reaches the boundary of O and that scales, metrics and transformations are chosen either implicitly or explicitly such that R is very approximately spherical and is centered at P."

Essentially all the designs whose development I have mentioned earlier were aimed at controlling the variance of the prediction based on the fitting of a polynomial of the first second or third degree. There had been some attention to the bias in estimated polynomial coefficients from higher polynomial terms that were ignored in the fitting. Box and Draper (1959) made a direct attack on the problem of bias, within the framework of previous developments. The situation considered was that a function $f(x_1, x_2, \ldots, x_k)$ is fitted, when the true functional dependency is $g(x_1, x_2, \ldots, x_k)$. The mean square error of a prediction consists of the variance plus the square of the bias. Box and Draper consider the average over a region of interest R in the $(x_1, x_2, \ldots, x_k)$ space of these two components, for the particular case when $f(x_1, x_2, \ldots, x_k)$ is linear and $g(x_1, x_2, \ldots, x_k)$ is quadratic. They conclude that the optimal design is very nearly that which would be obtained if variance is ignored and only bias is considered. If this conclusion is accepted, it would appear that the whole class of rotatable designs based on variance considerations, need careful re-examination from the viewpoint of bias. The development depends strongly, it would appear, on the choice of the region of interest as being spherical in the standardized variables, and on equal weighting over the interior of the "sphere" of interest. The reasons for choosing this framework appear to be mathematical, in that with this framework, integrals can be evaluated. Box and Draper prove a theorem that is highly indicative of the nature of the problem. The theorem states that if a polynomial of degree d_1 is fitted by least squares over any region of interest R in the k variables, when the true function is of degree d_2, greater than d_1, then the average squared bias over R is minimized by making the moments of order up to $d_1 + d_2$ equal to the corresponding moments of a uniform distribution over R. So if one knew nothing about the true function except that it can be represented by a polynomial of indefinitely large degree one should spread the observations evenly over the region R. Clearly the definition of the region R should be made in terms of variables for which one could hope that a low degree polynomial would give a good fit.

The whole line of development appears, however, to suffer from some defects which are illustrated by the simplest designs that were developed -- the simplex first order designs. For the case of 3 variables with 4 observations, Box exhibited two designs which he claims to be equally good:

$$D_a = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \qquad D_b = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -\sqrt{2} & \frac{-\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ \sqrt{2} & \frac{-\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ 0 & 2\frac{\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ 0 & 0 & \frac{3}{\sqrt{3}} \end{bmatrix} \end{array}$$

with

$$D_a'D_a = D_b'D_b = 4I$$

where I is the 3 x 3 identity matrix. These two designs have the same center and have equal spread with the definition of Box. However, if design D_b can be used, x_1 can be varied between $-\sqrt{2}$ and $\sqrt{2}$, x_2 can be as large as $2\frac{\sqrt{2}}{3}$, and x_3 can be as large as $3/\sqrt{3}$, whereas in design D_a the limits for each x are from -1 to +1. If however, the situation is such that one can vary the x's over the ranges specified in design D_b, one would be foolish in not varying them over the same range, with the first order design D_a, and if one does, the resultant design D_a^*, say, is clearly better as a first order design than the design D_b. The same criticism has been made by Kiefer (1961b).

This simple example brings to light one of the basic problems of exploration, as opposed to optimum seeking, namely, that the region of possible experimentation must be defined if one is to attempt to develop a good design. The simple example above shows that the standardization of variables in terms of root mean square deviation of levels results in peculiar restrictions. It would seem more natural and appropriate to define the region of possible experimentation in terms of the original unstandardized variables. If one is exploring the relationship of a yield variable y to a single control variable X, a natural restriction would be that one can experiment at X values in a prechosen interval of X, say from $X = a$ to $X = b$. If one has two control variables X_1 and X_2, a possible specification of the region of permissible experimentation would be X_1 in the interval (a_1, b_1), and X_2 in the interval (a_2, b_2). It is, of course, quite likely that as soon as one has more than one variable, the region of possible experimentation will not be rectangular in the variables originally thought of. It is inconceivable that one will be able to develop a useful theory of experimentation for an arbitrary region of possible experimentation. It does, however, seem reasonable that one can choose "new" control variables that are functions of the originally thought of variables so that the region of possible experimentation in the "new" variables is approximately either a hypercube or a hypersphere. At least in this way one can set up a mathematically defined problem for which one can hope to get an answer. One might hazard the guess with the emphasis on sphericity that results from considerations of rotatability, that the rotatable designs will prove to be good designs in the case when the region of possible experimentation can be defined to be spherical. Some problems of allocation for polynomial regression within a spherical region have been considered by Kiefer (1961b) and are discussed below. It appears that a few of the Box-Hunter rotatable designs of very specialized nature are optimal with respect to two of the possible criteria. However the implications of the scaling in the Box-Hunter rotatable designs are obscure.

It appears, then, that a more fundamental approach to the problem of design would take as its base a definition of region of possible experimentation, provided by the experimenter. It is then necessary to formulate the aims of the experiment, and it is at this point that one opens a Pandora's box, because of the multiplicity of partially conflicting aims that always occurs.

Since the beginning of the formal development of designs there has been some attention to optimality of design. In the simple case of linear regression on an interval it has been known for decades that the best disposition of resources for estimation of the slope is to place half of the observations at each end of the interval. In the case of comparisons of two groups it is obvious that for maximum precision of the group difference one should have equal numbers of observations in the two groups. It is also obvious that if one has several groups, and one has the same interest in all possible differences of pairs of groups, one should, with homoscedasticity, have each group equally represented. Indeed the requirement of equal interest forces equality of representation. The classical symmetrical designs for error control, such as randomized blocks, Latin squares, balanced incomplete blocks, were considered good, because the prime interest of the experimenter was considered to be estimation, with equal interest in all the treatments, which were taken to be fixed. They were also based on the idea that the main difficulty of experimentation was to control variability between experimental units, and that variability within a group of experimental units was a monotonic function of group size.

Work on optimality of design was done early by Plackett and Burman who showed that the orthogonal 2^n plans or fractions of these, such as those based on Hadamard matrices were optimal in a useful sense for qualitative main effects of two-level factors. Indeed they resulted in as efficient estimation for each single parameter, as one could obtain if one used the whole of the experimental resources just to estimate that single parameter, and this, really, is much more than one was ever entitled to hope for. A few years later optimality of design was attacked frontally by Elfving (1952), Chernoff (1953) and Ehrenfeld (1955). The topic was taken up very extensively by Kiefer and Wolfowitz (1959) and Kiefer (1958, 1959, 1961a, b, 1962).

The whole problem of optimal design is of course, to decide what to optimize for. Kiefer (1959) lists several possibilities:

(a) maximizing the infimum of power of test of a null hypothesis against a class of alternatives (M-optimality),

(b) maximizing the limiting power of test in the neighborhood of the null hypothesis (L-optimality),

(c) minimizing generalized variance of estimates of parameters (D-optimality),

(d) minimizing the maximum eigenvalue of the variance-covariance matrix of estimates, used by Wald (1943) and Ehrenfeld (1955) (E-optimality),

(e) minimizing the trace of the variance-covariance matrix of estimates (A-optimality),

and

(f) minimizing the maximum variance of prediction over the experimental region (G-optimality).

These criteria can be applied to the totality of parameters or to a chosen subset of the parameters.

It needs to be emphasized, I think, that all these criteria are related to the problem of control of error with a model which is assumed to be true. It is not clear that designs which are good for error control are also good for detection of bias of model, as Box and Draper showed in work that I mentioned earlier. In the incomplete block problem, for instance, I am inclined to the view that designs which have some repetition of treatments within blocks are desirable. Such designs will be inefficient with regard to any of the above optimality criteria, if balanced incomplete block designs are possible, but will enable better examination of the adequacy of the usual additive model.

Kiefer (1958, 1959) has proved that balanced block designs, Latin squares, Youden squares, orthogonal arrays, are optimal with regard to criteria A, D, E and L. These results are, I suppose, of some mathematical interest, and suggest that if one has a balanced array of experimental units one should try to use the restrictions of the array. However they do not answer questions like whether one should use a Latin square design rather than a complete block design. The Latin square result states that if one is going to use the Latin square model for analysis one should use the Latin square design, and as such is not at all surprising.

Kiefer (1958, p. 676) characterizes M-optimality as "the strongest and least artificial of the four" criteria, D, E, M and L, and it was attention to testing of hypotheses that led Kiefer to give the examples which generated, apparently, much unnecessary heat at the Royal Statistical Society meeting. Kiefer pointed out that if one had 6 observations to be split among three populations which are $N(\theta_i, \sigma^2)$, $i = 1, 2, 3$, then different designs were optimal for the three problems:

(a) point estimation of θ_1, θ_2, θ_3

(b) testing the hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$

(c) testing the hypothesis $\theta_1 = \theta_2 = \theta_3$,

where in (b) and (c) one is interested in alternatives near the null hypothesis. For problem (a) one should take 2 observations from each population, for problem (b) one should take one of the populations at random and use all 6 observations on it, while for problem (c), one should take two of the three possible populations at random and then take 3 observations from each. This example shows very clearly that different criteria of optimality can give radically different designs.

The work of Kiefer and Wolfowitz is more informative, I think, in the area of polynomial regression than in the area of qualitative experimentation. The history of optimum allocation for polynomial regression appears to be as follows. In the one-dimensional case for which the units can be chosen so that the interval of experimentation is $(-1, 1)$, Guest (1958) considered the G criterion above, the maximum variance of a prediction, and showed that this was minimized by placing $\frac{1}{k+1}$ of the points at each end of the interval and at the zeros of the derivative of the k-th degree Legendre polynomial. Hoel (1958) showed that if one wishes to minimize the generalized variance of the coefficients of a k-th degree polynomial the optimum allocation was the same as that obtained by Guest. Kiefer and Wolfowitz (1959) showed that the best estimate of the coefficient of x^h, when a polynomial of degree h was required for the x-interval $(-1, 1)$, was to place $1/2h$ of the observations at each end of the interval and $1/h$ at the points $\cos(j\pi/h)$, $1 \leq j \leq h - 1$, which may be termed Chebychev spacing. In experimentation on the square $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$, the best test of interaction term x_1x_2 is obtained by placing 1/4 of the observations at each corner. Of course all the above solutions depend on the total number of observations being appropriately divisible. Kiefer (1959) gives the example that with 4 observations, the best placement for cubic regression on the interval (-1, 1) is at the values ± 1, $\pm 1/\sqrt{5}$, and with 5 observations the best placement is at the values 0, $\pm$ 0.511, ± 1. The dependency of optimum design on the specific value of N is avoided by Kiefer and Wolfowitz who consider how best one would place an infinite number of observations. Such placements can be regarded as approximate designs, and they proved (1960) a rather remarkable theorem that the design using a large number of observations which minimizes the generalized variance of the coefficients of a polynomial fitting would also minimize the maximum variance of a predicted value over the experimental region. It is not clear just how useful this result is for reasonable numbers of observations, and how one should use the approximate placing given by the theorem to arrive at a placement for a reasonable number of observations.

With this proviso, however, this later work of Kiefer and Wolfowitz gives an indication for the choice of design in "response surface exploration," at least if one views the matter as a problem of polynomial approximation. The fact that the generalized variance of coefficients is minimized would tend to indicate (though it does not guarantee) that all the coefficients of a polynomial are being estimated with reasonable precision, and the fact that the maximum variance of a prediction is minimized should to a moderate extent permit the discovery of lack of fit by the polynomial.

In the case of quadratic regression on a hypercube bounded by -1 and 1 in each direction, in q(= 2, 3, 4, or 5) dimensions, Kiefer (1961) shows that the best "infinite" design is to assign a proportion α of the experimental points to each of the 2^q corners, a proportion β to the mid point of each of the $q2^{q-1}$ edges, and a proportion γ to the center of each of the $q(q-1)2^{q-3}$ 2-dimensional faces of the hypercube. In the case of q equals 5, the values of α, β, γ are

$$\alpha = .01928$$

$$\beta = .0003125$$

$$\gamma = .004475$$

However, in view of the fact that the α set contains 32 points, and the β and γ sets contain 80 points each, this "infinite resources" answer is not really useful. It does not tell us, for instance, how we should place say 50 or 60 observations. It does appear to indicate, however, that if the G criterion, which seems a somewhat superior one for exploration, is adopted, then the experimental points should be placed near the corners and edges of a rectangular experimental region. This is in considerable contrast to the rotatable designs discussed earlier, which seem to devote much attention to the center and interior of the region.

Later Kiefer (1961b) examined polynomial regression when the region of experimentation and interest is the hypersphere or "ball," $\Sigma x_i^2 \leq 1$. It might be expected that the designs he would get would be related to the rotatable designs in that the latter seem to be aimed at a spherical region of interest. Kiefer considers the approximate case, that is, the "infinite resources" case, so that D-optimality and G-optimality are equivalent. He was able to characterize partially the approximate optimal design, and showed that it is rotatable. In the case of linear polynomial fitting, the best design has equal weight at the vertices of an inscribed regular simplex or the vertices of any other inscribed regular polygon. So for this case the maximally spread simplex design of Box (1952) is optimum with these criteria. Also in two-dimensions with quadratic regression, the design with one observation at the center and one at each vertex of an inscribed regular pentagon is D-optimal and hence G-optimal. However, apparently most of the rotatable designs do not have these optimality properties. I cannot regard the lack of optimality properties as seriously as apparently Kiefer does. Kiefer (1961b, p. 398), feels that he justified for the first time the use of rotatable designs but I regard his results as mathematically rather elegant, and not totally relevant to the problems of the experimenter. The representation of yield as a polynomial in the control variables is unaesthetic and uneconomical of parameters, except in the optimization problem. Even in the optimization problem it is highly questionable whether one should do local experimentation other than to get gradients. I agree with Kiefer that the framework within which Box and his associates have worked has serious logical deficiencies, but also have the view that they developed some very useful designs and design ideas.

CONCLUSIONS ON THE EXPLORATION PROBLEM. The problem of studying the dependence of a yield variable on control variables is not well-defined. Experimenters with this problem will have a multiplicity of aims, such as to obtain reasonably precise estimates, reasonable strength of evidence against particular null hypotheses of interest, ability to select a functional form that represents the data well and is economical of parameters, and so on.

The theoretical statistician can obtain optimal designs only by forcing the problem into a highly idealized simplified form, and there is a tendency to regard the optimal design for idealized simplified form as the design the experimenter should use. This attitude seems to be exemplified by Kiefer's remark (1959, p. 316), "Why not think in terms of the right space of decisions from the outset?" I have yet to meet an experimenter whose aims can be represented by a space of decisions, which is sufficiently well-defined to be susceptible to such an attack.

The work of the optimizers is, however, valuable, because it gives us suggestions of respects in which a plan may be weak. The upshot for me of the work I have reviewed is exemplified by the following cases. In the case of 3 factors in a cubic region (-1, 1), I would do the following:

(i) with 4 observations I would take a 2^{3-1} factorial at the corners;

(ii) with 9 observations I would use a 1/3 replicate of the 3^3 with levels -1, 0 and 1 for each factor;

(iii) with 15 points I would use the corners and center of each face and the center which is essentially a central composite design but not rotatable;

(iv) with 27 points I would use the full 3^3 factorial with levels -1, 0, and 1.

If the problem is really one of studying the dependence I would try to persuade the experimenter to do the full factorial (iv), because it would enable me to think, to some advantage, about representations other than by a polynomial. In the case of 4 factors, I would think with a low number of possible observations in terms of main effect plans with observations at the corners. If more were possible I would consider the sets of points:

S_1 : $(\pm 1, \pm 1, \pm 1, \pm 1)$

S_2 : $(\pm 1, \pm 1, \pm 1, 0)$ with permutations

S_3 : $(\pm 1, \pm 1, 0, 0)$ with permutations

S_4 : $(\pm 1, 0, 0, 0)$ with permutations

and S_5 : $(0, 0, 0, 0)$.

I would take a combination of these sets. For instance, if I were allowed 24 points, I would use S_1 and S_4, and with 40 points I would use S_1 and S_3 and so on [cf. De Baun, 1959]. Obviously my views have been influenced by both Box's work and by Kiefer's work.

It is, however, also obvious that a realistic procedure should take account of sequential plans. Consider, for example, the investigation of the dependence of a yield variable y on a control variable x in (-1, 1). Suppose that the information on y for each chosen x is available as soon as the experimental run has been made. A rational procedure is not to use Chebychef spacing or Legendre spacing, but to take an observation at x=-1 and at x=+1. One would then take one at x=0, and try to connect three points by a quadratic, or seek a reasonable transformation (non-linear) of the x scale so that the 3 observations fell on a line. One would then probably take additional observations in the middle of the gaps of the best picture one has obtained up to the time of planning new observations. One would, of course, have prefaced the whole matter by obtaining a rough idea of experimental error. It is very difficult to see how the concepts of decision theory and testing of hypotheses can be brought to bear on such a process.

It is clear that practical optimum designing depends on more ingredients than have so far been incorporated in the theory. What one should do depends crucially on:

(a) what use will be made of incomplete information?

(b) what is the rate of feed-back of experimental information?

(c) will the experimenter be able to do additional experiments to fill in gaps in information?

(d) how valuable is information to the experimenter in relation to time? [What is the present value of future information? This will of course depend on what the future information is.]

(e) what is the cost of experimentation? The simple idea of a fixed cost per observation appears to be relevant at best only in some technological studies.

The difficulties of constructing a theory which incorporates these aspects appear to be very great, but should not dissuade us.

FINAL NOTE. It is unavoidable that I cannot describe the results of every paper in the field. The reference list gives only papers referred to and much good work is not discussed. A notable example is the work of Scheffé (1963) on experimentation on a simplex.

REFERENCES

Addelman, S. (1961). Irregular fractions of 2^n factorial experiments. Technometrics 3: 479-496.

__________ (1962). Orthogonal main effect plans for asymmetrical factorial experiments. Technometrics 4: 21-47.

__________ and O. Kempthorne. (1961). Some main effect plans and orthogonal arrays of strength two. Ann. Math. Stat. 32: 1167-1176.

Anscombe, F. J. (1961). Examination of residuals. Proc. fourth Berkeley Symposium on Mathematical Statistics and Probability 1,: 1-36.

__________ and J. W. Tukey. (1963). The examination and analysis of residuals. Technometrics, 5: 141-160.

Blum, J. (1954). Multidimensional stochastic approximation methods. Ann. Math. Stat. 25: 737-744.

Bose, R. C. and N. R. Draper. (1959). Second order rotatable designs in three dimensions. Ann. Math. Stat. 30: 1097-1112.

Box, G. E. P. (1952). Multifactor designs of first order. Biometrika, 39: 49-57.

__________ and D. W. Behnken. (1960a). Simplex sum designs: A class of second order rotatable designs derivable from those of first order. Ann. Math. Stat. 31: 838-864.

__________ and __________ (1960b). Some new 3 level designs for the study of quantitative variables. Technometrics 2: 455-476.

__________ and N. R. Draper. (1959). A basis for the selection of a response surface design. J. A. S. A. 54: 622-654.

__________ and J. S. Hunter. (1957). Multifactor experiments for exploring response surfaces. Ann. Math. Stat. 28: 195-241.

__________ and __________ (1961a). The 2^{k-p} fractional factorial designs. Technometrics 3: 311-352.

__________ and __________ (1961b). The 2^{k-p} fractional factorial designs. Technometrics 3: 449-458.

__________ and K. B. Wilson. (1951). On the experimental attainment of optimum conditions. J. R. S. S., B. 13: 1-45.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Stat. 24: 586-602.

Daniel, C. (1959). Use of half-normal plots in interpreting two-level experiments. Technometrics 1: 311-341.

DeBaun, R. (1959). Response surface designs for three factors at three levels. Technometrics 1: 1-8.

Dempster, A. P. (1960). Random allocation designs I: On general classes of estimation methods. Ann. Math. Stat. 31: 885-905.

__________ (1961). Random allocation designs II: Approximate theory for simple random allocation. Ann. Math. Stat. 32: 387-405.

Draper, N. R. (1960a). Second order rotatable designs in four or more dimensions. Ann. Math. Stat. 31: 23-33.

____________ (1960b). Third order rotatable designs in three dimensions. Ann. Math. Stat. 31: 865-874.

____________ (1961a). Third order rotatable designs in three dimensions. Ann. Math. Stat. 32: 910-913.

____________ (1962). Third order rotatable designs in three factors analysis. Technometrics 4: 219-234.

Dvoretsky, A. (1956). On stochastic approximation. Proc. third Berkeley Symposium on Mathematical Statistics and Probability. 439-455.

Ehrenfeld, S. (1955). On the efficiency of experimental designs. Ann. Math. Stat. 26: 247-255.

____________ and S. Zacks. (1961). Randomization and factorial experiments. Ann. Math. Stat. 32: 270-297.

____________ and ________ (1963). Optimal strategies in factorial experiments. Ann. Math. Stat. 34: 780-791.

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Stat. 23: 255-262.

Finney, D. J. (1945). The fractional replication of factorial experiments. Ann. Eug. 12: 291-301.

Fletcher, R. and M. J. D. Powell. (1963). A rapidly convergent descent method for minimization. The Computer Journal 6: 163-

Gardner, D. A., A. H. E. Grandage and R. J. Haden. (1959). Third order rotatable designs for exploring response surfaces. Ann. Math. Stat. 30: 1082-1096.

Guest, P. G. (1958). The spacing of observations in polynomial regression. Ann. Math. Stat. 29: 294-298.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Stat. 29: 1134-1146.

Hooke, R. and T. A. Jeeves. (1962). Direct search solution for finding stationary values of a function of several variables. J. Assoc. Comp. Mach., 8: 212-229.

Kesten, H. (1958). Accelerated stochastic approximation. Ann. Math. Stat. 29: 41-59.

Kiefer, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs. Ann. Math. Stat. 29: 675-699.

________ (1959). Optimum experimental designs. J. Roy. Stat. Soc. B, 21: 272-319.

________ (1961a). Optimum designs in regression problems II. Ann. Math. Stat. 32: 298-325.

________ (1961b). Optimum experimental designs V with applications to rotatable and systematic designs. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability 1 : 381-405.

________ (1962). Two more criteric equivalent to D-optimality of designs. Ann. Math. Stat. 33: 792-796.

________ and J. Wolfowitz. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Stat. 23: 462-466.

________ and ________. (1959). Optimum design in regression problems. Ann. Math. Stat. 30: 271-294.

________ and ________. (1960). The equivalence of two extremum problems. Canad. J. of Math. 12: 363-366.

Kurkjian, B. and M. Zelen. (1962). A calculus for factorial arrangements. Ann. Math. Stat. 33: 600-619.

Rao, C. R. (1947). Factorial arrangements derivable from combinatorial arrangements of arrays. J.R.S.S., B, 9: 128-139.

Robbins, H. and S. Monro. (1951). A stochastic approximation method. Ann. Math. Stat. 22: 400-407.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Stat. 29: 373-405.

Satterthwaite, F. E. (1959). Random balance experimentation. Technometrics 1: 111-138.

Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. J. Roy. Stat. Soc. B, 25, No. 2: 235-263.

Shah, B. V., R. J. Buehler and O. Kempthorne. (1964). Some algorithms for minimizing a function of several variables. J. SIAM 12: 74-92.

Tukey, John W. (1959). Discussion of papers by Messrs. Satterthwaite and Budne. Technometrics 1: 166-174.

__________ (1962). The future of data analysis. Ann. Math. Stat. 33: 1-67.

Wald, A. (1943). On the efficient design of statistical investigations. Ann. Math. Stat. 14: 134-140.

Whitwell, John C. and Graham K. Morbey. (1961). Reduced designs of resolution five. Technometrics 3: 459-478.

Wilde, D. J. (1964). Optimum seeking methods. Prentice Hall Inc.

Wilk, M. B. and R. Gnanadesikan. (1963). Graphical analysis of multiresponse experimental data using ordered distances. Technometrics 4: 1-20.

________ and __________. (1964). Graphical methods for internal comparisons in multiresponse experiments. Ann. Math. Stat. 35: 613-631.

Zacks, S. (1963). On a complete class of linear unbiassed estimators for randomized factorial experiments. Ann. Math. Stat. 34: 751-768.

APPLICATIONS OF DIMENSIONAL ANALYSIS TO MULTIPLE REGRESSION ANALYSIS

David R. Howes
U. S. Army Strategy and Tactics Analysis Group
Bethesda, Maryland

INTRODUCTION. The theory of dimensions which I will discuss, is concerned with the relations that may be found between quantities occuring in nature as a result of the operations which must be performed in order to measure them. Dimensions are things like inches, pounds, minutes, or volts, or rather, the characteriestics which standard measurement units such as inches, pounds, minutes, or volts characterize; namely, length, mass, time, or electrodynamic potential. Physicists and engineers have been making an analysis of these dimensions, as a phase of every problem for many years. The point I want to make today is that a dimensional analysis of a problem should be even more important to a statistician, since such an analysis can reduce both the size of an experiment and the work required to analyze it. As it is not hard to show, a dimensional analysis could, in a given problem, reduce the sample size by more than half. In fact, in the present stage of development of the design of experiments, dimensional analysis offers greater hope for reducing the cost of experiments than any further refinements in construction of blocks, replicates, and so forth. In addition to its promise toward reducing the cost of an experiment, dimensional analysis has another virtue almost equally important. That is, a dimensional analysis carefully conducted can yield a great deal of information, which would otherwise be unobtainable, about the type of model which should be adopted in planning and analyzing an experiment.

Although the basic ideas in dimensional analysis have been in use among physicists and engineers for over a century, they are apparently almost unknown among statisticians; at least there is no reference to the subject in the index of the Journal of the American Statistical Association or any other statistical publication or textbook that I am acquainted with.

However, the theory of dimensions has profound implications in the study of statistical problems. The theory, originated by Joseph Fourier [1] is based upon the observation that; to quote Fourier:

> "Every undetermined magnitude or constant has one dimension proper to itself, and the terms of one and the same equation could not be compared if they had not the same exponent of dimension."

Thus, if a group of variables are connected in a linear equation involving coefficients to be determined by a multiple regression those coefficients must represent quantities whose dimensions are such as to give the same overall dimension to every term in the equation. Similarly also, for equations of higher degree.

Therefore, when linear or polynomial expressions are selected as models for the design or analysis of an experiment, it should be required that any coefficients postulated in these expressions have a dimensionality which bears a reasonable interpretation in context. However, one might justly criticize a model in which one of the coefficients were required to assure the dimension of cubic tons per square degree dollar (and I have seen such an example). If we apply the theory in a more detailed way we can arrive even more exactly at the type of model which should be appropriate, and obtain information concerning those interactions which are to be expected and which can be ruled out.

An example will serve to illustrate what dimensional analysis can provide the statistician. In Duncan, 2, one finds an experiment in which cotton yarn specimens are tested for yarn strength, yarn length, fiber tensile strength, and fineness. Slide No. 1.

X_1: Yarn Strength, Pounds

X_2: Fiber Length, Inches

X_3: Fiber Tensile Strength, Pounds per square inch

X_4: Fiber Fineness, Micrograms per inch.

This problem is discussed and analyzed as one involving one dependent, and three independent variables. However, as a result of dimensional analysis, one is able to postulate:

$$f\left(\frac{X_1 X_3}{X_4^2}\right) \approx \frac{X_2 X_3}{X_4}$$

where an univariate relationship exists between the quantities on the right and left. An analysis of the data is shown in figure 1. Using the method of least squares, and the data on page 674, one obtains the regression line:

$$X_1 X_3/X_4^2 = .05872 \ (X_2 X_3/X_4) - 3.90$$

with a coefficient of correlation of r = .955. Applying this formula to another set of data from the same source given on page 699, and comparing predicted with actual values of X_1, one has a sum of residuals of 107, and a standard error of 9.86. Comparable results using the multiple regression equation given on page 693 are 114 for the sum of residuals and 8.22 for the standard error.

The value of the dimensionless equation is appreciated by considering that it contains only two fitted constants as against four for the multiple regression equation and yet predicts approximately as well. Moreover, the calculations were vastly simplified. Finally, by keeping the number of fitted constants to a minimum, one avoids the danger in complex predictive hyper surfaces that wild contortions may occur in regions which do not happen to be represented in the data, yet which are superficially interpolative. This simplifies and improves the situation from every point of view. In general, the insights provided by dimensional analysis are valuable, and the method is easy.

THEORY OF DIMENSIONS. As is shown in Murray [3], any primary dimension which is effectively present in an experiment or process can be used to reduce the number of variables by one. This fact is explained as follows: External standards of measurement, such as an international metric unit are not necessary to describe a process. Any quantity within the process itself could serve as a standard of measurement for other variables measured in the same dimension. In any formulae, tables or charts describing a process measured in this way, the symbol of the variable taken as the mensurator would not occur, since, being the standard, its value would always be unity. However, an outside observer could convert these same formulae, tables or charts for use with external measurement units, by supplying the symbol of the mensurator as a denominator under the symbol of each variable to measured.

The ratio of a simple or compound variable to its mensurator is referred to as a dimensionless term. Since we reduce the number of variables by one for each primary dimension, m variables in n dimensions can be represented in the form of m-n dimensionless terms provided an adequate system of mensuration can be found.

Each variable may be said to have a vector of dimension

$$P = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{bmatrix}$$

where each U_i represents the exponent taken by the i th dimension in the dimensionality of the variable as a whole. Thus, if i_1 = mass, i_2 = length and i_3 = time, the dimensional vectors of speed (meters^1 min.$^{-1}$) and pressure (KG^1 meter^{-2}) are:

$$\text{Speed} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad \text{Pressure} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

The dimensional vectors of all variables that can be relevant to a problem forms a set which has the property that if a vector P belongs to the set so does CP where C is selected arbitrarily, and if P_1 and P_2 belong to the set, so does $P_1 + P_2$, the vector of the product of the variables. These properties define a linear vector space which is a closed set.

If we can find n variables with linearly independent vectors in this space, these variables are said to span the vector space. The vector of any variable can be duplicated from the n independent vectors by scalar multiplications and vector additions. These independent vectors are a basis for the vector space and a mensurator for any variable can be constructed by combining the variables having these vectors. Any n vectors can be tested for independence by forming the determinant which has these vectors as columns. If it is not zero, they are independent.

Provided then, that a basis of n independent vectors exists, all m-n variables can be measured by mensurators constructed from the n variables having those vectors. Thus, the process can be represented

(1) $$f(\Pi_1, \Pi_2, \ldots, \Pi_{m-n}) = 1$$

where each term is composed of the ratio of a variable to its mensurator.

The theorem above is referred to as the Buckingham Theorem after Buckingham [4]. For practical methods of constructing sets of terms see Langhaar [5], or Murphy [6].

The completely general functional expression (1) is as far as the theory of dimensionality can take us. The explicit function must be determined by experimentation and statistical analysis, or from subject matter theory, or both. When $m - n = 1$, the problem is solved by dimensional analysis alone, and when $m - n = 2$, simple statistical techniques will usually suffice.

MIXED DIMENSIONAL AND DIMENSIONLESS EXPRESSIONS. Previous texts have considered only completely dimensionless representations and have ignored the possibility of a partially dimensionless formulation. Under these circumstances no guidance was provided for the analysis of problems in which the vectors of the variables given are insufficient to span the vector space. This occurs when a complete specification of the forces acting in a process cannot be made. Such incomplete dimensional specifications do not necessarily negate the advantages of dimensional analysis. Some of the variables may still be reduced to a common mensurator, thus, permitting some reduction in the number of variables. For example, consider a chemical experiment with the following variables:

X_1 Amount of Yield, Mols

X_2 Amount of Reactant, Mols

X_3 Amount of Acid, Mols

X_4 Temperature, Degrees, C

X_5 Length of Reaction, Minutes.

Obviously, no mensurator can be found for X_4 or X_5. Therefore, a completely dimensionless expression is impossible - unknown forces have been omitted from the specification. However, X_3 can serve as a mensurator for X_1 and X_2 permitting the formulation

$$X_1 = f\left(X_2, X_4, X_5\right)$$

where the unit of length is the length of X_3, or

$$\frac{X_1}{X_3} = f\left(\frac{X_2}{X_3}, X_4, X_5\right)$$

in any units.

Therefore, an incomplete dimensional specification reduces our ability to condense the number of variables. If the variables are all incomensurable we can make no condensation. If, however, some of the variables are commensurable, we can reduce their number to the extent that commensurability exists.

A CHEMICAL WARFARE EXAMPLE. Thus far, we have described a theory which offers a clear-cut reduction in the number of variables required in an experiment. Its implications are so plain that only skepticism concerning its validity would be grounds for ignoring the theory and benefits to be derived from Dimensional Analysis.

In order to dispel skepticism concerning the theory, I have applied Dimensional Analysis to a number of problems in various fields from which data was available; problems in Chemical Engineering, Agricultural Economics, and Quality Control. In every case, the Dimensional Analysis accomplished a successful reduction in the number of variables with a predictive value equal or superior to any previous analysis made using the raw variables.

One application was in the field of assessment of the coverage capability of toxic chemical ammunition against military targets. I am gratified by the results obtained so far, since for many years I was active in this field and am aware of the high potential savings that would result from any simplification in the problem; especially any model which would eliminate or reduce requirements for testing ammunition over wide ranges of meteorological conditions.

I am aware that much theory has been evolved which purports to describe behavior of toxic clouds in the atmosphere, but also am aware that the mathematical complications of these theories are such that actual models for purpose of prediction rest on approximations whose accuracy is uncertain, and which do not, in my experience, match up with test data obtained in the field, Dimensional Analysis cuts across this theory and leads to an empirical model which accounts for meteorological factors more satisfactorily than existing models.

To illustrate this analysis, Figure No. 1 shows the variables generally agreed to be pertinent to the problem under the assumption of isotropic diffusion. You will note that n, the Sutton turbulence parameter enters into the problem not as a variable, but as the exponent of dimension in which the diffusivity is expressed.

The temperature is omitted from this list since there is no completely agreed manner for considering it and it does not fit into the dimensional picture here. Sutton's theory ignores it and it is customary to consider it as a component of source strength; varying the effective source strength.

Figure 1 shows a set of three dimensionless Π terms which according to our theory should be able to replace the six variables shown on the previous slide. A study of these terms shows that the data from one experiment in the field could be used to predict the results of other experiments under different meteorological conditions. Also, it implies that the results of all conceivable experiments could be represented by a single surface in three dimensional space, or as a family of curves in two dimensional space.

Figure 2 shows the results of two field trials plotted in the space of the dimensionless variables shown previously. The two trials were conducted with the same type of shell, and at approximately the same temperature. However, the wind and stability conditions were considerably different, and therefore, the coverage figures obtained were also considerably different. In the 202 trial on the left the wind speed was 1 meter/sec as correspond with 3.23 meters/sec for the trial 203 the right. Stability was moderate inversion for the trial on the left and moderate lapse for that on the right.

The Sutton parameters n, and C were calculated form from the wind-height profiles given in the test reports using the Barad-Hilst equations.

As the chart shows, the two trials were sufficiently different to prevent any overlap between the two families of curves. However, the critical point is that the two sets of observations are recognizably members of the same family and that the curve - 8.6, which occurs in both sets of data matches up very well: in fact, a line projected through the two points obtained in trial #202 passes exactly through 4 of the 6 points shown for trial #203. This is highly encouraging since it was only in meteorological conditions that the trials were different, implying that the analysis given did, in fact, satisfactorily account for the changes in the area dosage curve, and did so for every time interval.

We infer from this example that additional tests could be analyzed to fill in the blank spots on our chart and an empirical equation fitted to this data with ease, since only three variables are involved, and the curves obtained are approximately colinear.

DIMENSIONAL ANALYSIS AND MULTIPLE REGRESION ANALYSIS. Dimensional Analysis is a great help in solving the difficulties encountered in multiple regression analysis. It has several advantages:

a. The number of variables, and therefore the extent of the calculations required, is reduced.

b. The freedom with which alternative representations of the data can be formed facilitates the discovery of collinear representations which simplify the analysis.

c. The predictive equation partakes of a structural validity not entirely dependent on statistical estimation.

The value of the dimensional approach may be appreciated in relation to the Bean Ezekiel graphical method of multiple curvilinear regression analysis, [7]. In that procedure, no explicit mathematical form need be ascribed to the relationship among the variables but by an iterative graphical process an increasingly accurate approximation to the curves involved is obtained, and the result is a set of charts which can be used directly for predictive purposes, or, if desired, converted to tables, nomographs or slide rules. A scatter plot of residuals is also obtained, for an estimate of error. The principal drawback of the method was the frequent inability of the analyst to isolate recognizable "draft lines" at the outset due to non-collinearity of response. The freedom of dimensional representation should largely overcome this difficulty and increase the scope of the method.

CONCLUSION. The foregoing exposition has shown that the application of dimension theory to statistical problems can result in valuable insight and savings in experimental design and analysis and should, therefore, become part of the equipment of statisticians generall. Objections to the theory have at times been advanced, usually on the basis of special examples wherein functional invariance under change of units prevails without dimensional homogeniety (see Bridgman, [8]). However, in its favor, the results obtained by Dimensional Analysis are obtained also from the theory of partial differential equations as applied to physical problems (see Langhaar, Chapter 10); the theory has successfully supported the researches of Maxwell, Rayleigh, Helmholtz, and others, and neither the literature nor the experience of the present writer offers an instance wherein the supposed relationships have been found absent in fact.

It is also unclear to what extent the standard statistical designs, tests, and techniques customarily applied to dimensional variables can be applied to dimensionless variables. Thus, it is recognized that many developments in error analysis and the theory of sampling will be required to exploit Dimensional Analysis to its fullest (a recent paper by Halperin and Mantel, [9] would appear to be of value in this connection). An obvious case requiring attention is that of setting confidence limits on a dependent variable which is a constituent of one or more terms, although setting limits for the term themselves would be straightforward.

It is hoped that being made aware of the advantages of Dimensional Analysis, statisticians will bend it to their needs with the necessary developments.

REFERENCES

1. Fourier, J. The Analytical Theory of Heat Dover Press, 1955 Ch. II, Sec. 9.

2. Duncan, A. J. Quality Control and Industrial Statistics Richard Irwin Co. 1959, p. 674.

3. Murray, F. J. Mathematical Machines Vol. II, Columbia University Press 1961, p. 229.

4. Buckingham, E. On Physically Similar Systems Phys. Rev. Vol. IV, No. 4, 1914, p. 345.

5. Langhaar, H. L. Dimensional Analysis and the Theory of Models Wiley & Sons 1951, Ch. 4.

6. Murphy, G., Similitude in Engineering Ronald Press, New York 1950. p. 23.

7. Ezekiel, M. and Fox, K. Methods of Correlation and Regression Analysis Wiley and Sons, 1951. Ch. 16.

8. Bridgman, P. W. Dimensional Analysis Yale University Press, 1922, pp. 41-43.

9. Halperin, M. and Mantel, N. Interval Estimation of Non-Linear Parametric Functions, JASA, Vol. 58, No. 303, 1964, pp. 611-627.

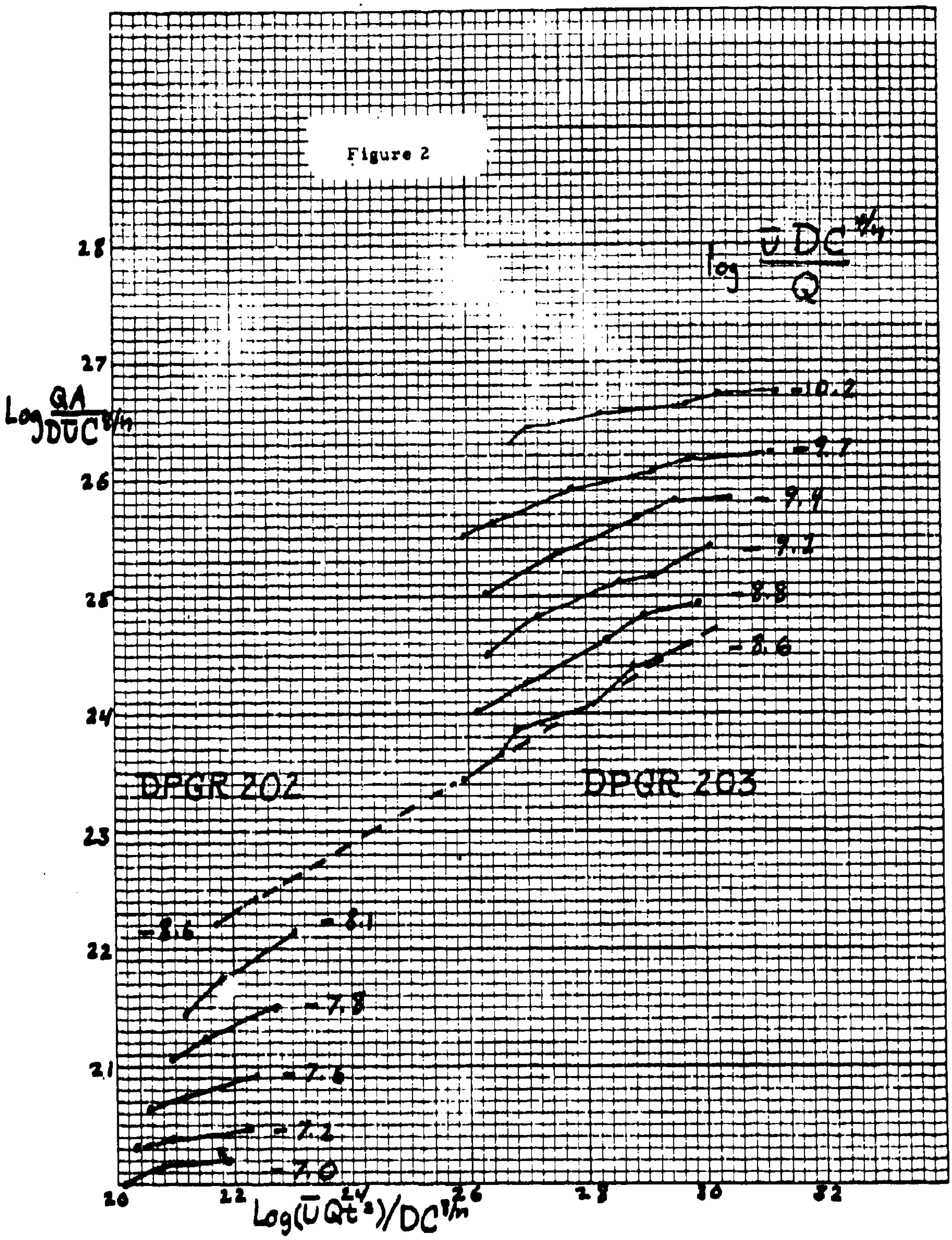
CHEMICAL WARFARE PROBLEM

Variable	Dimension
AREA	$MTRS^2$
DOSAGE	$MG\text{-}MIN/MTR^3$
SOURCE STRENGTH	MG
TIME	MINUTES
WIND SPEED	MTRS/MIN
DIFFUSIVITY	$(MTRS)^{n/2}$

$$\frac{QA}{DUC^{8/n}}, \quad \frac{\bar{U}DC^{4/n}}{Q}, \quad \frac{\bar{U}Qt^2}{DC^{8/n}}$$

Figure 1

Figure 2



28
27
26
25
24
23
22
21
20
22
24
26
28
30
32
Log(ŪQt²)/DC^1/n
Log QA/DŪC^1/n
log ŪDC^1/4/Q
-10.2
-9.7
-9.4
-9.2
-8.8
-8.6
DPGR 202
DPGR 203
-8.6
-8.1
-7.8
-7.6
-7.2
-7.0

THE USE OF REGRESSION ANALYSIS FOR CORRECTING FOR MATRIX EFFECTS IN THE X-RAY FLUORESCENCE ANALYSIS OF PYROTECHNIC COMPOSITIONS

Raymond H. Myers
Virginia Polytechnic Institute: Blacksburg, Virginia

and

Bernard J. Alley
U. S. Army Missile Command: Redstone Arsenal, Alabama

I. INTRODUCTION. X-Ray fluorescence methods are widely used in industry for the analysis of a variety of materials. The non-destructive nature and exceptional speed of these methods are largely responsible for their widespread use and increasing acceptance. The direct analyses of many materials, for example can be accomplished 20 to 50 times faster than by conventional chemical procedures. This allows sufficient time after an analysis to permit the correction or rejection of a substandard batch of material before processing is completed.

The actual X-Ray fluorescence method of analysis may be briefly described as follows: the primary beam from an X-Ray tube impinges on the surface of a specially prepared sample. The components in the sample surface are immediately excited and emit their characteristic emission lines in all directions. Qualitative analyses are made by determining the angles at which the characteristic emission lines from the sample occur. Quantitative analyses can in general be performed on a particular component of a mixture, of say K components by positioning the radiation detector at an angle which corresponds to the characteristic emission line for that component and measuring the emission line intensity. The intensity is then related to the component percentage by a suitable calibration procedure.

The intensity of a component's characteristic radiation is not a simple function of the concentration of that component alone in the sample. The intensity depends also on the concentrations of the other components. This is caused by the absorption and enhancement among the components, of the primary and excited radiation. The existence of these interelement or "matrix" effects is one of the more serious problems encountered in X-Ray fluorescence analysis and hence inhibits, to a great extent, the use of this technique as a quantitative analytical tool.

Many non-mathematical methods have been devised to either minimize or correct for these matrix effects. However, they have been found to be either too costly or too time consuming on samples from large scale production of multicomponent mixtures. It is the purpose of this paper to discuss the use of regression analysis in the correction of these interelement effects for the estimation of concentration of individual components in a mixture and to emphasize the application to a particular solid rocket propellant mixture in current use by the U. S. Army at Redstone Arsenal, Huntsville, Alabama.

Effect of Solid Particle Size

A problem which may be encountered when one is analyzing slurry mixtures containing solid constituents is the influence of solid particle sizes on the X-Ray intensities. It might be necessary that any analytical procedure contain some type of correction for this effect, unless of course the individual particle sizes always remain constant throughout production. Part I of this paper gives the analytical technique for the situation in which the particle sizes were experimentally held constant. Part II extends the analytical procedure to the case of variable particle size.

II. ESTIMATION OF CONCENTRATION (Particle Size Constant). Samples of a five component solid propellant mixture were prepared and analyzed for four of the components. (The actual ingredients are classified and hence we shall denote them in the text as components 1, 2, 3, and 4 respectively). These samples were taken from the twelve different batches in a narrow concentration range in which the product is usually manufactured. The particle sizes of the solids in the slurry mixture were held essentially constant. The number of seconds for a fixed count intensity measurements were recorded in rapid succession for each component. The same was done for a synthetic standard sample. The response variable used was $R = t_s/t_c$, where t_s is the number of seconds for the standard and t_c the number of seconds for the component in question. This is standard procedure used in this type of X-Ray work. The purpose of the standard and the subsequent use of the ratio of the standard reading to the unknown reading is to correct for electronic and mechanical changes in the spectrograph. The data is found in Table I.

Consider the model;

(1) $R_{ij}=B_{i,0}+B_{i,1}X_{1j}+B_{i,2}X_{2j}+B_{i,3}X_{3j}+B_{i,4}X_{4j}+\epsilon_{ij}$, (i=1, 2, 3, 4)

where R_i is the intensity ratio for component i, X_1, X_2, X_3, and X_4 are the concentrations of the individual components, the B's are regression coefficients, and ϵ_{ij} is the random error associated with R_{ij}. Note that the concentrations of each component appear in the model despite which of the four ingredients is being detected. Least squares estimates of the regression coefficients were found for the four regression lines. These estimates are shown in Table II along with the error mean squares for the regression lines. The intensity measurements are not in general linearly related to concentration but in the reasonably narrow range of interest shown in Table I, a linear relationship appears to hold quite well.

TABLE I. Intensity Ratio Measurements and Composition of Mixtures

(Compositions in weight percent)

Batch	R_1	R_2	R_3	R_4	X_1	X_2	X_3	X_4
1	1.1240	0.8980	0.8219	0.9906	0.5514	70.18	12.53	15.04
2	0.9285	0.8872	0.9308	0.9944	0.4426	68.84	14.26	14.75
3	1.1214	0.8030	0.7668	1.1221	0.5631	67.51	12.79	17.39
4	1.1635	0.8706	0.9272	0.9832	0.5624	67.52	14.83	15.34
5	0.9415	0.8064	0.9026	1.1127	0.4505	66.10	14.52	17.03
6	0.9039	0.8314	0.7596	1.0994	0.4425	68.86	12.30	16.72
7	1.0712	0.8404	0.8662	1.0836	0.5290	67.34	13.95	16.35
8	0.9561	0.8731	0.8206	1.0290	0.4702	69.00	13.07	15.68
9	1.0186	0.8431	0.8346	1.0591	0.5001	68.07	13.51	16.02
10	1.0744	0.8124	0.7432	1.0967	0.5379	68.52	12.24	16.64
11	0.9005	0.8320	0.8606	1.0798	0.4321	67.26	13.93	16.34
12	0.9318	0.8913	0.8126	0.9880	0.4498	69.96	12.49	14.99

TABLE II. Regression Coefficients and Error Root Mean Squares

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
s_e=0.00768	s_e=0.00776	s_e=0.01130	s_e=0.01265
$b_{1,0}$=0.15411	$b_{2,0}$=-1.4370	$b_{3,0}$=-1.51670	$b_{4,0}$=0.60788
$b_{1,1}$=1.8573	$b_{2,1}$=0.01832	$b_{3,1}$=-0.07426	$b_{4,1}$=-0.13257
$b_{1,2}$=-0.00074	$b_{2,2}$=0.03020	$b_{3,2}$=0.02008	$b_{4,2}$=-0.00442
$b_{1,3}$=0.00919	$b_{2,3}$=0.02561	$b_{3,3}$=0.08024	$b_{4,3}$=-0.00641
$b_{1,4}$=-0.00832	$b_{2,4}$=-0.00790	$b_{3,4}$=-0.00328	$b_{4,4}$=0.05605

We can use the equations in (1) to develop a set of working expressions for estimating the concentrations, i.e.,

$$\underline{R}=\underline{b}+B\underline{\hat{X}} \tag{2}$$

where $\underline{R}$ represents the vector of intensity ratios and $\underline{b}$ the vector of intercept terms. The b_{ik} element of B is the coefficient of X_k in the ith regression line. $\underline{\hat{X}}$ is the vector of unknown concentrations that one seeks to estimate in practice. Inverting (2), we have:

$$\underline{\hat{X}}=B^{-1}(\underline{R}-\underline{b}) . \tag{3}$$

Here we have a case of the use of a set of simultaneous multiple linear regression lines in reverse, i.e., inverting the regression lines to estimate the X's i.e., the concentrations. Williams [3] gives a discussion of the general problem. It might be noted that the concentrations were used as the independent or concomitant variable since the error in the X's is very small as compared to that for the X-Ray intensity ratios.

Equation (3) represents a working set of equations for estimating the concentration from samples from running production. The four equations given by the matrix expression in (3) are as follows:

$\hat{X}_1=-0.14381+0.54061\ R_1\ +0.07935\ R_2\ -0.08034\ R_3\ +0.08670\ R_4$

$\hat{X}_2=38.2619-0.5767\ R_1\ +42.5690\ R_2\ -13.1116R_3\ +5.1478\ R_4$

$\hat{X}_3=8.9016+0.6984\ R_1\ -10.4829\ R_2\ +15.6926\ R_3\ -0.4547\ R_4$

$\hat{X}_4=-7.1523+1.3131\ R_1+2.3448\ R_2\ +0.5705\ R_3\ +18.4010\ R_4$

The residual errors of estimation, calculated from the original data, are shown in Table III.

TABLE III. Residual Errors of Estimation of Concentration

Batch	$X_1-\hat{X}_1$	$X_2-\hat{X}_2$	$X_3-\hat{X}_3$	$X_4-\hat{X}_4$
1	-0.0035	0.02	-0.19	-0.09
2	0.0026	0.43	-0.14	-0.23
3	0.0013	-0.01	0	0.10
4	-0.0026	-0.04	0.14	0.30
5	-0.0026	0.16	-0.24	0.07
6	-0.0026	0.03	0.06	0.07
7	0.0027	-0.30	0.01	-0.31
8	0.0046	-0.42	0.24	0.13
9	0.0016	0	0.12	-0.11
10	0.0011	0.39	-0.06	-0.13
11	-0.0014	-0.17	0.11	0
12	-0.0012	-0.14	-0.02	0.19

CONFIDENCE INTERVAL ESTIMATES ON THE CONCENTRATIONS. Box and Hunter [1] discuss the problem of joint confidence interval estimates on the solution of a set of simultaneous equations when the coefficients are subject to error. Their work was actually a part of a more specific problem of finding a confidence region for a stationary point in response surface analysis. However, the procedure also applies to our problem of attaching confidence limits to concentrations. Suppose that in general we have K simultaneous equations of the type;

$$\sum_{j=0}^{K} b_{ij} X_j = 0 \qquad (i=1, 2, \ldots K) \tag{4}$$

where the b_{ij} are subject to error (for our case $X_0=1$). Consider the quantities,

$$\sum_{j=0}^{K} b_{ij} \xi_j = \delta_i \ (i=1, 2, \ldots K),$$

where the ξ are the values of the X's that would satisfy (4) if the actual regression coefficients were used in place of the b_{ij}. If we consider a vector of the δ's, say $\underline{\delta}$ as having a multivariate normal distribution with mean vector $\underline{0}$ and variance-covariance matrix $E(\underline{\delta}\,\underline{\delta}')=V$, then the expression $\underline{\delta}'V^{-1}\underline{\delta}$ follows a X^2 distribution [2] with K degrees of freedom. For our case, the ith element of $\underline{\delta}$ can be written as $R_i-\hat{R}_i$, where $\hat{R}_i$ is the estimate of the intensity ratio in the ith regression line. For estimates of the elements of V, we can write

$$\mathrm{Var}\,(R_i-\hat{R}_i)=s_{ii}\,[1+\frac{1}{n}+\sum_h\sum_l C_{hl}\xi_h\xi_l]$$

$$=s_{ii}.H.$$

and $$\hat{\mathrm{Cov}}\,(R_i-\hat{R}_i,\ R_k-\hat{R}_k)=s_{ik}[1+\frac{1}{n}+\sum_h\sum_l C_{hl}\,\xi_h\xi_l]$$

$$=s_{ik}.H.$$

where:

s_{ii}= sample estimate of the variance of R_i for particular values of $\xi_1, \xi_2, \xi_3, \xi_4$.

s_{ik}= sample estimate of the covariance between R_i and R_K.

C_{hl}=(hl) element of the inverse of the matrix of corrected sums of squares and products of the X's for the calibration sample.

If we replace the elements in V by their corresponding estimates and divide by the appropriate degrees of freedom we arrive at the ratio

$$\frac{n-8}{4} \sum_i \sum_k \frac{\delta_i \delta_k w^{ik}}{H}$$

which is distributed as F with 4 and n-8 degrees of freedom, where w^{ik} is the (ik) element of the inverse of the matrix W, the matrix of residual sums of squares and products of the R's. We can write

$$\begin{aligned} \delta_i &= R_i - \hat{R}_i \\ &= \sum_j b_{ij} \hat{X}_j - \sum_j b_{ij} \xi_j, \end{aligned} \tag{5}$$

where the $\hat{X}_j$ are the estimates of the concentration obtained from equation (3). If we replace δ_i by the expression in (5), we have

$$\begin{aligned} F_{4,n-8} &= \left(\frac{n-8}{4}\right) \frac{\sum_i \sum_j \sum_k \sum_l (\hat{X}_j - \xi_j)(\hat{X}_l - \xi_l) b_{ij} b_{kl} w^{ik}}{H} \\ &= \left(\frac{n-8}{4}\right) \frac{\sum_j \sum_l (\hat{X}_j - \xi_j)(\hat{X}_l - \xi_l) q_{jl}}{H} \end{aligned} \tag{6}$$

where q_{jl} is the (jl) element of the matrix;

$$Q = B' W^{-1} B.$$

Here b_{ij} is the (ij) element of the matrix B.

Equation (6) represents simultaneous joint confidence interval estimates of the actual concentrations ξ_1, ξ_2, ξ_3, and ξ_4. Thus if we are given values of the estimates $\hat{X}_1$, $\hat{X}_2$, $\hat{X}_3$, and $\hat{X}_4$, we can substitute particular values of the concentrations ξ_1, ξ_2, ξ_3, and ξ_4 into equation (6) and if the resulting expression is less than $F_{\alpha, 4, n-8}$ (upper tail), then those values of the ξ's fall inside the $100(1-\alpha)\%$ confidence band.

The elements of the W^{-1} and Q matrices are:

$$W^{-1} = \begin{bmatrix} 7214.8162 & 2554.8459 & -3201.5790 & 3439.8046 \\ & 4014.0983 & -1714.2325 & 2122.4663 \\ & & 2679.7650 & -1867.4456 \\ & & & 2825.4942 \end{bmatrix}$$

$$Q = \begin{bmatrix} 24274.424 & -15.343 & -274.4115 & 219.582 \\ & 2.4899 & 2.8058 & 0.77389 \\ & & 10.8662 & -3.6781 \\ & & & 5.3264 \end{bmatrix}$$

III. VARIABLE PARTICLE SIZE. An experiment was conducted in a manner similar to that described in II except that the particle size was allowed to vary. Components 2 and 4 are the only ones for which the particle size is an important factor in its effect on the intensity ratio measurement. The point should be made here that it is assumed that the particle sizes are known in a practical situation, i.e., for a sample of the propellant from running production one can determine, from the physical source of components 2 and 4, at least the mean particle size. The degree of difficulty here would depend upon the precision with which these two components are manufactured. No attempt was made here to consider such problems as particle size distribution. Likewise no attempt was made to consider the degree to which the particle sizes of components 2 and 4 are altered by the mixing process itself.

A 1/8 fraction of a 2^6 factorial design was used with four replications at each point and in the center of the design. The factors are the concentrations X_1, X_2, X_3, X_4, and particle sizes W_2 and W_4. Table IV gives the design matrix and the defining contrasts.

TABLE IV. Design Data and Defining Contrasts

Batch	Treatment Combination	X_1	X_2	X_3	X_4	W_2	W_4
1	abef	1	1	-1	-1	1	1
2	cdef	-1	-1	1	1	1	1
3	(1)	-1	-1	-1	-1	-1	-1
4	ace	1	-1	1	-1	1	-1
5	bde	-1	1	-1	1	1	-1
6	abcd	1	1	1	1	-1	-1
7	adf	1	-1	-1	1	-1	1
8	bcf	-1	1	1	-1	-1	1
9	midpoint	0	0	0	0	0	0

Defining contrasts: I, ADE, BCE, ACF, BDF, ABCD, ABEF, CDEF. (Particle Size Units are per cent fine fraction on total ingredient basis)

A set of multiple regression equations of the type

(8) $$R_{ij} = \sum_{k=0}^{4} \left[B_{ik} X_{kj} \right] + B_{i5} W_{2j} + B_{i6} W_{4j} + \epsilon_{ij} \quad (i=1, 2, 3, 4)$$

were fit to the design data, where as before $X_0=1$. Table V shows the estimates of the coefficients of the regression line in (8). (8) can be written as

$$\underline{R} = B_1 \underline{\hat{X}} + B_2 \underline{W} .$$

We can then "correct" the intensity ratio vector for particle size and solve for the vector $\underline{\hat{X}}$;

(9) $$\underline{\hat{X}} = B_1^{-1}(\underline{R} - B_2 \underline{W})$$

This results in the following set of equations

$$\hat{X}_1 = 11.998725R_1 + 598.526R_2 + 82.076R_3 + 395.848R_4 - 988.676 - 13.5897W_2 - 2.2816W_4.$$

$$\hat{X}_2 = 8.84359R_1 + 3207.192R_2 + 439.287R_3 + 2109.9897R_4 - 5226.75124 - 74.7551W_2 - 11.4927W_4.$$

$$\hat{X}_3 = 1.653744R_1 + 867.2777R_2 + 137.368R_3 + 578.007R_4 - 1437.2059 - 19.37799W_2 - 3.02207W_4$$

$$\hat{X}_4 = 3.0437R_1 + 1258.126R_2 + 175.7089R_3 + 847.6307R_4 - 2073.6127 - 28.46016W_2 - 5.32504W_4.$$

The equation in (8) could also be used to estimate particle size when either the particle size cannot be determined or one feels that the mixing process has caused sufficient "grinding" that there has been a change from the particle sizes of the pure components. Of course this would require a chemical analysis of two of the components of the mixture, which of course, is time consuming.

TABLE V. Estimates of Regression Coefficients and Error Root Means Squares for Equation (8)

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
$s_e = 0.02005$	$s_e = 0.01199$	$s_e = 0.00830$	$s_e = 0.02298$
$b_{10} = -4.8413$	$b_{20} = 2.82710$	$b_{30} = -8.4503$	$b_{40} = -8.19590$
$b_{11} = 1.9320$	$b_{21} = -0.03948$	$b_{31} = 0.11398$	$b_{41} = 0.08438$
$b_{12} = 0.05104$	$b_{22} = -0.01436$	$b_{32} = 0.09337$	$b_{42} = 0.08200$
$b_{13} = 0.06237$	$b_{23} = -0.02355$	$b_{33} = 0.15847$	$b_{43} = 0.08462$

TABLE V
(cont'd.)

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
b_{14}=0.05010	b_{24}=-0.05424	b_{34}=0.07888	b_{44}=0.14812
b_{15}=-0.00582	b_{25}=0.01072	b_{35}=-0.00682	b_{45}=-0.00815
b_{16}=0.00024	b_{26}=-0.00218	b_{36}=-0.00245	b_{46}=0.00417

Table VI shows the residual errors in estimation of the concentration using equation (9).

TABLE VI. Residual Errors in Estimation of the Concentration Using Equation (9) (Units in wt. %)

Batch	Ingredient 1 $\hat{X}_1 - X_1$	Ingredient 2 $\hat{X}_2 - X_2$	Ingredient 3 $(\hat{X}_3 - X_3)$	Ingredient 4 $(\hat{X}_4 - X_4)$
1	-.006332	.2216	-.1006	-.0196
2	-.008813	.3308	-.1491	-.0373
3	-.013621	.5426	-.2567	-.0684
4	.003560	-.0692	.0417	-.0485
5	.008362	-.2675	.1365	-.0293
6	-.001603	.0227	-.0061	.0097
7	.004032	-.0966	.0539	-.0513
8	.007943	-.2583	.1234	-.0280
9	.007812	-.8720	.3427	.3712

SUMMARY. A set of equations is given for estimating the component concentration in a certain solid propellant mixture in terms of the X-Ray intensity readings of each component. The method used involves inverting a set of simultaneous multiple linear regression equations. The concentration of each ingredient appears in each equation in order to correct for "matrix" conditions which do effect the X-Ray intensities. The significance tests on

individual components indicate that these interelement conditions do, in fact, exist for the mixture in question. Joint confidence regions were developed for the concentrations.

Since it was suspected that the particle size of pure components 2 and 4 also effect the X-ray intensity, a linear model involving particle size was fit to the data from a 1/8 fraction of a 2^6 factorial design. This did indicate that particle size was in fact a necessary consideration and resulted in a set of equations for estimating the concentration of each component in terms of an intensity reading which is adjusted for particle size.

ACKNOWLEDGEMENTS. Thanks is given to Mr. Robert Davis who helped with the calculations and Mr. Robert Tankersley of the U. S. Army Missile Command Computation Center, Redstone Arsenal who performed the necessary regression analyses.

BIBLIOGRAPHY

(1) Box, G. E. P. and Hunter, J. S. (1954). A Confidence Region For the Solution of a Set of Simultaneous Equations With an Application to Experimental Design. Biometrika, 41; 190-199.

(2) Graybill, Franklin A. (1961). "An Introduction to Linear Statistical Models." McGraw-Hill Book Company, Inc.

(3) Williams, E. J. (1959) "Regression Analysis". John Wiley and Sons Inc.

SAMPLING FOR DESTRUCTIVE OR EXPENSIVE TESTING

Joseph Mandelson
Quality Evaluation Division, Quality Assurance Directorate
Edgewood Arsenal, Maryland

INTRODUCTION. In recent years the engineer has been impressed with the fact that the principles of sampling are essentially statistical in character because the effect of sampling can only be appraised in terms of operation of the laws of chance. Consistent with this revelation, the engineer by and large has been content to retire from the field of sampling and abdicate his responsibilities in this area to the statisticians. A few hardy souls, confirmed do-it-yourselfers, took it upon themselves to invade the statistical field and learned to acquit themselves creditably in the area of sampling. They even branched out into other aspects of statistics germane to engineering. However, the influx of engineers into the statistical preserve was not sufficiently large to be able to handle the relatively heavy volume of activity required. Then, too, a number of working tools were prepared by statisticians, presumably for use by quality engineers and inspection personnel, to cover a multitude of sampling problems as these occur in quality assurance. Some of these tools are quite complicated; for their complete comprehension they demand more in the way of statistical knowledge on the part of the would-be user than the authors are prepared to admit. As a consequence there is a degree of obscurity in the field. The engineer is urged to consult the statistician whenever his state of confusion or the importance of the matter in hand appears to warrant. However, the engineer should long algo have risen in wrathful protest against statistical tools supposedly prepared for his use but which he finds slippery and elusive to the point of unintellgibility.

Actually, is it so important that comprehension of the mathematical derivation of statistical methods be made an essential prerequisite to their efficient use? Would not an explanation of the basic factors, in non-mathematical terms followed by a detailed by-the-numbers procedure to use in the given context suffice? At any rate, I propose to try this approach.

SAMPLING RISKS. The layman has long regarded the field of sampling with healthy suspicion; he has felt in his bones that sampling is a risky business at best. The well-publicized failures of public opinion polls in predicting the results of crucial election weaken his confidence in statistical methods. His instinct in regard to risks is, of course, entirely correct as his everyday experience with matters governed by the law of chance illustrates. Allusion may be made to games of change, insurance and the like. Much can be learned from pertinent analogies. Let us consider examples from games of chance such as bridge, etc.

A well shuffled card deck is analogous to a lot of material from which a sample is taken, with one important difference: the exact composition of the deck is known, that of the lot is not. Each hand in bridge is a sample of 13 from a lot of 52. A hand of exactly average strength would contain one card of each of the 13 values and a 4-3-3-3 distribution in suits. Our experience tells us that such a hand is almost never observed. Instead we find that some hands are stronger and, by the same token, others are weaker than the average. This should teach us that a sample is very rarely truly indicative of the composition of the lot. Instead, we find that the sample sometimes appears to be better, sometimes worse than the average, if by average we mean a sample whose composition is exactly proportionate to that of the lot. Further, we find that small variations from the average-strength hand are quite frequently encountered, large variations are relatively rare.

In real life, the composition of the lot is almost never known, the purpose of the sample is to permit us to make inferences and decisions regarding the acceptability of the lot sampled. Since we recognize that the sample rarely reveals precisely what the true quality is, it must be accepted that some of the decisions at which we arrive, based on results obtained in testing the sample, may be in error. There are two types of such error.

PRODUCER'S RISK. The Type I error, so called, is the decision to reject a lot which is really acceptable. This occurs when the sample, through chance variation, indicates a larger proportion of defectives than that which is really present in the lot. It is the equivalent of the bridge hand which contains almost no strength. These hands occur occasionally, with predictable frequency. In the same way lots of acceptable quality will produce a sample of given size which, with predictable frequency, will indicate the lot to be unacceptable. It should be noted that, while the frequency of such occurrences can be predicted (say once in twenty samples) the actual event (which one, if any, of the twenty) cannot be foreseen; it occurs at random intervals. In any case, the rejection of an acceptable lot occurs with a certain probability equivalent numerically to this predicted frequency of the Type I error. Since a rejected lot will require 100% inspection of the lot, rework or scrapping of the material, it is plain to see that the risk of this unfortunate occurrence is one which will cause the producer some economic loss. For this reason this is called the Producer's Risk.

CONSUMER'S RISK. On the other side of the coin we have the Type II or beta error which occurs when we decide to accept a lot which is really unacceptable. This occurs when the sample, by chance, yields results which happen to conform to the requirements which decide the acceptability of material offered him. This situation is analagous to the bridge hand which is abnormally strong. The comments already made with respect to the Type I error are also applicable to the Type II error, viz., the frequency of such occurence can, within reason, be predicted if certain information, normally not available, is at hand or can be assumed. The effects of the Type II error are quite different, of course, since the material now becomes the property of the user and the excessively high proportion of defectives it possesses will undoubtedly cause him to sustain certain kinds of loss. The Type II error gives rise to the Consumer's Risk.

EFFECT OF SAMPLE SIZE ON RISK. Both types of error and the associated risks may be reduced by using larger samples. It can be shown that the amount of information concerning the quality of the lot, available from the sample, varies as the square root of the numerical size of the sample. Consequently, if one wishes to double the information in the sample he must multiply his sample size by four. Clearly, this can soon become an expensive business and leads to diminishing returns.

It must ever be kept in mind that the risks we have considered have substantial significance, economic and otherwise. Both risks lead to various types of loss, many (but not all) of which can be measured in monetary terms and all of which must be assumed either by the supplier or the consumer. Whether these costs will weigh more heavily on the former or the latter is determined by the quality of the lot, the sampling plan and the level of quality specified. The risks and, therefore, their cost can be reduced by increasing the sample size but this, in turn, raises the cost of sampling and test which is customarily borne by the consumer. We are reminded that raising the sample size to effect an arithmetic increase in information will necessitate a geometric increase in the costs associated with the sample size.

TOTAL COST OF SAMPLING. If one is realistic he will recognize that the total cost of sampling includes not only the cost of taking and testing the sample but also the losses occasioned by the operation of the risks already discussed. It may appear strange, perhaps unbelievable, that there should be any who will not accept the fact that there are risk losses to evaluate and will not agree to include these in the reckoning. But these doubting Thomases are like their predecessor - unless they see little green bills passing over a counter from one hand into another they cannot agree that a cost or loss has been sustained. It is particularly unfortunate when such short-sighted persons get into a position where they are able to influence the sampling plan to be used. When, in consequence, losses are sustained from defectives regarding which complaints are received from users, and from lots unnecessarily screened or reworked, such people eloquently display newly washed hands as tokens of their freedom from sin and learnedly discuss the poor inspection job turned out by that overly-large and over-paid staff of inspectors. Now, say these management experts, if we really want to save money, here is some fat which can be advantageously trimmed. It will never occur to them that insistance on minimum sample sizes reduces a relatively small cost but incurs much larger risks which require the piper to be paid in large and repeated installments.

The true total cost of sampling is determined by several parameters, chief among which are the sample size, the specified quality level, and the consumer's and producer's risks. There are other parameters involved in the final result such as the cost of making a test, the true quality of the lot, the cost of reworking an item declared defective, etc. For our purpose, it is desirable to search out the interrelationships among the four parameters first mentioned.

Clearly, the larger the sample, the more costly the test. At the same time, the risks and their attendant costs are reduced by large samples. This situation leads naturally to the supposition that there may be some point at which the size and cost of the sample are so happily related to the costs of the corresponding risks that the over-all cost is a minimum. The size of sample which, within the stated conditions, brings about such a desirable result may, with propriety, be designated the optimum sample size. The existence of such an optimal solution can easily be demonstrated arithmetically (2). However, there are some matters which we should clarify before venturing further. These include the meaning of and ways to handle the cost of the risks.

COSTING THE PRODUCER'S RISK. The producer's or alpha risk has already been described as the risk that the sample may indicate the lot to be unacceptable when it is, in fact, quite acceptable. If the test is non-destructive or the cost of making the test is not prohibitively high, it is economically possible to test or examine each item in all rejected lots. In this way the original erroneous decision will be corrected at a price - the cost of such test or inspection is the cost of rejecting the lot and, under these circumstances, the price paid for the Type I error is relatively low. But if the test is quite expensive, particularly if it damages or destroys the item tested, it is not feasible to test each item in the lot. Hence a rejection, whether right or wrong, is practically an order to scrap the lot or rework it. In this case, the cost of the producer's risk is painfully evident especially when one recalls that the producer's risk causes rejection of acceptable lots which, due to a sampling quirk, give the false impression of being rejectable. In any case, the cost of rejecting a lot is easy to calculate and it is given in the following symbolic form: (The meaning of the symbols is provided in the Glossary appended hereto.)

$$C_R = (N - n)(C_U - V_S)(P_P)$$

It should be obvious that C_R, the cost of rejection, can be computed to the last penny; very few approximations are necessary.

COSTING THE CONSUMER'S RISK. It is otherwise with the task of calculating the cost of the consumer's risk in dollars and cents. We will recall that the consumer's risk is the chance he takes that the sample may represent an unacceptable lot as acceptable material. This causes him to pay for and take possession of merchandise which contains an undesirably high proportion of defective material. There the difficulty begins; to assess the cost of accepting a defective lot one must solve the problem of fixing the cost of a single defective item and follow this by discovering the actual percentage of defectives in the lot. If the latter information were at hand, it would have been unnecessary to test the lot for acceptability in the first instance and, had the test revealed the true percent defective in the lot, it would never have been accepted. This difficulty pales to insignificance compared with the problem of determining the cost of an item found to be defective when it is used. This is particularly true of exotic items such as space rockets and military material where failure in use may have strong adverse effect on national prestige and/or security, may cause casualties or even lead to tactical defeat in situations of various degrees of significance. Almost always the loss due to the defective unit depends upon the circumstances surrounding the malfunction. These are unpredictable. Thus, a premature shell burst may cause no casualties or damage in certain situations or it may result in several deaths and a ruined gun. Chance, completely unforeseenable, will determine the loss in each case. Again, how can we compare the cost of a dud hand grenade on the practice field with the loss sustained when a grenade, thrown into an enemy machine gun emplacement, is a dud and the brave soldier who had to expose himself to the gun to make the throw, is cut down? Someone else will have to make that throw and who can tell how many casualties will be sustained to silence the gun which would have been destroyed had the grenade functioned in the first place? The additional casualties are part of the loss associated with the dud. How can anyone predict the course of such events? If one wishes to dramatize this problem he may say that his objective is to put a price on human blood and look into his crystal ball to determine, on the average, how much will be poured out on each defective item.

We must not take the attitude that the cost of the beta risk can never be ascertained. If the item involved is a component and the defect is one that will be caught in attempting to assemble it in the end item then the nuisance loss of this type of defect can be determined. In that case, the method described in Reference (1) can be used for determining sample size while minimizing the total cost of both risks and of sampling.

As we shall see later, the cost of the two risks strongly influence the sample size determined to the optimum in the sense of reducing the total cost to a minimum. If the cost assessed therefore is very high, the optimum sample size calculated to reduce the total cost to a minimum will be unrealistically high as will the minimum total cost computed in these circumstances. In a democracy such as ours, great value is placed on human life. It is commonly regarded as priceless and any attempt to set a monetary value on blood or on life itself is considered a particularly obnoxious form of sacrilege. Yet if such matters are to enter in to the calculation of optimum sample size in a specific case, a monetary value must be set. The engineer seems to be impaled on the horns of an insoluble dilemma.

HOW TO HANDLE THE CONSUMER'S RISK. Yet a solution is possible. The price of blood or life must simply be equated to zero. In other words, it must be eliminated from consideration in monetary terms as suggested in Reference (3). Such a step makes the problem soluble. In this case, the casualty-producing defective can better and more appropriately be handled by prescribing a suitable quality level for acceptance. To adopt this course is equivalent to a decision to eliminate the casualty-producing defective in its role of a sample size determinant and to direct its influence into another path, so that it will act to determine the pertinent quality level instead.

LOT TOLERANCE. One way to handle the problem of determining the optimum sample size for destructive tests, without assessing any cost for the consumer's risk (this is the same as ignoring it or setting it equal to zero) is provided in Reference (2). There the required quality level is set at a figure appropriate to the protection desired as an LTFD (= Lot Tolerance Fraction Defective; see Glossary) which is a level of quality so poor that the engineer would take to his sick bed at the thought of having to accept consistently material of LTFD quality though, once in a long while, to prevent shutting down the line or for some other noble purpose, he might be willing to accept such a lot. By setting the Consumer's risk at some low figure (e.g. 0.10 or 0.05) the engineer insures that only one lot of LTFD quality out of 10 or 20 submitted will be accepted, the others being rejected. Obviously no producer can stand the economic pressure of wholesale rejection, so the quality he must produce to stay in business will have to be a good deal better than the LTFD, which is what our engineer wants. Having decided on a proper LTFD the paper goes on to show how the optimum sampling plan is computed which will yield the desired protection against material of LTFD quality.

Reference (1), on the other hand, is a much more sophisticated approach. However, as has already been noted, it can be applied only where the cost of the beta risk can be computed with reasonable correctness, at least to the extent of knowing in what ball park the doubleheader will be played. Our concern, however, is with the area within which the cost of the beta risk cannot be approximated. It is interesting that the solution herein delineated can be used equally well whether one can or cannot estimate the beta risk cost because in either case the cost can be ignored, if desired, and the acceptance or surveillance quality level may be set at a figure which will keep the outgoing lot percent defective at some desired limit with given probability given some information as to distribution of lot quality. That is, we set the level to take a calculated risk. Then we figure the sampling plan that will insure that outgoing material accepted thereby will conform to that level within the specified risk.

Now we shall consider how this purpose may be accomplished by the engineer without the need to become a statistician, amateur or professional. To do this, we propose to outline the procedure "by the numbers" and ask the engineer to accept as an article of faith that the procedure is, in fact, valid and will do the things and afford the protection attributed to it. It is not our purpose to provide mathematical theory or proofs here and demand that you grasp them before we will permit you to touch the procedure. Rather, we want to present a method which you can grasp in hands grimy from contact with your work and responsibilities and from a knowledge of your problem and needs, proceed to calculate a sampling plan tailored to do what you want it to do.

COMPUTING ACTUAL COSTS. If we consider the case of single sampling (see Glossary) wherein we fix the consumer's risk (i.e. by establishing some desired lot tolerance fraction defective with a 10% chance of acceptance - the consumer's risk), the total cost of the inspection is expressed by the equation

$$T = n(C_U + C_T) + (N - n)(P_P)(C_U - V_S)$$

Since this equation is basic to understanding what we are about to do, it is well to explain it without going to the Glossary. T is the total cost of testing including the Producer's Risk the cost of which is the expression to the right of the central plus sign. To the left of that sign is the cost of testing: n, the sample size, times the sum of the cost of one unit (which the test will destroy) and the cost of testing it. Thus, if the sample size is 35 and we shall destroy an item costing \$3 and spend \$2 to do it, then the test alone will cost 35 x (3 + 2) = \$175. Now, as for the Producer's Risk, the rest of the lot, N - n, is subject to the probability (P_P) that it will be rejected even though the lot is really acceptable. The symbol P_P is the Producer's Risk; it is computed as $1 - L\bar{p}$ by subtracting from unity the chance, $L\bar{p}$, that a lot of process average quality ($\bar{p}$), presumably better than LTFD, will be accepted. If unity represents all possibilities and $L\bar{p}$ is calculated as a decimal fraction, say, 0.95 then $1 - L\bar{p}$ is the chance of rejection; in this case 1 - 0.95 = 0.05. Now $(N - n)(P_P)$ gives the number, on the average, which we will lose from the lot by the action of the Producer's Risk. We may not lose this lot but when we do lose a lot and its N - n is prorated over all the lots we do not lose, each lot will lost about (N - n) (P_p). It remains only to cost this loss. This is done by multiplying (N - n) (P_p) by the cost of one item less its salvage value, if any, $C_U - V_S$. If an item costs $10 and can be reworked for $3, then $C_U - V_S$ = $3 so that ($C_U - V_S$) may also be called the cost of reworking the item. When the appropriate values are filled in, the total cost T of using any proposed sampling plan against material of the quality being produced ($\bar{p}$) may be calculated. A bit laborious but, as you can see, not too difficult.

The calculation, from scratch, of an optimum sample size would require quite a bit of work. First, as indicated in (2), one would have to determine a succession of different sample sizes and an associated allowable number of defects (c) for each. Each plan must be designed to furnish the same protection (same Consumer's Risk) against material of lot tolerance (LTFD) quality. Then, the total cost of each plan would be computed, using the above equation. It would require facility in using a table of probabilities. While this would not be difficult to learn, such a table is, after all, a statistician's reference. Happily, Ellner and Savage (4) have developed short-cut methods for calculating optimum single and double (see Glossary) sampling plans utilizing graphical methods and graphs developed by Dodge and Romig (5). These graphs are reproduced and appended hereto with the kind permission of the originators and publishers and, in any case, can be consulted in (5).

THE WORK OF DODGE AND ROMIG. It is generally acknowledge that Dodge and Romig are the fathers of statistical sampling as used in quality assurance work. It is astonishing to see how sophisticated their thinking was, even in its earliest published form in the Bell Telephone Technical Journal. Their methods are intensely practical but that should not surprise anyone since they were engineers faced with the eminently practical problem of sampling. While their rejected lots would be inspected 100%, they recognized that the cost of such 100% inspection is an economic loss. Their sampling plans were calculated to minimize the over-all cost of the inspection operation including the 100% inspections caused by the Producer's Risk. Therefore the idea of optimizing sample size for minimum cost originated with Dodge and Romig. The use of the same principle for destructive or expensive testing where 100% inspection of rejected lots was patently impracticable was urged by (2) and (4), substituting $C_U - V_S$ for the Dodge and Romig's 100% inspection of rejected lots. With this great similarity in basic ideas, it is not too surprising that we can use Dodge and Romig's graphical methods to avoid a good deal of computational work which might be not only laborious but confusing to the non-statistician. To avoid the latter, we propose to develop single and double sampling plans using the Dodge-Romig graphs and to proceed step by step explaining only as required to facilitate achievement of the final objective - the sampling plan.

CONTROLLING THE PROCESS. In their eagerness to insure receipt of high quality material, engineers can easily fall into the trap of specifying acceptance criteria so high as to increase production and inspection costs beyond reason and hamper production of a smooth flow of acceptable material. For the dubious advantage of an exceedingly low outgoing proportion of defective material, the consumer pays through the nose. There are other ways to do this without incurring prohibitive costs and strangling production. Perhaps the most effective way is to engineer production and establish effective quality controls at the right points on the production line so that production of the most critical or significant types of defects will be almost impossible. Another way, not as effective and more costly, but easier and more convenient for the purchaser is to establish an LTFD at such a level that, to avoid a costly high proportion of rejections, the producer will have to maintain an average quality output well above the LTFD.

ESTABLISHING THE LTFD. In establishing the LTFD we shall assume a Consumer's Risk of 10% or 0.10 for two reasons. First, ever since Dodge and Romig first calculated their tables this has been the risk conventionally accepted for the LTFD. Second, their graphs are based on an 0.10 risk. The engineer should set his LTFD at some fraction defective such that, even if a lot of LTFD quality were accepted on rare occasion, it would cause no insurmountable problem in the field. Since sampling plans developed by our method with reject lots of LTFD quality nine times out of ten, if the contractor would regularly produce material of this quality he would surely face economic disaster. If the supplier's Producer's Risk is to be at a tolerable level he must produce material by a process which is statistically controlled to give a process average ($\bar{p}$) proportion defective very roughly 1/3 or 1/4 of the LTFD. Thus, if the LTFD is 0.08, the supplier should produce a $\hat{p}$ of about 0.02 or 0.03 to avoid excessive loss due to the Producer's Risk. If the supplier's $\hat{p}$ is much lower than the LTFD the optimum sample size will be relatively low.

The engineer should choose an LTFD that will give him what he needs at an acceptable price. From the facts already indicated, he must have a reasonable expectation that the supplier will be able to produce a controlled $\bar{p}$ which is 1/3 LTFD. If he cannot, his prices will have to be raised to cover the excessive rejections he is sure to experience. The engineer must avoid demanding material of prohibitively high quality solely for the purpose of bolstering his reputation for designing items which work all the time. He must remember that, if the supplier is trying to make material at a controlled $\bar{p}$ = 1/3 LTFD, very rarely will the process make a lot of LTFD quality and, even if it does, the chance of its being accepted is only one in ten, so the engineer can rest assured that, for practical purposes, almost all accepted lots will be much better than LTFD quality. With this in mind, he can afford to be fairly generous in setting the LTFD.

Perhaps as good a way as any is to assume some realistic $\bar{p}$ which the engineer feels a qualified supplier can maintain under statistical control when producing the item in question. Then the engineer multiplies $\bar{p}$ by 4 and 3 and asks whether a product of quality $4\bar{p}$ or $3\bar{p}$ can, on rare occasion, be accepted without causing excessive trouble to the user. Using this as a criterion he sets his LTFD at $4\bar{p}$ if possible, at $3\bar{p}$ otherwise. The engineer should realize that, if the supplier maintains control over his quality a lot of LTFD quality will almost never be produced, much less accepted. The supplier should recognize that if a sampling plan is computed on an LTFD basis he would be well advised to get his process under statistical control at a $\bar{p}$ no greater than 1/3 LTFD and keep it there. If, for some reason, the LTFD must be set at some figure noticeably less than $3\bar{p}$, the engineer should expect higher prices, uncertain deliveries or repeated requests for waivers or changes in contract requirements. The supplier can anticipate occasional, even frequent rejections and organize with this possibility in mind. The above procedure is only a useful rule-of-thumb. By making a number of trial calculations, the engineer can satisfy himself that when $\bar{p}$ is very small compared with LTFD, the sample size required will be relatively small and rejections will be few. As $\bar{p}$ approaches the LTFD, sample size will be at a high and rejection will tend to occur in 9 cases out of 10.

<u>DESIGNING THE OPTIMUM SINGLE PLAN (EXAMPLE 1)</u>. To illustrate how to design a single sampling plan, we shall use the example furnished in (4). First we shall list by symbols the things we need to know

quantitatively. If any of this information is lacking, it is advisable to use your best guess and make any correction which later information indicates to be suitable.

$N = 5000$ $\qquad C_U = \$5$

$LTFD = p_t = 0.07$ $\qquad C_T = \$10$

$\bar{p} = 0.02$ $\qquad V_S = \$3$

We calculate the qualtities $A = C_U + C_T = \$15$ and $B = C_U - V_S = \$2$. Usually A and B can be determined quite accurately but they are not as important as the ratio $\frac{B}{A}$. Using these figures, we calculate the following:

$$p_t N = 0.07 \times 5000 = 350$$

$$\frac{B}{A}N = \text{approximate equivalent lot size} = \frac{2}{15} \times 5000 = 667$$

$$\frac{\bar{p}}{p_t} = \frac{0.02}{0.07} = 0.29, \text{ and}$$

$$p_t \frac{B}{A}N = (p_t)\,(\text{approximate equivalent lot size}) = 0.02 \times 667 = 46.7.$$

We enter Figure 2 with $p_t \frac{B}{A}N = 46.7$ and $\frac{\bar{p}}{p_t} = 0.29$ and get an acceptance number $c = 4$. Now going to Figure 3, we follow the curve for an acceptance number of 4 and we find it leaves the chart at p_tN of 200. Our p_tN is 350 but since the curves for $c = 0$ to $c = 10$ remain parallel to the horizontal axis past $p_tN = 200$, we read (p_t) (sample size) or $p_tn = 8$. Since $p_t = 0.07$ we find $n = \frac{8}{0.07} = 114$. We substitute 114 in the expression for the exact equivalent lot size, $p_t[\frac{B}{A}N + (1 - \frac{B}{A})n]$ which converts to $0.07[\frac{2}{15} \times 5000 + (1 - \frac{2}{15})114] = 53.6$. We could not calculate the exact equivalent lot size before this because we need to know n, the sample size. That we obtained by first using the approximate equivalent lot size. We re-enter Figure 2 with the new estimate of n_t (equivalent lot size) = 53.6 and $\frac{\bar{p}}{p_t} = 0.29$ and get c = 5. Now we re-enter Figure 3 with c = 5 and $p_t'N = 350$ and read 9.2 (by using dividers and a scale). Since $p_t = 0.07$, 0.07n = 9.2 whence n = 131. The optimum single sampling plan, then, is n = 131, c = 5. We can check this by recalculating the long expression above and getting 54.5 which when used to enter Figure 2 again with $\frac{\bar{p}}{p_t} = 0.29$, finds c = 5 unchanged. That is all there is to it.

INFLUENCE OF THE PROCESS AVERAGE, $\bar{p}$. To insure that the optimum in sampling economy is maintained, the process average should be recomputed every 5 or 10 lots. If any sizeable change is noted, it would be wise to recompute the sampling plan, which is not an onerous task as you have seen. The question may be put as to what value to use for $\bar{p}$ when calculating the original sampling plan, when no quality history exists for the production line. At such a time, your best guess as to the average quality the line is expected to produce is adequate or you may prefer to estimate $\bar{p}$ conservatively at about $0.3p_t$. It probably will not make too much difference either way since, even if the estimate is off somewhat, it will not be too far away and will be changed as soon as a quality histroy becomes available. As an exercise, one might vary the process average, using some figures much higher and much lower than $\bar{p} = 0.02$ and notice the effect on the sample size which results from the change.

DOUBLE SAMPLING. Some time ago, double sampling and the related multiple sampling were regarded as ways to reduce the over-all cost of sampling since, for sampling plans giving the same protection the total number of sample items needed for single sampling was normally noticeably more than what double sampling demanded which, in turn, was greater than what multiple sampling required. Thus, if the amount of retesting could be kept down, as when quality is either very good or very poor, appreciable savings appear possible. Since the system for calculating optimum single sample plans takes into account changes in sample sizes when $\bar{p}$ changes, it possesses some of the advantages of double and multiple sampling without the disadvantages. Again, many like the idea of getting a second chance with double sampling, several chances with multiple sampling. One does not feel so tied down to the one chance of the single sample. This is, of course, purely psychological for, mathematically, there is a price to pay. Additional costs must be borne in selecting second and other samples that are used only infrequently. There is the physical burden and inconvenience of handling more sample items and of returning unused samples to the parent lots. Then, too, when retests become more frequent than originally anticipated, heavy work loads are experienced leading to over-work, fatigue and, eventually, to error. These factors have caused double and multiple sampling to lose some of their popularity and led to greater dependence upon and use of single sampling plans. Nevertheless, we shall include a method for computing optimum double sampling plans.

DESIGNING THE OPTIMUM DOUBLE SAMPLING PLAN (EXAMPLE 2). For this example we shall use the figures used in Example 1. To spare you the trouble of looking them up they are listed below:

$N = 5000$	$C_U = \$15$
$LTFD = p_t = 0.07$	$C_T = \$10$
$\bar{p} = 0.02$	$V_S = \$3$

Again we calculate $A = C_U + C_T = \$15$ and $B = C_U - V_S = \$2$. Using these figures we calculate

$$p_t N = 0.07 \times 5000 = 350$$

$$\frac{B_N}{A} = \text{approximate equivalent lot size} = \frac{2}{15} \times 5000 = 667$$

$$\frac{\bar{p}}{p_t} = \frac{.02}{.07} = 0.286 \text{ and}$$

$$p_t\, B_N/A = (p_t)\,(\text{approximate equivalent lot size}) = 0.02 \times 667 = 46.7.$$

To determine the respective c numbers for our double sampling plan we use Fig 2-7 which is analagous to Fig 1-2. We enter Fig 2-7 with $p_t\, B_N/A = 46.7$ for the ordinate or vertical component and $\bar{p}/p_t = 0.286$ for the horizontal component or abscissa. We find $c_1 = 1$ and $c_2 = 7$, almost inside $c_2 = 8$.

Now we use Fig 2-8 and, at $p_tN = 350$, the curve for $c_1 = 1$ gives a reading of 4.5 on the ordinate which represents p_tn or p_t times the first sample size. Since $p_tn_1 = 4.5$ and $p_t = 0.07$, $n_1 = \frac{4.5}{0.07} = 64$. Similarly we look up c_2 for $p_tN = 350$ and we find an ordinate of 12.8 which now represents $p_t(n_1 + n_2)$. Now if $p_t(n_1 + n_2) = 12.8$ and $p_t = 0.07$ then $n_1 + n_2 = \frac{12.8}{0.07} = 183$. Since $n_1 = 64$, $n_2 = 183 - 64 = 119$. As before, this is a first approximation to the sampling plan we want. Substituting in the expression

$$\frac{BN}{A} + (1 - \frac{B}{A})(n_1 + n_2)$$

we get $667 + (1 - 2/15)(183) = 826$. Back we go to Fig 2-7, using $p_t(826) = 0.07 \times 826 = 57.8$ and we get $c_1 = 1$ and $c_2 = 8$. Again we enter Fig 2-8 with $p_tN = 350$ as the abscissa and for $c_1 = 1$ we get $p_tn_1 = 4.5$ so that $n_1 = 64$ as before. However for $c_2 = 8$, we get $p_t(n_1 + n_2) = 14.0$ whence $n_1 + n_2 = \frac{14.0}{0.07} = 200$, from which $n_2 = 200 - 64 = 136$. The sampling plan then is $c_1 = 1$, $c_2 = 8$, $n_1 = 64$, $n_2 = 136$. If desired, the sample sizes can be rounded to $n_1 = 65$, $n_2 = 135$ without too great a change in the effect of the plan. As you can see, the calculations are a bit more involved for the double sampling plan as compared with the single sampling plan but the principle is the same.

The desire to keep the presentation simple requires omission of several facets which might be useful such as an easy way to calculate the expected total cost of a given sampling plan if β is known. However, if this information is required it can be obtained from other graphs in (5).

In all the previous discussion, it was assumed that the only information available regarding the quality of the lot to be tested was that developed from the sample. In an actual production situation a substantial amount of engineering information is developed during the production cycle which, properly interpreted, can indicate whether the process is in statistical control and, therefore, may be considered to be producing substantially homogeneous material. If the material is homogeneous from lot to lot then the results of tests generated in previous lots may be considered to have significant bearing on the results expected in the latest lot. Hence when statistical control has been established, the sample size, lot by lot, can be reduced substantially and remain reduced provided no evidence is obtained indicating loss of control.

Basically, if advantage is taken of available engineering knowledge of previous experience with the process sampling, testing, and their attendant costs may be reduced. This notion lends itself readily to statistical ingenuity but the engineer will require the assistance of a statistician to take advantage of the possibilities. A number of ingenious schemes to permit useful employment of existing engineering data can be devised to reduce the sample size and test costs below the "optimum" solution just described.

The author desires to express his appreciation and gratitude to Ellner and Savage for permission to use the results of their research and most particularly to Professor Harold F. Dodge, Dr. Harry G. Romig, and John Wiley and Sons, Inc. for their unselfish generosity in allowing reprinting of their graphs without which this work would have been impossible.

REFERENCES

(1) Barnard E. Smith, "Some Economic Aspects of Quality Control", Applied Mathematics and Statistics Laboratories, Stanford University, Technical Report No. 53, 3 July 1961.

(2) Joseph Mandelson "Estimation of Optimum Sample Size in Destructive Testing by Attributes", Industrial Quality Control, November 1946.

(3) E. G. D. Paterson "Quality Control Engineering in Product Evaluation", Industrial Quality Control, May 1960 wherein the author indicates that cost cannot intelligently be assigned to the beta risk and that this factor can best be governed by "...the employment of acceptance criteria and procedures which will, to the extend practicable, obviate their presence in the accepted product." Paterson was vice-president of Bell Laboratories in charge of quality control.

(4) H. Ellner and I. R. Savage "Sampling for Destructive or Expensive Testing by Attributes" presented at the Second Engineering Statistics Symposium at Army Chemical Center, Md., in April 1956 and at the Army Science Conference, West Point, N. Y., in June 1957.

(5) H. F. Dodge and H. G. Romig, Sampling Inspection Tables, 2nd Ed., John Wiley and Sons, New York, 1959.

(6) Joseph Mandelson "Lotting", Industrial Quality Control, May 1962.

GLOSSARY

C_R = Cost of rejection

N = Lot size

n = Sample size

C_U = Cost of a single unit

V_S = Salvage value of a single unit or its value as rework material

P_P = Producer's risk: probability (expressed as a decimal fraction) that the sample will, on test, represent the lot to be unacceptable when it is, in fact, quite acceptable

C_S = Cost of sample item

C_T = Cost of testing a single unit

$A = C_U + C_T$ = The cost of destroying one item in testing

$B = C_U - V_S$ = The value of one rejected item

c = Acceptance number, the maximum number of defectives that will be permitted in a sample of size n from an acceptable lot. If more than c defectives are observed in the sample of n items the lot will be rejected.

DOUBLE SAMPLING SYMBOLS

n_1 = Size of first sample

n_2 = Size of second sample

$n_1 + n_2$ = Size of combined first and second samples

c_1 = Acceptance number for first sample, n_1. If c_1 or fewer defectives are found in n_1, the lot is accepted straight-away. If the number of defectives found in n_1 is greater than c_1 but equal to or less than c_2, the second sample, n_2, is tested and the number of defectives in n_1 and in n_2 is totalled. If that number is greater than c_2 (the number of defectives permitted in $n_1 + n_2$) the lot is rejected. If c_2 or less defectives are found in $n_1 + n_2$ on retest, the lot is accepted.

DEFINITIONS

Single Sampling - A system of sampling whereby a single sample is drawn from a lot and the acceptability of the lot is determined from the results obtained in testing the sample. No retest is permitted if results are unfavorable.

Process average ($\bar{p}$) - The apparent proportion of percent of defectives manufactured by the production process. It is generally computed by dividing the total number of defectives found in the samples taken from the last few lots tested (5 or 10) by the sum of the sample sizes. This gives $\bar{p}$ as a decimal fraction.

Lot Tolerance Fraction Defective (LTFD or p_t) - Lot quality, expressed as a decimal fraction defective, so poor that we want to permit only a small chance or probability (the Consumer's Risk, say one chance in 10 = 10% = 0.10 probability) that the sampling plan will permit acceptance if such a lot is submitted.

Double Sampling - A system of sampling wherein two samples are taken and one set of acceptance and rejection criteria are furnished for each sample. If the results obtained in testing the first sample meet neither the acceptance nor the rejection criterion for that sample, the second sample is tested (called the retest) and the decision is made using the second set of criteria. A decision is always possible using the second set of criteria after the retest.

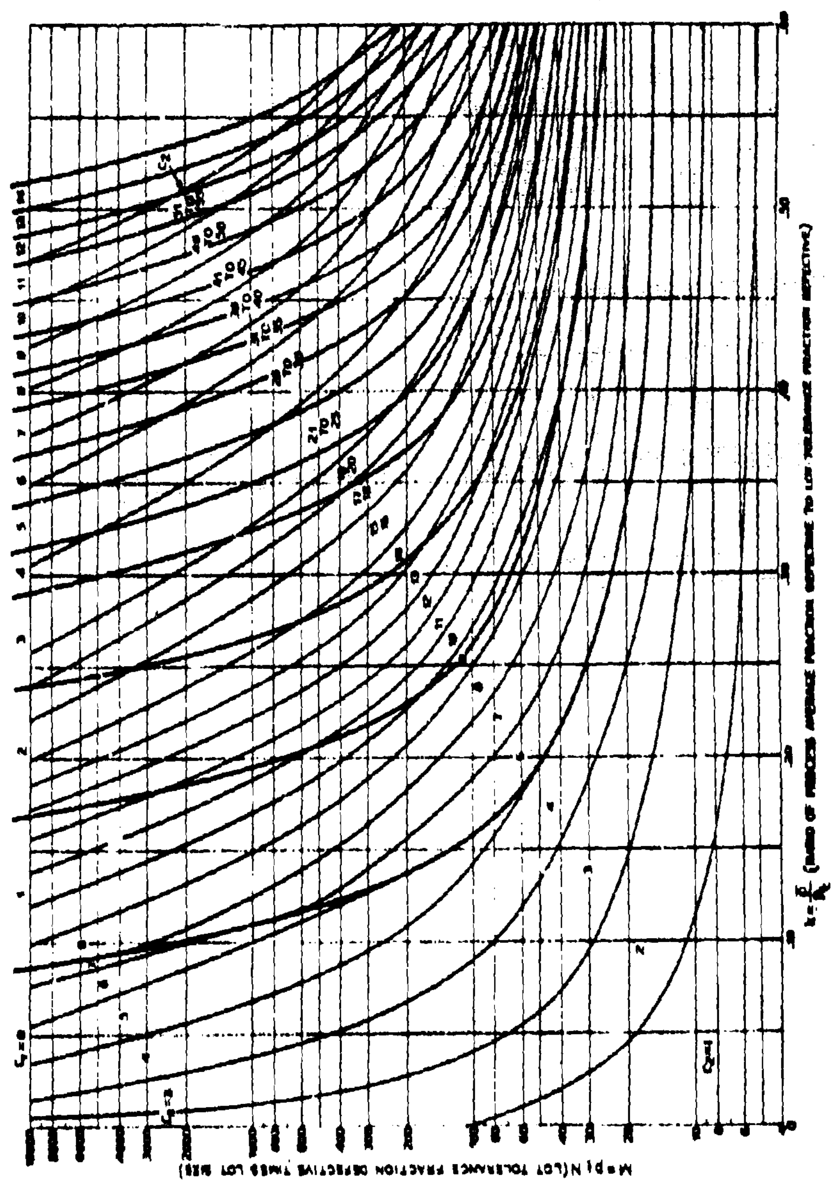


Fig. 2-7 Chart for determining acceptance numbers c_1 and c_2 for lot tolerance protection, Consumer's Risk 0.10

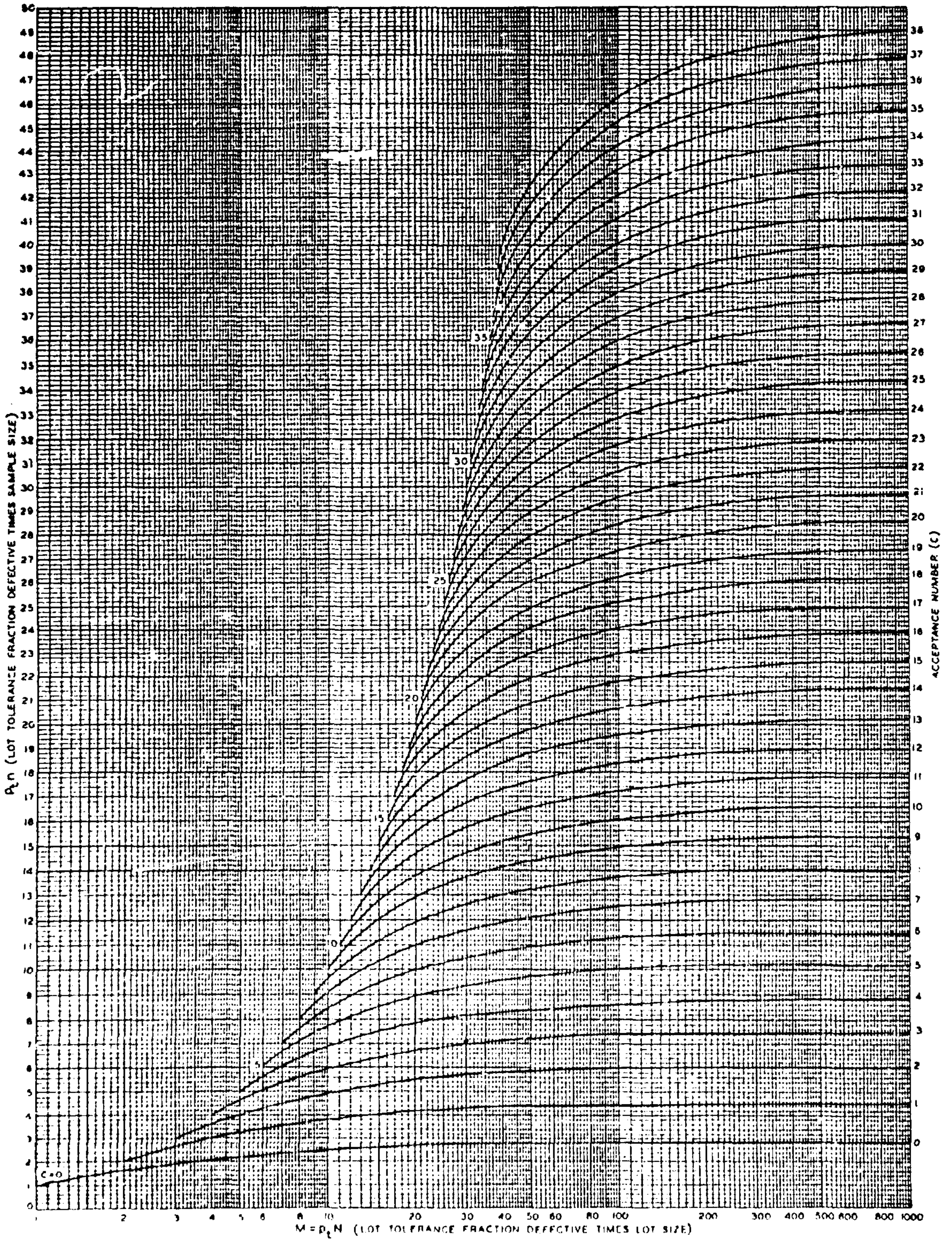


Fig. 2.8 Chart for determining sample sizes n_1 and n_2; lot tolerance protection, Consumer's Risk 0.10

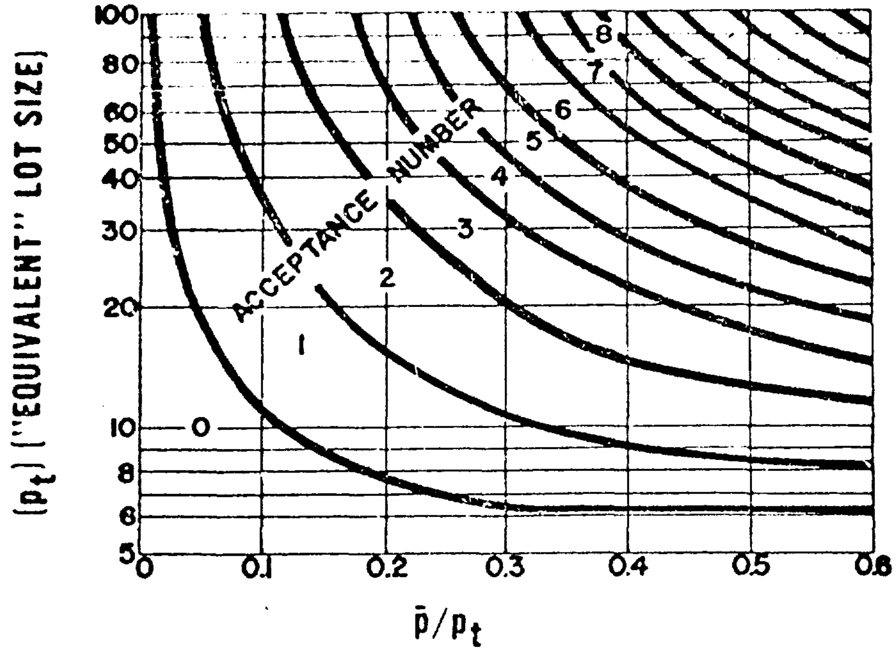


FIG. 2.* CHART FOR FINDING ACCEPTANCE NUMBER OF SINGLE SAMPLING PLAN. (CONSUMER'S RISK, 0.10).

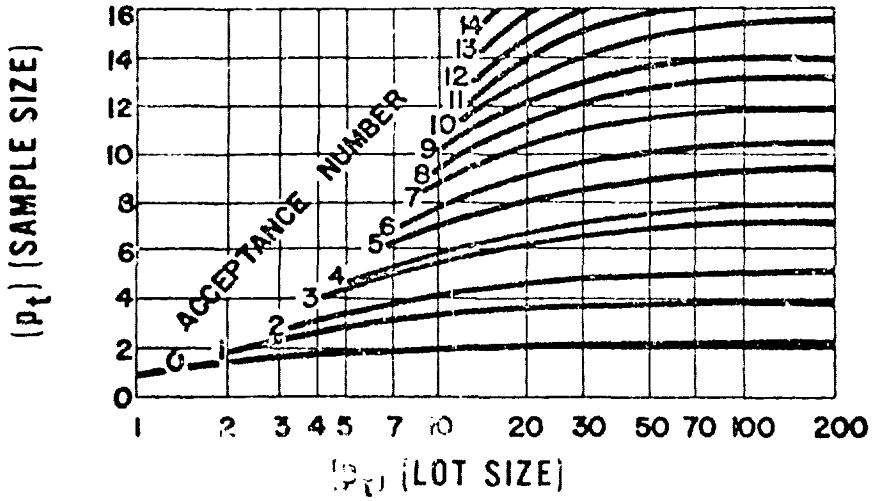


FIG. 3.* CURVES FOR FINDING SIZE OF SINGLE SAMPLING PLAN. (CONSUMER'S RISK, 0.10).

*Reproduced in part, by permission, from "Sampling Inspection Tables" by Dodge & Romig, published by John Wiley & Sons, Inc.

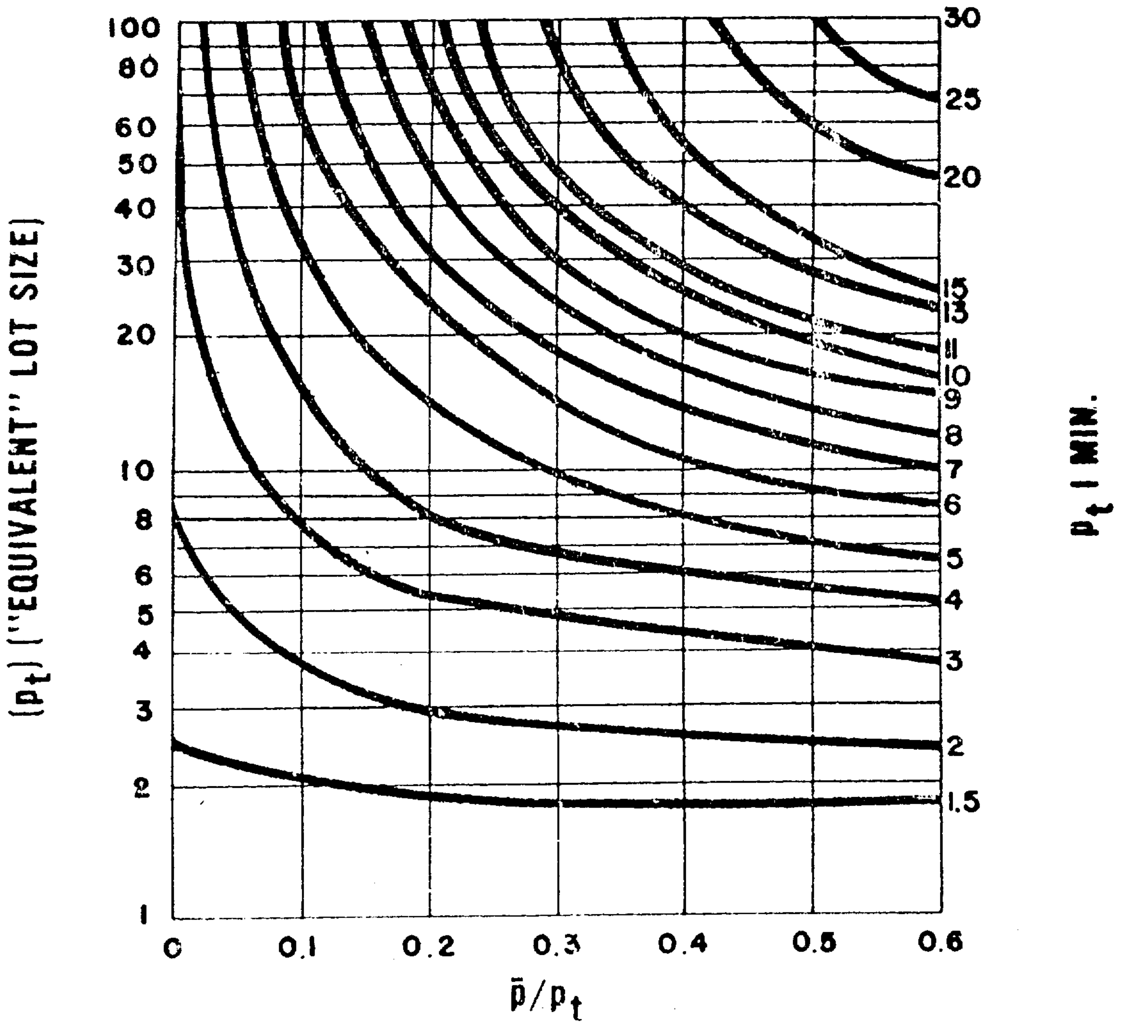


FIG. 4.* CURVES FOR FINDING THE MINIMUM COST OF INSPECTION PER LOT (SINGLE SAMPLING PLAN - CONSUMER'S RISK, 0.10)

*Reproduced in part, by permission, from "Sampling Inspection Tables" by Dodge & Romig, published by John Wiley & Sons, Inc.

PROCEDURES FOR FINDING TOTAL SAMPLE STATISTICS FROM SUBSAMPLE STATISTICS

Paul C. Cox
Reliability and Statistics Division
Army Missile Test and Evaluation Directorate
White Sands Missile Range, New Mexico

ABSTRACT. While procedures for obtaining the variance for a total sample from subsample statistics is fairly well known, there appear to be very few instances in which such procedures are found in print. Therefore, twenty-five formulas are presented which are in one way or another, related to obtaining the mean and variance for a total sample from subsample statistics. In addition, techniques are demonstrated for using these formulas to determine the mean and variance for a sample in which a portion of the observations have been modified, some have been added, or a few have been deleted.

The discussion includes: applications of these formulas; precautions which should be observed; methods for deriving the formulas; and, procedures for their use.

I. INTRODUCTION. This report presents techniques and formulas for determining the mean and variance of a total sample if this sample has been partitioned into a set of non overlapping and mutually exhaustive subsamples; and the mean, variance, and sample size are known for each subsample.

Similarly, techniques are discussed for changing the variance when observations are added to, deleted from, or changed in a sample. Procedures for deriving these formulas are discussed and some of the derivations are included in this report.

Most people know these formulas exist, and they are not, for the most part, difficult to derive. However, they are often useful and it is usually difficult to find them in print. To illustrate this point, a total of eighty-six statistics, design of experiments, probability, sampling, and quality control texts were reviewed and of that number, only two* included a discussion

*(1) Sampling Inspection by Variables, Bowker and Goode, pp. 62, 63, and 92.
(2) Techniques of Statistical Analysis, Eisenhart, Hastay, and Wallis, pp. 42-43.

on how to determine the total variance from subsample statistics. From this, it appears that while the formulas and procedures which are presented here may be well known, few authors seem to have bothered to put them in print. Forthermore, it has been observed that many people have needed certain of these formulas and not being able to locate them in print have found it necessary either to spend considerable time deriving them or simply to do without.

One obvious method for obtaining the mean and variance for a total sample is to gather the raw data from all the subsamples and compute these statistics by conventional procedures. It is equally clear that use of raw data will be unsatisfactory if the subsamples are quite large bacause of the amount of work involved; and the raw data certainly cannot be used in those frequent cases in which it is no longer available.

II. <u>APPLICATIONS AND PRECAUTIONS</u>. The following are uses of the procedures and formulas of this section:

A. After estimating the mean and variance for a number of different populations, a research worker may want to know the mean and variance for a population composed of a combination of these populations. This would be accomplished by combining the samples from the sub-populations to obtain a total sample.

(1) An example of this would be the case of production lots. The mean and variance will be known for a sample from each lot, but an estimate of the mean and variance for the entire production may be desired. To obtain this it would be necessary to combine the lot samples to obtain a total sample.

(2) A second example: After conducting an analysis of variance to determine the effect of certain treatments, the research worker may want to estimate the mean and variance for a population composed of several subpopulations, each identified by a certain treatment level. For this, subsamples could be combined to form a total population.

B. Sample data may come from many sources, for example, from several parts of the country, from several agencies, or from several periods of time, and it may frequently be desirable to combine the data to form one total sample. Obviously, it may be that only the mean, variance and sample size for each subsample are available or can easily be transmitted rather than the complete raw data.

C. Frequently, sample data has been completely analyzed when it becomes evident that a few observations must be added, certain observations should be deleted, or a few should be corrected. The procedures of this report may be very useful in changing or correcting the original estimates of the mean and variance as a result of changing or correcting the basic data.

In this connection, these formulas may be useful in computing statistics associated with moving averages.

D. As a final application, those who teach statistics at the Sophomore or Junior level might find the derivation and application of some of these formulas an interesting assignment.

The main precaution to observe when using these formulas is that the total sample may represent a population with such strange or unknown characteristics that an estimate of the variance would be useless when obtained. For example, a total population composed of k normal sub-populations, each with a different mean and variance, is not likely to be normal or even close to normal.

On the other hand, it is quite possible that the characteristics of the total population will be known and the estimates of its parameters useable. For example, the sub-populations may not be normal, but it may be possible to combine them to form a normal total population. Similarly, the variance for the total population may be needed to describe the distribution of sample means, and this distribution should approach normality regardless of the distribution of the total population.

Another precaution is that one should observe whether the ratio of each subsample size to the total sample size is about the same as the ratio of the corresponding sub-population. If this is not the case, weighting factors should be introduced to obtain the correct ratios.

As a final precaution, before combining subsamples to form a total sample, one should always observe whether it is inherently reasonable to combine such data. That is to say, the subsamples may contain such different types of observations that combining them would be nonsense.

The actual differences between a total estimate and a pooled estimate of the variance should be discussed at this point.

A total variance is the variance of one complete sample, which has been broken down into two or more subsamples. No assumptions are made concerning the populations corresponding to each subsample. More specifically, no assumption is made concerning the variances of these populations. However, it is assumed that when the total variance has been obtained, its corresponding total sample corresponds to a population with known characteristics. If this were not so, there would be little purpose in a total variance.

The pooled estimate of the variance can be obtained from subsample statistics, just as a total variance. It differs, however, in that it is in no way related to a total sample or a total population. Therefore, no assumptions need be made concerning a total population. The assumption is made, however, that all subsamples come from the same population, or at least from populations which have equal variances. The pooled estimate is then an improvement over each of the estimates obtained from any single subsample.

III. DEFINITIONS.

A. k = Number of subsamples.

B. n_i = Size of the i^{th} subsample ($i = 1, 2, \ldots k$).
(If all n_i are equal, use n)

C. N = Size of the complete sample.

(1) $N = \Sigma n_i$ ($i = 1, 2, \ldots k$).

(2) $N = kn$ if all n_i are equal.

D. $\bar{x}_i$ = Mean for the i^{th} subsample.

E. $\bar{\bar{x}}$ = Mean for the total sample.

F. s^2 = Variance for the complete sample.
s = Standard deviation for the total sample.

G. s_i^2 = Variance for the i^{th} subsample.

H. s_p^2 = Pooled estimate of the variance.

IV. PROCEDURES.

A. The overall mean $\bar{\bar{x}}$:

(1) If the n_i are unequal:

$$\bar{\bar{x}} = \frac{\Sigma n_i \cdot \bar{x}_i}{N} . \qquad \text{Formula (I)}$$

(2) If the n_i are equal:

$$\bar{\bar{x}} = \frac{n\Sigma \bar{x}_i}{N} = \frac{\Sigma \bar{x}_i}{k} . \qquad \text{(II)}$$

B. Pooled Estimate of the Variance s_p^2.

The pooled estimate of the variance is actually an average of the subsample variances, and should be computed only if there is reasona e assurance that all subsamples were selected from populations with equal variances.

(1) If the n_i are unequal:

$$s_p^2 = \Sigma (n_i - 1) \cdot \frac{s_i^2}{N-k} \qquad \text{(III)}$$

(2) If the n_i are all equal:

$$s_p^2 = \frac{(n-1)\Sigma s_i^2}{N-k} = \frac{\Sigma s_i^2}{k} . \qquad \text{(IIIa)}$$

C. Determining the Variance from an Analysis of Variance Table.

One method for computing both the total variance and the pooled estimate of variance is by preparing a single variable analysis of variance table. In addition to determining the variances, it will also be possible to test the null hypothesis of the equality of subsample means. This is described by Table 1.

TABLE 1 - The Analysis of Variance Method

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	k-1	TR	tr	F
Error	N-k	E	s_p^2	
Total	N-1	T	s^2	

Table 1 is completed as follows:

(1) Complete all entries under degrees of freedom.

(2) Compute and enter:

$E = \Sigma(n_i - 1)s_i^2$, or if all n_i are equal $E = (n-1)\Sigma s_i^2$.

(3) $TR = \Sigma n_i \cdot \bar{x}_i^2 - N\bar{\bar{x}}^2$, or if all n_i are equal $TR = n\Sigma \bar{x}_i^2 - N\bar{\bar{x}}^2$

$= n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)$.

(4) $T = E + TR$.

(5) $s^2 = \frac{T}{N-1}$ <u>This is the desired solution.</u>

(6) The pooled estimate $s_p^2 = \frac{E}{N-k}$.

(7) If it is desired to test the null hypothesis for equality of means:

$tr = \frac{TR}{k-1}$, and

$F = \frac{tr}{s_p^2}$ with (k-1) and (N-k) degrees of freedom.

D. Formulas for the Variance of a Total Sample.

It is simple to obtain the desired formulas for the variance (and standard deviation) for the total sample by following the procedures of the analysis of variance given in the previous section. These formulas are given below:

(1) The general formula:

$$s^2 = \frac{\Sigma(n_i-1)s_i^2 + \Sigma n_i \cdot \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} . \qquad \text{(IV)}$$

(2) If all n_i are equal:

$$s^2 = \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} . \qquad \text{(V)}$$

(3) If a pooled estimate of the variance is available and the n_i are unequal, formula IV may be written thus:

$$s^2 = \frac{(N-k)s_p^2 + \Sigma n_i \cdot \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} . \qquad \text{(VI)}$$

(4) If a pooled estimate of the variance is available and the n_i are all equal, formula V may be written thus:

$$s^2 = \frac{(N-k)s_p^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} . \qquad \text{(VII)}$$

(5) If k = 2 and $n_1 = n_2$, formula V may be further simplified:

$$s^2 = \frac{N-2}{2(N-1)} \cdot (s_1^2 + s_2^2) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2. \qquad \text{(VIII)}$$

(6) If all n_i are equal and n is large, the following approximation may be used for formula V:

$$s^2 \cong s_A{}^2 = \frac{\Sigma s_i{}^2 + \Sigma \bar{x}_i{}^2}{k} - \bar{\bar{x}}^2 = s_p{}^2 + \frac{\Sigma \bar{x}_i{}^2}{k} - \bar{\bar{x}}^2. \qquad \text{(IX)}$$

In Appendix II, it is shown that the error in formula IX is as follows:

$$\text{Error} = s_A{}^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i{}^2 - s^2) . \qquad \text{(X)}$$

The error described in formula X is always positive.

(7) Formula XI is offered as a substitute for formula IV and formula XII as a substitute for formula V. Actually, formulas XI and XII may require more labor than the original formulas, but they will usually involve smaller numbers and may frequently result in greater accuracy.

$$s^2 = \frac{\Sigma(n_i-1)s_i{}^2 + \Sigma n_i(\bar{x}_i - \bar{\bar{x}})^2}{N-1} \qquad \text{(XI)}$$

$$s^2 = \frac{(n-1)\Sigma s_i{}^2 + n\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N-1} . \qquad \text{(XII)}$$

E. Formulas Associated with Changes in Data.

(1) Frequently, after computing the desired statistics for a sample of size n_1, the worker is faced with the necessity of adding an extra group of n_2 observations to the sample. If n_1 is large and n_2 relatively small, it would appear to be desirable to compute the mean and variance for the n_2 additional observations and determine the statistics for the entire sample from formulas I and IV. This technique is illustrated in Appendix I, Section D.

(2) In the event only one new observation (y) has been added to the sample, formulas XIII and XIV offer a simple procedure for obtaining the desired mean and variance. Similarly, formulas XV and XVI may be used if two observations, (y) and (w) are to be added.

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + y}{n_1 + 1} \qquad \text{(XIII)}$$

$$s^2 = \frac{n_1 - 1}{n_1} \cdot s_1^{\ 2} + \frac{(\bar{x}_1 - y)^2}{n_1 + 1} \qquad \text{(XIV)}$$

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + y + w}{n_1 + 2} \qquad \text{(XV)}$$

$$s^2 = \frac{(n_1 - 1)s_1^{\ 2} + y^2 + w^2 + n_1 \cdot \bar{x}_1^{\ 2} - (n_1 + 2)\bar{\bar{x}}^2}{n_1 + 1} \qquad \text{(XVI)}$$

(3) Similarly, after computing the mean and variance for a sample of size n_1, it may be necessary to discard n_2 observations. If the mean and variance are computed for the n_2 observations which have been discarded, formulas XVII and XVIII may be used to obtain the mean and variance for the remaining observations.

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 - n_2 \bar{x}_2}{n_1 - n_2} \qquad \text{(XVII)}$$

$$s^2 = \frac{(n_1 - 1)s_1^{\ 2} - (n_2 - 1)s_2^{\ 2} - (n_1 - n_2)\bar{\bar{x}}^2 - n_2 \bar{x}_2 + n_1 \bar{x}_1^{\ 2}}{(n_1 - n_2 - 1)} \qquad \text{(XVIII)}$$

This is illustrated in section E of Appendix I.

(4) Discarding One Term

If there is only one term (y) to be discarded, formulas XIX and XX may be used.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y}{n_1 - 1} \qquad \text{(XIX)}$$

$$s^2 = \frac{n_1 - 1}{n_1 - 2} \cdot s_1^{\,2} - \frac{n_1 \cdot (\bar{x}_1 - y)^2}{(n_1 - 1)(n_1 - 2)} \;. \qquad \text{(XX)}$$

(5) Replacing Observations

If a group of n_2 observations in a sample of size n_1 should by changed, one may follow the steps discussed in sections (1) and (3). If it is only one observation, formulas XXI, XXII and XXIII may be used. Assume y is the value to be removed and replaced by w.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y + w}{n_1} \qquad \text{(XXI)}$$

$$s^2 = s_1^{\,2} + \frac{w^2 - y^2 - n_1(\bar{\bar{x}}^2 - \bar{x}_1^{\,2})}{n_1 - 1} \qquad \text{(XXII)}$$

or:

$$s^2 = s_1^{\,2} + \frac{(w-y)\cdot[(n_1-1)w + (n_1+1)y - 2n_1 \cdot \bar{x}_1]}{(n_1)(n_1-1)} \;. \qquad \text{(XXIII)}$$

F. Variance and Mean for a Total Population Composed of k Normal Populations

It appears appropriate to conclude with a brief discussion of population parameters. Assume a total population is composed of k normal sub populations, with mean μ_i and variance σ_i; and each contributing to the total population in the proportion f_i, with $\Sigma f_i = 1$. Formula XXIV gives the mean (μ) for the total population and formula XXV for the variance (σ^2) of the total population.

$$\mu = f_1\mu_1 + f_2\mu_2 + \ldots + f_k\mu_k \qquad \text{(XXIV)}$$

$$\sigma^2 = \sum_1^k t_i \sigma_i^2 + \sum_1^k t_i \cdot (\mu_i - \mu)^2, \qquad (XXV)$$

The derivation of these formulas is given in Appendix III. The chief reason for including this section is to point out the similarity between formulas XXIV and I and between XXV and XI, which is just as would be expected.

APPENDIX I - Examples

A. Example One - (All n_i equal)

(1) Consider the example given by Table 2 in which there are four equal subsamples, each of size ten.

TABLE 2

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	300	305	290	305
	325	325	285	290
	330	285	295	260
	335	310	300	325
	325	325	305	290
	335	330	290	280
n	10	10	10	10
$\bar{x}_i$	334.00	310.00	340.50	291.50
s_i^2	243.33	205.56	319.17	361.39
				N = 40
				$\bar{\bar{X}}$ = 310

(2) Using the raw data in this example, the value $s^2 = 503.85$ may be easily computed. However, it is the purpose of this example to demonstrate techniques for obtaining s^2 if the raw data is unavailable or if N is so large that it would not be feasible to use the raw data. The first step will be to use formula II to obtain $\bar{\bar{x}}$.

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{k} = \frac{1240}{4} = 310.$$

(3) Table 3 demonstrates the application of the analysis of variance, as described by Table 1, to obtain s^2.

TABLE 3

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	$(k-1) = 3$	TR = 9,485	tr = 3161.67	11.20
Error	$(N-k) = 36$	E = 10,165	$s_p^2 = 282.36$	
TOTAL	$(N-1) = 39$	T = 19,650	$s^2 = 503.85$	

Where:

$$E = (n-1)\ (\Sigma s_i^2) = 10{,}165$$

$$TR = (n)\ (\Sigma \bar{x}_i^2) - N\bar{\bar{x}}^2 = 9{,}485$$

$$T = E + TR = 19{,}650$$

$$s^2 = \frac{T}{(N-1)} = 503.85$$

$$s = \sqrt{503.85} = 22.45.$$

If a pooled estimate of variance is desired:

$$s_p^2 = \frac{E}{(N-k)} = 282.36.$$

If it is desired to test for the equality of means:

$$tr = \frac{TR}{(k-1)} = 3{,}161.67$$

$F = \dfrac{tr}{s_p^2} = 11.20$ with 3 and 36 degrees of freedom. The value of F indicates that the difference in means is highly significant.

(4) Applying the formulas from Section III-D, one obtains:

(a) $$s^2 = \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} = \qquad \text{(V)}$$

$$\frac{(9)(1129.45) + (10)(385{,}348.50 - 4 \times 96{,}100)}{39} = 503.85$$

(b) If a pooled estimate of the variance is available, one may use formula VII.

$$s^2 = \frac{(N-k)s_p^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} = \qquad \text{(VII)}$$

$$\frac{(36)(282.36) + (10)(385{,}348.50 - 4 \times 96{,}100)}{39} = 503.85.$$

(c) If it is desired that the numbers be kept smaller, formula XII may be used.

$$s^2 = \frac{(n-1)\Sigma s_i^2 + n\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N-1} = \qquad \text{(XII)}$$

$$\frac{(9)(1129.45) + (10)(948.50)}{39} = 503.85.$$

(d) If an approximation is desired, one may use formula IX.

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 =$$

$$\frac{1129.45 + 385,348.50}{4} - 96,100 = 519.49 \quad ;$$

giving a positive error of 15.64, exactly what formula X would indicate the error to be.

B. Example Two - (n_i unequal)

(1) Consider the following example in which there are four subsamples and a total sample size of 32; Table 4.

TABLE 4

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	355	305	290	305
	300	325	285	290
	325	285	300	
	330	310		
	325	325		
		330		
n_i	9	10	7	6
$\bar{x}_i$	333.89	310.00	307.86	293.33
s_i^2	273.61	205.56	415.48	196.67
				N = 32
				$\bar{\bar{x}}$ = 313.125

(d) If an approximation is desired, one may use formula IX.

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 =$$

$$\frac{1129.45 + 385,348.50}{4} - 96,100 = 519.49 \;;$$

giving a positive error of 15.64, exactly what formula X would indicate the error to be.

B. Example Two - (n_i unequal)

(1) Consider the following example in which there are four subsamples and a total sample size of 32; Table 4.

TABLE 4

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	355	305	290	305
	300	325	285	290
	325	285	300	
	330	310		
	325	325		
		330		
n_i	9	10	7	6
$\bar{x}_i$	333.89	310.00	307.86	293.33
s_i^2	273.61	205.56	415.48	196.67
				N = 32
				$\bar{\bar{x}}$ = 313.125

(2) From the sample of 32, the value $s^2 = 452.82$ can easily be computed.

(3) Table 5 gives the analysis of variance.

TABLE 5

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	(k-1) = 3	TR= 6529.68	tr=2176.56	8.11
Error	(N-k) = 28	E= 7515.15	s_p^2= 268.40	
TOTAL	(N-1) = 31	T=14,044.83	s^2 = 453.06	

Where:

$$E = \Sigma(n_i-1)\, s_i^2 = 7515.15$$

$$TR = \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2 = 3,144,042.18 - 3,137,512.50 = 6529.68$$

$$T = E + TR = 14,044.83$$

$$s^2 = \frac{T}{(N-1)} = 453.06.$$

(Note that there is a slight difference between this estimate and the one obtained from the basic data, due to rounding errors.)

$$s = \sqrt{453.06} = 21.29.$$

If a pooled estimate of variance is desired:

$$s_p^2 = \frac{E}{(N-k)} = 268.40.$$

If it is desired to test the equality of the means for the four subsamples:

$$tr = \frac{TR}{(k-1)} = 2176.56$$

$$F = \frac{tr}{s_p^2} = 8.11 \text{ with 3 and 28 degrees of freedom.}$$ This

indicates that the difference in means is highly significant.

(4) Applying the formulas from Section III-D, one obtains:

(a) $$s^2 = \frac{\Sigma(n_i-1)s_i^2 + \Sigma n_i\bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} = \qquad \text{(IV)}$$

$$\frac{7515.15 + 3,144,042.18 - 3,137,512.50}{31} = 453.06,$$

(b) If a pooled estimate of the variance is available, formula VI may be used.

$$s^2 = \frac{(N-k)s_p^2 + \Sigma n_i\bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} = \qquad \text{(VI)}$$

$$\frac{(28)(268.40) + 3,144,042.18 - 3,137,512.50}{31} = 453.11.$$

(c) If it is desired that the numbers be kept small, formula XI may be used.

$$s^2 = \frac{\Sigma(n_i-1)s_i^2 + \Sigma n_i(\bar{x}_i - \bar{\bar{x}})^2}{N-1} = \qquad \text{(XI)}$$

$$\frac{7515.15 + 6523.42}{31} = 452.86.$$

(Note that this is much closer to the true value than those listed under a or b).

C. Example Three - ($k = 2, n_1 = n_2$)

(1) To illustrate formula VIII, the first two columns from Table 2 will be used. From this:

$n = 10, N = 20$

$\bar{x}_1 = 334, \bar{x}_2 = 310$

$s_1^2 = 243.33, s_2^2 = 205.56$

$\bar{\bar{x}} = 322, s_p^2 = 224.44, s^2 = 364.21.$

(2) Applying formula VIII:

$$s^2 = \frac{(N-2)}{2(N-1)} \cdot (s_1^2 + s_2^2) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2 =$$

$$\frac{9}{19} (448.89) + \frac{5}{19} (576) = 364.24.$$

D. Example Four: (Add n_2 observations to a sample of size n_1)

(1) Consider the sample of 40, given by Table 2. It may be observed that:

$\bar{x}_1 = 310, s_1^2 = 503.85.$

(2) Suppose it is necessary to add the five additional items:

310, 293, 314, 280, and 300

$n_2 = 5, \bar{x}_2 = 305.40, s_2^2 = 374.80.$

(3) One may proceed by using formula I and then either IV or XI. For this formula, XI was used.

$N = 40 + 5 = 45$

$$\bar{\bar{x}} = \frac{(40)(310.00) + (5)(305.40)}{45} = 309.49$$

by formula I.

$$s^2 = \frac{(39)(503.85) + (4)(374.80) + (40)(.51)^2 + (4)(4.09)^2}{44}$$

$= 482.42$

by formula XI.

(4) Actually, the value for s^2 using raw data is 482.80.

E. Example Five: (Remove n_2 observations from a sample of size n_1.)

(1) The data of example four will be used for this.

$n_1 = 45$

$\bar{x}_1 = 309.49$

$s_1^2 = 482.80.$

(2) Remove the 5 observations which were added in example four.

$n_2 = 5$

$\bar{x}_2 = 305.40$

$s_2^2 = 374.80.$

(3) Use formulas XVII and XVIII, giving:

$$\bar{\bar{x}} = \frac{(45)(309.49) - (5)(305.40)}{40} = 310.00$$

$$s^2 = \frac{(44)(482.80) - (4)(374.80) - (40)(310)^2 - (5)(305.40)^2 + (45)(309.49)^2}{39}$$

$= 504.64.$

APPENDIX II - DETERMINATION OF THE ERROR IN FORMULA IX

Using formulas V and IX, the following error is observed:

$$E = s_A^2 - s^2 = \left\{ \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 \right\} - \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1}$$

$$= \frac{(N-n)\Sigma s_i^2}{N(N-1)} - \frac{n\Sigma \bar{x}_i^2}{N(N-1)} + \frac{\bar{\bar{x}}^2}{N-1}$$

$$= \frac{\Sigma s_i^2}{N} - \frac{(n-1)\Sigma s_i^2}{N(N-1)} - \frac{n\Sigma \bar{x}_i^2}{N(N-1)} + \frac{kn\bar{\bar{x}}^2}{N(N-1)}$$

$$\text{Error} = s_A^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i^2 - s^2).$$

Inasmuch as Σs_i^2 is larger than s^2, the error will always be on the positive side.

APPENDIX III - DERIVATION OF THE MEAN AND VARIANCE FOR A POPULATION COMPOSED OF k NORMAL POPULATIONS

A. Assume each of the k normal populations have a mean μ_i, variance σ_i^2, and contributes to the total population in the proportion f_i, with $f_1 + f_2 + \ldots + f_k = 1$.

B. $$y = \frac{f_1}{\sqrt{2\pi}\sigma_1} \cdot e^{\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right)} + \ldots + \frac{f_k}{\sqrt{2\pi}\sigma_k} \cdot e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}.$$

C. $$m(\theta) = f_1 e^{\left(\frac{1}{2} \cdot \theta^2\sigma_1^2 + \theta\mu_1\right)} + \ldots + f_k e^{\left(\frac{1}{2} \cdot \theta^2\sigma_k^2 + \theta\mu_k\right)}.$$

D. $$\frac{\partial m(\theta)}{\partial \theta} = f_1(\theta\sigma_1^2 + \mu_1) \cdot e^{\left(\frac{1}{2} \cdot \theta^2\sigma_1^2 + \theta\mu_1\right)} + \dots + f_k(\theta\sigma_k^2 + \mu_k) \cdot e^{\left(\frac{1}{2} \cdot \theta^2\sigma_k^2 + \theta\mu_k\right)},$$

E. The mean of the total population:

$$\mu = f_1\mu_1 + f_2\mu_2 + \dots + f_k\mu_k .$$

F. $$\frac{\partial m_{(y-\mu)} \cdot \theta}{\partial \theta} = f_1 \left[\theta\sigma_1^2 + (\mu_1-\mu)\right] \cdot e^{\left(\frac{1}{2} \theta^2\sigma_1^2 + \theta(\mu_1-\mu)\right)} + \dots,$$

G. $$\frac{\partial^2 m_{(y-\mu)} \cdot \theta}{\partial^2 \theta} = f_1 \left[\theta\sigma_1^2 + (\mu_1-\mu)\right]^2 \cdot e^{\left(\frac{1}{2} \theta^2\sigma_1^2 + \theta(\mu_1-\mu)\right)} + f_1\sigma_1^2 \cdot e^{\left(\frac{1}{2} \theta^2\sigma_1^2 + \theta(\mu_1-\mu)\right)} + \dots,$$

H. The variance for the total population:

$$\sigma_y^2 = f_1(\mu_1-\mu)^2 + f_1\sigma_1^2 + \dots + f_k(\mu_k-\mu)^2 + f_k\sigma_k^2 = \Sigma f_i\sigma_i^2 + \Sigma f_i(\mu_i-\mu)^2.$$

APPENDIX IV - SUMMARY OF FORMULAS

A. $n_1 = n_2 = \dots n_k$

(1) The Mean for the total sample

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{k} .$$ Formula II

(2) Pooled Estimate of the Variance

$$s_p^2 = \frac{\Sigma s_i^2}{k}. \qquad \text{Formula III (a)}$$

(3) Variance for the total sample

$$s^2 = \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} \qquad \text{Formula V}$$

$$s^2 = \frac{(N-k)s_p^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} \qquad \text{Formula VII}$$

$$s^2 = \frac{(n-1)\ \Sigma s_i^2 \ + \ n\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N-1}, \qquad \text{Formula XII}$$

(4) Approximation (Use only if n is large.)

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 = s_p^2 + \frac{\Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2. \qquad \text{Formula IX}$$

The error in Formula IX

$$\text{Error} = s_A^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i^2 - s^2) \geq 0. \qquad \text{Formula X}$$

(5) $k = 2,\ n_1 = n_2$

$$s^2 = \frac{N-2}{2(n-1)} \cdot (s_1^2 + s_2^2) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2. \qquad \text{Formula VIII}$$

B. The n_i are unequal

(1) The mean for the total sample

$$\bar{\bar{x}} = \frac{\Sigma n_i \bar{x}_i}{N}. \qquad \text{Formula I}$$

(2) Pooled estimate of the variance

$$s_p^2 = \frac{\Sigma (n_i - 1) s_i^2}{N-k} . \qquad \text{Formula III}$$

(3) Variance for the total sample

$$s^2 = \frac{\Sigma (n_i - 1) s_i^2 + \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} \qquad \text{Formula IV}$$

$$s^2 = \frac{(N-k) s_p^2 + \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} \qquad \text{Formula VI}$$

$$s^2 = \frac{\Sigma (n_i - 1) s_i^2 + \Sigma n_i (\bar{x}_i - \bar{\bar{x}})^2}{N-1} . \qquad \text{Formula XI}$$

C. Formulas associated with changes in data

(1) Add an observation y to a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + y}{n_1 + 1} \qquad \text{Formula XIII}$$

$$s^2 = \frac{n_1 - 1}{n_1} \cdot s_1^2 + \frac{(\bar{x}_1 - y)^2}{n_1 + 1} . \qquad \text{Formula XIV}$$

(2) Add observations y and w to a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + y + w}{n_1 + 2} \qquad \text{Formula XV}$$

$$s^2 = \frac{(n_1 - 1) s_1^2 + y^2 + w^2 + n_1 \bar{x}_1^2 - (n_1 + 2)\bar{\bar{x}}^2}{n_1 + 1} . \qquad \text{Formula XVI}$$

(3) Discard n_2 observations from a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - n_2\bar{x}_2}{n_1 - n_2} \qquad \text{Formula XVII}$$

$$s^2 = \frac{(n_1-1)s_1^{\ 2} - (n_2-1)s_2^{\ 2} + n_1(\bar{x}_1^{\ 2}-\bar{\bar{x}}^2) - n_2(\bar{x}_2^{\ 2}-\bar{\bar{x}}^2)}{(n_1 - n_2 - 1)}, \qquad \text{Formula XVIII}$$

(4) Discard the observation (y) from a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y}{n_1 - 1} \qquad \text{Formula XIX}$$

$$s^2 = \frac{n_1-1}{n_1-2} \cdot s_1^{\ 2} - \frac{n_1(\bar{x}_1-y)^2}{(n_1-1)(n_1-2)} . \qquad \text{Formula XX}$$

(5) Replace the observation y by w

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y + w}{n_1} \qquad \text{Formula XXI}$$

$$s^2 = s_1^{\ 2} + \frac{w^2-y^2-n_1(\bar{\bar{x}}^2-\bar{x}_1^{\ 2})}{n_1 - 1} \qquad \text{Formula XXII}$$

or

$$s^2 = s_1^{\ 2} + \frac{(w-y)\left[(n_1-1)w + (n_1+1)y - 2n_1\bar{x}_1\right]}{(n_1)\ (n_1-1)} . \qquad \text{Formula XXIII}$$

D. Formulas associated with a total population composed of k normal populations

$$\text{Let } y = \frac{f_1}{\sqrt{2\pi}\sigma_1} \cdot e^{\left(-\frac{1}{2\sigma_1^{\ 2}} \cdot (x-\mu_1)^2\right)} + \ldots + \frac{f_k}{\sqrt{2\pi}\sigma_k} \cdot e^{\left(-\frac{1}{2\sigma_k^{\ 2}} \cdot (x-\mu_k)^2\right)}$$

Where $f_1 + f_2 + \ldots + f_k = 1$, then:

$$\mu = f_1\mu_1 + f_2\mu_2 + \ldots + f_k\mu_k$$ Formula XXIV

$$\sigma^2 = \Sigma f_i\sigma_i^2 + \Sigma f_i(\mu_i - \mu)^2.$$ Formula XXV

SYSTEM CONFIGURATION PROBLEMS
AND ERROR SEPARATION PROBLEMS*

Fred S. Hanson
Plans and Operations Directorate
White Sands Missile Range, New Mexico

ABSTRACT. Practical geometric criteria and optimization methods are needed for laying out, or selecting, multi-instrument configurations for flight measurement. The problem is to discover - and demonstrate - some principles that are at least in the right direction. A general solution should be possible for the variation of uncertainty of intersection location as a function of angles-of-intersection of lines-of-sight. It might also be possible to calculate the optimum ground-pattern for a given station density and missile trajectory. The second problem is to develop - in detail - analytical tools for separating position-measurement error, time-measurement error, and lack-of-fit of a given polynomial -- as these errors exist in undesigned, but redundant, data. Questions concern: the validity of linearization of data for this purpose; procedures for calculating lack-of-fit of polynomials of degrees greater than one; limitations in conversion of regressions to analyses of variance.

INTRODUCTION. This paper is clinical -- especially in the sense that it is not completed work.

BACKGROUND. Figure 1 is a White Sands Missile Range briefing chart. It shows: the principal Range (heavy line); the part-time extension (at the top); and the White Sands Monument (small internal area). Headquarters - and the main launch areas - are at the lower end of the Range.

The distinction between optical and electronic tracking instruments has been lost in this black-and-white print. Optical instruments include: cinetheodolites, telescopes, fixed cameras, and ballistic cameras. Not every station is shown. For instance, there are several hundred prepared sites where fixed cameras can be set up. Electronic tracking instruments include: radars, dopplers, and miss-distance systems. Again, not every station is shown. (There are several hundred prepared sites where DOVAP receivers can be set up.) The gray - and part-gray - dots are telemetry receivers.

*Comments on this paper by some of the panelists can be found following the figures at the end of this article.

It may be apparent that the systems in Figure 1 were not laid out on any rigorous basis.

CONFIGURATION HYPOTHESES. More than three years ago (Ref. 1), the writer asserted two hypotheses about instrument layout, or selection -- to initiate action toward solution.

First, it was asserted - intuitively - that the most favorable elevation angle for observing a missile is 45°. Second, the writer stated an optimum ground-configuration - for each integral number of stations - with respect to a single point in space. This was done on the assumption that the best intersection of lines-of-sight from two stations is - when considered by itself - 90°. Conversely, it was assumed that the worst intersection occures when one station looks over another's shoulder, or they look down each others throats -- 0° or 180°, parallel. Referring to Figure 2, the most favorable ground-configuration for optical stations was asserted - without proof - to be: two-station - right-isosceles triangle with missile at apex; three-station - equilateral triangle with missile at center; (in all subsequent cases, missile at center) four-station - any four corners of equilateral pentagon; five-station - said pentagon; six-station - any six corners of equilateral heptagon; seven-station - that heptagon; etc. The (corresponding) intersection angles are: 90°, 120°, 72°, and 51.4°. For twelve or thirteen stations - a tridecagon - the angle would be down to 27.7°.

DEMONSTRATION OF HYPOTHESES. After proposing this paper, the writer made a crude approach to demonstrating (the validity of) these simple hypotheses.

Figure 3a shows the asserted two-station optimum. This can be any plane through both stations and the missile. The diagram represents the 90° intersection - together with some dispersion index, such as the standard deviation.

Figure 3b is an enlargement of the area of uncertainty. We are assuming the two instruments are equally precise. Let's approximate the actual error-ellipse by the almost-square in Figure 3a - and approximate that by the square in Figure 3b. The horizontal diagonal is a measure of the combined error-variance. If we increase the intersection angle, by moving the stations farther apart - or by lowering the missile - the horizontal diagonal will lengthen. Of course, the vertical diagonal will shorten, correspondingly. In general, it's not sound practice to improve data in one coordinate by making it worse in another. (If we decrease the intersection angle - below 90° -, the horizontal diagonal gets smaller, at the expense of the vertical diagonal.) So, we may conclude 90° is the practical optimum.

Now, we have shown that 90° is the optimum intersection in any plane thru both stations and the missile. The plan for which the degradations from this optimum will be the same in its horizontal and vertical projections is the 45° plane. On the basis that there is no preferred coordinate, we have demonstrated the hypothesis regarding the optimum elevation angle.

If we choose to take our geometry in algebraic form, we can use the law of cosines to calculate the horizontal diagonal (Figure 3b):

$$a^2 = b^2 + c^2 - 2bc \cos \theta$$

where b and c are measures of the two observational variances. θ is approximately 90°. To see the effect of changing the intersection from 90°, let's replace θ by $90^\circ \pm \alpha$:

$$a^2 = b^2 + c^2 - 2bc \cos(90^\circ \pm \alpha)$$

In our case, b and c are equal, so:

$$a^2 = 2b^2 - 2b^2 \cos(90^\circ \pm \alpha)$$

$$= 2b^2 \; [1 - \cos(90^\circ \pm \alpha)]$$

Substituting,

$$a^2 = 2b^2 (1 \mp \sin \alpha).$$

So, approximately, if the intersection angle is changed, the combined variance in one coordinate increases as the sine of the angular deviation from 90°.

A similar exercise can be gone thru for the 3-station equilateral triangle. In that case, the error-ellipse is approximated by an almost-equilateral hexagon.

MORE ANALYTICAL SOLUTION. W. F. Mammack (Ref. 2) has furnished the writer a solution which does not depend on approximating the almost-square -- or on testing a hypothesis.

Referring to Figure 4a - the trigonometry for the general two-station case yields:

$$x = \frac{b}{2}\,\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \qquad y = b\,\frac{\sin\theta_1 \sin\theta_2}{\sin(\theta_1 + \theta_2)}$$

Applying the standard error-propagation formula:

$$\epsilon_x^2 = \left(\frac{\partial x}{\partial \theta_1}\right) \epsilon_{\theta_1}^2 + \left(\frac{\partial x}{\partial \theta_2}\right) \epsilon_{\theta_2}^2$$

(and similarly for y) yields:

$$\epsilon_x^2 = \frac{b^2}{\sin^4(\theta_1 + \theta_2)} \left(\frac{1}{4}\sin^2 2\theta_2 \epsilon_{\theta_1}^2 + \frac{1}{4}\sin^2 2\theta_1 \epsilon_{\theta_2}^2\right)$$

$$\epsilon_y^2 = \frac{b^2}{\sin^4(\theta_1+\theta_2)} \cdot \left(\sin^4\theta_2 \epsilon_{\theta_1}^2 + \sin^4\theta_1 \epsilon_{\theta_1}^2\right) .$$

Simplifying to the equidistant, equal-precision case (Figure 4b):

$$\mu^2 = \epsilon_x^2 + \epsilon_y^2 = b^2 \frac{(1 - \cos 2\theta)}{\sin^4 2\theta} .$$

If this total error is minimized with respect to θ, the minimum is found to occur at:

$$\cos 2\theta = 1/3$$

$$2\theta = 70.5^\circ .$$

So, Mimmack's optimum intersection angle is 109.5°.

In R. C. Davis' NOTS report on his cinetheodolite-reduction method (Ref. 3), he minimized the observational error-ellipse of the two-station-missile triangle, by a matrix process. With the stations fixed and the missile altitude allowed to vary, Davis found the optimum intersection to be 120°. He theorized this was the result of compromise between the most favorable intersection and the decrease in the linear error (corresponding to a given angular error) as the missile moves closer to the stations. Mimmack's solution represents this same case. So, there is an apparent discrepancy in their results.

With the missile altitude fixed and the stations free to move, Davis found the optimum intersection to be 60°. He theorized this was the result of compromise between most favorable intersection and moving the stations closer to the missile. The present writer thinks Davis' explanations are correct.

However, it appears that the optimum ground-configurations hypothesized in this paper are still optimum when the effect of slant range is included. Also, 45° planes are the only ones for which the degradations (of coordinate projections) from the optimum intersection will be the same - whatever the optimum may be. So, we have "demonstrated" a simple set of rules for laying out, or selecting, a group of stations - for any given point on a missile trajectory -, and for determining the optimum scale of their configuration. The point used could be the mid-point of a trajectory segment.

MINIMUM BIAS CONFIGURATION. The demonstration based on Figure 3 treated error as a dispersion index (or precision index). Let's consider (it as) a discrete, or net, error. Then, in Figure 3a, if we increase θ above 90°, the horizontal (error-) resultant - corresponding in size to the smaller almost-square - will lengthen if the (discrete angular) errors happen to have the same sign (Figure 5a); if the errors have opposite signs (Figure 5b), their (vertical) resultant will shorten correspondingly. (Of course - in the equal-accuracy case - there will be only a horizontal, or only a vertical, resultant.) In general, it's not sound practice to (set out to) improve data in one coordinate by taking an even chance that we will, instead, make it worse in another. (Even chance, because - to the extent that a given-type instrument consistently has the same sign, it is more likely to be adjusted, or corrected for.) If we decrease θ (below 90°), the possible homopolar (horizontal) error-resultant gets smaller, at the expense of the possible heteropolar (vertical) error-resultant gets smaller, at the expense of the possible heteropolar (vertical) error-resultant. So, we may conclude - 90° is the practical optimum. The rest of the writer's geometric and algebraic demonstrations apply similarly. Summary: perpendicular intersection (per se), 45° elevation, the right-isosceles triangle for the two-station case, etc. are all optimum for accuracy as well as precision.

PATTERN HYPOTHESES. How does one generalize from a single group of stations to a larger area -- for (several segments of) a family of trajectories? What sort of patterns can we construct with our optimum figures? In Figure 6, what is wrong with a grid built up of optimum three-station configurations? Equilateral triangels form hexagons, which violates our odd-sided rule. Each station is in line with all the other stations. Continuing in Figure 6, pentagons seem to form a desirable pattern - leaving a few gaps of isosceles-triangle pairs. (Four stations are in line across each triangle pair.) Heptagons might do as well.

In determining the optimum layout, the decisive constraint could be the number of stations needed to meet requirements (for precision). Or, it could be budgetary (the number of stations permitted per hundred sq. mi.). Or, it could be the effective range of a station - as a configuration radius.

Perhaps someone can demonstrate that the optimum pattern is random. Or, that a random pattern is not optimum. A random pattern might have the minimum percent of stations in line with each other - but it wouldn't be the most efficient dispersion. Mimmack (Ref. 2) notes that it is desirable for a position measurement to be independent of any coordinate system; that this implies the station geometry should be free of symmetries; that the symmetry of being in the same ground-plane is largely unavoidable.

DISCUSSION OF CONFIGURATION. The optimum configuration would maximize: accuracy, precision, versatility, reliability, and economy. Flight-measuring instruments exist in three conditions: fixed, (self-contained) mobile, transportable (to prepared sites).

The writer chose to start with the precision of a single point-in-space, because this is WSMR's operating standard - and because it lends itself to an analytical approach which proceeds from the simple to the complex. The Range's instrumentation plans are prepared per segment of a trajectory. The present standard seems to be the best (single) compromise between an operating viewpoint and a missile-engineer viewpoint. Aside from having a consistent benchmark, the important question is: "What aspect of a given missile-performance variable is most significant to a particular missile project?"

This is, after all, a clinical paper. The writer's aim is not - necessarily - to solve the whole problem by an analytical approach. (It is to increase understanding of the subject.) We "demonstrated" the "90° - optimum" intersection in any plane - for observing a point-in-space. We found a (limited) approximate solution, in two dimensions, for the variation of uncertainty-of-intersection-location as a function of angle-of-intersection-of-lines-of-sight. Mimmack (Ref. 2) obtained a general solution (to this problem) for two dimensions; his method could be extended to three dimensions. It may be that an optimum ground-pattern can be constructed with pentagons.

The optimum-overall-pattern problem could be stated: "Is there a unique solution for the most efficient layout, for a given optical-station density - or for a given effective station-range - and for the Range's total trajectory-volume?" It seems clear that any thorogoing analysis of this problem must be made in three dimensions.

Reference 4, revised annually, discusses computer programs for propagating "typical" errors-of-observation thru the (trigonometric) equations relating coordinates of any given point-in-space to the (angular, etc.) "observations" of the point by stations-of-known-location. These are essentially the same programs used for trial-and-error simulation at White Sands. AMR (now ETR) calls the - a priori - error estimates so obtained "a geometric dilution of precision (GDOP)". Properly, this term should be reserved for the geometric component of position-measurement variance.

ERROR SEPARATION PROBLEM. The second problem is this: "Can we determine (by statistical methods) - qualitatively and quantitatively - how much of the error-variance in our (final) missile-position data is position-error, and how much is time-error?" For velocity and acceleration (or smoothed position data), we would also like to know the relative magnitude of a third variance component - the lack-of-fit of the polynomial which we use to obtain (smoothed and) derivative data.

The jitter (and wander) of time-signal generators is small. Propagation- and receiver-delays are appreciable - different for each station - somewhat variable - and partly compensated for. Recording delays for: time-code marks, missile image, (angular) dial readings, etc. are appreciable, different, and somewhat variable. Overall time-measurement error includes errors in synchronizing: timing, missile position, and mount position -- physically, on the record, in conversion, in computing, and in reporting.

For a Mach 10 missile, a millisecond overall time-measurement error would be equivalent to a position error of 10 ft. A recent figure for the speed of an ICBM warhead is 26,400 ft/sec (Ref. 5); in that case a millisecond is 26.4 ft.

Actual requirements - and capabilities - for instrumentation timing-and-synchronization should be known - in specifiable terms. A complete description of position accuracy - or precision - would include a separate specification of time accuracy - or precision. If time-measurement error is ignored, it shows up as position error - but, it cannot be decreased by improving the position-measuring device (as such). If time-measurement error is appreciable, these two components of position error should be separated before calculating velocity (or acceleration) error. We don't know that time-measurement error is an appreciable part of the whole - but we can't afford not to know how much it is.

This paper presents problems -- not solutions. But - in presenting this problem - let's review the approaches the writer has already considered.

SEMI-QUANTITATIVE SEPARATION. About four years ago, the writer suggested a semi-quantitative method for "separating" time error from position error - in final data. Let's look at the three types of "regression" (correlation) of a position coordinate and time (Figure 7).

Figure 7a shows regression of x as a function of t - in which time is assumed to be exactly measured, and that curve is fitted which minimizes the (sums of the squares of) the deviations in position. This is the one WSMR uses, in its data reduction.

Figure 7b shows regression of t as a function of x - in which position is assumed to be exactly measured, and that curve is fitted which minimizes the (sums of squares of) the deviations in time. From a mathematical standpoint, this is as logical as the first.

Figure 7c shows simultaneous regression of x and t - in which they are assumed to be measured equally well, and that curve is chosen which minimizes the (ss of) the deviations. This is sometimes called the "best fit".

If measurements of x and t are about equally in error, curve c will (tend to) fall about halfway between a and b - and is the best choice, in this case.

If one variable is badly measured, the curve which minimizes the variability of the badly measured variable will (tend to) deviate the most from the other two -- but will (tend to) be closest to the (physically) true relationship. This justifies use of method a (by WSMR) - if the assumption that position is (always) much more poorly measured proves correct. The curve of "best fit" - c - best represents the data, as such, in any case.

By comparing these three types of regression - and taking into account any knowledge of the (physically) true curve from independent data, and/or physical theory -- it is possible to obtain semi-quantitative estimates of how relatively well two variables are measured. The writer knows from experience this works in applying linear regression to rather poor data. It may be an even sharper tool in applying curvilinear regression to rather good data.

QUANTITATIVE SEPARATION. On the basis of redundancy in measuring missile position, these three regressions can be converted to corresponding analyses of variance. This should permit quantitative separation of time error and position error. Procedures are available for analysis of variance of types a and b regression. Type c regression could be handled - for the linear case - by these same (single-fixed-variate) methods, by a rotation of axes. It may also be possible to discover (or devise) a bivariate analysis - at least for the linear case. If necessary curvilinear data can be transformed to linear.

Such analyses of variance include a lack-of-fit term, which is available for the linear fixed-variate case in Reference 6. It appears to be available for the curvilinear fixed-variate case from (such sources as) References 7 and 8.

The usual procedure at WSMR is to fit a second-degree polynomial. If our lack-of-fit proves to be appreciable compared to position-error, it will follow that we need to improve our data-reduction procedure.

The writer's questions with regard to the above analyses of variance are these:

1. What analysis-of-variance components can we get from linear fixed-variate regressions of types a and b if we have (apparent) redundancy in (a given) position (coordinate) at (equally-spaced) apparent times -- and (if we) convert these assumed-x redundancies to assumed-t redundancies by (means of) the reciprocal-of-the-slope of the type a regression (i.e., if we multiply by the corresponding value of $\Delta t/\Delta x$). Specifically, can we separate timing-error, position-error, and (two) lack-of-fit terms? As a working reference for this would the Panel recommend Reference 9 - or some other? Same questions for curvilinear case -- using the reciprocal of the type a slope at each point to convert - and substituting Reference 10 as a working source.

2. Suppose we apply this fixed-variate analysis to type c linear regression by a rotation of axes -- and calculate the assumed-normal redundancies by interpolating between the assumed-x and the (corresponding) assumed-t redundancies, above (in proportion to the ratio of the angle-between-the-x-axis-and-normal to 90°). Can we get anything out of this transformed type c analysis of variance?

3. Can the Panel give a reference which shows how to calculate lack-of-fit for type c linear regression?

4. Can the Panel give a reference to - or device - a bivariate analysis of variance for linear regression if we have (apparent) redundancy in (a given) position (coordinate) at (equally-spaced) apparent times? Same question for curvilinear regression.

5. Suppose we transform a variable to linearize a (curvilinear) regression -- and then perform the (linear) analysis of variance under question 1. Is it necessary to leave the result in the transformed state? Is it valid to "untransform" the variance of the transformed variable? Can the Panel give a reference on estimating the error due to "untransforming"?

6. Does Reference 7, 8, or 10 clearly give a procedure for calculating lack-of-fit for curvilinear single-fixed-variate regression? If not, can the Panel give a reference which does?

SEPARATION AT A POINT. So far we've taken a time-varying look at the flight-measurement process. White Sands is also interested in (knowing) the uncertainties associated with single values of unsmoothed data. It should be possible to make a hypothetical - if inconclusive - analysis of the errors of a single point (in space and time) by looking at the error as all (in) position, all (in) time, all tangential, or all normal. An additional approach to the "instantaneous" aspect might be to consider (two) successive data-points as observations of their mean point. Can we get any - qualitative or quantitative - separation of timing and position error out of these approaches? Can the Panel suggest any further approach to analysis of the errors of single-values-of-unsmoothed-data?

REFERENCES

1. WSMR Study Group Report, "Formulating Requirements for Range Instrumentation": DF, Supplement to Part b., 1962.

2. Letter from Wm. E. Mimmack, White Sands Missile Range - on leave to Graduate School, University of Rochester - 1965.

3. Davis, R. C. "Techniques for the Statistical Analysis of Cinetheodolite Data", NAVORD Report 1299 (NOTS 369), China Lake, Cal., 1951.

4. Mann, H. P. "The Accuracy of AMR Instrumentation", RCA Systems Analysis Technical Report 23, Patrick Air Force Base, Fla., 1962.

5. Drewry, I. O. "Hot Rod Missile", Army Information Digest, May 1965, pp. 22-56.

6. Anderson and Bancroft, Statistical Theory in Research, McGraw-Hill, 1952, pp. 156-158.

7. Schultz, H., "The Standard Error of a Forecast from a Curve", J. Amer. Stat. Assoc., 25, 139-185 (1930).

8. Ezekiel and Fox, Methods of Correlation and Regression Analysis, (3rd ed.) John Wiley & Sons, 1959.

9. Acton, Analysis of Straight-Line Data, John Wiley & Sons, 1959.

10. Brownlee, Statistical Theory and Methodology in Science and Engineering, John Wiley & Sons, 1960.

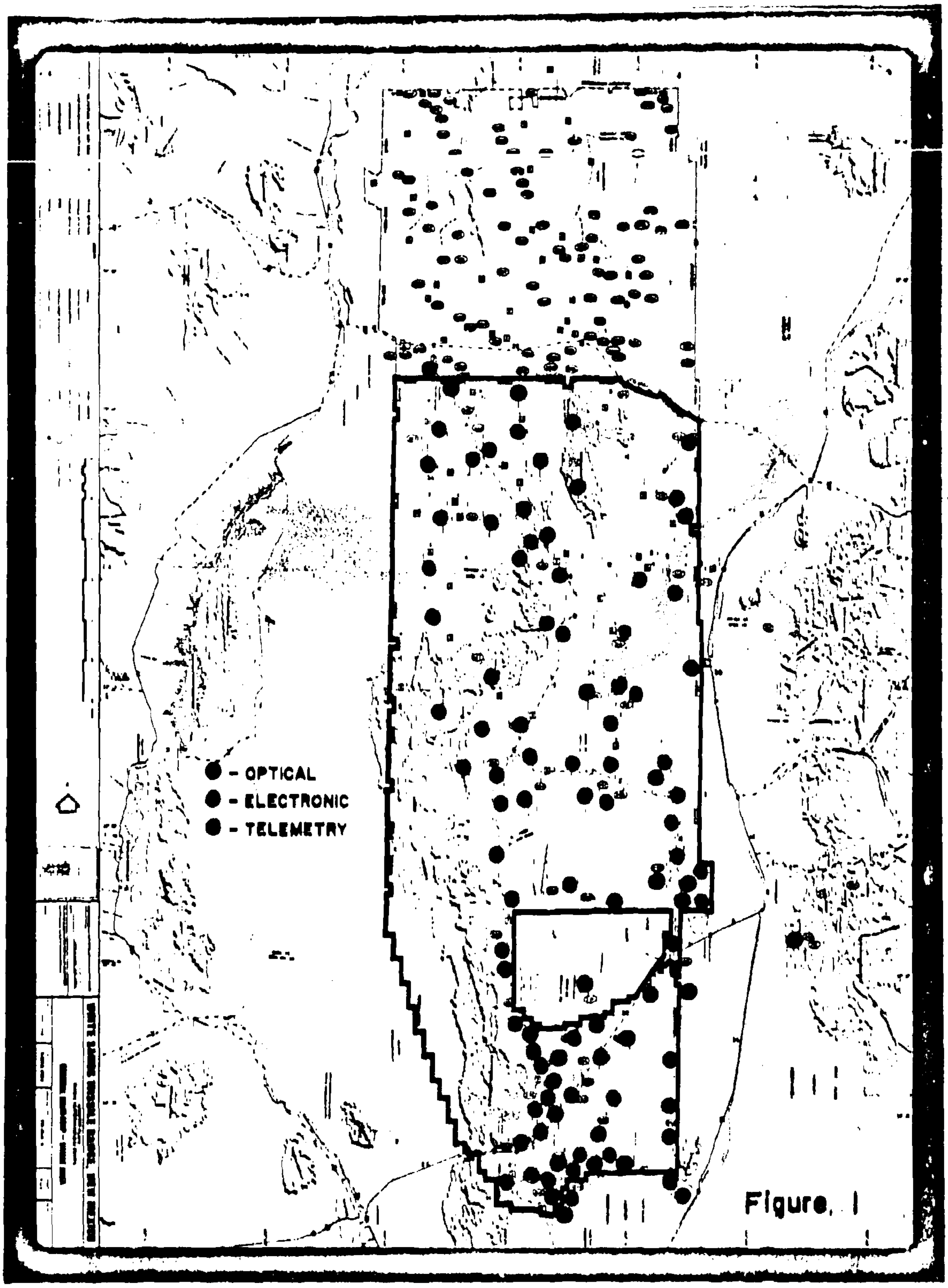


Figure. 1

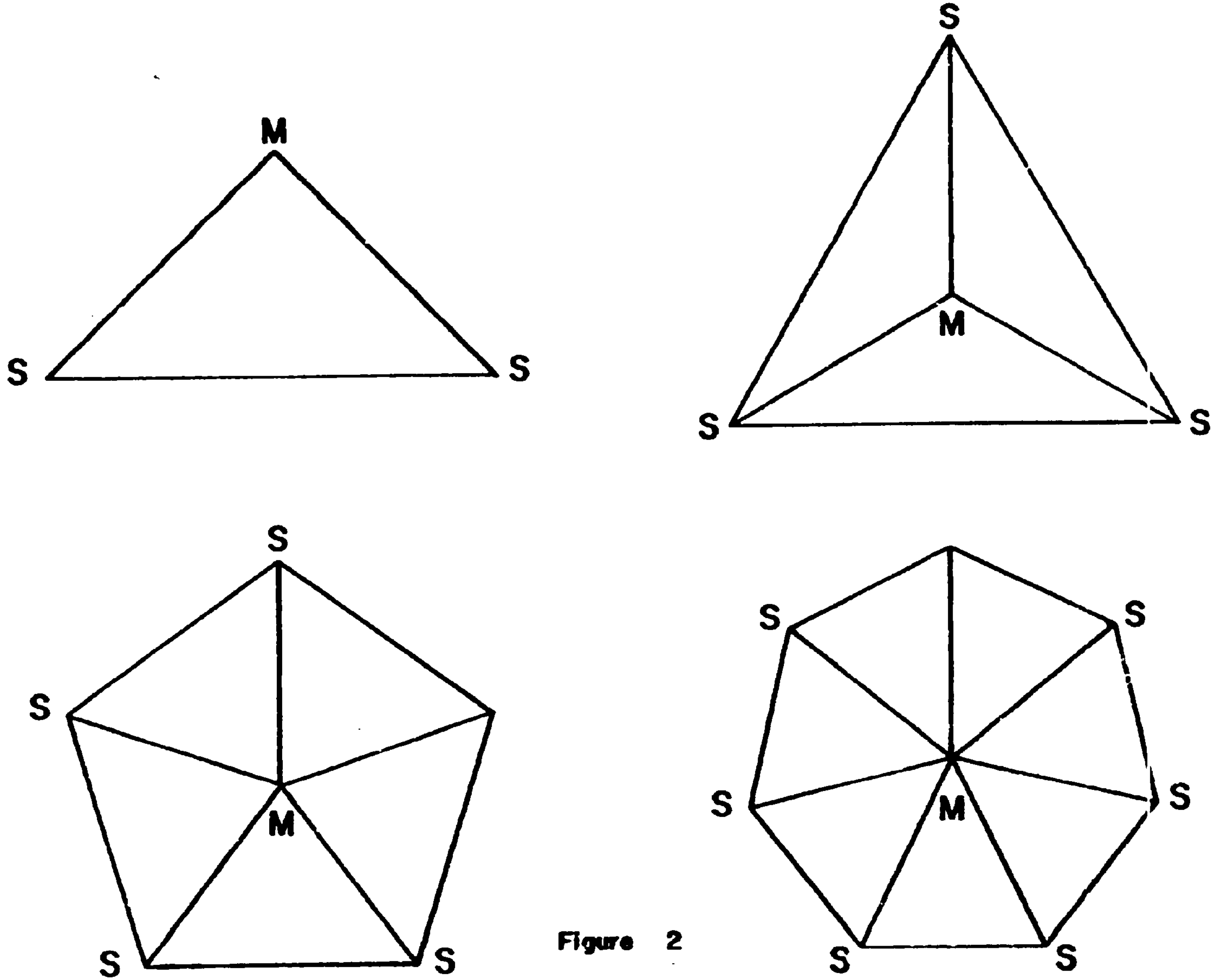


Figure 2

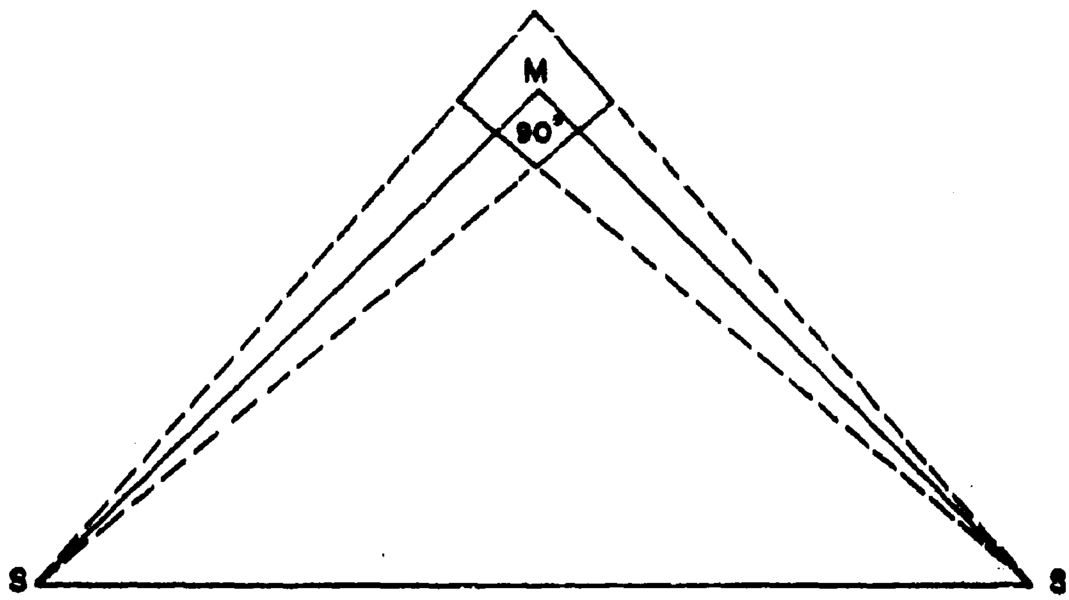


Figure 3a

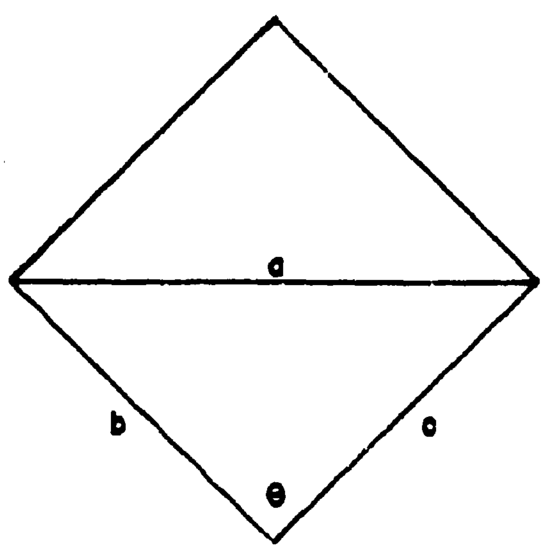


Figure 3b

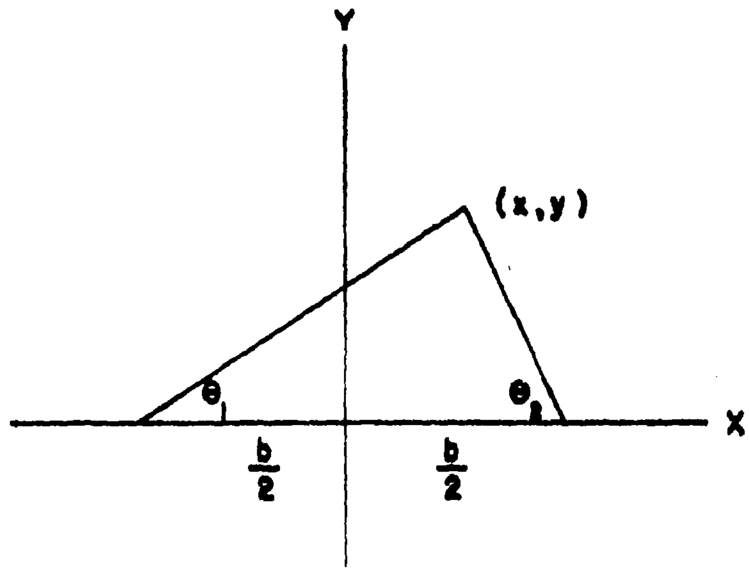


Figure 4a

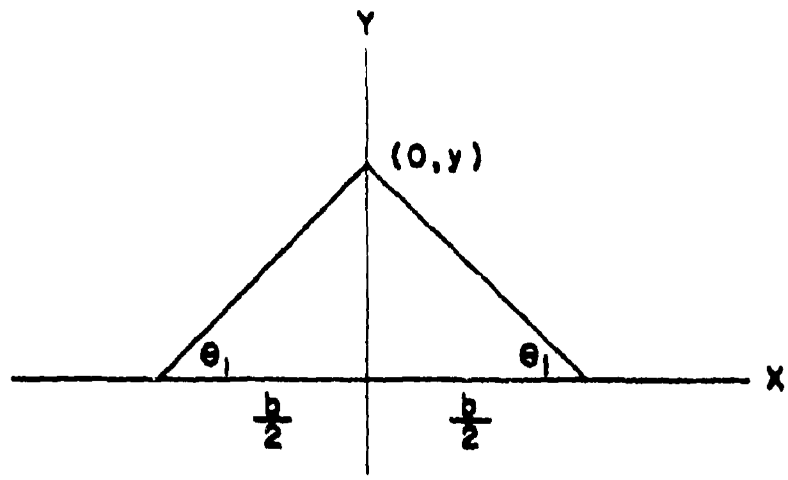


Figure 4b

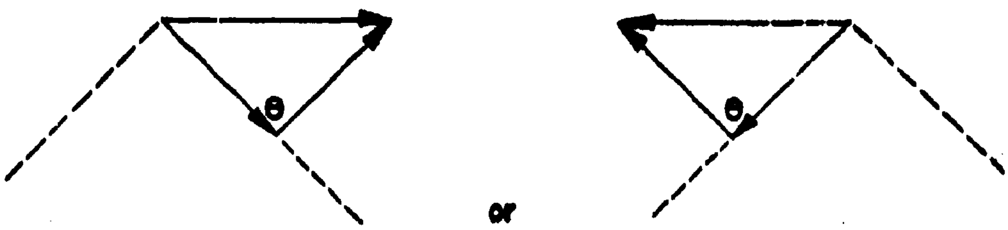


Figure 5a. Same sign

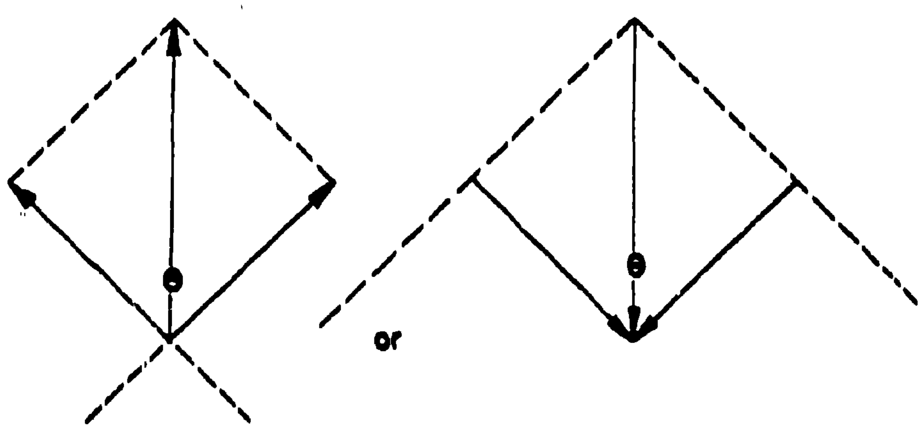


Figure 5b Opposite signs

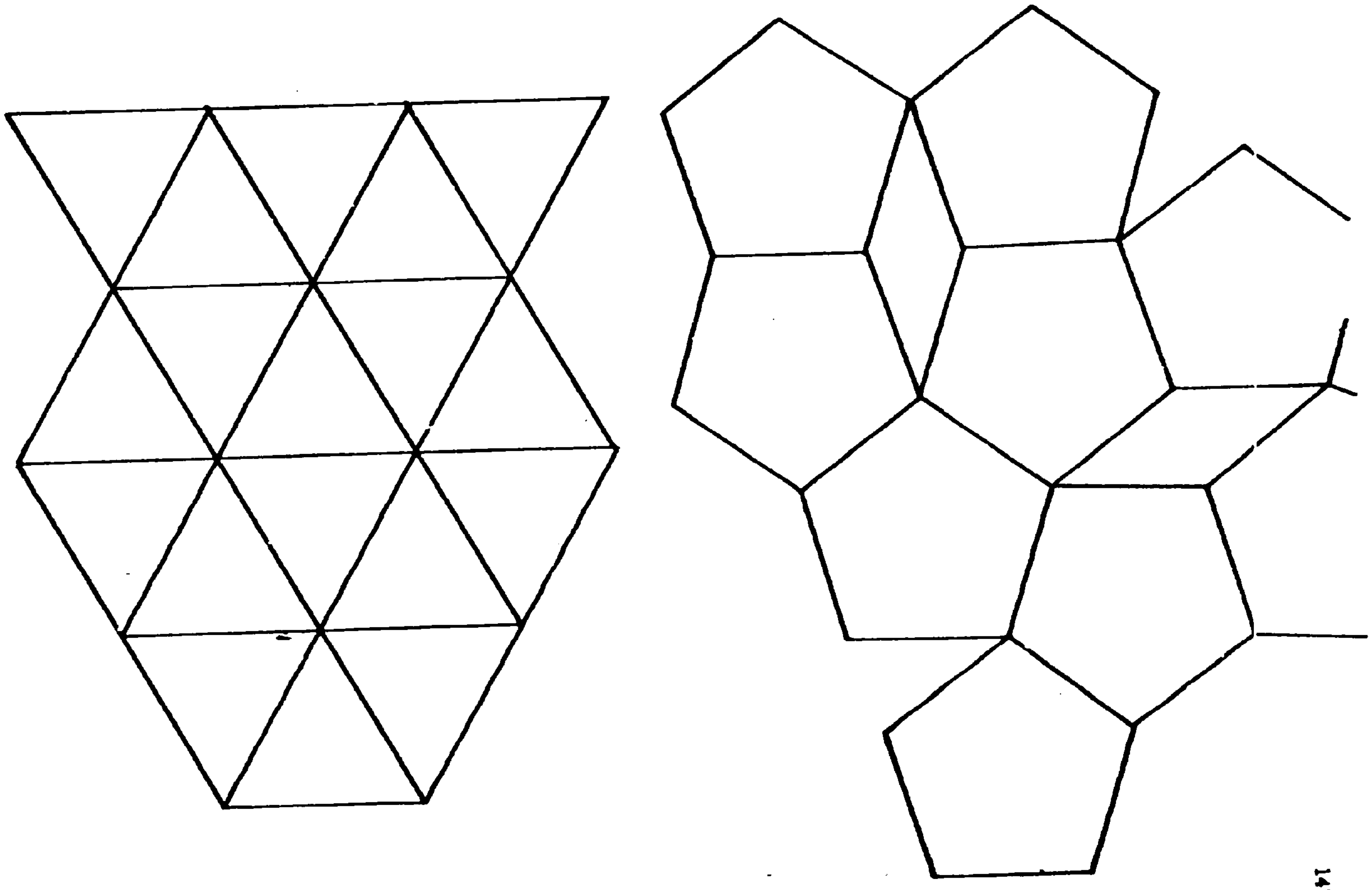


Figure 6 POSSIBLE SYSTEM CONFIGURATIONS

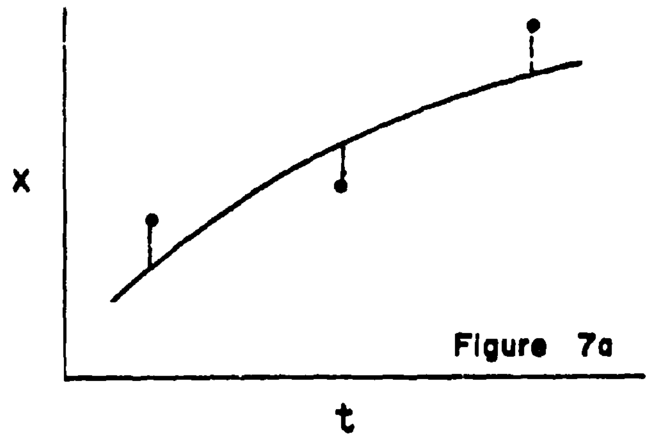


Figure 7a

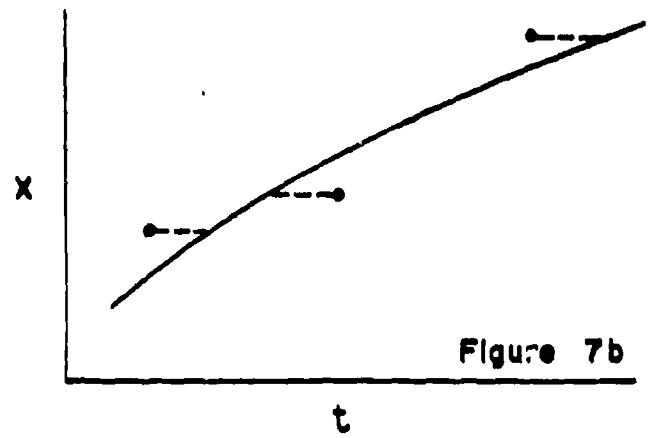


Figure 7b

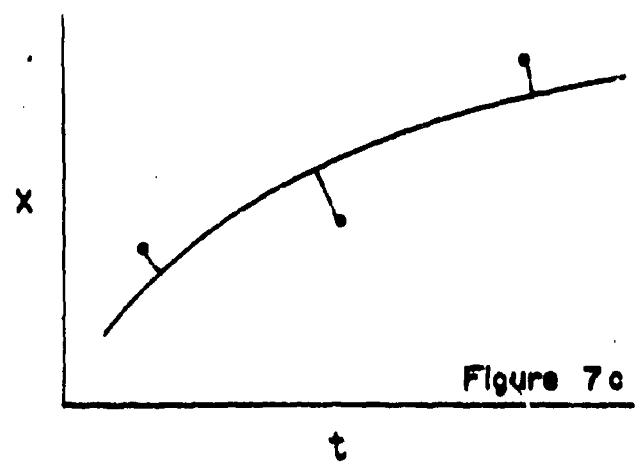


Figure 7c

COMMENTS ON PRESENTATION BY FRED HANSON

Frank E. Grubbs
Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

In my opinion the problems and questions Dr. Hanson raised can be solved satisfactorily only by competent personnel working rather full time on the overall problem! I say this because the problem is so involved from both the physical and the analytical standpoints that it is easy to overlook the importance of all of the "errors" operating simultaneously, so to speak.

Concerning station location geometry, I think that something can indeed be done on this and Dr. Hanson's ideas may be near enough the optimum, considering other involved difficulties. I can see that White Sands might decrease position estimation errors, etc., by optimum station locations, whereas the Atlantic Missile Range cannot really do this.

Just what sums of squares must be minimized, as Dr. Hanson points out, involves considerable study. From my limited experience, I have the feeling that relative time is quite good but that position data is not so good because of intersection geometry, and the errors which creep into this depending on unexplainable biases for the missile flight, calibration, refraction and other corrections, etc. Of course, all of these things vary with the type of instrumentation, etc.

Power spectral density type analyses, are certainly being looked into by many people now and this work is no doubt paying off as many of the problems involved necessarily fall in this area, even though this is an added dimension of complication.

The nearest publication, as Dr. Hanson is aware, which I think is beginning to approach methods required to settle some of the questions Dr. Hanson is raising is the annual report, "Accuracy of AMR Instrumentation", by H. P. Mann. The latest version, as Dr. Hanson knows, does contain a lot of good material and attempts to cover most of the important viewpoints, but still doesn't go far enough.

I think the tracking data analysis problem is by far the most interesting overall one I have been introduced to in recent years, but unfortunately it is something that does not carry the proper priority with many of us in spite of its great importance. Our Panel on Tracking Data Analysis is quite inactive now but if anything comes up on this in the future, I would hope to be in touch with Dr. Hanson.

COMMENTS ON PRESENTATION BY FRED HANSON

Emil H. Jebe
Institute of Science and Technology
The University of Michigan
Ann Arbor, Michigan

Before commenting on Dr. Hanson's two problems, I will first take up the matter of references. I certainly recommend F. S. Acton and K. A. Brownlee (titles Dr. Hanson mentioned). Dr. Hanson has also used Anderson and Bancroft, which is good. Further, I will mention E. J. Williams' "Regression Analysis", J. Wiley & Sons, and Plackett's "Regression Analysis", Oxford Press. Also, O. Kempthorne's "Design and Analysis of Experiments" and H. Scheffe's "Analysis of Variance" may prove useful. There is a book by an Australian, P. G. Guest, "Numerical Methods of Curve Fitting", Cambridge University Press, 1961. Perhaps Dr. Hanson should look at the symposium publication, "Time Series Analysis", SIAM Series in Applied Mathematics, J. Wiley & Sons, 1963.

Now, to Dr. Hanson's problems, Number 1 first. Certainly, I must comment that my experience with the NORC project at Ft. Monroe, 1941-42, and with the Anti-aircraft Artillery Board, Camp Davis, 1942-44, is ancient history by comparison with the state of the art in the 60s. Generally, I agree with Dr. Hanson's analysis of the geometry of the situation, i.e., 45 degree elevation for line-of-sight and nearly orthogonal to missile path for a "reasonable" interval of time. From the algebra associated with the geometry one should be able to work out the error propagation for the position determinations. Of course, one must keep in mind the "best" physical model for the flight path of the missile in using the observed data to obtain <u>best</u> <u>apparent</u> position of missile at a given time.

I tend to think of this first problem more in practical considerations, given that the technical problem of determining location has been resolved to a useful accuracy and precision. Some method of assigning priorities to each day's or each week's missions must be worked out. Then with the resources at hand, an allocation must be made of stations to be manned with selected equipments. Consider Figure 1 for Mission A (highest priority). Enough paired stations, a and a', b and b', etc., must be manned to keep this missile path under adequate surveillance. Now, if a, b, c and d, etc., are too far apart, there will be too much uncertainty in the computed positions in the halfway-between regions. Next, Mission B (second priority) has to be similarly supported at a desired minimum level. If launch times can be programmed to some extent, it may be that some manned stations can support more than one mission. Continue for say two more Missions C and D. If any resources are left over, consider increasing density of manned pairs for Missions A, B, C and D in that order to shore up obvious weaknesses in trajectory assessment. These practical considerations seem much more relevant to me than going into geometrical considerations beyond the triangle. If is recognized that my sketch implies using rectangles or quadrilaterals in assessing position. When launch times are adequately separated so that all manned stations for each of the four missions can track each launch, then further geometrical considerations may be taken into account along the lines Dr. Hanson has discussed.

Now I turn to the second problem of analysis. Yes, one would like to have variance components for timing error and for position-measuring error. But how can one separate them? Without considerable study, more than I can give at this time, I have no direct suggestion. It is hoped that Dr. Hartley has given Dr. Hanson some useful direct suggestions. I use the term indirect for my ideas because I wish to lean on "design of experiments" considerations. By direct suggestions I mean extracting from present method of collecting data, components of variance of the two kinds desired.

In directing Dr. Hanson's attention to design of experiments concepts, I believe WSMR is in an outstanding position to carry out some special studies. Of course, these activities must be budgeted, but it does not seem unreasonable to program some percent of the WSMR annual budget for R&D on its own job. What the percent should be, I don't know, but 2%, 5% or 7% seems reasonable. Electronics and A/C firms do better. What kinds of experiments one asks? On some missions WSMR may have enough spare resources so that it can double up on position measurements, i.e., re Figure 1, again, put two equipments at each location b, b', c, c', say. I assume that timing errors would be nearly equal at any single location. The smoothed apparent position data (after averaging) should then indicate something about possible "timing component" of error. If a competent person in design of experiments were to spend 3-6 months at WSMR, it seems reasonable that other experiments with useful treatment combinations could be suggested and suitably designed within WSMR's resource frame work.

With respect to the orthogonal regression line, there is nothing in the literature that I am aware of on sampling theory for the regression coefficient or for predicted points. A general reference I recommend is J. B. Coleman, Annals of Math. Stat. 3, 79 (1932). In 1963, I did some work on the design of a flight program carried out in Arizona. By flight replication, we were able to obtain sampling error information about the orthogonal regression coefficient and, thus, overcome the lack of sampling theory based on an internal estimate of error.

Further, as both Prof. Lieberman and I have pointed out, there are no difficulties in obtaining an analysis of variance including a goodness-of-fit term even though the regression fitted is polynomial or otherwise non-linear, so long as the least squares equations are linear in the unknown parameters to be estimated. For the non-linear least squares equations cases, which might arise from a physical model of the missile flight path, I suggest Prof. Hartley's recent paper in Biometrika, 51, 347 (Dec. 1964).

At IST, we have a quite general purpose regression program which is due to Dr. Wyman Richardson. Also, Robert O. Bennett, Jr. and myself are working on a packaged set of sub-routines which can be used for doing Analysis of Variance type calculations. Perhaps, Dr. Hanson should visit us to get information on these programs. Both programs operate on IBM 7090 within University of Michigan Computer Center Executive System.

No doubt WSMR is studying the application and use of the newer high accuracy oscillators for its timing standards. Could not these "atomic clocks" help resolve some of its "timing error" problems? Any WSMR comment on the use of these oscillators will be of interest to us at IST, since we are studying their employment for networks even more widely distributed than those in the WSMR systems.

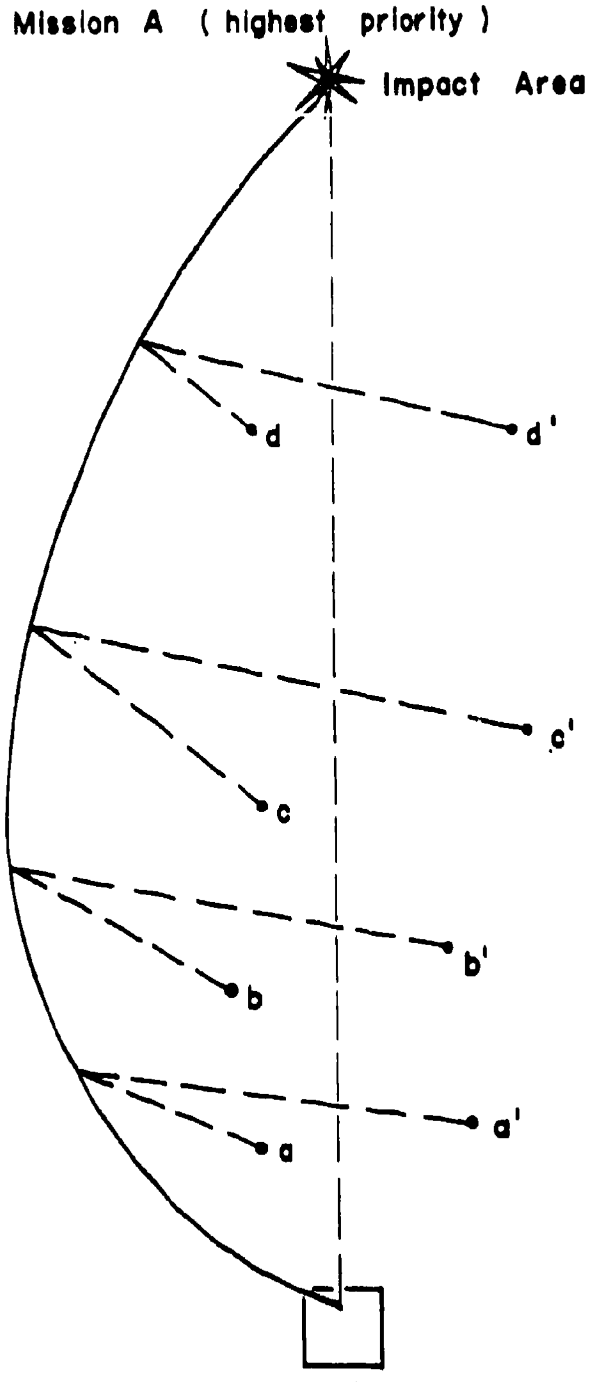


Figure 1
E. H. Jebe

AN EXPERIMENT IN MAKING TECHNICAL DECISIONS USING OPERATIONS RESEARCH AND STATISTICAL METHODS

Andrew H. Jenkins
U. S. Army Missile Command, Directorate of Research and Development
Physical Sciences Laboratory, Redstone Arsenal, Alabama

and

Edwin M. Bartee
University of Alabama in Huntsville,
College of Engineering, Huntsville, Alabama

ABSTRACT. This paper presents a case where decisions are reached and recommendations were made on a multi-disciplined technical research program. The decisions were made on the basis of a technical survey using operations research techniques and statistical methods for evaluation rather than a rigorous technical evaluation of all disciplines. The paper presents the technique used and discusses the practical limitations of the method.

I. INTRODUCTION. The engineer and scientist in government research programs are often required to make decisions and/or recommendations on programs involving advanced technology. Decisions may be required from the individual engineer or a group of engineers. Frequently, the decisions must be made in a minimum of lead time.

The tremendous advances in technology have precipitated a situation where very few research programs are of a single technical discipline. They are usually related either directly or indirectly to other technical disciplines and cannot be treated singularly. A research program, regardless of the number of technical disciplines involved, is an effort to explore and determine the unknown and because of the unknowns is not always conducive to rigorous technical evaluation by an individual or quite often a small group. Certainly, as the number of disciplines increase, the more complex the evaluation becomes.

The engineer, no matter how competent he may be in one discipline, often finds himself making decisions intuitively rather than by rigorous analysis of technical facts. This is so because quite often he does not have the necessary facts, he does not have the time; or he does not have the necessary capability in many disciplines. When the decisions are made intuitively, they are shaded and toned by the engineer's biases, preconceived notions, and past experiences. As the amount of information increases in a multi-disciplined problem so do his vacillations between biases and preconceptions in the process of making a decision. This condition is accentuated where the research program is such that the technical opinions of others must be considered.

Therefore, what is needed is a systematic approach to the problem, consideration of as many technical factors which may affect the decision as possible, and a method of weighting the factors and quantifying the opinions. In other words, a set of rules are determined and followed systematically until a decision can be reached.

The authors were recently involved in a problem of making a decision and recommendations on certain research programs. The purpose of this paper is to present the approach taken and the use of statistics in the decision making process for an actual case. None of the government agencies or research groups are identified except the U. S. Army Missile Command since the information is for government program planning.

II. BACKGROUND. The U. S. Army Missile Command (USAMICOM) is the technical director of a research program being performed by a research group for the U. S. government. This research program was a multi-disciplined program in missile phenomenology involving theory and experimentation in such disciplines as electromagnetics, optics, plasma diagnostics, microwave-plasma interactions, aerothermochemistry, thermodynamics, fluid dynamics, experimental techniques and instrumentation. This program was one of several similar programs of an overall research program.

The group directed by USAMICOM (identified as Establishment 7) proposed the development and utilization of a larger, much improved hypervelocity launcher of projectiles for research purposes. This among other things precipitated a review of overall research effort in missile phenomenology. In view of this, USAMICOM was requested to give recommendations on the following categorical questions:

I. The past and future utilization of Establishment 7.

II.. The need for a large caliber, light gas gun and possible uses in missile phenomenology research.

III. The desirability of building such a gun at some establishment other than 7.

The experimental approach taken by the authors is included except for the coding of all agencies and research groups.

III. THE EXPERIMENT.

A. Design Approach

The purpose of this effort is to provide recommendations in three categories which are of concern to missile phenomenology research programs.

The three categories are as follows:

Category I: The past and future utilization of Establishment 7.

Category II: The need for a large caliber, light-gas gun in missile phenomenology research.

Category III: The desirability of building such a gun at some other establishment.

Due to USAMICOM's close association with past programs and in an effort to carry out this task with minimum bias and maximum objectivity, it was considered appropriate to conduct a technical survey of theoretical and experimental groups associated with such programs. Time limitations permitted only a representative sample of such groups. These groups are known to have knowledge pertinent to all of the above categories.

It was anticipated that a wide variation of data and opinions would be obtained from these groups making orderly, efficient, and unbiased analysis of the survey results difficult. It was decided that a method of analysis based on quantifying of data and opinions must be used. The method selected is the "Case Institute Method of weighting objectives" and is described in Reference 1 in detail.

It was decided to send four engineers as interviewers to visit the selected theoretical and experimental groups. The groups were selected as a representative cross section of those familiar with aeroballistic range techniques and associated research programs, and therefore able to contribute to the resolution of the three categorical problems. The groups were allowed to comment on or off the record to increase responsiveness.

The establishments were visited as shown in Table 1. It can be seen that Interviewer 1 visited Establishments 2, 5, 7, and 11; Interviewer 2 Establishments 3, 4, 9, and 10; Interviewer 3 Establishments 1 and 8; and Interviewer 4 Establishments 6, 12, and 13.

For consistency of the interviews, a master list of questions considered pertinent to the categories was provided to each interviewer and discussed at each establishment. The interviewers recorded a summary of facility data and opinions for use during rating of the factors. Thereby, each interviewer obtained sufficient technical background information upon which he could quantitatively rate ten factors considered pertinent to each category. The ten rating factors for Categories I, II, and III are shown in Tables 2, 3, and 4 respectively. The ten factors were selected as a representative sample which were required to make a systematic evaluation of each category.

The ten factors in Category I were designed to rate Establishment 7 against other establishments. The establishments chosen for 7 to be rated against were 1, 4, 6, 9, 12, and 13. These represented establishments similar to 7 and operated by all government agencies of the Department of Defense, private corporate facilities and an educational institution.

The ten factors in Category III were designed to rate establishments 1, 4, 6, 9, 12, and 13 against 7.

The ten factors in Category II were designed to rate the opinions of both theoretical and/or experimental groups on the need for a large light gas gun.

Each interviewer, after discussion of the factor with the principle investigatory, numerically rated each factor in each category for the establishments visited. These ratings were between 0 and 4. In the selection of a quantitative rating, if the rating was not clearly and easily differentiated from the mean value of 2, the rating was established at that level. This procedure tends to minimize individual bias and enables the survey to approach a truly unbiased conclusion.

B. Factor Rating Criteria

The discussion is confined to the types of information, data and comments obtained for use as a basis for rating the ten factors of each category.

Category I

In Category I the first factor was rated on the basis of the information received on program objectives, types of models required, instruments required, and types of data collected. Also considered was reporting in journals or at symposiums, the opinion of the reporting by other groups, and the degree of success of the program. The rating of the second factor was based on the overall instrumentation capability in flow field visualization, optical radiation, and microwave diagnostic instruments, as well as special instrumentation. The third factor was rated on such criteria as complexity of model shapes, velocity, and data gathering and launching problems. The fourth factor was rated on the basis of type of gun, launch weights, velocities, repeatability, and freedom from malfunction. The fifth factor was rated on the basis of comments of professionals who have had close or personal contact with professionals of Establishment 7. The sixth factor was rated on the basis of the number of available ranges, guns, standard and special instruments, and utilization factor of the facilities. A criteria of minor consideration was estimated capital investment. The seventh factor was rated on a basis of some of the same criteria as factor six plus the ability to initiate programs of widely varying experimental parameters on short notice. The eighth factor was rated on a basis of such things as available space, facility cooperativeness, and facility workloads. Most establishments have existing funded programs planned and limited staff level responsiveness. The ninth factor was rated on relative defense efforts of the establishment. The tenth factor was included on the premise that accomplishments are often proportional to support received.

Category II

In this category an attempt was made during the survey to establish the need for a large caliber gun in missile phenomenology research and to define a large caliber gun. In regard to the large gun proposed reactions varied from "it is feasible" to "it can't be done". Others stated a preference for approaching the possibilities of designing such a gun in small diameter phases, e.g., 2.5 in., 4 in., then perhaps 6 in. It appears from comments obtained that a 3 or 4-inch gun may be the optimum size. A 4-inch gun capable of velocities of 25,000 feet per second would be a size large enough to allow for expansion of the types of experiments which could be performed on an aeroballistic range. A 4-inch gun would also be more easily fabricated, handled, operated, maintained and be capable of a reasonable firing rate. However, definition of a large caliber gun was a secondary issue, the prime factor being the determination of the real need for a large caliber gun. Factors one and two were rated on the basis of the capibility of a large bore gun to expand the types of experiments and measurements that may be effectively executed under simulated conditions. These factors were most heavily weighted in Category II. The concensus is that this is the foremost justification for a large gun. However, those who expressed this opinion could suggest few programs but some examples are: (1) launching complex geometrical shapes, (2) blast vulnerability studies, and (3) on-board-model telemetry measurements. The fact that new programs cannot currently be suggested does not exclude many suggestions when such a device is available. New types of measurements will be developed in parallel with new types of experiments with larger models. Factor three was only a rating of the opinions of the interviewers on the need for a large bore gun. These opinions vary strongly from favor to disfavor and are reflected in column 3 of Table 6. The X_i column reflects the composite of all factors for each establishment. Factors four and five sought to determine if, in the opinions of others, larger models would improve the thresholds of measurements made by current instruments at a given simulated altitude or provide equal thresholds at a higher simulated altitude. Some respondents indicated that, on a quantitative analysis, significant improvements would not be obtained. Other respondents feel that larger guns would improve thresholds and resolution significantly, especially in optical measurements but not on microwave measurements. Respondents generally agree that simulated data can be more easily utilized in theoretical modeling and computations than in full scale. Some respondents did not feel that this was particularly true to the point of justifying a larger gun than is nominally used, e.g., 1-1/2 inch gun. Some of the respondents to factor seven could not comment, especially if this factor is viewed from the standpoint of a large gun reliability, capital cost, and useful life. Other respondents, even in view of these criteria, feel that more usable data can be obtained at less expense on ballistic ranges than under full scale conditions. The overall response to factors eight and nine varied from neutral on eight to slightly negative on nine. One respondent described quantitatively that examinations of scaling limit increases show that from 10,000 to 20,000 feet of altitude may be obtained by a fivefold increase in size for binary scaling of wake electron densities. Also, only a 20 percent increase in wake lengths that could be scaled would be obtained.

Factor ten was included, at very low weight, merely to emphasize this advantage of ballistic range data gathering when contrasted to full-scale data. While full scale does represent the real case, for purposes of study repeatability is highly desirable. In view of the fact that such diverse opinions and wide variations in responses were obtained, the analysis was made easier by use of the Case Institute Method approach.

Category III

This category assumes that a large caliber gun is needed. It is, therefore, important to determine the best places that such a device should be installed and operated.

The installation of a large caliber gun, which would be heavy, long and cumbersome, would require that the establishment have the necessary heavy moving equipment, transfer locations, and housing to properly operate and maintain it. Factor one considers these present capabilities without new construction.

The installation of a large caliber gun would necessitate increasing the number of persons required to operate and maintain it in a data-gathering program. The operation and maintenance necessitates handling and storage of large amounts of munitions and H_2 of He gas, fabrication of larger models and sabots, telementry packages, and other incidental items required to effectively pursue such a program. In establishments where programs are presently funded to accomplish a mission, such a large program would perhaps overload their present capability. In view of this, the desire of an establishment to participate in a program utilizing a large caliber gun is important. This, in turn, is a function of their interests in the experimental programs to be pursued with a large caliber gun.

The ratio of chamber diameter to model diameter for good compatibility has been estimated between 20 and 30. Therefore, a 5-inch model would require (taking the average) a chamber of 125 inches (approximately 10 feet). Some establishments would require additional chambers for 4-or 5-inch models if this ratio is accepted. Therefore, some establishments may have the desire and interest but not adequate facility and personnel capability or range compatibility.

Other important considerations are the attitude of the establishment to the full - or part-time participation of contractors in data gathering on the range and the participation of contractors intermittently to obtain a few data points of a specific interest. This requires that a certain amount of space on the range for instrumentation be available. Quite often the data can be gathered on shots of opportunity.

In anticipation of research contractor participation, the accessibility of the facility is important to maximum utilization of the facility. In conjunction with this will be the ability to control and direct programs and program changes. Program orientation is also important. It may be desired to pursue a basic long term program with short specific tasks overlaid, the results of which may on occasion change the basic program orientation.

Finally, the cost of a large gun is considered. The overall opinion is that the costs will probably not differ greatly between government establishments. However, an industrial or corporate facility may be more economical than the government facilities.

C. Numerical Analysis

The Case Institute Method of weighting objectives (1) was selected for use in weighting the factors and quantifying the respondent's comments and opinions.

The lack of a universal standard deviation and the small sample dictated the use of Student's 't' distribution for test of significance of the results.

In the Case Institute Method, the ten factors are weighted as follows:

1. One factor in each category is rated most important and given a value of 1.00. Each of the other nine factors are then rated between 0 and 1.00 according to its relatively judged importance.

2. After all factors in a category are rated, the most important factor is compared to the other nine collectively as to importance in the category. If it is judged more important than the other nine collectively, the value of 1.00 first assigned is changed to a value larger than the sum

of the other nine values. If the most important factor is considered to be of the same importance as the other nine, the value for the most important factor should be equal to the sum of the other nine factors. If it is considered to be of less importance than all the other nine, then its value is adjusted to some value less than the sum of the other nine.

3. The most important factor and its weight are established. Next, the second most important factor is compared to the remaining eight. Its weight is established in the same manner as described in 2 above. When the factor's weight is established, the procedure continues to the third, fourth, etc. most important factor until all 10 factors are weighted.

4. This procedure is followed for all three categories.

A composite of the weighting for all categories is shown in Table 8 in order of descending weight. The factors for all three categories can be seen in Tables 2, 3, and 4.

The method of rating the factors was to use the five discrete numerical levels 0, 1, 2, 3, 4. In Category I, each establishment contrasted with 7 was set at level 2 and 7 rated below or above at 0 or 1 and 3 or 4 respectively. In Category II, a neutral position on each factor by the respondent was set at 2 and the degree of disfavor or favor of Category II at 0 or 1 and 3 or 4 respectively. In Category III, 7 was set at 2 for each factor, and each establishment was rated below or above with 0 or 1 and 3 or 4 respectively.

The rating established for each factor in each category was multiplied by the corresponding factor weight and is recorded in Tables 5, 6, and 7. The values are summed for each establishment. In order to normalize the range of response for each establishment in each category, the following equation is used:

$$X_i = \frac{\Sigma(\text{factor wt} \times \text{factor rating}) - 2 \times \Sigma(\text{factor wt})}{2 \times \Sigma(\text{factor wt})}$$

For Category I:

$$X_i = \frac{\Sigma(w_f \times R_f) - 2 \times 3.12}{2 \times 3.12} .$$

For Category II:

$$X_i = \frac{\Sigma(W_f \times R_f) - 2 \times 3.75}{2 \times 3.75}.$$

For Category III:

$$X_i = \frac{\Sigma(W_f \times R_f) - 2 \times 1.49}{2 \times 1.49}.$$

For all categories:

The limits for each X_i in all categories becomes

for $R_f = 0 \quad X_i = -1$

$R_f = 4 \quad X_i = +1$

$R_f = 2 \quad X_i = 0 = \bar{X}'$ (hypothesis value)

$$\bar{X} = \frac{\Sigma X_i}{N}$$

where

X_i = establishment computed response

W_f = factor weight

R_f = factor rating

N = number of establishments.

The sample deviation (S) for each category is

$$S = \left[\frac{\Sigma(X_i - \bar{X})^2}{N - 1}\right]^{1/2}.$$

Student's 't' test for significance is

$$t = \frac{\overline{X} - X'}{S/\sqrt{N}} \quad .$$

Using the data from Tables 5, 6, and 7 for Categories I, II, and III respectively, we calculate the sample standard deviations:

$$S_I = \left[\frac{0.1115}{5}\right]^{\frac{1}{2}} = 0.149$$

$$S_{II} = \left[\frac{1.71}{12}\right]^{\frac{1}{2}} = 0.378$$

$$S_{III} = \left[\frac{.328}{5}\right]^{\frac{1}{2}} = 0.256 \; .$$

Before the 't' tests are made, a confidence level of 70 percent is set, which is considered appropriate for research (i.e., risk of first kind* α = .30) and the following hypotheses are made on each category:

Category I: There is no significant difference in utilization of 7 and other establishments (i.e., $\mu = 0$).

Category II: There is no significant need for a larger caliber gun in the missile phenomenology research program (i.e., $\mu = 0$).

Category III: There is no significant difference between establishments where a large gun should be built (i.e., $\mu = 0$).

The 't' tests are computed for each category:

$$t_I = \frac{.067 - 0}{.149/\sqrt{6}} = 1.105$$

*The risk of rejecting a hypothesis when it is true. Also called the producer's risk.

$$t_{II} = \frac{.148 - 0}{.378/\sqrt{13}} = 1.41$$

$$t_{III} = \frac{.214 - 0}{.256/\sqrt{6}} = 2.05 \; .$$

The computed values are compared with Student's 't' table values as shown below:

Computed Value for Categories	Degrees of Freedom	Table Value: Percentile Point 70	80	90	95
I = 1.105	5	0.56	0.92	1.48	2.01
II = 1.41	12	0.54	0.87	1.36	1.78
III = 2.05	5	0.56	0.92	1.48	2.01

It can be seen that the tests for all three categories are significant at the original level of confidence of 70 percent which is considered appropriate for advanced research projects. As the tests are significant at this level (the computed value is greater than the table value), all three hypotheses are rejected. The highest level at which the tests are significant and the hypotheses rejected are Category I ≈ 80 percent, Category II ≈ 90 percent, Category III ≈ 95 percent.

If, however, it is considered that the level of confidence should be 95 percent, then the tests for Categories I and II are not significant and the hypotheses accepted. Category III is still significant but inconsequential. For the purposes of decision making in this type research and development programs, the 95-percent level of confidence is considered excessively high by the investigator.

D. Summary and Conclusions

This task is one which is highly complex. Many technical areas of an advanced nature are involved. An honest and sincere effort has been made to reach an unbiased and technically sound solution. The groups queried have provided comments which are spontaneous and which instinctively draw on years of technical experience pertinent to the problem. Therefore, considerable intellectual attention and technical capability have been concentrated on the three categories. It is not supposed or proposed that every facet has been considered and explored, nor has a rigorously technical approach been used as this would be a formidable task. However, a representative sample of the foremost factors has been considered, and the technical analysis was performed mentally by the respondents.

A systematic approach tothe analysis of a highly complex problem has been used as shown in the numerical analysis. The importance of this approach is the capability to make a decision in the realm of uncertainty and random variation.

Review of the results of the ratings of Category I presented in Table 5 shows that (considering all factors) 7 rates below 9 at -0.178 (or 17.8%) and slightly above all others with 4 and 13 closest with a +0.008 or (0.80%) and +0.024 (or 2.4%), respectively. Comparing 7 to all other establishments for all factors 7 rated at +0.0665 (6.65%) which is significant when compared to the sample standard deviation by the 't' test.

Review of the results of the ratings of all factors for Category II, presented in Table 6, shows that 2 was strongly not in favor of a large gun by a value of 0.701, followed by 10 and 8. Seven was strongly in favor of a large gun with a value of +0.948, followed by 5, 4, 6, and 9. Twelve and 1 were slightly in favor, with values of +0.040 and +0.041, respectively. On an overall comparison of all factors and all establishments there was a favorable response of +0.148 (14.8%). This evaluation does not include the exact launch tube diameters.

Review of the results of rating the factors in Category III, presented in Table 7, shows that 9 with a value of -0.0067 and 13 with a value of -0.0436 compare closest with 7 as the place to build a large gun. Twelve was least favorable with a value of -0.711.

Therefore, on the basis of the analysis of the overall results shown and within the limits of this study the following conclusions were drawn:

Category I

There is an apparent difference in the overall usefulness of 7 compared to other facilities. There is a significantly positive opinion that 7 may be effectively utilized in the future.

Category II

There is an apparent need for a large caliber gun in the missile phenomenology research program. There is a significantly positive opinion that such a device is needed presently and in the future.

Category III

There is an apparent difference between establishments where a large gun could be built and utilized. Establishment 7 is a foremost contender as a desirable establishment for developing the large caliber gun. Recommendations on program continuation together with suggested experiments were made based on these conclusions.

IV. DISCUSSION. The preceding case is a real-world example of how operation research and statistical methods can be utilized to assist in the process of making technical decisions. The particular features of this approach are:

1. An inter-disciplinary team is utilized to bring a variety of technical viewpoints to bear upon the problem.

2. The results of such a team effort are quantified to make it possible for analysis to be made at optimum objectivity.

3. Statistical techniques are applied to evaluate the quantified results.

The key feature of such methods is the concept of risk and probabilistic conditions. Such an approach is particularly useful in the realm of decision-making since the risks are often great and the probabilistic environment is every present. Under such conditions there is no opportunity for drawing a definite conclusion. A decision can only be made at a given level of confidence. The risk of a decision being wrong becomes a calculated part of the problem.

The use of quantitative methods for expressing the results of the experiment can often lead to a process of over interpretation of results often to the neglect of sound technical judgement. Obviously, the decision cannot be made solely with such methods. At best, the decision-maker can be fortified with certain analyses of the experimental results that will provide a statement of the risk he would take if he should make a decision in one direction or another. Such factual data can often be provided with a minimum of bias from lower echelons so that the decision-maker can benefit from it while exercising his best judgement in the problem.

The experiment was basically concerned with the determination of technical facts that existed within each of the installations. To obtain such facts required us to go through several "bias filters" such as:

1. The ability and willingness of the installation representative (the interviewee) to state the true facts that exist in his group as free of bias and inaccuracies as possible.

2. The ability of the interviewer to gather and transmit the data to the investigator with a minimum of his own personal bias involved.

3. The ability of the investigator to compile the final data as free of his own personal bias as possible.

To accomplish the above purposes is obviously no easy task under any circumstances. The problem was faced in the investigation by utilizing these basic techniques:

1. A multiple of closely related questions were used to conduct the interviewers with each installation representative.

2. The interviewee bias was observed and evaluated by the interviewer in each case.

3. The data was transmitted to the investigator and a concerted attempt was made on the part of the investigator to balance the bias of the interviewer and interviewee through the conduct of an extensive "debriefing" procedure.

4. The bias of the investigator was controlled by both the influence of the interviewer in the debriefing sessions and the systematized method of quantifying the results.

Obviously, the efforts just described could never hope to eliminate all bias and inconsistencies. The recognition of this fact leads us to evaluate the final results with techniques that have been developed for such situations.

We have, in effect, produced quantified results within an enviroment of uncertainty. Such uncertainty is made up of two basic elements. That is, the observed differences in results between installations can be attributed to:

1. Differences that are explained by residual errors and biases that still remain in the experiment in spite of the procedures that were established to eliminate them.

2. Differences that are explained by real effects of the installation on the category in question as far as the study can determine.

The test of hypothesis used in the analysis served to partition these two basic causes of observed differences. To say that a resulting effect was signigicant is to say that, within the limits of this investigation, the observed differences between the selected installations cannot be attributed merely to experimental error. The conclusion is therefore drawn that a real difference exists and a positive conclusion is therefore drawn. It is important to note that for each conclusion there is a comparable level of confidence. Within the realm of an environment of uncertainty all conclusions or decisions must carry this element of risk.

V. REFERENCE

1. Churchman, Ackoff, and Arnoff, "Introduction to Operation Research," John Wiley and Son, New York, New York.

TABLE 1

Establishments Visited by Interviewers

Establishment	Interviewer			
	1	2	3	4
1			X	
2	X			
3		X		
4		X		
5	X			
6				X
7	X			
8			X	
9		X		
10		X		
11	X			
12				X
13				X

TABLE 2

Establishment Nr. 7 Utilization Evaluation Factors

Category I

Factors are listed in descending order of established weights. Each factor rated with 2 representing each establishment against which Establishment 7 is rated. The rating levels are chosen by this interviewer and the chairman of the survey committee.

1. How do 7's past program results compare to other establishments?
2. How does 7's past instrumentation rate in comparison to other establishments?
3. How did 7's program rate with other ranges in degree of difficulty to perform?
4. How does 7's past gun performance rate in comparison to other ranges?
5. How do 7's professionals compare with professionals of other ranges?
6. How does 7's past facility development rate in comparison to other ranges?
7. How does 7's utility as a data gathering facility in future compare with other ranges?
8. How does future possibility of contractors participation on ranges at 7 compare to other establishments?
9. How strong is 7's desire to continue participation in missile phenomenology research compared with other ranges?
10. How does 7's past funding compare to other range programs?

TABLE 3

Large Bore Gun Evaluation Factors

Category II

Factors are listed in descending order of established weights. Each factor rated at levels between 0, 1, 2, 3, and 4 on basis of data and opinions gathered with 2 representing neutral opinion. The rating levels are chosen by the interviewer and the chairman of the survey committee.

1. Will they expand the types of experiment that may be effectively executed under simulated conditions?

2. Will they open avenues of new types of measurements?

3. What is opinion of others doing theoretical work on need for large bore guns?

4. Will they increase observables levels at higher simulated altitudes significantly?

5. Will larger bore guns improve reliability and confidence in range measurements?

6. What is opinion of others on the value of simulated data vs full scale for utilization in theoretical modeling and computations?

7. How does cost of usable ballistic range data gathering compare with usable full scale data gathering?

8. Will they contribute significantly to scaling between theory and full scale?

9. Will they contribute significantly to the establishment of binary scaling limits?

10. What is the opinion of ballistic range data gathering capability from standpoint of repeatability?

TABLE 4

Large Bore Gun Location Evaluation Factors

Category III

Factors are listed in descending order of established weights. Each factor rated at levels between 0, 1, 2, 3, and 4 on basis of data and opinions gathered with 2 representing Establishment 7 against which each establishment is evaluated. The rating levels are chosen by the interviewer and the chairman of the survey committee.

1. To what degree are other establishments able to accommodate a large gun from standpoint of housing, operating, and maintenance without facility construction relative to 7?

2. What was capability of other establishments for taking on additional range measurements programs relative to 7?

3. How strongly do other establishments indicate they want to build a large bore gun relative to 7?

4. What was interest of other establishments in taking additional programs relative to 7?

5. To what degree is their present range chamber diameter compatible with large models relative to 7?

6. What is attitude of other establishments toward contractor participation in data gathering on their ranges relative to 7?

7. Is space presently more available on their ranges for contractors' utilization relative to 7?

8. How does accessibility of other establishments compare to 7?

9. How does the ability to control programs at other establishments compare to 7?

10. How will cost of large gun development at other establishments compare to 7?

TABLE 5. CATEGORY I FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.58	0.50	0.49	0.40	0.28	0.24	0.20	0.18	0.15	0.10	3.12	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	2	2	3	3	2	2	2	3	3	2		
	Factor Wt x Level	1.16	1.0	1.47	1.20	0.56	0.72	0.40	0.54	0.45	0.20	7.70	+0.234
4	Factor Level (R_f)	1	2	3	1	2	3	3	2	2	3		
	Factor Wt x Level	0.58	1.00	1.47	0.40	0.56	0.72	0.60	0.36	0.30	0.30	6.29	+0.008
6	Factor Level (R_f)	2	3	2	2	2	3	2	2	2	2		
	Factor Wt x Level	1.16	1.50	0.98	0.80	0.56	0.72	0.40	0.36	0.30	0.20	6.98	+0.119
9	Factor Level (R_f)	1	2	2	0	1	3	2	4	1	3		
	Factor Wt x Level	0.58	1.00	0.98	0	0.28	0.72	0.40	0.72	0.15	0.30	5.13	-0.178
12	Factor Level (R_f)	1	2	3	3	2	4	4	4	1	0		
	Factor Wt x Level	0.58	1.00	1.47	1.20	0.56	0.96	0.80	0.72	0.15	0	7.44	+0.192
13	Factor Level (R_f)	2	2	2	2	2	2	2	2	3	2		
	Factor Wt x Level	1.16	1.00	0.98	0.80	0.56	0.48	0.40	0.36	0.45	0.20	6.39	+0.024
ΣX_i													+0.399
N													6
$\bar{X} = \frac{\Sigma X_i}{N}$													0.0665
$\bar{X}' = 0$ (be design)													0

TABLE 6. CATEGORY II FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	3	3	1	0	1	3	2	1	2	4		
	Factor Wt x Level	2.76	2.22	0.56	0	0.29	0.87	0.36	0.17	0.22	0.36	7.81	+0.041
2	Factor Level (R_f)	1	0	0	1	0	1	2	0	0	3		
	Factor Wt x Level	0.92	0	0	0.40	0	0.29	0.36	0	0	0.27	2.24	-0.701
3	Factor Level (R_f)	3	3	1	2	2	1	3	1	3	2		
	Factor Wt x Level	2.76	2.22	0.56	0.80	0.58	0.29	0.54	0.17	0.33	0.18	8.43	+0.124
4	Factor Level (R_f)	3	3	3	1	2	3	3	3	3	3		
	Factor Wt x Level	2.76	2.22	1.68	0.40	0.58	0.87	0.54	0.51	0.33	0.27	10.16	+0.355
5	Factor Level (R_f)	3	3	4	3	4	3	2	3	1	4		
	Factor Wt x Level	2.76	2.22	2.24	1.20	1.16	0.87	0.36	0.51	0.11	0.36	11.79	+0.572
6	Factor Level (R_f)	2	3	3	2	3	3	2	3	2	2		
	Factor Wt x Level	1.84	2.22	1.68	0.80	0.87	0.87	0.36	0.51	0.22	0.18	9.55	+0.273
7	Factor Level (R_f)	4	4	4	4	4	4	4	3	2	4		
	Factor Wt x Level	3.68	2.96	2.24	1.60	1.16	1.16	0.72	0.51	0.22	0.36	14.61	+0.948
8	Factor Level (R_f)	3	1	1	0	0	3	4	1	2	4		
	Factor Wt x Level	2.76	0.74	0.56	0	0	0.87	0.72	0.17	0.22	0.36	6.40	-0.146

TABLE 6. CATEGORY II FACTOR RATING (Concluded)

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75	
Establishment No.												Sum of Wt x Level	
9	Factor Level (R_f)	3	3	1	2	3	3	3	1	1	3		
	Factor Wt x Level	2.76	2.22	0.56	0.80	0.87	0.87	0.54	0.17	0.11	0.27	9.17	+0.223
10	Factor Level (R_f)	2	1	1	2	2	1	3	3	3	2		
	Factor Wt x Level	1.84	0.74	0.56	0.80	0.58	0.29	0.54	0.51	0.33	0.18	6.37	-0.151
11	Factor Level (R_f)	2	2	3	2	2	3	4	2	1	4		
	Factor Wt x Level	1.84	1.48	1.68	0.80	0.58	0.87	0.72	0.34	0.11	0.36	8.78	+0.171
12	Factor Level (R_f)	2	2	1	3	2	3	2	3	2	2		
	Factor Wt x Level	1.84	1.48	0.56	1.20	0.58	0.87	0.36	0.51	0.22	0.18	7.80	+0.040
13	Factor Level (R_f)	2	2	3	2	3	3	2	3	2	2		
	Factor Wt x Level	1.84	1.48	1.68	0.80	0.87	0.87	0.36	0.51	0.22	0.18	8.81	+0.175
ΣX_i													+1.924
N													13
$\bar{X} = \frac{\Sigma X_i}{N}$													+0.148
$\bar{X}' = 0$ (by design)													0

TABLE 7. CATEGORY III FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.33	0.22	0.21	0.20	0.14	0.12	0.09	0.08	0.06	0.04	1.49	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	2	1	0	2	2	3	2	3	3	2		
	Factor Wt x Level	0.66	0.22	0	0.40	0.28	0.36	0.18	0.24	0.18	0.08	2.60	-0.128
4	Factor Level (R_f)	2	1	1	1	2	2	2	2	2	2		
	Factor Wt x Level	0.66	0.22	0.21	0.20	0.28	0.24	0.18	0.16	0.12	0.08	2.35	-0.211
6	Factor Level (R_f)	2	2	1	1	2	1	1	3	2	2		
	Factor Wt x Level	0.66	0.44	0.21	0.20	0.28	0.12	0.09	0.24	0.12	0.08	2.44	-0.181
9	Factor Level (R_f)	2	2	3	3	0	1	1	3	1	3		
	Factor Wt x Level	0.66	0.44	0.63	0.60	0	0.12	0.09	0.24	0.06	0.12	2.96	-0.0067
12	Factor Level (R_f)	0	1	0	1	0	0	0	3	2	2		
	Factor Wt x Level	0	0.22	0	0.20	0	0	0	0.24	0.12	0.08	0.86	-0.711
13	Factor Level (R_f)	2	2	1	2	2	2	2	3	2	2		
	Factor Wt x Level	0.66	0.44	0.21	0.40	0.28	0.24	0.18	0.24	0.12	0.08	2.85	-0.0436
ΣX_i													-1.282
N													6
$\bar{X} = \frac{\Sigma X_i}{N}$													-0.214
$\bar{X}' = 0$ (be design)													0

TABLE 8

Factor Weights by Category

Factor No.	1	2	3	4	5	6	7	8	9	10	Σ
Category I	0.58	0.50	0.49	0.40	0.28	0.24	0.20	0.18	0.15	0.10	3.12
Category II	0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75
Category III	0.33	0.22	0.21	0.20	0.14	0.12	0.09	0.08	0.06	0.04	1.49

IMPROVEMENT CURVES: PRINCIPLES AND PRACTICES

Jerome H. N. Selman, Stevens Institute of Technology,
Rep. the U. S. Army Munitions Command, Dover, N. J.

To build the 1000th B-29 Aircraft took only 3% of the time required to build the first. To build your first window screen or dog house will take you more time than each succeeding one--unless you are a professional window screen or dog-house maker. This feeling is intuitive. The estimation of time reduction for each succeeding item, based upon judgment and experience, is attributed to a human "learning" effect. Mathematically, the way to express this condition would be to use a reduction-type function: A straight line equation with constant negative slope for a constant linear reduction of cost with quantity; a hyperbolic equation with negative exponent for rapid initial reduction of cost with quantity, then slowing down to a limit; more complex equations which are designed to reflect the phases of the specific learning situation.

Models of the cost-quantity relationship, as a predictive technique, came into general use in the airframe industry during World War II after their development in the 1930's. T. P. Wright's pathfinding article* hyperbolically related the average direct man-hour cost to the number of airframes produced. Others have modified Wright's model to show the inverse relationship between the direct labor hours per unit versus quantity produced; this latter formulation being known as the Unit (Improvement) Curve. A linear improvement curve having linear component curves implies that the rate of learning is the same; intuitively, again, the assumption of constant learning rate in all operations is open to question. Wright was of the opinion that different rates of learning are found in the airframe manufacturing process, but he did not inquire into the implications.

Studies in the then-new airframe industry for sub-sonic, reciprocating engine, electrically simple aircraft indicated that although the precentage slope of the improvement curve varied, for every doubling of successive quantities of aircraft, the percentage value was a constant percentage of the unit value of the quantity immediately prior to doubling. The percentage reduction was approximately 80%. This meant that each time the quantity was doubled, the man-hours required to make that designated aircraft was 80% of the man-hours required immediately prior to doubling. Plotting the improvement curve on logarithmic grids gives a "straight line curve",

---

*T. P. Wright, "Factors Affecting the Cost of Airplanes," Journal of the Aeronautical Sciences, Vol. 3, February, 1936, pp. 122-128.

as the grids are so scaled that the interval between doubled quantities are equal; i.e., the distance between one and two is the same as the distance between two and four, or four and eight, or eight and sixteen, etc. Of course, the linear hypothesis should be discarded whenever the unit curves of man-hours and cost depart significantly from linearity--"significant departure" being determined from the slopes of the parallel linear component curves, based on the error permissible in the problem in hand.

Improvement curves are expressed in terms of percentages, such as 80% Curve, 90% Curve, 92% Curve, etc. The percentage figure referring to the fact that man-hours tend to decrease by a definite amount each time the quantity produced is doubled. By correlation and other statistical techniques it has been shown that a graph of the actual performance data (cost, as inferred by man-hours per unit versus quantity produced, or tasks accomplished) may be approximated by a hyperbolic function of the form $y=ax^b$, with a relatively high degree of significance. The fundamental hyperbolic shape is postulated rather than tested (for linearity on double-log scales versus some alternate non-linear functional form for comparison), as a descriptive device for accumulated data. In Improvement Curve terminology, y, is in direct man-hour cost, a, is the direct man-hour cost for "unit Number one", and b defines the "slope" of the curve---"slope" being the ratio of the unit (or average) man-hour cost at two cumulative outputs that differ by a factor of two (2), so that the slope is 2^b. Wright's empirical data on unspecified aircraft yielded a "b"-value of -.322, giving the popular "80% Curve". On arithmetic grid the 80% Curve with a unit one cost of 1000 man-hours is shown in Figure 1, the equation being $y = 1000x^{-.322}$.

To illustrate the mechanics of constructing improvement curves, the 80% Curve will be done in three parts; as shown in Figure 2:

<u>The Unit Time Line</u>: Given a value for any unit P and the slope of the Improvement Curve in percentage form, draw a line from point P through a point X so that it will be twice the unit number of P, i.e., P equals twice X; and the value of X will be the value of P, multiplied by the percent slope of the curve. Equation: $y_i = ax_i^b$ for unit curve.

<u>The Average Time Line Per Cumulative Unit</u>: The Cumulative Average line is drawn in two steps:

1. The Asymptote. The Cumulative average line approaches a straight line which is parallel to (after about the 15th unit) and higher than the unit line. To construct the asymptote, obtain the "b" for the improvement curve in question. Draw the asymptote parallel to the unit line so that the values of all points on the asymptote are equal to 1/(1+b) times the values on the unit line. For the 80% Curve, the conversion factor for (1+b) is 0.687, as given on Table I, giving each point on the asymptote a value of 1.475 the corresponding value on the unit line. Equation:

$$\bar{\bar{y}} = \frac{a\,N^b}{1+b} \quad .$$

2. The Cumulative Average Line. As an approximation for values between 2 and about 15, the cumulative average values for any unit X is approximately equal to the value shown on the asymptote for, X+3. That is, the average cost of the 4th unit is approximately equal to the value of the asymptote at unit 7. For practical purposes, the average line for units 16 and above may be considered to equal the values of the asymptote. Equation:

$$\bar{\bar{y}} = a \sum_{i=1}^{n} \frac{x_i^b}{N} \quad .$$

<u>The Total Line</u>: Draw a line from the value of unit number one to a point at, say, unit number 10, which has a value equal to 10 times the cumulative average value of unit number 10. It is logical that the total time for the first ten units is equal to 10 times the average time (cost) of the first ten units. Equation:

$$Y = a \sum_{i=\frac{1}{2}}^{N+\frac{1}{2}} x_i^b \quad ;$$

the corresponding asymptote is N times the cumulative average asymptote, just as the Total line is N times the cumulative average line.

Improvement Curves have been utilized in the Aerospace Industries for Cost estimates, scheduling, efficiency comparisons, procurement and subcontracting, facilities planning, personnel planning, long-range forecasting, etc., and was proposed for various industries such as home appliances, electronics, construction, machine shops, ship building, etc. The accuracy of the Improvement Curve function as an estimating device is dependent upon a number of factors, including:

Accuracy of Basic Estimate
Choice of the Improvement Rate exponent "b"
Non-linear elements is the real world
Changes in the output rate
Design Changes in product
Influx of "green" manpower
Exit of skilled manpower.

The basic tenet of Improvement Curve philosophy is where there is life (people) there can be learning, the more man-oriented the work, the more learning potential possible. Figure 3 illustrates the generally accepted improvement curve percentages for various man-machines mixes: 75% Man-25% Machine for the 80% Improvement Curve; 50% Man-50% Machine for about 85% Improvement Curves; 25% Man-75% Machine for the 90% Improvement Curves, etc.

Munitions Command Regulation 715-1 requires thorough justification where "program costs are not reduced in accordance with expected learning curve costing." The techniques of the learning or improvement curves can set realistic management goals for setting expected rates of improvement in reducing operating expenses in the Army "Five-Year Cost Reduction Program".

Operations develop trends that are characteristic of themselves. Projecting such established trends is more valid than assuming level performance, or no learning effect. The Improvement Curve function which has remained parochial to the aerospace industries has been presented with the same motive as the rooster who showed his hen an ostrich egg--"It's not that I'm complaining, it's just that I'd like you to see what others are doing!"

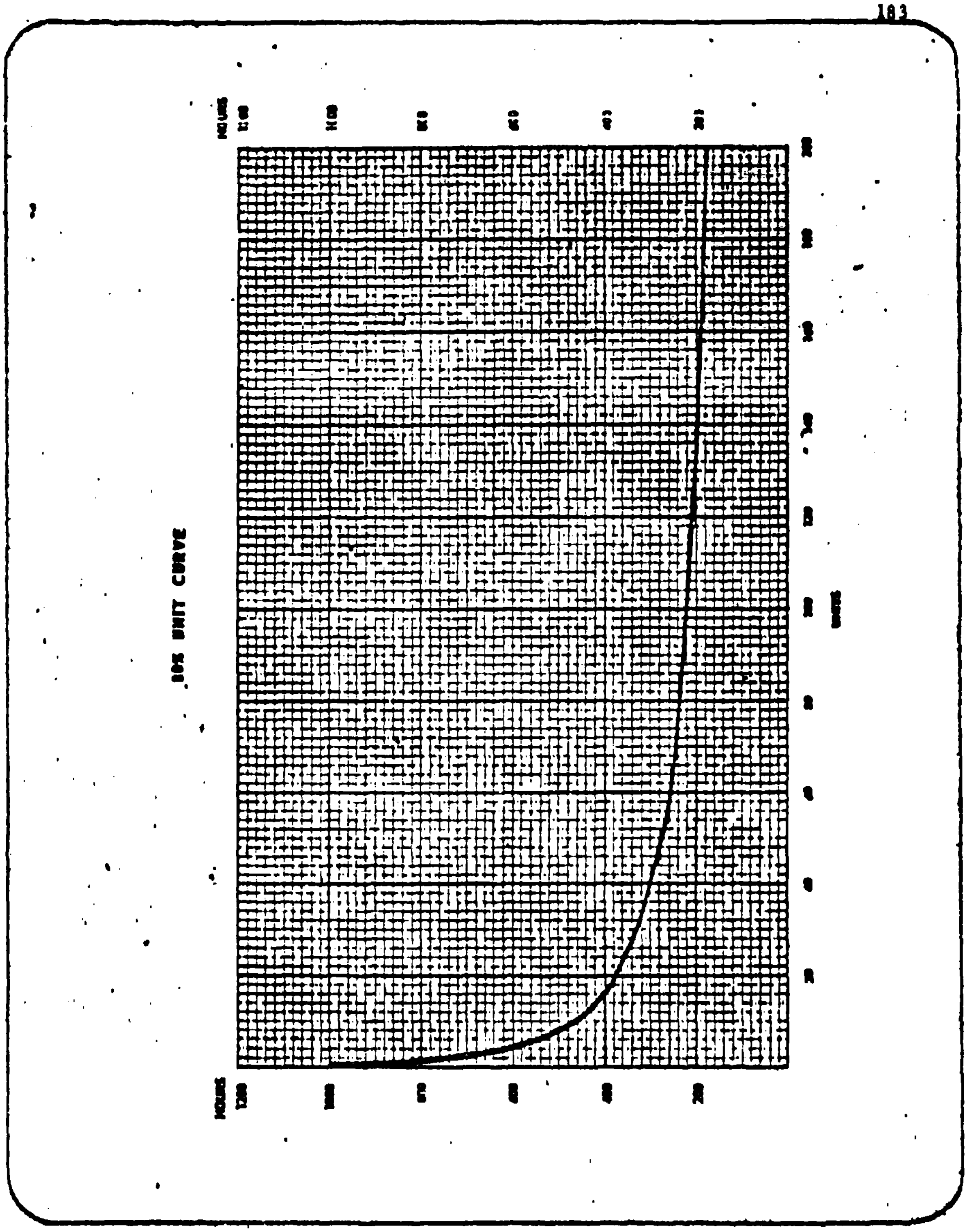


Figure 1

GRAPHICAL CONSTRUCTION OF IMPROVEMENT CURVES
80% IMPROVEMENT

("SLOPE" b = tan^{-1} 17°51' = -0.3219)

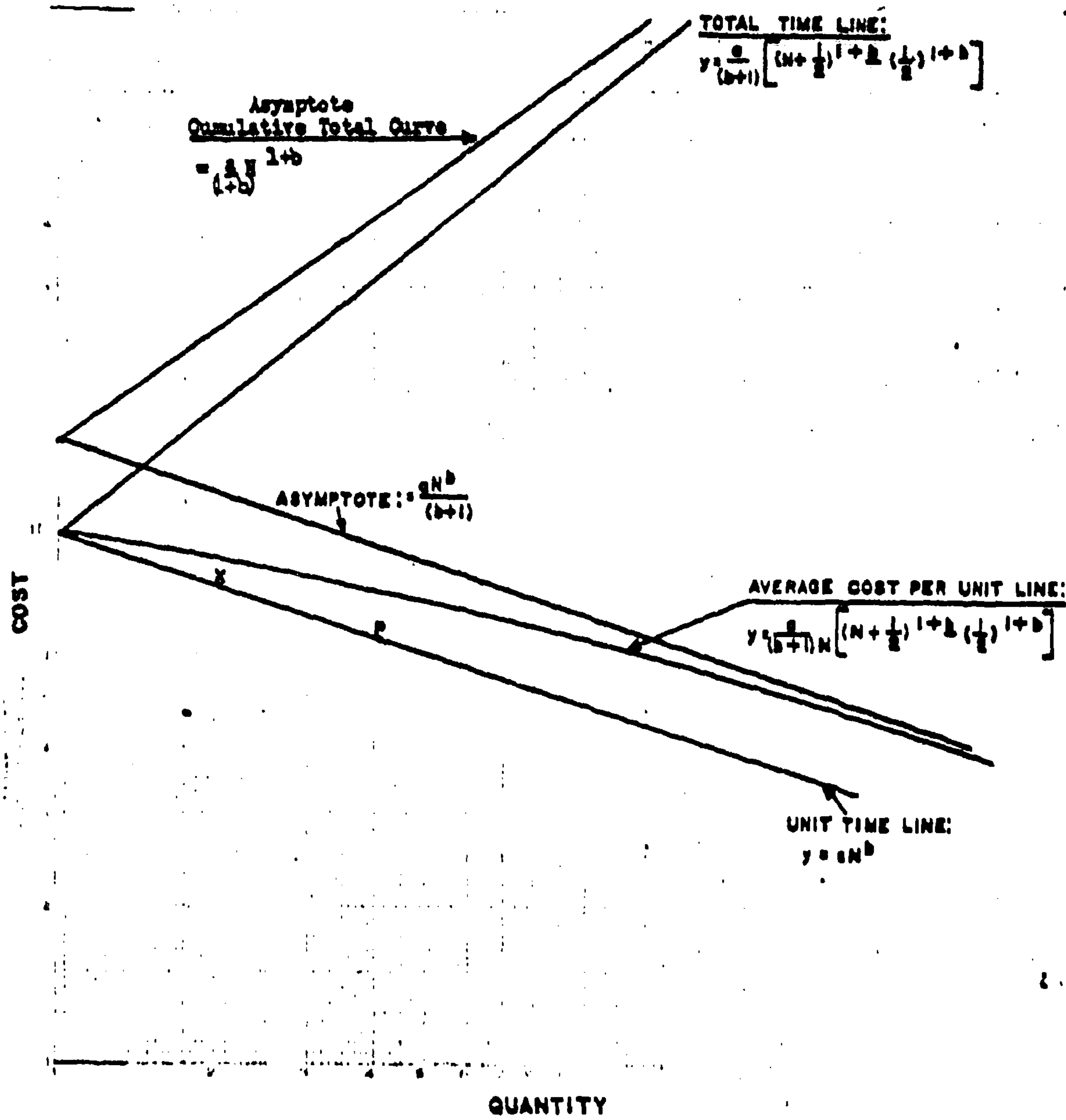


Figure 2

IMPROVEMENT CURVE FACTORS

IMPROVEMENT CURVE, %	b=TANGENT θ	(1+b) CONVERSION FACTOR	θ
50	-1.000	----	45°
55	- .863	.137	40° 47'
60	- .737	.263	36° 24'
65	- .622	.378	31° 52'
70	- .515	.485	27° 14'
75	- .415	.585	22° 32'
80	- .322	.678	17° 51'
81	- .304	.696	
82	- .286	.714	
83	- .269	.731	
84	- .252	.748	
85	- .235	.765	13° 12'
86	- .218	.782	
87	- .201	.799	
88	- .184	.816	
89	- .168	.832	
90	- .152	.848	8° 38'
91	- .136	.864	
92	- .120	.880	
93	- .105	.895	
94	- .089	.911	
95	- .074	.926	4° 14'
99	- .015	.985	0° 50'

Table I

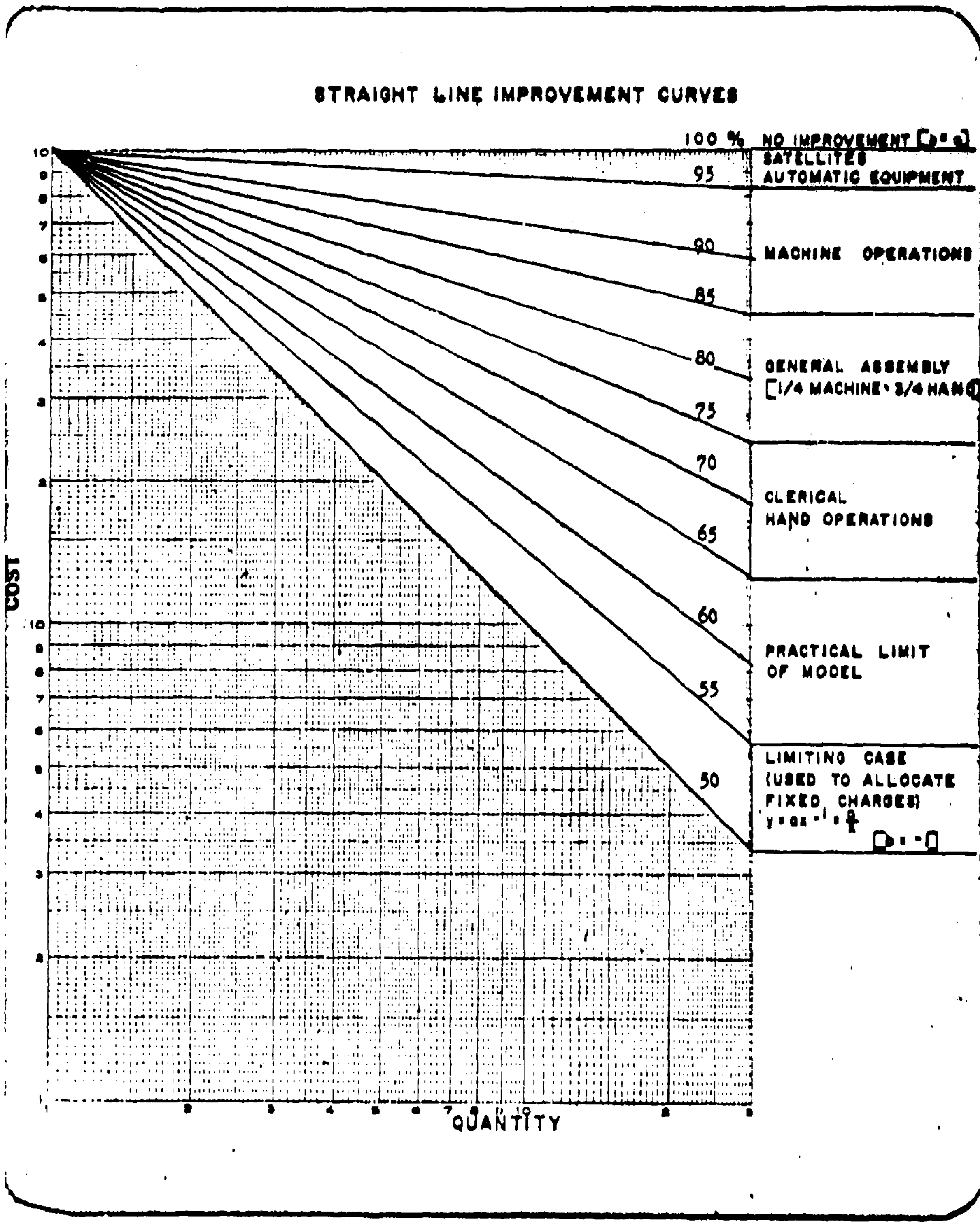

STRAIGHT LINE IMPROVEMENT CURVES
100 %
NO IMPROVEMENT
95
SATELLITES
AUTOMATIC EQUIPMENT
90
MACHINE OPERATIONS
85
80
GENERAL ASSEMBLY
[1/4 MACHINE + 3/4 HAND]
75
70
CLERICAL
HAND OPERATIONS
65
60
PRACTICAL LIMIT
OF MODEL
55
50
LIMITING CASE
(USED TO ALLOCATE
FIXED CHARGES)
COST
QUANTITY

Figure 3

THE EFFECT OF RELIABILITY, LENGTH, AND SCORE CONVERSION ON A MEASURE OF PERSONNEL ALLOCATION EFFICIENCY

Richard C. Sorenson and Cecil D. Johnson
U. S. Army Personnel Research Office
Washington, D. C.

Within the United States Army it has been realized for many years that an effective military organization must have the right kind of men as well as the most advanced and effective equipment. Of course this does not mean that the Army must have only the 'best' of the personnel pool, but does mean that those men taken from the personnel pool must be matched with jobs in a way that facilitates maximum manpower utilization. There are two sides to this task of manpower utilization: 1) the various functions performed within the Army must be analyzed to determine the different skills needed to perform those functions, and 2) the individual differences within the personnel pool must be analyzed to find those different abilities that can be reliably measured. At this point we are left with the problem of developing effective measuring instruments and of devising ways and means of assigning men to jobs on the basis of the measure of abilities. This whole attack on manpower utilization rests on the realization that while few men can be trained--no matter how extensive and careful the training--to do all the Army jobs as well as those who do them best, most men accepted by the Army can be trained such that they are effective in performing those skills for which they are most apt, and when properly assigned, will be an asset to the Army.

Thus the solution of the problem rests on successfully accomplishing the following: 1) identifying job families within the Army that require personnel with different ability, 2) identifying and measuring these abilities within the personnel pool, 3) estimating the performance on the job on the basis of measures of ability related to job requirements and 4) assigning men to jobs so as to maximize overall performance.

The first of these steps has been treated in the establishment of the Army occupational areas. Ten occupational areas have been identified and shown to be satisfactory in classifying the various Army functions assigned to enlisted men (EM) [10]*. Recent research indicates that nine

*The numerals in brackets indicate numbered references listed at end of paper.

categories of training schools within Army Advanced Individual training may be differentiated [5]. It may be assumed that continuing research will be required to evaluate the constantly changing functions performed by Army EM as new methods and procedures are introduced.

The Army Classification Battery (ACB) has been developed to measure aptitudes related to Army jobs [4]. An important research mission of USAPRO is to introduce new measuring devices, and to revise and/or validate present tests [7].

The eight current Aptitude Areas are functions of the eleven tests within the ACB and serve as performance estimates for the Military Occupational Specialties (MOS) in one or two occupational categories. These Aptitude Area Scores are currently used for differential classification [10]. (See Figure 1.)

The benefits inherent in differential classification using Aptitude Area Scores stem from the fact that information is obtained relative to the differences in ability between individuals and to differences within the individual. Thus EM may be assigned to jobs for which their probability of success may be a good deal greater than that for Army jobs in general.

The technical gain is twofold. First, a given level of aptitude for a given job can be assured by a lower score on the specific selector highly related to the job than would be required to maintain the same standard of excellence if the selection were based on an instrument less valid for the purpose at hand. Secondly, when recruits are taken above a given cutting score on a general selector, they are removed from that score interval of the aptitude pool for all other jobs as well. However, when recruits are taken above a given cutting score on a specific selector, they come from a much broader range of scores as far as the pool for another specific selector is concerned. To the extent that one specific selector is uncorrelated with a second, the entire range of scores is still available on the latter after selection has been accomplished on the first selector.

Thus we see that for a particular sample of 1800 individuals drawn for the purpose of standardizing a subsequent version of one of the tests 56% were above average on the Armed Forces Qualification Test (AFQT) relative to the original standardization population. Of this same sample, however, 91% were above the average for the Aptitude Area in which they scored highest. (See Figure 2.)

One further operational gain was investigated. Under the former system in which a single test--the Army General Classification Test (AGCT) -- was practically the sole determinant of Army classification, selection for one set of jobs automatically gave those jobs the upper segment of the distribution of test scores. The lower segment was left for the remaining jobs. In the operational problem filling the manpower requirements of an infantry division, approximately one half of the men were combat infantrymen . If, as happened at times during the war, a test were used to select primarily for the noncombat specialties, these jobs would be filled by using the upper half of the distribution . In such a case only the lower half of the distributions of test scores would be available for the combat jobs as indicated in Figure 3. However, when the results of the distribution of men into aptitude areas corresponding to job families for the infantry division used in the standardization study mentioned above are viewed, the distribution of AGCT scores for the non-priority or combat jobs is seen to be almost equal to the distribution of AGCT scores for the priority jobs. This is shown graphically in Figure 4.

A great deal of research has been undertaken to make optimal allocation feasible. Various versions of the optimal regions and other methods are now available for operational use [1] . In the research reported in this paper a routine derived from the Hungarian solution to the transportation problem was used [8] .

In this paper we will be concerned with investigating characteristics of performance estimates (and the test battery from which they were derived) as they relate to the criterion of personnel allocation efficiency as measured by the average performance under conditions of optimal allocation. This measure of performance is the objective function to be maximized in the transportation problem. Many relationships involving this objective function and the variables of this study may easily be calculated analytically assuming ideal conditions, e.g., continuous normally distributed psychological test scores. For instance Brogden [2, 3] has shown that when other factors are held constant and certain conditions assumed, the efficiency of allocation is directly proportional to the validity of the performance estimate, and that one may determine by analytic means the allocation efficiency for given numbers of jobs, percent of personnel pool rejected, and intercorrelation of performance estimates. In reality, however, we are not dealing with continuous variables and frequently other assumptions are not met. Also, in practice the scores are often transformed in such a way that considerable information is lost. It is less easy to investigate the more realistic situations analytically. Thus we have embarked on a program to study by a Monte Carlo approach the general relationship between amount of information in a distribution of discrete performance estimates and the performance level it is possible to achieve by the most efficient pattern of personnel assignments.

The basic step in the implementation of a statistical experiment is the generation of uniformly distributed random numbers. We have used computer routines which generate pseudo-random numbers by the power residue method [9]. These distributions of uniform variables are then transformed to distributions of normal variables. This transformation results in a matrix, X, of order n by k, i.e., n entities are represented each by a vector of k simulated scores:

$$(1) \qquad X = \begin{bmatrix} X_{11}, X_{12}, \ldots X_{1k} \\ X_{21}, X_{22}, \ldots X_{2k} \\ \vdots \\ X_{n1}, X_{n2}, \ldots X_{nk} \end{bmatrix}$$

where

$$(2) \qquad \left.\begin{array}{lll} & X'X & \rightarrow nI \\ \text{and} & 1'X & \rightarrow 0 \\ \text{when} & n & \rightarrow \infty. \end{array}\right\}$$

We see then that for each sample we generate a matrix that has an expectation for its covariance matrix of the identity matrix.

Now we desire to further transform the matrix X by post multiplication by a matrix T such that the resulting matrix has for its expected covariance matrix a given matrix C:

$$(3) \qquad \left.\begin{array}{lll} & XT & = Y \\ \text{where} & Y'Y & \rightarrow nC \\ \text{when} & n & \rightarrow \infty \end{array}\right\}$$

The matrix C is specified as a function of the desired standard deviation and intercorrelation of the variables:

$$(4) \qquad C = s\, R\, s,$$

Where R is the desired correlation matrix and s is the diagonal matrix of standard deviations.

We wish to find the matrix T such that the conditions in (3) will hold. From these equations we may write the requirement that:

$$(5) \qquad \left(\frac{1}{n}\right) Y'Y = \left(\frac{1}{n}\right) T'X'XT \rightarrow C$$

when $n \rightarrow \infty$.

From (2) we see that

$$(6) \qquad \frac{1}{n} X'X \rightarrow I \quad \text{when} \quad n \rightarrow \infty$$

and from (5) and (6) we have

$$(7) \qquad T'T = C\,.$$

We may represent the matrix C in terms of its basic structure:

$$(8) \qquad C = Q \Delta Q'$$

where $QQ' = Q'Q = I.$

We know that the matrix C to any power e.g., ℓ, may be formed by raising the eigen values of C to that power, premultiplying by Q and post-multiplying by Q′[6]:

$$(9) \qquad C^{\ell} = Q \Delta^{\ell} Q'$$

$$(10) \qquad \text{thus} \quad C^{\frac{1}{2}} = Q \Delta^{\frac{1}{2}} Q'.$$

Formula (10) could be demonstrated as follows:

$$C^{\frac{1}{2}} C^{\frac{1}{2}} = Q\Delta^{\frac{1}{2}} Q' Q\Delta^{\frac{1}{2}} Q' = Q\Delta Q' = C. \qquad (11)$$

We will let

$$T = C^{\frac{1}{2}} \qquad (12)$$

We see that

$$T'T = C^{\frac{1}{2}} C^{\frac{1}{2}} = C \qquad (13)$$

Hence a transformation solved for by equation (11) meets the requirement of (7) and while there are an infinite number of transformations that meet this requirement the one indicated is by far the most advantageous since it provides for uniformity of rounding errors and impartially improves normality of the transformed scores.

Thus we may simulate samples of personnel by building into the score distribution characteristics of performance estimates in which we are specifically interested. These performance estimates may in turn be a function of such test characteristics as length, reliability and validity. The effectiveness of a test or of the resulting performance estimation is determined by its potential contribution to the optimal allocation average, that is, the average estimated performance of men on the jobs to which they are assigned.

Let us first consider one of these characteristics of a distribution of performance estimates: namely the standard deviation. Often times, in the course of personnel operations where men are actually being assigned to jobs on the basis of measured attributes, distributions of scores are transformed from distributions in which there are two or three significant digits to distributions in which there is only one significant digit. This is the case in assigning men to jobs in the Army. The three digit Army Aptitude Area Score is coded according to AR 611-259 to a score taking on the values ranging from zero to nine. The questions we ask are: 1) What loss of information occurs when scores are coded to a one digit scale, and 2) What affect does this loss of information have on average performance when these scores are used to assign men to jobs?

In Figure 5 we demonstrate the effect of coding the scores of a continuous distribution centered at 50 into nine score scales, e.g., entities with scores less than 11.5 were given a coded score of 1, entities with scores 11.5 or greater but less than 22.5 were given a coded score of 2, ... entities with scores 88.5 or greater when given a score of 9. The upper portion is the resulting distribution when the original distribution has a standard deviation of 20. The information measure, H, has an intuitive appeal because it is sensitive to both the size of the coded interval and the spread of scores. For the above distribution H may be calculated by

$$H = \sum_{i=1}^{9} (p_i \log p_i) \tag{14}$$

where p_i is the proportion of the entities in the ith interval and $\log p_i$ is the natural logarithm of p_i. The information measure corresponding to the distribution represented in the top of Figure 5 is 1.991. In the lower figure, a similar transformation was performed on a continuous distribution, where the original distribution has a standard deviation of 10. We see here that the cases are primarily distributed in intervals 4, 5, and 6, that they are much more closely grouped together. That much more information is lost is indicated by the corresponding information measures which is 1.372. We may note that the maximum value for the information measure corresponding to a nine score scale is 2.197 which occurs when the distribution is uniform.

Now we can easily see that information is lost when we go from several significant digits to one significant digit. We also see that more information is lost when the standard deviation of the parent distribution is small than when it is large. We desire to investigate the degree to which such information loss affects the optimal allocation average.

Another variable of interest is the quota restriction places on the optimal allocation. A natural quota is defined as the number of men that would be assigned to a job if everyone were assigned so as to maximize his individual performance without regard to quotas. In the case of equal variances and intercorrelations among performance estimates, the natural quotas are equal, i.e., uniform. On theoretical grounds we can conclude that the degree to which the quotas are perturbed from the natural is related to the allocation average. However, the effect of this quota factor on the other relationships must be studied empirically. We see in Figure 6 the percentage quotas imposed on optimal allocation for the situation where we have 16 jobs and where we simulate only 4 jobs. Note that the natural or uniform quota for 16 jobs is .0625. That is the proportion of the total personnel pool that would be allocated to each job. For 4 variables it is .25. There are two considerations that determined the perturbed quotas. The first was that we wanted at least one individual to be assigned to each job, for both the 16 and 4 variables for each of the sizes of samples. The second was that we wanted the ratio of the information measure that was found to exist between the 4 and 16 variable situation, for the natural quotas, to exist also for the perturbed quotas. We required that the uncertainty of assigning men to jobs with 16 variables be twice that for assignment with 4 variables for both the natural and perturbed quotas. The resulting proportions indicated in the table were the result of the two considerations mentioned above. We feel that in imposing these quota restrictions in our experiment we are being realistic, in that the necessary perturbations in the quotas in the actual operational conduct of the Army personnel system would not be greater than this.

In order to study these effects, a 2^5 factorial experiment using simulated performance estimates was designed. The five factors were: (1) standard deviation of the estimated performance; (2) number of cases in the sample; (3) number of variables; (4) number of score intervals; and (5) quota restriction. Figure 7 indicates the various levels of the five factors that were used. The performance estimate variables were generated such that they had an expectation of .70 for their intercorrelation. For those samples that were randomly assigned to Level a of Factor 1, the parent distribution was generated to have a standard deviation of 10; for those assigned to Level b, the standard deviation was 20. Similarly, those samples assigned to the first level of Factor 4 were transformed to have 9 score intervals, while those assigned to Level b were transformed to have 99 score intervals. The number of cases and variables represented correspond to the level of Factors 2 and 3 to which the sample was assigned. Those samples assigned to Level a of Factor 5 were allocated with uniform quotas. Those samples assigned to Level b were allocated with perturbed quotas. Thus we have a 2^5 factorial experiment in which there are 32 cells. The experiment was initially replicated 10 times. Three hundred and twenty samples were generated from a simulated personnel pool and allocated optimally to either 4 or 16 job categories. Figure 8 is a flow diagram indicating the five steps in this experiment. In step 1, the matrix X of normally distributed random numbers, was generated. In the second step, the matrix Y of continuous performance estimates, was derived by multiplying the matrix X by the transformation matrix. The continuous performance estimates were used in evaluating the allocations under the various experimental conditions by averaging the estimated performance of men on the jobs to which they were assigned. In doing this, we used the continuous performance estimates, since continuous performance estimates yield an unbiased estimate of the actual performance of men on the job, whereas discrete performance estimates would have introduced a slight bias. As may be seen from the arrow going from step 2 to step 5 in the graphical presentation, the continuous performance estimates were used in the calculation of the allocation average. In step 3, the matrix $\tilde{Y}$ was derived by forming a discrete performance estimate from the continuous performance estimate. This was done simply by forming the scores into either 9 or 99 score intervals. Step 4, the allocation step, was accomplished by a computer program which optimally allocates men to jobs by a linear program derived from the Hungarian Solution to the transportation problem [8]. The average performance for men who are thus allocated is then calculated. It is these allocation averages which are subjected to the analysis of variance in this experiment.

We have put the analysis of variance to a slightly different use in our experiment than is the usual case. Theoretical considerations in this experiment dictate that we should expect significant differences between the two levels of each of these five factors. We are not testing to see if the null hypothesis should be rejected, but we are performing the analysis of variance so that in the event that the main effects are not significant, we can evaluate our simulation for its adequacy with regard to the number of replications. Thus, the purpose of the analysis of variance in this experiment is primarily that of evaluating the number of replications that we used in our simulation. With 10 replications, four of the five factors were highly significant at the .001 level or less. However, the effect of Factor 2, the number of cases in each simulated personnel sample, was not significant. We then repeated the experiment using as the level of Factor 2 different sizes of samples: 32 and 192. We found that while there was a small difference, this difference was insignificant both statistically and practically. We conclude that when allocating large quantities of men to jobs under the conditions specified above, we are justified in sub-optimizing (random sampling the overall sample into several subsamples and allocating each of the subsamples optimally). In so doing, we may operate with less computer space with little concern for the loss in allocation average.

In Figure 9 we have shown the mean performance for the levels of those factors that were found to be statistically significant. The results indicated that the number of variables is the most important of the factors of the experiment. We could increase the gain over random allocation by 72% by increasing the number of criterion variables from 4 to 16. This indicates that one of the most promising avenues of psychometric and personnel research is to differentially predict more job categories or job families than we are now doing. The number of score intervals factor was a significant one as was the quota factor. However, the latter was of no practical significance. We feel that we may continue to use natural (or uniform) quotas in our research work and generalize our interpretation of results to realistic situations where the quotas are not uniform.

The interactions of Factor 1 with Factor 4, and Factor 3 with Factor 4 were both significant at the .01 level. The cell means for these two interactions are found in Figures 10 and 11. It appears that the information loss is considerably more crucial when we are dealing with 16 differential job predictions than when we are dealing with only 4. The significant interaction between Factor 1 and Factor 4 indicates that the loss in the allocation average going from 99 score intervals to 9 score intervals is much greater when the standard deviation is 10 than when it is 20. (Recall that this was predicted from considerations of the amount of information in the respective distributions.) The results thus far indicate that: (1) mean performance may be increased by increasing the number of differential performance estimates, (2) when attempting to do # 1, it is important that all the information possible be retained in the score distribution by using as many score intervals as is meaningful, and (3) in going from a 99 interval distribution to a 9 interval one, the loss is doubled if the original standard deviation is 10 rather than 20.

These results may be evaluated from at least two points of view: first, from that of an agency dealing with actual score distributions, and second, from the point of view of the test constructor. He looks at our number of intervals factor as the number of items in a test, since the number of meaningful score intervals is related to the number of test items. Furthermore, he may consider our standard deviation factor in terms of the relationship between the standard deviation of a test and the reliability and number of items in the test.

Upon consideration of the factors mentioned above, an additional experiment was designed. The factors to be studied and their levels are indicated in Figure 12. Ten samples of 200 entities were assigned to each of the eight cells of the design formed by the first three factors. Each sample was optimally allocated and evaluated at each level of Factor 4. For each sample, vectors of test scores were generated and transformed to represent perfectly valid performance estimates.

Figure 13 represents by a flow diagram the steps followed in the experiment. First, the matrix of normal random numbers, X, was generated. In step 2, X was transformed to a matrix of continuous test variables. In step 3 the continuous test variables were formed which were to be used in the evaluation of our allocations in step 8. In step 4, the discrete test variables, $\widetilde{G}$, were formed from the continuous test variables, matrix G, by creating either 20 or 40 discrete score intervals. From $\widetilde{G}$, the performance estimates, $\widetilde{Y}$, were formed by the appropriate regression equation. These performance estimates were used in allocating the men to jobs in step 7. In step 6, the performance estimates were transformed to stanine form and again the men were allocated to jobs and the allocation was evaluated.

Note that this analysis of variance is a split plot analysis of variance in which we can analyze the between-samples variance and the within-samples variance. First, let us look at Figure 14, which reports the results of the between-samples variance. The effect of intercorrelations, reliability, and the inter-action between intercorrelations and reliability, were all significant. The number of items was significant only at the .25 level, with 10 replications. We see from the analysis of the within-samples variance (see Figure 15) that the score conversion factor was significant and the score conversion-reliability interaction was significant as were the three factor interactions of score conversion, intercorrelation, reliability and score conversion, reliability, number of items. Let us now look at the difference in the mean job performance for the two levels of each of the four factors as indicated in Figure 16. It is of interest to note that by reducing the intercorrelation among the test variables, a great increase can be brought about in the allocation average (i.e., mean job performance). We see also, that the test reliability is an important consideration. Let us note that the difference in mean performance for the two different levels of number of items, apart from validity, intercorrelation, and reliability, was in the direction that the larger the number of items, the higher the allocation average. The difference across the two levels of score conversion (i.e., no conversion vs. a conversion from the score to the stanine) was also a significant one. As we look at the interaction between the intercorrelation among the test variables and the reliability (see Figure 17), we see that the reliability is a more crucial consideration when high intercorrelations prevail than when they are low.

Inasmuch as we did not find the number of items to be a significant consideration, we replicated the experiment for crucial cells 20 more times. In Figure 18, we see the results of that analysis of variance. We see that the number of items is significant, and that the score conversion as well is statistically significant. In looking at the means for that experiment, we find that as we go from 40 items to 20 items, that is, when we cut the length of the test in half, even if we would keep the reliability of the test the same and the validity of the test the same, we would lose approximately 8% of our gain over random allocation of men to jobs.

The results of this work indicate that the use of caution is warranted in advocating the use of shorter tests in optimal differential classification, even if the shorter tests retain the reliability and validity of the longer tests, especially if the reliability of the tests is closer to .7 than to .9. This and other research currently in progress has impact on the planning of further test development research and on the operational handling of test scores and performance estimates. Furthermore, it demonstrates that simulated experiments can yield information concerning possible trade-off between allocation average, testing costs, and the relative costs of test development. Even more efficient experiments could be done to estimate the magnitude of differences by employing variance reduction methods. One, the regeneration of the same sample transformed for each cell in the design, would be especially appropriate for this type of study. It was not used in this project because the model for analysis of variance does not provide for a residual estimate of variance. Future projects will employ variance reduction techniques.

REFERENCES

1. Boldt, Robert F. Development of an Optimum Computerized Allocation System. Technical Research Report 1135, U. S. Army Personnel Research Office, OCRD, DA, 1964.

2. Brogden, Hubert E. Increased Efficiency of Selection Resulting from Replacement of a Single Predictor with Several Differential Predictors. Educ. and Psychol. Measurement, Vol. 11, No. 2, 1951.

3. Brogden, Hubert E. Efficiency of Classification as a Function of Number of Jobs, Percent Rejected, and the Validity and Intercorrelation of Job Performance Estimates. Educ. and Psychol. Measurement, Vol. 19, No. 2, 1959.

4. Helme, William H. Differential Validity of the ACB for Courses in Seven Job Areas. PRB Technical Research Report 1118, Personnel Research Branch, R&DD, TAGO, 1960.

5. Helme, William H., and Fitch, D. J. Grouping Army Training Courses by Army Classification Battery Factors. Technical Research Note 128, U. S. Army Personnel Research Office, OCRD, DA, 1962.

6. Horst, Paul. Matrix Algebra. Holt, Rinehart and Winston, Inc., New York, 1963 (see Section 17.7).

7. Johnson, C. D., Waters, L. K., Helme, W. H. Factor Analysis of Experimental Noncognitive Measures of Combat Potential. Technical Research Note 147, U. S. Army Personnel Research Office, OCRD, DA, 1964.

8. Kuhn, H. W. The Hungarian Method for the Assignment Problem. Naval Research Logistics Quarterly, Vol. 2, Nos. 1 and 2, 1955.

9. Shreider, Yu A. Statistical Testing. Elsevier Publishing Co., New York, 1964 (p. 199).

10. Zeidner, Joseph, Harper, B. P., and Karcher, E. K. Reconstruction of Aptitude Areas. PRB Technical Research Report 1095, Personnel Research Branch, PR&PD, TAGO, 1956.

ARMY CLASSIFICATION BATTERY

TEST	SYMBOL
VERBAL	VE
ARITHMETIC REASONING	AR
PATTERN ANALYSIS	PA
CLASSIFICATION INVENTORY	CI
MECHANICAL APTITUDE	MA
ARMY CLERICAL SPEED	ACS
ARMY RADIO CODE	ARC
GENERAL INFORMATION	GIT
SHOP MECHANICS	SM
AUTOMOTIVE INFORMATION	AI
ELECTRONIC INFORMATION	EI

ARMY APTITUDE AREAS

TITLE	SYMBOL	FORMULA
INFANTRY - COMBAT	IN	$\frac{AR + 2CI}{3}$
ARMOR, ARTILLERY, ENGINEERS-COMBAT	AE	$\frac{GIT + AI}{2}$
ELECTRONICS	EL	$\frac{MA + 2EI}{3}$
GENERAL MAINTENANCE	GM	$\frac{PA + 2SM}{3}$
MOTOR MAINTENANCE	MM	$\frac{MA + 2AI}{3}$
CLERICAL	CL	$\frac{VE + ACS}{2}$
GENERAL TECHNICAL	GT	$\frac{VE + AR}{2}$
RADIO CODE	RC	$\frac{VE + ARC}{2}$

Figure 1. Army Classification Battery (ACB) tests and Army Aptitude Areas as functions of ACB variables.

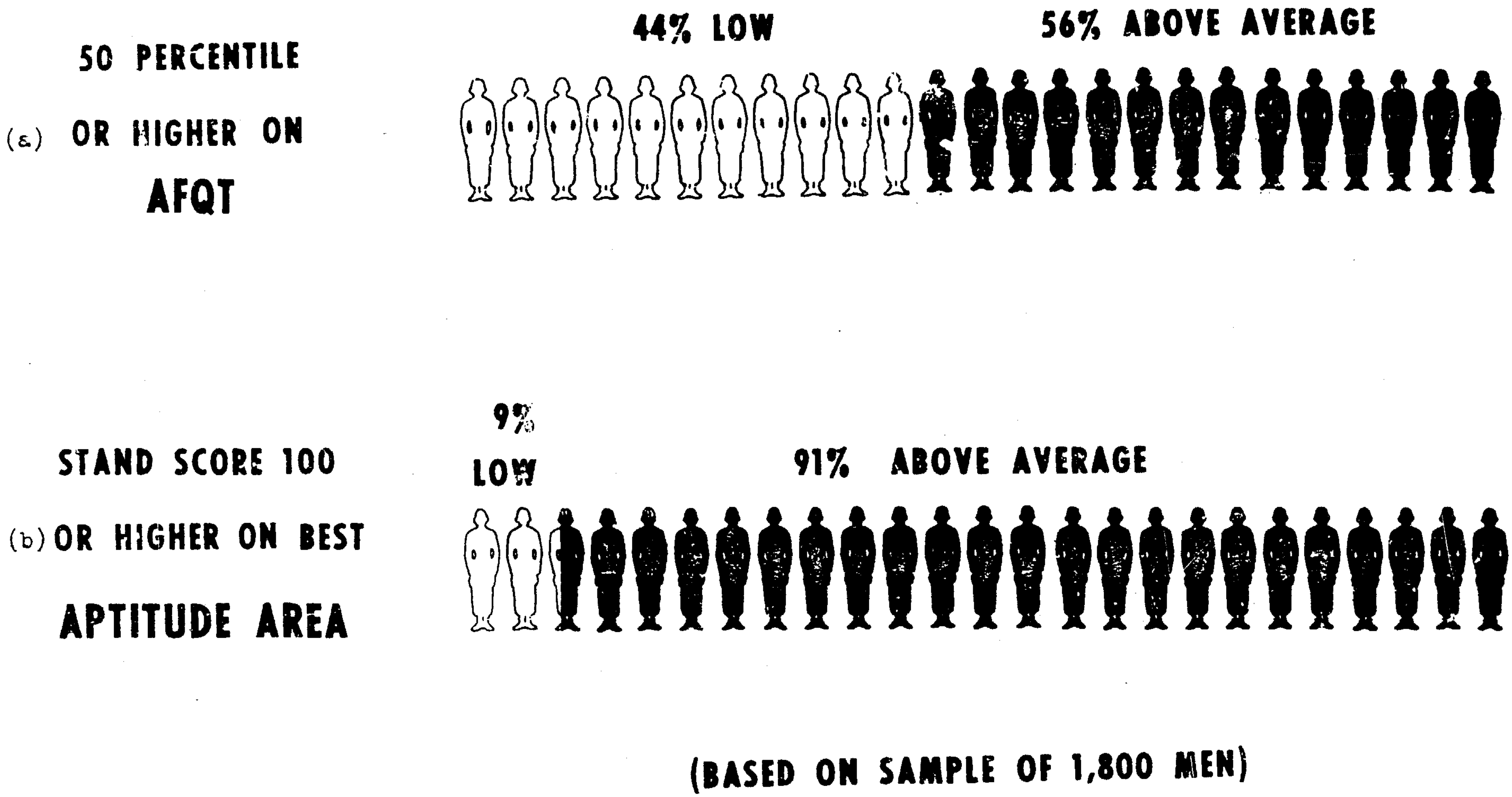


Figure 2. Proportion of sample of 1800 from input population of enlisted men scoring above 50th percentile (a) on AFQT and, (b) on their highest aptitude area.

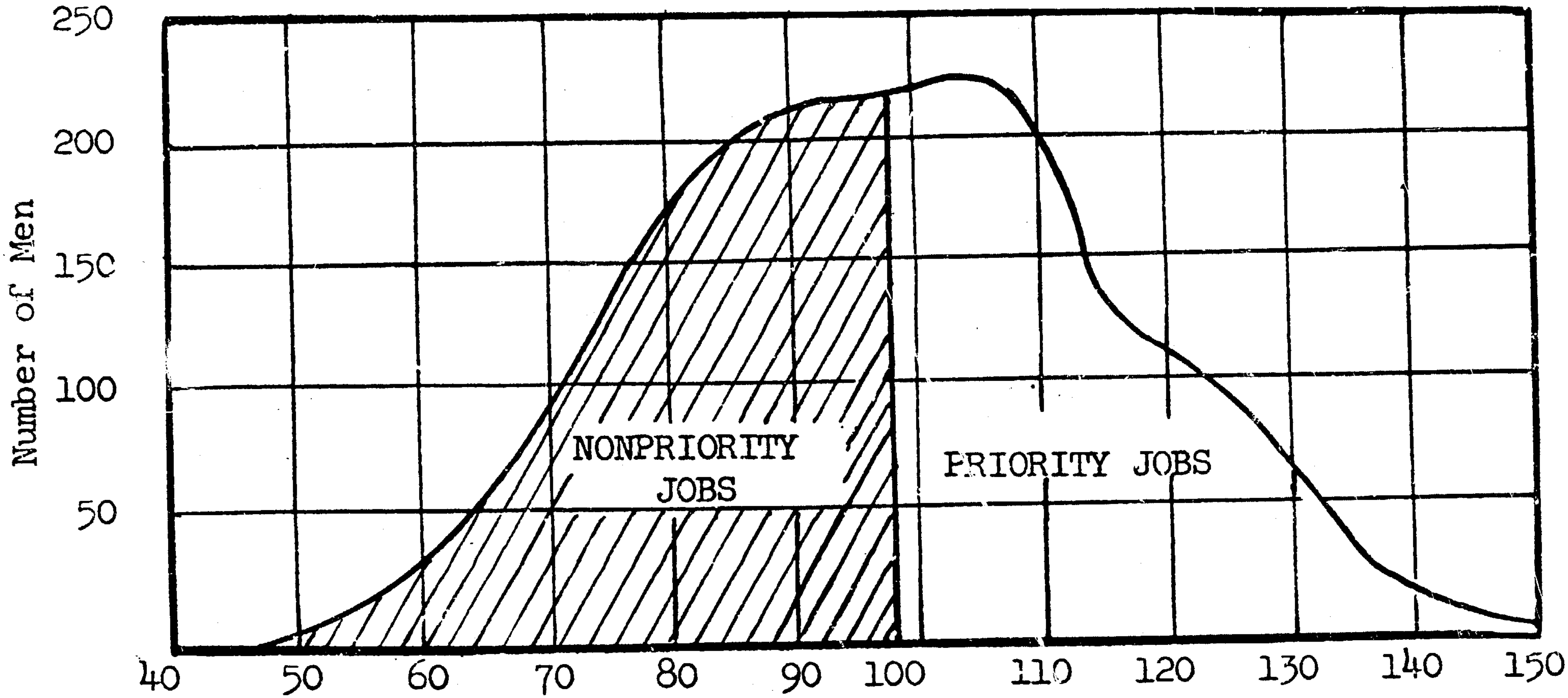


Figure 3. Distribution of Army Standard Scores on overall general ability for priority and nonpriority jobs, when assignment is based on a single measure.

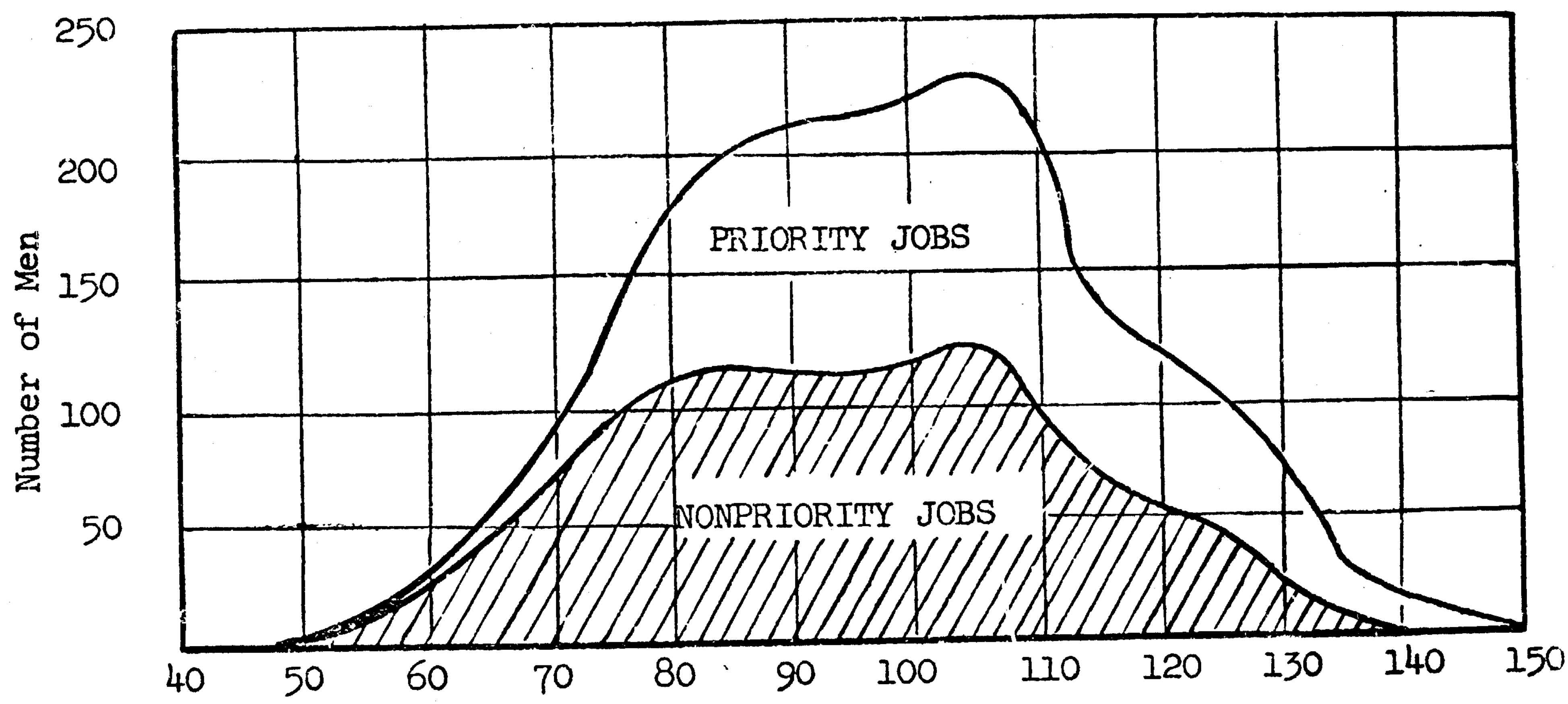


Figure 4. Distribution of Army Standard Standard Scores on overall general ability when assignment is based on battery of tests.

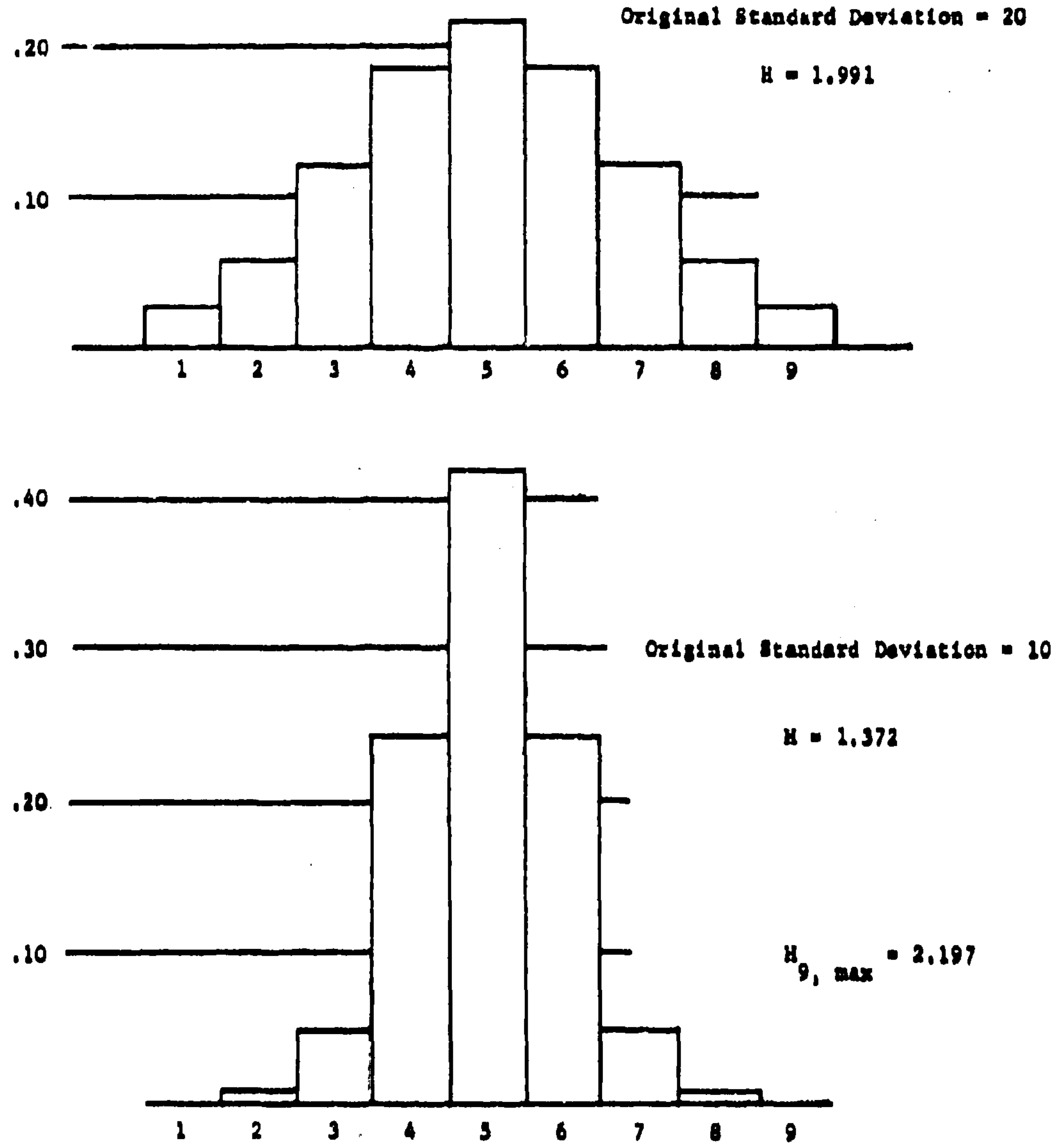


Figure 5. Discrete distributions resulting from continuous distributions with standard deviation of 10 and of 20.

Job	16 Variables Natural	16 Variables Perturbed	4 Variables Natural	4 Variables Perturbed
1	.0625	.0062	.2500	.1141
2	.0625	.0137	.2500	.2047
3	.0625	.0212	.2500	.2953
4	.0625	.0287	.2500	.3859
5	.0625	.0362		
6	.0625	.0437		
7	.0625	.0512		
8	.0625	.0587		
9	.0625	.0662		
10	.0625	.0737		
11	.0625	.0812		
12	.0625	.0887		
13	.0625	.0962		
14	.0625	.1037		
15	.0625	.1112		
16	.0625	.1187		

Figure 6. Job quotas, expressed as proportions, imposed as constraints on personnel assignment procedure.

Factorial Design for Experiment Using
Simulated Performance Estimates

Factor 1: Standard deviation
Level a: S = 10
Level b: S = 20

Factor 2: Number of cases
Level a: N = 160
Level b: N = 320

Factor 3: Number of variables
Level a: V = 4
Level b: V = 16

Factor 4: Number of intervals
Level a: I = 9
Level b: I = 99

Factor 5: Quota
Level a: Perturbed quotas
Level b: Natural quotas

Figure 7. Experimental conditions used in the five-factor experiment.

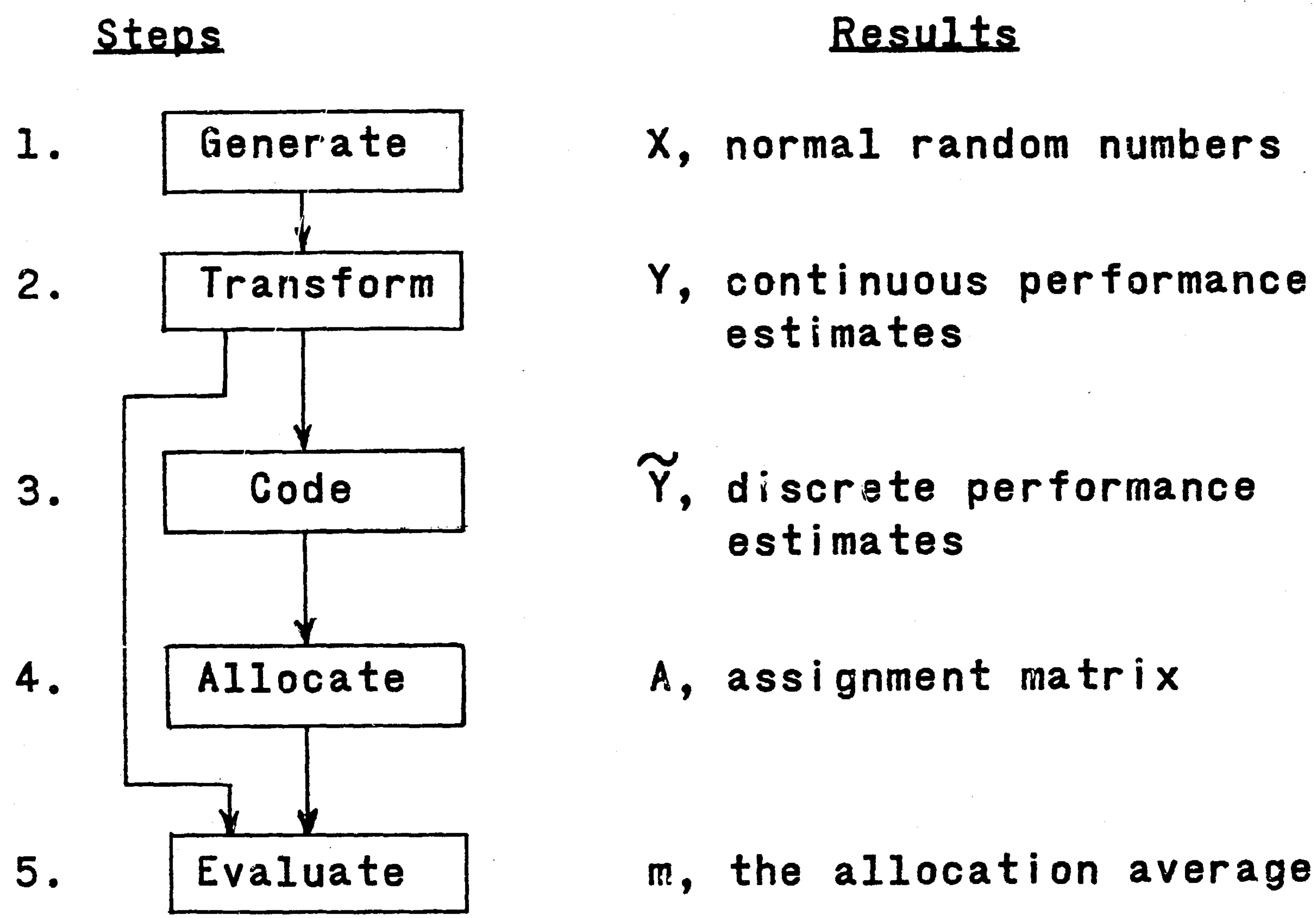


Figure 8. Flow diagram of experiment using simulated performance estimates.

Results of Experiment Using Simulated Performance Estimates

Factor	Level	Mean Performance: Standard Units	Mean Performance: Army S. Scores
Standard deviation	10	.65	113
Standard deviation	20	.70	114
Nr. of variables	4	.50	110
Nr. of variables	16	.86	117
Nr. of score intervals	9	.62	112
Nr. of score intervals	99	.74	115
Quota	Perturbed	.67	114
Quota	Natural	.69	114

Figure 9. Mean performance for selected factors.

Standard Deviation

Number of intervals		10	20
	9	.58 [111.5]	.67 [113.4]
	99	.73 [114.7]	.74 [114.8]

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 10. Mean performance for selected cells: First order interaction terms for standard deviation and number of intervals.

Number of Variables

		4	16
Number of intervals	9	.46 (109.2)	.78 (115.7)
	99	.55 (110.9)	.93 (118.5)

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 11. Mean performance for selected cells: first order interaction terms for number of variables and number of intervals.

FACTORIAL DESIGN FOR EXPERIMENT USING SIMULATED PSYCHOLOGICAL TEST VARIABLES

Factor 1: Test intercorrelation

Level a: $r_{ij} = .4$

Level b: $r_{ij} = .6$

Factor 2: Test reliability

Level a: $r_{tt} = .7$

Level b: $r_{tt} = .9$

Factor 3: Number of items

Level a: $n = 20$

Level b: $n = 40$

Factor 4: Score conversion

Level a: Score

Level b: Stanine

Figure 12. Experimental conditions used in split-level experiment.

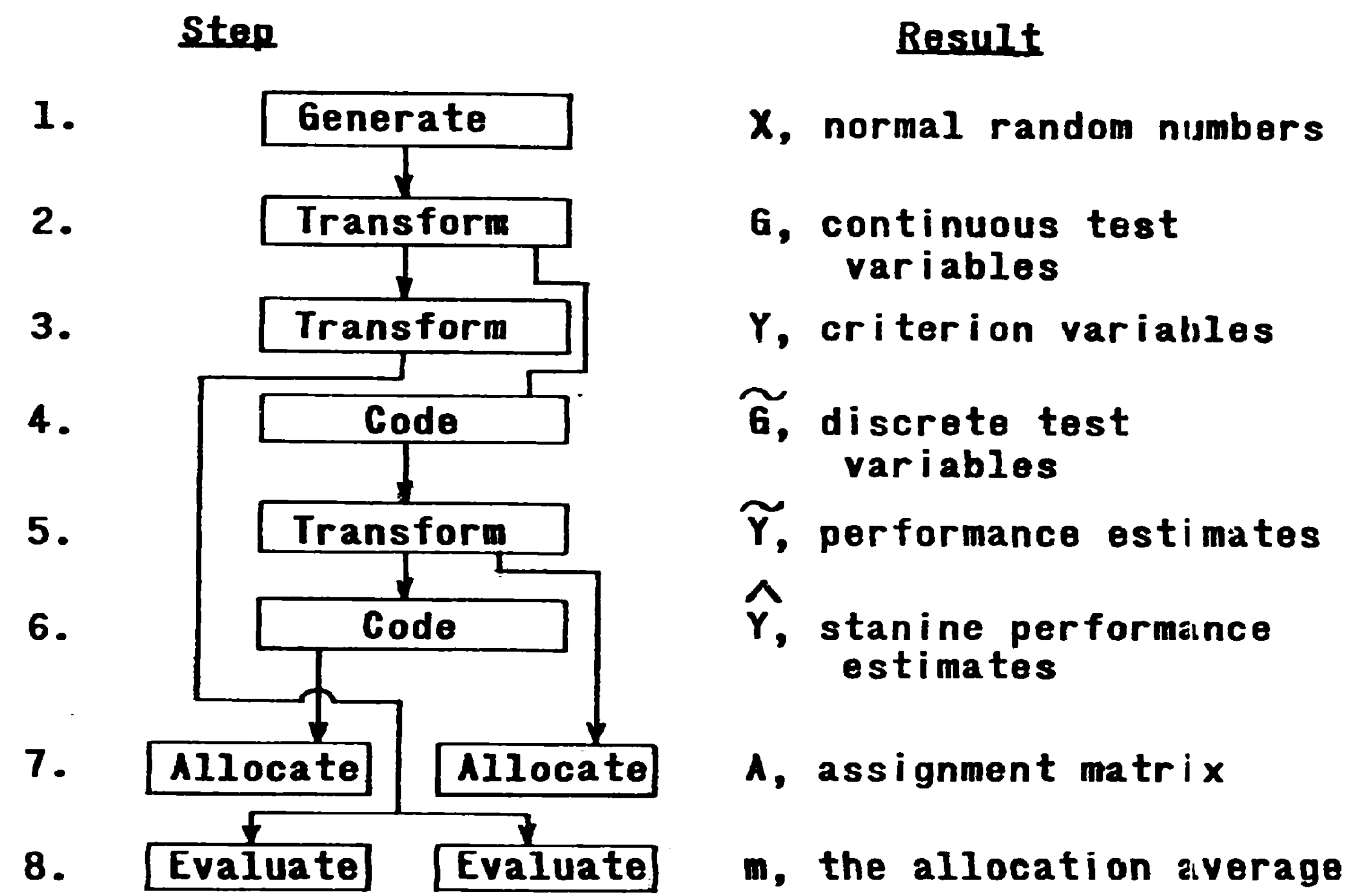


Figure 13. Flow diagram of experiment using simulated psychological test variables.

Source of Variation	Sum of Squares	d.f.	Mean Square	F	P
Intercorrelations	6.2960855	1	6.2960855	784.97	.001
Reliability	5.8434368	1	5.8434368	728.54	.001
Nr. of Items	.0133628	1	.0133628	1.67	.25
Intercorrelations x Reliability	.3419087	1	.3419087	42.63	.001
Intercorrelations x Nr. Items	.0087315	1	.0087315	1.09	---
Reliability x Nr. of Items	.0228176	1	.0228176	2.84	.10
Intercor. x Reliability x Nr. Items	.0002393	1	.0002393	---	---
Samples in Same Exp. Condition	.5774944	72	.0080208	---	---
Total between Samples	13.1040766	79			
Total within Samples	.0560724	80			
Total	13.1601490	159			

Figure 14. Results of analysis of variance applied to the average performance for independent variables: intercorrelation, reliability and number of items (between sample variance).

ANALYSIS OF VARIANCE OF ALLOCATION AVERAGE

Source of Variation	Sum of Squares	d.f.	Mean Squares	F	P
Score Conversion	.0441774	1	.0441774	777.78	.001
Score x Intercorrelation	.0000479	1	.0000479	---	---
Score x Reliability	.0061563	1	.0061563	108.39	.001
Score x Nr. of Items	.0001583	1	.0001583	2.79	.10
Score x Intercor. x Reliability	.0003036	1	.0003036	5.35	.025
Score x Intercor. x Nr. of Items	.0000131	1	.0000131	---	---
Score x Reliability x Nr. of Items	.0011273	1	.0011273	19.85	.001
Score x Intercor. x Reliab. x Nr. Items	.0000003	1	.0000003	---	---
Sample x Score	.0040882	72	.0000568	---	---
Total within Samples	.0560724	80			
Total between Samples	13.1040766	79			
Total	13.1601490	159			

Figure 15. Analysis of variance of average performances for the correlated variable: type of score conversion (within sample variance).

Results of Experiment Using Simulated Test Scores

		Mean Performance	
Factor	Level	Standard Units	Army S. Scores
Test intercorrelation	.4	.83	117
Test intercorrelation	.6	.43	109
Test reliability	.7	.45	109
Test reliability	.9	.82	116
Nr. of items	20	.62	112
Nr. of items	40	.64	113
Score conversion	Score	.65	113
Score conversion	Stanine	.61	112

Figure 16. Mean performance for the four factors.

INTERACTION TERMS

Test intercorrelation x test reliability

		Intercorrelation .4	Intercorrelation .6
Reliability	.7	.68 [113.7]	.20 [103.9]
Reliability	.9	.97 [119.5]	.67 [113.4]

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 17. Mean performance for selected cells: first order interaction terms for intercorrelation and reliability factors.

ANALYSIS OF VARIANCE OF ALLOCATION AVERAGE

(30 Replications)

Source of Variation	Sum of Squares	d.f.	Mean Square	F	P
Number of Items	.0727871	1	.0727871	6.77	.025
Samples in Same Exp. Cond.	.6237790	58	.0107548	----	----
Between Samples	.6310577	59			
Score Conversion	.0179231	1	.0179231	25.21	.001
Score x Number of Items	.0009848	1	.0009848	1.39	.25
Score x Sample	.0412308	58	.0007109	----	----
Within Samples	.0601387	60			
Total	.6911964	119			

Figure 18. Results of analysis of variance contrasting number of items at a fixed level of reliability and intercorrelation (.7 and .4, respectively).

Means for Experiment Using Additional Replications

	Nr. of Items		
	20	40	Total
Score	.676 [113.5]	.731 [114.6]	.703 [114.1]
Stanine	.657 [113.1]	.701 [114.0]	.679 [113.6]
Total	.667 [113.3]	.716 [114.3]	.691 [113.8]

Entries are in terms of standard units; bracketed values are in terms of the equivalent Army Standard Scores.

Figure 19. Mean performance for replicated cells. (reliability = .7, intercorrelation = .4)

QUANTITATIVE ASSAY FOR CRUDE ANTHRAX TOXINS*

Bertram W. Haines, Frederick Klein, and Ralph E. Lincoln
U. S. Army Biological Laboratories
Fort Detrick, Frederick, Maryland

ABSTRACT. The whole crude toxins of *Bacillus anthracis*, although apparently responsible for the death of animals with anthrax, had never been quantitated. A total of 14 lots of the toxic culture filtrate of *B. anthracis* were pooled into one large lot of crude anthrax toxins. An extensive assay of this reference material was conducted in four laboratories by use of the time-to-death of the intravenously challenged Fischer 344 rat as the response variable. Doses of the material were varied factorially by concentration, dilution, and volume. The data from this study were used to define a potency unit of the crude anthrax toxins. Procedures were developed and illustrated for the assay of unknown lots of the toxins by comparing the rate time-to-death response to the unknown with either (i) the responses reported in this study, or (ii) directly with the rat responses to a new sample of the reference toxins. The possibilities and limitations of this standardization and of the statistical procedure through which it was developed are discussed.

INTRODUCTION. The excellent work of Smith, Keppie, and Stanley (1955a), demonstrating the toxins of *Bacillus anthracis* organisms in the blood from guinea pigs in the terminal stages of anthrax, rekindled interest in the disease, particularly its toxins. (The toxic methabolic by-products of the growth of *B. anthracis* are composed of components with different biological or chemical properties. Naturally produced combinations of these components in unknown proportions will be referred to in this paper as "toxins.") To date, valid comparisons of results among the several experimenters (Smith et al., 1955a, b, 1956; Smith and Gallop, 1956; Thorne, Molnar, and Strange, 1960; Stanley and Smith, 1961; Beall, Taylor, and Thorne, 1962; Klein et al., 1962; Keppie, Smith, and Harris-Smith, 1955; Eckert and Bonventre, 1963; Harris-Smith, Smith, and Keppie, 1958; Sargeant, Stanley, and Smith, 1960; Stanley, Sargeant, and Smith, 1960) who have reported work with the toxic materials produced by *B. anthracis* have been difficult, because either whole crude toxins or the several components have been assayed by different methods, in different assay animals, and with no reference standard of the toxins.

*This paper appeared previously in the Journal of Bacteriology, Volume 89, No. 1, pages 74-83. Permission of the editors of this journal to publish this paper in these Proceedings is appreciated.

This paper presents the results of studies to quantitate, in terms of defined potency units, the lethality of anthrax toxins in Fischer 344 rats. The authors developed a reference lot of stabilized freeze-dried crude anthrax toxins. This reference material was used in the study described here, and is available for other studies against which samples of anthrax toxins of unknown concentration can be assayed.

MATERIALS AND METHODS. Animals. Fischer 344 albino rats weighing 200 to 300 g were obtained from the Fort Detrick colonies of Frank Beall and Frederick Klein. Both colonies are maintained through brother-sister matings descended from the colony described by Taylor, Kennedy, and Blundell (1961). This weight range was chosen, because preliminary data indicated that the response time of rats that weigh more than 300 g was significantly greater than that of rats weighing more than 200, but less than 300, g. Further study on rats, carefully selected for weight, revealed no significant difference within the weight range of 200 to 300 g (Table 1). The analysis of variance is presented in Table 2.

TABLE 1
Response time in minutes of 27 rats injected with 1 ml of crude anthrax toxins by weight of rat.

Weight (g) of rat		
200	250	300
99	102	100
97	81	94
96	80	88
94	79	105
93	78	90
92	114	101
89	76	78
88	102	82
87	71	86
835*	783	824
92.6**	84.9	90.7

* Totals
** Harmonic means.

TABLE 2
Analysis of variance of reciprocal response times recorded in Table 1

Source	df*	Sum of squares	Mean square	F
Between weights	2	.0485	.0242	1.50**
Within weights	24	.3859	.0161	
Total	26	.4344		

* Degrees of freedom.
** Not significant.

Rat lethal test. Toxins of B. anthracis were injected into the dorsal vein of the penis of the Fischer rat. In describing this test, Beall et al. (1962) noted a definite relationship between the dose of the toxins injected and time-to-death.

Antiserum. Equine hyperimmune serum (DH-1-6C) prepared by repeated injections of spores of the Sterne strain of B. anthracis, was used (Thorne et al., 1960).

Preparation of anthrax toxins. The medium used was described by Thorne et al. (1960), and was made with triple-distilled water. Subsequent to his original description, Thorne (personal communication) has suggested some changes. The medium used in this study was as follows.

Nine stock solutions (A, B, C, D, E, F, G, H, and I) were prepared. All stock solutions may be stored at 4 C for indefinite periods of time. Solution A contained $CaCl_2 \cdot 2H_2O$, 0.368 g/500 ml of water; B contained $MgSO_4 \cdot 7H_2O$, 0.493 g/500 ml of water; C contained $MnSO_4 \cdot H_2O$, 0.043 g/500 ml of water; D contained adenine sulfate, 0.105 g, and uracil, 0.070 g (both solids were dissolved in 100 ml of water, and the total volume was made up to 500 ml).

Solution E contained thiamine HCl, 0.025 g/500 ml of water; F contained tryptophan, 2.600 g; cystine, 0.600 g; and glycine, 0.750 g. The solids in solution F were dissolved as follows. Tryptophan was dissolved in 6 ml of 6 N HCl. Cystine was dissolved in 100 ml of water. Glycine was dissolved in 150 ml of water. These three solutions were combined, and water was added to bring the total volume up to 500 ml.

Solution G contained KH_2PO_4, 34.0 g/500 ml of water; H contained K_2HPO_4, 43.6 g/500 ml of water; I contained charcoal (Norit A), 3.75 g/500 ml of water.

A 10-ml amount of each stock solution, except that containing charcoal, was added to a suitable container; and 3.6 g of Casamino Acids (Difco) were added. The volume was brought up to 1 liter with triple-distilled water, and the pH of the medium was adjusted to 6.9 with 1 N H_2SO_4 or 1 N NaOH as needed. A 460-ml amount of this preparation was dispensed into a 3-liter Fernbach flask; 2 ml of charcoal suspension were added, and the preparation was autoclaved for 20 min at 15 psi.

Inoculation procedure. A 5-ml amount of 20% glucose (sterilized by filtration) was added to the Fernbach flask containing 460 ml of sterilized basal medium. Each flask of final medium was inoculated with 2×10^6 Sterne strain spores. The inoculated flasks were incubated statically for 23 to 27 hr at 37 C; 4 hr after inoculation 55 ml of 9% $NaHCO_3$ were added to each flask.

This final culture was centrifuged at 3,000 X g for 30 min. The supernatant fluid was decanted, and 10% horse serum was added. The solution was then sterilized by filtration through an ultrafine glass filter.

A preliminary test, to determine the potency of each of 14 toxic filtrates, was done by injecting 1-ml samples of each filtrate intravenously into two rats. The response (death) times of the rats were considered as indications of the toxicity of each batch. The total volume per batch and the response times of the test rats are given in Table 3.

The 14 toxic filtrates were combined, and a second preliminary test was conducted on the pooled material. The two rats used in this test died in 104 and 117 min, with a mean response time of 110.5 min. Both response times are within one standard deviation of the mean of all batches.

The pooled toxins were dispensed into 600 drying ampoules (40 ml), each containing 10 ml of toxins. Ampoules were shell-frozen in Dry Ice and alcohol (-79 C). Frozen ampoules were placed on an Aminco Dryer (American Instrument Co., Silver Spring, Md.), and dried under vacuum

TABLE 3
Volume per batch and response time of rats challenged with toxins by batch

Batch	Total volume	Response time (min)		
	ml	Rat A	Rat B	Mean
1	450	97	92	94.5
2	450	107	91	99.0
3	450	97	96	96.5
4	460	95	—*	95.0
5	420	122	124	123.0
6	450	114	125	119.5
7	510	116	90	103.0
8	410	121	120	120.5
9	370	88	82	85.0
10	510	90	94	92.0
11	465	106	94	100.0
12	425	106	92	99.0
13	425	117	121	119.0
14	300	100	117	108.5
Total	6,095			103.9**

* Missed the vein.
** SD = 12.14.

of 10 to 30 μ of mercury for 18 to 25 hr. Ampoules were sealed under vacuum, packed in cardboard containers, and stored at -20 C. A third preliminary test was conducted at this point. One randomly selected ampoule was reconstituted with 10 ml of triple-distilled water. A 1-ml amount of this toxic material was assayed in each of five rats. Their mean response time was 117.2 min. To further test the toxicity, 0.2 ml of undiluted and of serial twofold dilutions of the reconstituted material was injected intradermally into the shaven sides of a guinea pig, and observed for edematous reaction. The material reacted at a dilution of 1:32, and can be expressed according to Thorne et al. (1960) as containing 32 toxic units. Additional vials were reconstituted to 4X concentration, and tested on immunodiffusion plates against the standard spore antiserum (Thorne et al., 1960). Three individual lines of precipitate appeared in parallel arrangement when tested with a linear pattern. The strongest precipitate line was identified as the protective antigen (factor II) component when compared with a standard (Beall et al., 1962). An undiluted sample of the resuspended material had a protective antigen titer of 1: 64 against the standard spore antiserum.

Reference toxins. These preliminary tests constituted quality control measures on the remaining 597 vials of dried toxic filtrate. As a result of these tests, it was known that these vials contained the known components of anthrax toxins.

Procedures. The toxins were assayed independently by each of four investigators. The procedures followed by each of the four were as similar as possible.

The characterization of the dose-response relationship of the toxins in Fischer rats was based on an assay in which the two dose factors of amount and concentration of toxins were each tested at several levels as follows: (i) five levels of the amount of toxins designated as 4 ml, 2 ml, 1.5 ml, 1 ml, and 0.5 ml; (ii) seven levels of the concentration of the toxins designated as 4X, 2X, 1X, 0.5X, 0.25X, 0.125X, and 0.0625X, where 1X is defined as the concentration resulting when 1 ampoule is reconstituted to 10 ml with a diluent of triple-distilled water. Dilutions beyond 1X were made with distilled water plus 10% normal horse serum.

The 7 X 5 factorial combinations of the several levels of these two factors, plus 19 control groups, were each tested in two Fischer rats by each of four investigators (Table 4). Three sets of control animals are not shown in Table 4. The first set included five pairs of rats. Each pair was inoculated with one of the five amounts of diluent along (i.e., triple-distilled water plus 10% normal horse serum) to provide assurance that their companion animals responded to toxins as opposed to the inoculation of the diluents. The second set included seven pairs of animals. Each pair in this set was inoculated with 1.5 ml of one of the seven concentrations of toxins mixed with 0.5 ml (1/3 by volume) of specific antiserum (Thorne et al., 1960). The seven pairs of animals in the third set of controls were inoculated with 1.5 ml of one of the seven concentrations of toxins mixed with 0.5 ml of normal horse serum. These animals provided assurance that the control no. 2 animals that lived were saved by the antiserum specific against anthrax toxins.

TABLE 4
Response times in minutes of 280 Fischer rats by dose, concentration, technician, and rat

Concn	Technician	4*		2*		1.5*		1*		0.5*	
		A**	B	A	B	A	B	A	B	A	B
4X	1	58	55	53	54	57	57	61	60	76	71
	2	53	61	54	52	64	63	64	63	85	70
	3	57	62	56	52	58	56	64	62	78	72
	4	60	52	448	53	59	123	63	59	81	82
2X	1	57	57	61	63	59	61	72	70	100	89
	2	57	55	65	62	74	65	84	77	119	94
	3	50	56	56	58	66	77	72	78	109	117
	4	67	56	55	65	67	S***	127	S	107	83
1X	1	53	55	70	69	119	70	90	91	127	159
	2	73	64	78	72	82	81	61	100	181	199
	3	65	62	77	80	89	83	107	97	293	483
	4	S	63	S	S	S	100	132	S	161	202
0.5X	1	70	77	153	143	129	134	145	148	S	S
	2	74	83	114	103	138	131	425	281	S	S
	3	75	69	113	118	137	151	1588	244	S	S
	4	74	94	S	139	149	S	S	400	S	S
0.25X	1	111	112	173	176	S	481	S	S	S	S
	2	136	176	295	274	S	S	S	S	S	S
	3	103	124	S	300	S	S	S	S	S	S
	4	S	118	S	S	S	S	S	S	S	S
0.125X	1	185	195	S	S	S	S	S	S	S	S
	2	253	588	S	S	S	S	S	S	S	S
	3	473	234	S	S	S	S	S	S	S	S
	4	S	S	S	S	S	S	S	S	S	S
.0625X	1	S	S	S	S	S	S	S	S	S	S
	2	S	S	S	S	S	S	S	S	S	S
	3	S	S	S	S	S	S	S	S	S	S
	4	S	S	S	S	S	S	S	S	S	S

* Dose expressed in milliliters.
** Rat A or B.
***S indicates survival.

Each investigator required 32 ampoules of dried toxins. Each of the 32 ampoules was opened, and reconstituted with 2.5 ml of diluent precooled to 4 C. The contents of all 32 ampoules were then pooled, providing a total of 80 ml of reconstituted toxins at a concentration of 4X (4 times the original). All concentrations of toxins were maintained continuously at 4 C. To make the next dilution, 40 ml of the pool (4X) were combined with 40 ml of diluent (triple-distilled water). This provided 80 ml of toxins at a concentration of 2X. Further serial twofold dilutions were made to 0.0625X (1/16 X original concentration) and inoculated as planned.

Each investigator required 108 rats. These rats were caged in 54 consecutively numbered cages, each containing two animals. Each of the 54 treatment combinations was given to the two animals in one cage at the same time. The order of the treatments was randomized for each investigator. Response times-to-death, in minutes, were recorded for each rat and constituted the basic data.

RESULTS. The response times for animals are presented in Table 4. Although none of the controls appears in this table, none of either the first or second groups of control animals died. Some animals in the third control group challenged with 1.5 ml of toxins plus normal horse serum responded nearly the same as test animals challenged with 1.5 ml of toxins. The mean response times, in minutes, of these control animals by concentration of toxins are recorded in Table 5. The pattern of responses by the controls provided the needed assurance that the response of the test animals was specifically to the toxins of *B. anthracis*.

TABLE 5
Mean response time by dose and concentrations of toxins

Concn of toxin	Dose (ml) 4	2	1.5	1	0.5	Mean	Control*
4X	57.5	53.5	59.0	62.3	75.0	60.7	60.0
2X	55.2	60.7	66.4	75.2	105.1	69.0	70.0
1X	61.3	74.1	85.1	88.0	198.7	86.3	134.0
0.5X	74.4	121.6	136.3	247.0	S**	151.3	154.0
Mean	61.3	70.3	78.3	89.4	143.5	91.3	

*Control was 1.5 ml of toxins plus normal horse serum.
**All animals survived.

In spite of carefully controlled procedures and techniques, the results from one laboratory (technician 4) were so erratic that they were disregarded in any further analysis. Inspection of these data showed that technician 4 was the only one having reversal of results; i.e., a greater amount of toxins not killing and lesser amounts killing, or only one of the two test animals responding (except at doses eliciting a response above 300 min). These extremely variable results indicated that adequate controls on technique and environment were not maintained in this laboratory.

The reciprocals of the response times were used for analysis, because reciprocal response times are nearly normally distributed with equal variances, whereas the untransformed response times are positively skewed with unequal variances (Finney, 1952). The analysis of variance on the reciprocal response times of 120 rats from the four highest concentrations and the five doses is shown in Table 6. From this analysis it was seen that both dose level and concentration had statistically significant effects on the response time of Fischer rats injected intravenously with anthrax toxins.

TABLE 6

Analysis of variance of reciprocal response times

Line no.	Effect	df	Sum of squares	Mean square	F*
1	Dose (D)	4	11.9272	2.9818	229.37 *
2	Concentration (C)	3	16.5629	5.5210	424.69**
3	Technician (T)	2	0.1543	0.0772	5.94***
4	D X C	12	1.7984	0.1499	11.53**
5	D X T	8	0.1485	0.0186	1.43
6	D X T	6	0.1180	0.0197	1.52
7	D X C X T	24	0.6452	0.0269	2.07
8	Error	60	0.7814	0.0130	
9	Total	119	32.1360		

* Error line 8 was used to test all effects.
** Approximate probabilities < 0.001.
***Approximate probabilities < 0.05.

The analysis further showed an interaction between dose and concentration to be statistically significant. The mean response times by does and concentration of toxins are given in Table 5. From the tabled means, it can be seen that the magnitude of this interaction is slight and had no practical significance in the further analysis and interpretation of these data.

The analysis also showed a statistically significant difference among technicians. Inspection of the data showed that mean response times for all rats responding for technicians 1, 2, and 3 were, respectively, 78, 83, and 83 min. This is a practically unimportant difference which we believe may in part be due to environmental factors, because genetic differences would be almost nil after 100 generations of inbreeding. The rats used by technician 1 came from the Beall colony, which was maintained in a different environment than the Klein colony animals used by the other two technicians. This raised the question as to the effect on this assay of Fischer rats procured from non-Detrick sources. To examine this effect, commercially available Fischer rats obtained from two breeders were tested and found to be suitable for this assay. In this study, 20 Fischer 344 rats from each of two suppliers (Microbiological Associates, Inc., Bethesda, Md.; and Charles River Breeding Laboratories, Inc., Brookline, Mass.) were challenged in each of two laboratories. The response times of all 80 rats are reported in Table 7. No statistically significant difference in times of response for animals from the two suppliers was observed. A difference between the two operators and the interaction of operator X supplier was statistically significant at the 5% level. The mean response time of three of the four groups differed by less than 1 min, and the fourth group differed by approximately 5 min. This difference of about 5 min between these two groups could be caused by a difference of about seven units of toxins, which is well within the 95% confidence limits of an estimated potency. Thus, this difference, although statistically significant, was considered of no consequence concerning this assay.

A test to determine the storage characteristics of the reference toxins was conducted on a vial of the toxins which had been stored for 36 months. The test vial was reconstituted with 10 ml of triple-distilled water. Six rats were then challenged with these reconstituted toxins, according to the protocol described in this paper.

TABLE 7
Response times in minutes by supplier, operators, and rats

Rats	Charles River Breeding Labs., Inc.		Microbiological Associates, Inc.	
	1*	2	1	2
1	83	87	91	85
2	88	84	84	89
3	86	86	91	89
4	83	82	88	85
5	91	84	89	92
6	87	89	88	84
7	94	88	90	101
8	88	83	92	87
9	87	83	96	102
10	91	86	77	87
11	105	83	89	93
12	94	85	94	79
13	92	79	90	107
14	90	81	91	88
15	98	81	91	83
16	91	85	77	90
17	82	83	97	89
18	90	87	89	88
19	83	85	82	75
20	88	83	90	86
Harmonic mean response time	89.28	84.10	88.50	88.42

*Operator number.

The estimate of potency from that test was 32.4 potency units per ml at the 1X concentration. This was essentially identical to the 32 units per ml set up in the definition. Therefore, it was concluded that the reference toxins had not changed with respect to potency during 36 months of storage.

Development of procedures for direct assay method. A potency assay should be based on dose expressed in terms of well-defined units. No such units have as yet been defined for anthrax toxins. Varying the amount of toxins by varying either dose or concentration would have a significant effect on the response time of rats; however, rats injected with 1 ml of toxins concentrated to 2X responded in about the same time (75 min) as rats injected with 2 ml of toxins concentrated at 1X (74 min). This relationship holds true for most other dose-by-concentration combinations for which the product of these two factors is a constant. If doses are converted into 0.5-ml units, and concentrations into 0.0625 units, then the doses and concentrations in Table 4 can be expressed as shown in Table 8.

TABLE 8
Derivation of potency units of anthrax toxins

Concn of toxins in 0.0625-fold units	Dose of toxins in 0.5-ml units				
	8	4	3	2	1
64	512	256	192	128	64
32	256	128	96	64	32
16	128	64	48	32	16
8	64	32	24	16	8
4	32	16	12	8	4
2	16	8	6	4	2
1	8	4	3	2	1

The products of the marginal numbers in Table 8 for any two equivalent dose-by-concentration combinations are the same; thus, the product of two dose units and 32 concentration units gives 64 total potency units of toxins. Similarly, four dose units of 16 concentration units also contain 64 total potency units of toxins. We define the potency unit of anthrax toxins to be expressed as these products of dose by concentration of this particular lot of toxins.

If we were to carry the definition of a potency unit no further, then 1 ml of 1X concentration of any anthrax toxins, regardless of its actual effect in animals, would have 32 potency units. To standardize a potency unit, it is necessary to describe the association between the dose, in units, and the potency, in terms of a biological response to this particular lot of anthrax toxins. The potency of any other lot of toxins may then be measured by comparing the response to a known amount of the test toxins with the response to the same amount of the reference toxins.

These response characteristics were described as the dose-response relationship when measured doses of these toxins were injected intravenously into Fischer 344 rats. The challenged rats responded by dying at a time that is shown here to be highly dependent on the dose measured in potency units of these toxins.

The regression of mean reciprocal response times on the $\log_2$ of the potency units of anthrax toxins is shown in Figure 1. The least squares line has the equation:

$$Y = b_0 + b_1X + b_2X^2 \quad (1)$$

where Y is the mean reciprocal response time, X is the potency of anthrax toxins in $\log_2$ units, and the b values are regression coefficients computed from the data of this test. The values of the coefficients, their variances and covariances, are: $b_0 = -2.591$; $b_1 = 0.959$; $b_2 = -0.051$; $V(b_0) = 0.077121$; $V(b_1) = 0.009514$; $V(b_2) = 0.000068$; $V(b_0b_1) = -0.026902$; $V(b_0b_2) = 0.002238$; $V(b_1b_2) = -0.000800$. This regression line represents a basis upon which comparisons of potency of anthrax toxins can be made. Thus, test toxins can be assayed either indirectly against this curve, or directly with parallel assays of the reference toxins.

Development of procedures for indirect assay method. To use the responses of 120 rats to the reference toxins [for which the slope of response from the regression data (Figure 1) has been calculated], we recommend use of the indirect method for standardizing unknown potencies of anthrax toxins. The regression was nearly linear for doses from 16 to 128 units, corresponding to response times from 240 to 65 min. Thus, although the concentration of test or unknown toxins is arbitrary, it should be of such concentration that 1 ml, injected intravenously, will kill a Fischer rat in not less than 65 min, nor more than 240 min.

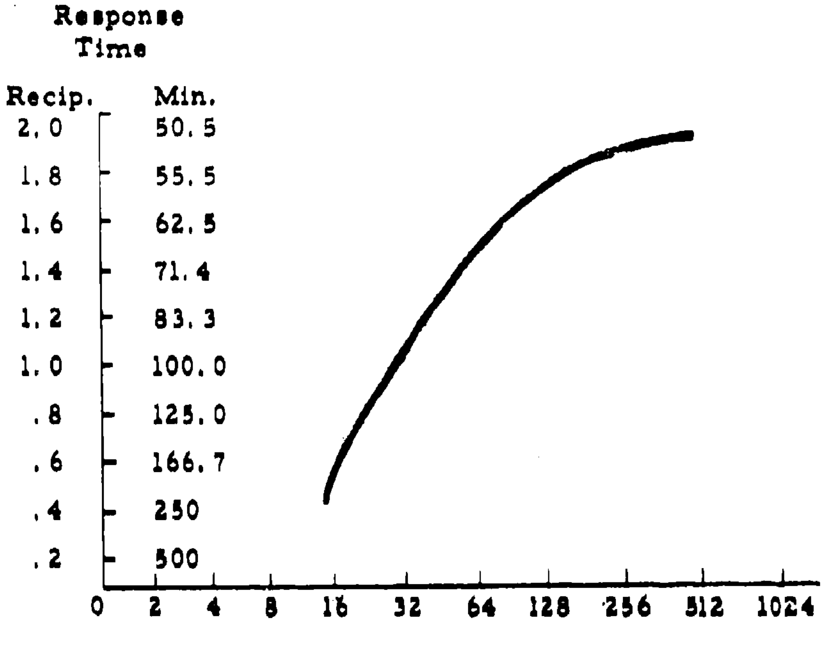


Figure 1. Regression of reciprocal response time of Fischer rats on log dose of anthrax toxins expressed in potency units.

To test the potency of test or unknown toxins, enough animals should be used so that the amount of variation in the final result, that can be attributed to the test rats, is at least no greater than the amount of variation contributed by the standard rats. Thus, at least six Fischer rats of 200 to 300 g from a suitable colony should be intravenously inoculated, three with 2 ml of the test toxins, and three with 1 ml.

The test is based on the mean reciprocal response times of the rats. (The rat response is very uniform; thus, any observed nonresponse must be considered the result of technique at some stage of the assay procedure.) This is simply the sum of reciprocal times-to-death of the rats in minutes (100/t) with the average time calculated. The reciprocal response times of the rats can be put in the following form:

Reference Toxins
$Y = 100/t$

	1 ml		2 ml	
Rat	1.	______	4.	______
	2.	______	5.	______
	3.	______	6.	______
	ΣY	______	ΣY	______
	$\bar{Y} = R_1$		R_2	______

$R_1 + R_2 =$ ______

Test Toxins
$Y = 100/t$

	1 ml		2 ml	
Rat	1.	______	4.	______
	2.	______	5.	______
	3.	______	6.	______
	ΣY	______	ΣY	______
	$\bar{Y} = T_1$	______	T_2	______

$T_1 + T_2 =$ ______

where R_1, R_2, T_1, and T_2 are mean reciprocal response times. This form for calculation can be used for either the direct or indirect assay method.

The estimate of the difference in potency (D) between the test toxins and the reference can be found as:

(2) $$D = \frac{(T_1 + T_2) - (R_1 + R_2)}{2L}$$

where the letters T and R represent the mean reciprocal response times from the table above, and L is the average slope of the reference dose-response curve at the two dose levels used in the test. This average slope may be calculated as:

$$L = b_1 + b_2 (X_1 + X_2) \quad (3)$$

where X_1 and X_2 are the dose levels of the reference toxins (in $\log_2$ potency units) that were used in the test, and b_1 and b_2 are the estimates of the regression coefficients from equation 1. When the test is run using 1- and 2-ml doses of toxins, then $X_1 = 5$ and $X_2 = 6$. Under these conditions $R_1 = 0.92$, $R_2 = 1.34$ from equation 1, and $L = 0.3985$ from equation 3, so that equation 2 becomes:

$$D = \frac{(T_1 + T_2) - 2.26}{0.7970} \quad (4)$$

where the letter D represents the amount of difference between the test and reference toxins in terms of $\log_2$ potency units. If D is positive, then the test toxins are more potent than the reference, whereas, if D is negative, the test toxins are less potent than the reference. The reference toxins have a potency of 5 $\log_2$ units per ml at a concentration of 1X; thus, the potency (P) of the test toxins in $\log_2$ units at the concentration tested will be found as:

$$P = 5 + D \quad (5)$$

To find the number of potency units per ml of the test toxins, its potency needs to be converted from $\log_2$ units to $\log_{10}$ units. The conversion formula is:

$$\log_{10} P = \log_2 P \log_{10} 2$$

The value of P in units is found by looking up the antilog of this product. This value will be the number of potency units per milliter of the test toxins at the concentration tested.

Estimation of variance. There is variation inherent in this assay system in addition to the variation between samples of toxins. Thus, the single estimates of the potency of any particular sample of an unknown toxin should be bounded by confident limits. To determine these limits it is necessary to calculate the variance (V) of the estimate D of the $\log_2$ of the difference in potency between the test and the reference. The variance of the estimate D will depend on the variances of the observed response times and of the regression.

If we express D as N/G where

(6) $$N = (T_1 + T_2) - (R_1 + R_2)$$

and

$$G = 2L$$

then the variance of D can be expressed as:

(7) $$V(D) = \frac{1}{4L^2} \left\{ V(N) + D^2 V(G) \right\}$$

which will apply, because N and G are estimated from independent observations (Finney, 1952). The four mean reciprocal response times are stochastically independent; thus, the estimate of V(N) can be expressed as:

(8) $$V(N) = V(R_1) + V(R_2) + V(T_1) + V(T_2)$$

where $V(T_1)$ and $V(T_2)$ are obtained directly from the data of the test, and $V(R_1)$ and $V(R_2)$ are calculated from the regression line as:

$$V(R_i) = V(\overline{Y}) + (X_i - \overline{X})^2 V(b_i) \tag{9}$$
$$+ (X_i^2 - \overline{X}^2)^2 V(b_2),$$

The variance of G is given by the equation:

$$V(G) = 4\{V(b_1) + (X_1 + X_2)^2 V(b_2) \tag{10}$$
$$+ (X_1 + X_2)\, V(b_1 b_2)\},$$

When the test is run using 1- and 2-ml doses of toxins, then $X_1 = 5$ and $X_2 = 6$. Under these conditions:

$$V(R_1) = 0.0134, \quad V(R_2) = 0.0018$$

and

$$V(G) = 0.0355$$

so that:

$$V(D) = \frac{1}{0.6352} \{V(N) + 0.0355D^2\} \tag{11}$$

and:

$$V(N) = 0.0134 + 0.0018 + V(T_1) + V(T_2). \tag{12}$$

Example. A sample of toxins of unknown potency was tested in this laboratory. It was known to kill Fischer rats in slightly more than 90 min when injected intravenously in doses of 1 ml at a concentration of 1X. The response of the unknown toxins was compared with the response curve described by equation 1. Each of three Fischer rats was injected with 1 ml of the test toxins, and their reciprocal response times in minutes were recorded (Figure 2). Three other Fischer rats were each injected intravenously with 2 ml of the test toxins. Their reciprocal response times were also recorded (Figure 2). From these six reciprocal response times, values of T_1 and T_2 were calculated. Corresponding values of R_1 and R_2 were obtained from the regression line by substituting, respectively, the values 5 and 6 for X in equation 1. The value of L was calculated from equation 3 by use of the values 5 and 6 for X_1 and X_2. The values 5 and 6 were used in these two cases, because they are the $\log_2$ of the number of units in 1 and 2 ml of the reference toxins.

The value of D was calculated by substituting the previously calculated values of R_1, R_2, T_1, T_2, and L in equation 2. This value of D was found to be 0.78. This indicates that the test toxins were 0.78 $\log_2$ unit more potent than the reference. A 1-ml amount of the reference toxins contains 5 $\log_2$ units, so the test toxins must contain 5.78 $\log_2$ units. Thus, the test toxins have 55.0 potency units per ml at the concentration tested (5.78 X .301 = 1.73978 $\log_{10}$ units).

The formulas for calculating the variance of the estimate D of the $\log_2$ of the difference in potency between the test and the reference are described above as equations 6 through 10. These calculations were made in this example, and it was found that SE (D) = 0.26. Using normal theory, the 95% confidence limits of D become UL(D) = 1.30, and LL(D) = 0.26. From these the 95% confidence limits of P were calculated as UL(P) = 79.4 units per ml, and LL(P) = 38.0 units per ml.

DISCUSSION. Anthrax toxins are composed of at least three factors, I, II, and III, by the classification of Stanley and Smith (1961, 1963) or, respectively, edema factor, protective antigen, and lethal factor according to Beall et al. (1962). Both in vitro-produced toxins, as used in this report, and in vivo toxins, as reported by Klein et al. (1963), may be quantitated accurately. The procedure further provides an effective reference for quantitating natural resistance or relative immunity as described by Klein et al. (1963), because the absolute dose of toxins required to elicit a given response will bear a definite relationship to host resistance or susceptibility.

	Reference Toxin $Y = 100/t$ 1 ml.	2 ml.		Test Toxin $Y = 100/t$ 1 ml.	2 ml.
Rat 1			Rat 1	1.39	1.67
Rat 2			Rat 2	1.25	1.56
Rat 3			Rat 3	1.15	1.59
ΣY			ΣY	3.79	4.82
$\bar{Y} = R_i$	0.92	1.34	$\bar{Y} = T_i$	1.26	1.61
$R_1 + R_2 =$	2.26		$T_1 + T_2 =$	2.87	
ΣY^2			ΣY^2	4.8171	7.7506
$V(R_i)$	.0134	.0018	$V(T_i)$	.0048	.0011

$b_0 = -2.5912$
$b_1 = .9592$
$b_2 = -.0510$

$V(b_0) = .07712089$
$V(b_1) = .00951356$
$V(b_2) = .00006804$
$V(b_1 b_2) = -.000800$

$L = b_1 + b_2(x_1 + x_2)$

$x_1 =$ 5 $x_2 =$ 6 $(x_1 + x_2) =$ 11 $(x_1 + x_2)^2 =$ 121

$b_1 = 0.9592$
$b_2(x_1 + x_2) = 0.5607$
$L = 0.3985$
$2L = 0.7970$
$4L^2 = 0.6352$

$$D = \frac{(T_1 + T_2) - (R_1 + R_2)}{2L} = \frac{2.87 - 2.26}{0.7970} = 0.78$$

$D^2 = 0.6084$

$\text{Log}_2 P = 5 \pm D = 5$

$\text{Log}_{10} P = 0.301 \times 5.78 = 1.74$ $P =$ **55.0 U/ml**

$V(G) = 4\{V(b_1) + (x_1 + x_2)^2 V(b_2) + (x_1 + x_2) V(b_1 b_2)\} =$ 0.0335

$V(N) = V(R_1) + V(R_2) + V(T_1) + V(T_2) =$ 0.0211

$V(D) = \frac{1}{4L^2} \{V(N) + D^2 V(G)\} =$ 0.0672

SE (D) = 0.26

UL (D) = 1.30 Log_{10} UL(P) = 1.90 UL(P) = 79.4

LL (D) = 0.26 Log_{10} LL(P) = 1.58 LL(P) = 38.0

FIG. 2. *Calculation form for potency of anthrax toxins.*

The biological activities of these compounds are numerous, and it is likely that some responses are still to be discovered. The problem of evaluating activity and mode of action of compounds which have a synergistic biological action is more difficult than for "single compounds." Quantitation, therefore, is important to allow the work of various investigators to be related more exactly to each other. The Fischer 344 rats are commercially available, and reference anthrax toxins will be provided for responsible investigators who desire to work with this material for use in establishing units. The methods used in this standardization of these toxins may be appropriate to the standardization of other biologically active toxins.

ACKNOWLEDGEMENTS. We thank Martha K. Ward, U. S. Public Health Service, and Dorothy M. Molnar for their cooperation and independent assay of the toxins, and Francis A. Beall who supplied his own animals and, in addition, independently assayed the toxins. Thanks are also extended to Bill G. Mahlandt and Charles C. Wigington for producing the anthrax toxins in vitro, and preparing them for storage.

LITERATURE CITED

Beall, F. A., M. J. Taylor, and C. B. Thorne. 1962. Rapid lethal effect in rats of a third component found upon fractionating the toxin of *Bacillus anthracis*. J. Bacteriol. 83:1274-1280.

Eckert, N. J., and P. F. Bonventre. 1963. *In vivo* effects of *Bacillus anthracis* culture filtrates. J. Infect. Dis. 103:226-232.

Finney, D. J. 1952. Statistical method in biological assay. Hafner Publishing Co., New York.

Harris-Smith, P. W., H. Smith, and J. Keppie. 1958. Production *in vitro* of the toxin of *Bacillus anthracis* previously recognized *in vivo*. J. Gen. Microbiol. 19:91-103.

Keppie, J., H. Smith, and P. W. Harris-Smith. 1955. The chemical basis of the virulence of *Bacillus anthracis*. III. The role of the terminal bacteremia in death of guinea pigs from anthrax. Brit. J. Exp. Pathol. 36:315-322.

Klein, F., B. W. Haines, B. G. Mahlandt, I. A. DeArmon, Jr., and R. E. Lincoln. 1963. Dual nature of resistance mechanisms as revealed by studies of anthrax septicemia. J. Bacteril 85: 1032-1038.

Klein, F., D. R. Hodges, B. G. Mahlandt, W. I. Jones, B. W. Haines, and R. E. Lincoln. 1962. Anthrax toxin: causative agent in the death of rhesus monkeys. Science 138: 1331-1333.

Sargeant, K., J. L. Stanley, and H. Smith. 1960. The serological relationship between purified preparations of factor I and II of the anthrax toxin produced in vivo and in vitro. J. Gen. Microbiol. 22: 219-228.

Smith, H., and R. C. Gallop. 1956. The chemical basis of the virulence of Bacillus anthracis. VI. An extracellular immunizing agressin isolated from exudates. Brit. J. Exp. Pathol. 37: 144-155.

Smith H., J. Keppie, and J. L. Stanley. 1955a. The chemical basis of the virulence of Bacillus anthracis. V. The specific toxin produced by B. anthracis in vivo. Brit. J. Exp. Pathol. 36: 460-472.

Smith, H., J. Keppie, J. L. Stanley, and P. W. Harris-Smith. 1955b. The chemical basis of the virulence of Bacillus anthracis. IV. Secondary shock as a major factor in death of guinea pigs from anthrax. Brit. J. Exp. Pathol. 36: 323-255.

Smith, H., D. W. Tempest, J. L. Stanley, P. W. Harris-Smith, and R. C. Gallop. 1956. The chemical basis of the virulence of Bacillus anthracis. VII. Two components of the anthrax toxin: their relationship to known immunizing aggressins. Brit. J. Exp. Pathol. 37: 263-271.

Stanley, J. L., K. Sargeant, and H. Smith. 1960. Purification of factors I and II of the anthrax toxin produced in vivo. J. Gen. Microbiol. 22: 206-218.

Stanley, J. L., and H. Smith. 1961. Purification of factor I and recognition of a third factor of the anthrax toxin. J. Gen. Microbiol. 26: 49-66.

Stanley, J. L., and H. Smith. 1963. The three factors of anthrax toxin: their immunogenicity and lack of demonstratable enzymatic activity. J. Gen. Microbiol. 31: 329-337.

Taylor, M. J., G. G. Kennedy, and G. P. Blundell. 1961. Experimental anthrax in the rat. I. The rapid increase of natural resistance observed in young hosts. Amer. J. Pathol. 38: 469-480.

Thorne, C. B., D. M. Molnar, and R. E. Strange. 1960. Production of toxin _in vitro_ by _Bacillus anthracis_ and its separation into two components. J. Bacteriol. 79: 450-455.

AN INVESTIGATION OF THE DISTRIBUTION OF DIRECT HITS ON PERSONNEL BY SELF-DISPERSING BOMBLETS*

David M. Moss and Theodore W. Horner
Booz·Allen Applied Research Inc.

ABSTRACT. The question has been raised concerning the lethal hazard to personnel from self-dispersing bomblets. The solution of this question involved the derivation of a distribution and the computation of parameters for a specific problem. The basic method used was to define a random variable, θ, the number of individuals which are hit;

$$\theta = \sum_{i=1}^{N} (1 - 0^{n_i}),$$

where N is total number of personnel and n_i is the number of bomblets striking the i^{th} individual. The moment-generating-function of this random variable was found and, hence, its distribution function. The distribution of casualties was found to be Poisson under the general assumptions of the problem.

* * * * * * * * * * * *

The question has been raised concerning the lethal hazard to personnel from self-dispersing bomblets by direct hits. In trying to determine the lethality of these bomblets many factors must be taken into account.

Among the factors which bear on this problem is that of protection. The flight of the bomblets might be intercepted by trees, buildings, or other natural or man-made obstructions, and would therefore descrease the chances of a lethal hit. In this study the interest is directed toward assessing the maximum hazard to personnel. It is, therefore, assumed that all personnel are completely exposed. It is also assumed that all personnel are in an upright position and no person provides any protection for another person. Thus, each person is completely and equally exposed to the possibility of a direct hit by a bomblet.

---

*Work on which paper is based was supported by contract with the U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland.

Other assumptions made in order to assess the maximum hazard are that all personnel are within the target area of interest and all bomblets hit somewhere within this area. It can also be assumed that the vulnerable portions of an individual are his head and neck. If other portions of the body are struck, it is assumed that lethal damage is not inflicted.

The objective here will be to determine the hazard to personnel on target resulting from a drop of self-dispersing bomblets. The distribution of the number of lethal hits resulting from such a drop will be determined and in addition the expected number of such hits and the associated variance will be found. The results found will reflect the maximum hazard involved.

In addition to the theoretical work done here, the results for a specific case will be given. This will be the case where 600 bomblets are dropped on a one square kilometer area which contains 4000 persons.

First it will be assumed that there are N individuals in the target area, A_T. There are n bomblets dropped, all of which land in the target area. Further it will be assumed that bomblets and individuals are uniformly and independently distributed in the target area; however, it will be shown later that the individuals may assume any distribution. It will also be assumed that individuals and bomblets can be represented by circles with areas given by

(1) $$A_p = \pi r_1^2$$

and

(2) $$A_b = \pi r_2^2,$$

where r_1 is the radius of the critical area of an individual and these areas for all individuals are considered to be the same, and r_2 is the radius of a bomblet. Now in order to produce a casualty, the center of a bomblet must fall within the circle with radius

(3) $$r = r_1 + r_2 \quad .$$

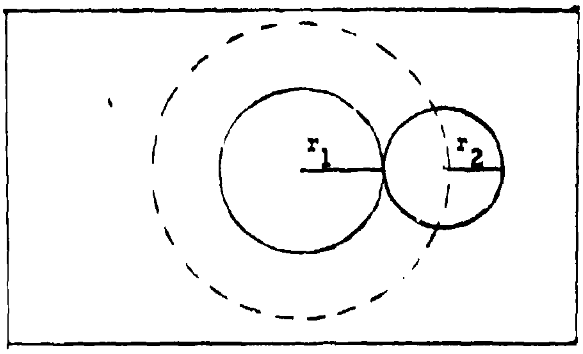


Figure 1
Casualty Radius Diagram

The target area can be divided up into N circular cells, each with radius $r = r_1 + r_2$ representing individuals, plus one cell which represents that part of the target in which there are no individuals. We can assign a value p_i to the probability that a bomblet falls in the i^{th} cell. Let n_i represent the number of bomblets that fall in the i^{th} cell. Then

(4) $$\sum_{i=1}^{N+1} p_i = 1$$

and

(5) $$\sum_{i=1}^{N+1} n_i = n ,$$

where n is the total number of bomblets.

The interest now is in the number of persons hit or the number of casualties, denoted here by θ. What is needed is a variable which will give the number of casualties, regardless of whether an individual is hit more than once. One such variable could be obtained by defining a variable which is either zero or one depending on whether an individual is missed or hit. If such a variable is then summed over all individuals, the result would be the total number of casualties, θ.

Note that

$$(6) \qquad 0^{n_i} = \begin{cases} 1 \text{ when } n_i = 0 \\ 0 \text{ when } n_i > 0 \end{cases}$$

and

$$(7) \qquad 1 - 0^{n_i} = \begin{cases} 0 \text{ when } n_i = 0 \\ 1 \text{ when } n_i > 0; \end{cases}$$

that is, if the number of hits of an individual is one or more, $(1-0^{n_i})$ will be one and will be zero otherwise. Thus, let us define our variable of interest as

$$(8) \qquad \theta = \sum_{i=1}^{N} (1-0^{n_i}) .$$

This variable tells us the number of individuals which are hit and it is about this random variable that we want more information.

Now, before going on, let's look more closely at our probabilities, where p_i $(i = 1, 2, \ldots, N)$ defines the probability of a hit of the individual in the i^{th} cell. Obviously, the probability that any particular bomblet hits any particular individual is the same for all bomblets and all individuals. Also it is quite clear that the probability of a randomly chosen bomblet from a uniform distribution of bomblets hitting any individual is equal to the ratio of the area, A_c, of the circle with radius r to the total target area, A_T.

Thus

$$(9) \qquad p = A_c/A_T ,$$

where

(10) $$A_c = \pi(r_1 + r_2)^2 .$$

Note that since

(11) $$\sum_{i=1}^{N+1} p_i = 1$$

and the p_i's, $i = 1, 2, \ldots, N$, are equal, we thus have

(12) $$p_{N+1} = 1 - Np .$$

What we have is essentially the probability of a randomly selected point being within a certain area. Note in Figure 2 that the probability that a randomly selected point lies in a given circle is the same in A and B and also that the probability of at least x of the n points lying within a circle is the same in both. Based on this it can be seen that our results will be independent of the distribution of personnel.

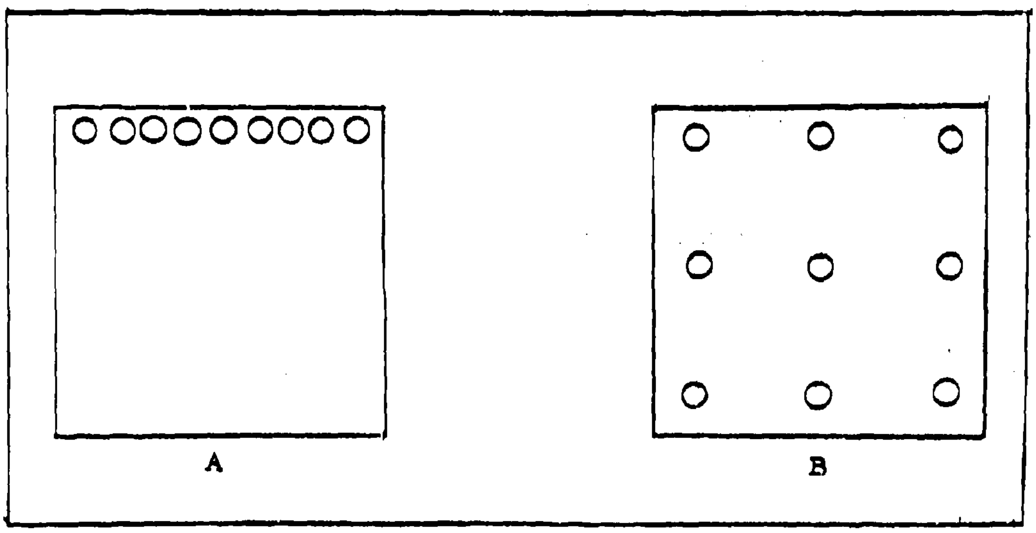


Figure 2
Possible Personnel Configurations

Now let us look at an analogous situation. Suppose that we have N + 1 cells into which we randomly throw n balls.

Table I
Distribution of Balls Falling into Cells

Cell	Probability	Number of Balls Falling Into Cell
1	$p_1 = p$	n_1
2	$p_2 = p$	n_2
.	.	.
.	.	.
.	.	.
N	$p_N = p$	n_N
N + 1	$p_{N+1} = 1-Np$	$n_{N+1} = n - \sum_{i=1}^{N} n_i$

This is a multinomial situation where

$$f(n_1, n_2, \ldots, n_{N+1}) = \left(n! \Big/ \prod_{i=1}^{N+1} n_i!\right) p_1^{n_1} p_2^{n_2} \ldots p_{N+1}^{n_{N+1}} . \quad (13)$$

Since it is θ in which we are interested, we need to discover the distribution of θ. The approach taken here will be to find the moment-generating-function of θ and from it the distribution of θ.

Recalling the definition of moment-generating-function from mathematical statistics and substituting for θ from equation (8), we have

$$M_\theta(t) = E\left\{e^{t\theta}\right\}$$

$$= E\left\{\exp\left[t\sum_{i=1}^{N}(1-0^{n_i})\right]\right\}$$

(14)

$$= e^{tN}E\left\{\exp\left[-t\sum_{i=1}^{N}0^{n_i}\right]\right\}$$

$$= e^{tN}E\left\{\prod_{i=1}^{N}e^{-t0^{n_i}}\right\}.$$

Now

(15) $$e^{-t0^{n_i}} = \begin{cases} e^{-t} & \text{when } n_i = 0 \\ 1 & \text{when } n_i > 0, \end{cases}$$

and equivalently

(16) $$e^{-t0^{n_i}} = 1 + 0^{n_i}(e^{-t}-1).$$

Note that (16) holds identically and that the right hand side is not part of a series expansion. Substituting back in (14), we have

(17)
$$M_\theta(t) = e^{tN} E \left\{ \prod_{i=1}^{N} \left[1 + 0^{n_i}(e^{-t} - 1) \right] \right\} .$$

Now let

(18)
$$b_i = 0^{n_i}(e^{-t} - 1)$$

and substitute in (17):

$$M_\theta(t) = e^{tN} E \left\{ \prod_{i=1}^{N} (1 + b_i) \right\}$$

(19)
$$= e^{tN} E \left\{ 1 + \sum_{i=1}^{N} b_i + \sum_{\substack{i \\ j>}} \sum_{j} b_i b_j \right.$$

$$+ \sum_{i} \sum_{j} \sum_{\substack{k \\ k>j>i}} b_i b_j b_k + \dots$$

$$\left. + \sum_{i} \sum_{j} \dots \sum_{\substack{m \\ m> \dots >j>i}} b_i b_j \dots b_m \right\} .$$

Now taking the expectation of a typical term, say the $g + 1^{st}$ term and substituting from (18), we have

$$T_{g+1} = E\left\{\underbrace{\Sigma\,\Sigma\,\ldots\,\Sigma}_{g}\; b_i b_j \ldots b_m\right\}$$

$$(20) \qquad = \Sigma\,\Sigma\,\ldots\,\Sigma\; E\{b_i b_j \ldots b_m\}$$

$$= \Sigma\,\Sigma\,\ldots\,\Sigma\; E\left\{(e^{-t}-1)^g 0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\}$$

$$= (e^{-t}-1)^g\, \Sigma\,\Sigma\,\ldots\,\Sigma\; E\left\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\}\;.$$

Now the expectation of the last factor in (20) is

$$E\,\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\} = \Sigma\; 0^{n_1} 0^{n_2} \ldots 0^{n_g} f(n_1, n_2, \ldots, n_{N+1})$$

which becomes upon substitution from (13)

$$E\left\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\} = \Sigma\; 0^{n_1} 0^{n_2} \ldots 0^{n_g} \frac{n!}{\Pi n_i!}\; p_1^{n_1} p_2^{n_2} \cdots p_{N+1}^{n_{N+1}}$$

$$(21) \qquad = \Sigma\,\frac{n!}{\Pi n_i!}\,(0 \cdot p_1)^{n_1}\,(0\, p_2)^{n_2} \ldots (0 \cdot p_g)^{n_g} \cdot p_{g+1}^{n_{g+1}} \ldots \ldots p_{N+1}^{n_{N+1}}$$

$$= \left(0 + 0 + \ldots + 0 + p_{g+1} + \cdots + p_N + p_{N+1}\right)^n$$

$$= (1 - gp)^n\;,$$

the last step following from the fact that

$$(22)\qquad \sum_{i=1}^{N+1} p_i = 1 .$$

Substituting the result from (21) back in (20) we get

$$(23)\qquad \begin{aligned} T_{g+1} &= (e^{-t}-1)^g \Sigma\Sigma \ldots \Sigma (1 - gp)^n \\ &= (e^{-t}-1)^g (1 - gp)^n \Sigma\Sigma \ldots \Sigma (1) \\ &= (e^{-t}-1)^g (1 - gp)^n \binom{N}{g} . \end{aligned}$$

Using this last result in equation (19), we now find the moment-generating-function to be

$$(24)\qquad \begin{aligned} M_\theta(t) &= e^{tN} \Big\{ 1 + N(e^{-t}-1)(1-p) \\ &+ \binom{N}{2} (e^{-t}-1)^2 (1-2p)^n + \binom{N}{3} (e^{-t}-1)^3 (1-3p)^n \\ &+ \ldots + (e^{-t}-1)^N (1-Np)^n \Big\} \\ &= e^{tN} \sum_{g=0}^{N} \binom{N}{g} (e^{-t}-1)^g (1 - gp)^n . \end{aligned}$$

The maximum value of gp is Np. However, Np is extremely small as seen from the example following the theory. Since, therefore, gp is extremely small,

(25) $$(1 - gp)^n \simeq e^{-npg};$$

which follows because

$$e^{-ngp} = (e^{-gp})^n$$

(26) $$= (1 - gp + \frac{(gp)^2}{2!} - \frac{(gp)^3}{3!} + \dots)^n$$

$$\simeq (1 - gp)^n.$$

Therefore

$$M_\theta(t) \simeq e^{tN} \sum_{g=0}^{N} \binom{N}{g} (e^{-t}-1)^g (e^{-np})^g (1)^{N-g}$$

$$= e^{tN} \sum_{g=0}^{N} \binom{N}{g} \left[e^{-np}(e^{-t}-1)\right]^g (1)^{N-g}$$

(27) $$= e^{tN} \left[e^{-np}(e^{-t}-1) + 1\right]^N$$

$$= \left[e^t e^{-np}(e^{-t}-1) + e^t\right]^N$$

$$= \left[e^{-np}(1-e^t) + e^t\right]^N$$

$$= \left[e^{-np} - e^t e^{-np} + e^t\right]^N$$

$$= \left[e^{-np} + e^t(1 - e^{-np})\right]^N.$$

Now in the above result let

$$Q = e^{-np} \tag{28}$$

and

$$P = 1 - e^{-np}$$

We then have

$$M_\theta(t) \simeq (Q + Pe^t)^N \tag{29}$$

which can be recognized as the moment-generating-function for the binomial distribution. Thus θ is approximately binomially distributed with parameters P, Q, and N. The expected value of θ or the mean number of casualties is given by

$$\begin{aligned} E\{\theta\} &= NP \\ &= N(1 - e^{-np}) \qquad (30) \\ &= N\left[1 - \left(1 - np + \frac{(np)^2}{2!} - \frac{(np)^3}{3!} + \ldots\right)\right] \\ &\simeq Nnp, \end{aligned}$$

the last step following since np is extremely small. Thus the $E(\theta)$ is small unless N is extremely large. Also because P is small, the distribution of θ can be approximated by a Poisson distribution and therefore the variance is also approximately Nnp. The distribution of θ, where θ is the number of casualties, is given by

$$p(\theta) = (Nnp)^\theta \; e^{-Nnp}/\theta! \tag{31}$$

Now let's look at the specific problem: namely that of dropping 600 bomblets on a one square kilometer target which contains 4000 personnel. It is given that:

a. $A_T = 10^{10}\ cm^2$ d. $n = 6 \times 10^2$

b. $A_p = 314\ cm^2$ e. $r_1 = 10\ cm$

c. $N = 4 \times 10^3$ f. $r_2 \approx 7\ cm$.

From these it is found that

$$p = A_c/A_T$$

$$= \pi (10 + 7)^2 \ / 10^{10}$$

$$= 9.1 \times 10^{-8} ,$$

and that

$$E \{\theta\} = Nnp$$

$$= (4 \times 10^3)(6 \times 10^2)(9.1 \times 10^{-8})$$

$$= 0.22$$

and

$$VAR \{\theta\} = 0.22.$$

Note that N_p, which is the maximum value of gp, is $Np = 3.64 \times 10^{-4}$, a very small quantity. Further, it is found that the probability of exactly x casualties under the given assumptions are as in Table II.

Table II
Casualty Distribution

Number of Casualties	Probability of Occurrence
0	0.80252
1	0.17655
2	0.01942
3	0.00142
4	0.00008
5	0.00000

Note that the expected number of casualties, 0.22, is approximately 0.0055 percent of the 4000 personnel or approximately one casualty in five similar drops.

The hazard to personnel resulting from a drop of self-dispersing bomblets was found to be very low. It was found that the number of casualties, θ, is Poisson distributed of form

$$p(\theta) = (Nnp)^{\theta} e^{-Nnp}/\theta! ,$$

where N is the number of personnel on target, n is the total number of bomblets, and p is the probability that an individual is hit by a particular bomblet. For the specific case of 600 bomblets and 4000 persons in a one square kilometer area, p is approximately 9.1×10^{-8} and the expected number of casualties is 0.22.

EXPLOSIVE SAFETY AND RELIABILITY ESTIMATES FROM A LIMITED SIZE SAMPLE

J. N. Ayres, L. D. Hampton, I. Kabik
U. S. Naval Ordnance Laboratory
White Oak, Silver Spring, Maryland

ABSTRACT. The problem of predicting, from small sample testing, high reliability and/or high safety for explosive items is becoming more acute. Often the available test sample is no greater than 200. Only a single test per item is allowable and the data is always of the go/no-go variety. Methods being used for making conservative extrapolations to the high and low probability of firing points are reviewed and illustrated. The question of how to do the job better is posed and left to the clinicians for answer.

INTRODUCTION. The problem which we wish to present is how to make, with small samples, reasonable estimates of the stimuli corresponding to the high and low probability of firing of electro-explosive devices (EED's).

A typical EED is shown in Fig. 1. Essentially, it consists of an insulator carrying two electrical conductors across which is attached a resistance wire. Surrounding the resistance wire is a sensitive explosive. When electrical energy is dissipated in the wire, the resultant temperature rise causes the explosive to heat and react chemically, and thus produce an explosion.

EED's are used by the military for a number of purposes; to cause detonation of explosive loaded shells, bombs, grenades, missiles, mines, etc., to ignite propellants for guns and rockets, to close switches such as in fuze arming circuits, to release stores from aircraft, to eject pilots from aircraft, and to separate missile stages. These are only some of the more common uses.

The designer of explosive ordnance has always been faced with the problem of estimating the safety and reliability of his explosive system. The safety and reliability associated with the EED of electrically operated explosive ordnance, are, of course, important links in this system. For reasons to be given, estimating the safety and reliability to be expected from an EED subjected to various stimuli is usually not simple. The ordnance designer in the past has often overcome lack of information on reliability at least, by the numbers of items strategically used, i. e., the number of shells fired or the number of bombs dropped, etc. Thus unreliability could be compensated for in actual field usage.

Modern weapons and warfare, however, have introduced new problems. It is too costly to fire large numbers of expensive ordnance devices: the catastrophic results of a safety failure of certain types of munitions are intolerable; the intensity of certain stimulis which may cause inadvertent firing (electro-magnetic radiation from radars for example) has increased tremendously in the last decade and is slated to increase further. These changes have made it virtually mandatory that reasonable estimates of response of EED's to electrical stimuli be made.

RELEVANT FACTS.

(a) For economic reasons it is impossible to make a direct demonstration of the response of interest. The stimulus for reliability of 99.9+% is usually desired at 95% confidence. Conversely, safety may demand estimates at 95% confidence of the stimulus at which no more than 1 in a million devices would be expected to fire. Funds are never available to run direct demonstration tests.

(b) The nature of EED's preclude repeated testing on a single device. Since these systems respond chemically to temperature elevation at the resistance wires, it is not known, once a single test at a given stimulus was large enough to have altered the EED's response characteristics. It must therefore be assumed that the possibility of alteration is great enough to preclude more than one test on a given EED. The only piece of information thus possible from each single test is either the EED fired or failed at that particular test stimulus.

(c) It has been found[1], from a large number of firings on EED's (approx. 10,000 firings of Squib Mk I), that no standard distribution function fits exactly the tails of the observed EED stimulus-response distribution. A number of distribution functions have been tested for their conformance to the experimental firing data. They all fail at the tails of the curve, see Fig. 2. But it is precisely these regions of the distribution which we must estimate.

(d) Usually no more than 200 test samples are available to make estimates on one side of the mean firing (50%) point, whether high or low. Even a sample size of 200 is sometimes very difficult to obtain and may be quite expensive.

(e) Popular test schemes, such as the "Bruceton" test[2], which are conservative of sample size, often give poor estimates because of long extrapolation, poor estimate of the standard deviation, and/or non-applicability of the selected underlying distribution[3,4].

THE PROBLEM. By now it should be obvious that we must make multi-million dollar estimates on tens or hundreds of dollars worth of data. We must design our experiments so that we most wisely expend our available samples so that we can minimize the error of making extrapolations to the desired answer. We realize that extrapolation is at best a risky business but: is there any other choice?

In the following section we will tell you what we think we know and the methods we are now using.

The basic problem is to collect data which will permit the computation of the variation of the probability of firing as a function of the firing stimulus. It is desirable to allocate the samples so that the data collected will be as close as possible to the functioning level(s) we wish to estimate. Ideally we should collect go/no-go data at a number of stimulus levels. As shown in Fig. 3, we wish to estimate the stimulus, Xe, at which we can expect a high level of response, Ye. We show data collected at five levels of stimulus X_1, $X_2 \ldots X_5$. A line has been fit to the observed data and at point Xe, Ye on this line, is the intersection which gives us the desired stimulus value.

The process of drawing the straight line shown in Fig. 3, and making the indicated prediction implicitly makes the following assumptions.

1. That there is no sampling error,
2. That the distribution function is chosen correctly, and
3. That there is no systematic error in the instrumentation or testing procedure.

But we know that there must be some sort of error simply because the data points do not fall on the line. By performing the "Chi-Square" statistical test on the data we can decide whether or not the observed variability (scatter) is what might be expected from sampling error alone. If this is the case, then we can draw an appropriate confidence band as in Fig. 3.

WHAT WE HAVE DONE. But rather than multi-point testing we have made what we believe to be conservative estimates of extreme probability of firing points by the test and extrapolation procedures given below.

To minimize the importance of assumptions regarding the frequency distribution it is again desirable to base these estimates on data taken as close as possible to the per cent point to be determined. The simplest such test would be one which calls for testing at two stimulus levels near the region in question. One of the two levels will be farther from the mean and closer to the desired point than the other. This will be designated the remote stimulus level. The data obtained can then be extrapolated to determine the stimulus associated with the desired per cent point. In planning such an experiment the following conditions should be met:

a. The difference between the stimuli used should not be small compared to the extrapolation distance (the difference between the desired point and the observed remote stimulus).

b. The number of trials at the remote stimulus level and the expected response at this level should be chosen so that the probability of observing a saturated level (either all-fires or all fails) is small*.

c. The number of trials made at the remote functioning level should be greater than the number of trials at the level closer to the mean in an attempt to obtain equal weighting of the two levels. A good choice is to take the number so that the product np (1-p) is the same for both levels, where n is the number of trials and p is the expected probability of fire.

---

*If a saturated level is observed, one trial can be converted to 1/2 fire + 1/2 fail. Or another, reversed, trial can be arbitrarily added to the data. Either method will give a conservative result.

It is assumed that only two hundred samples are available to estimate either an extremely high or else an extremely low probability of firing. The general procedure will be illustrated below for a high probability point; a numerical example is given in Appendix A.

a. Run a preliminary Bruceton type test on 20 samples using a log transform for the dosage*.

b. Use the Bruceton results to estimate the $\bar{X}+0.2s$, $\bar{X}+0.4s$, and $\bar{X}+1.3s$ levels**.

c. Test 50 EED's at the computed $\bar{X}+0.4s$ level.

d. If more than 5 fail, test 130 samples at the above calculated $\bar{X}+1.3s$ level.

e. If 5 or fewer failures occur, continue testing until 130 samples have been tested, and test 50 at the calculated $\bar{X}+0.2s$ level.

f. Using a log-logistic[5] probability space, plot the two points.

g. Extrapolate the straight line through the points so obtained to the desired probability or stimulus value.

By using only two points we have no way of applying the chi-square test. Nor can we draw the confidence band without a further assumption. To obtain more conservatism, two methods have been used.

Heterogeneity Assumption

We proceed as above but assume a heterogeneity factor*** of 1 in the equation for the confidence limit. This assumption allows computation and drawing of the confidence band as in Fig. 4. Implicit

*Considerable testing has led us to believe a logarithmic dosage to stimulus transform is of proper form.

**For low probability estimates these terms would be $\bar{X}-0.2s$, $\bar{X}-0.4s$, $\bar{X}-1.3s$, and following computations would be consistent.

*** $F = \frac{\chi^2}{n-2}$, where F = heterogeneity factor and n = the number of test levels.

in the assumption are the assumptions previously given also, i.e., we have chosen the correct distribution function; there is no systematic error in the instrumentation and test procedure; and only normal sampling error occurs.

Binomial Method

Using the second method of gaining conservatism, rather than plotting the measured points directly, calculate, at a desired confidence level (say 75%), the one-sided lower value of the higher percentage firing point, and the one-sided upper value of the lower percent firing point. Plot these points in a log-logistic probability space. Draw the straight line through these points and extrapolate to the desired value. See Fig. 5.

It is, of course, possible that if too conservative a value be set for the confidence limits of the upper one-sided, lower and the lower one-sided, higher per cent firing points, the slope of the line drawn through these limits will be negative. Such a situation, when it occurs, is not realistic and this more conservative estimating technique should be abandoned.

Our experience has shown us that although the logistic distribution function does not give an accurate fit to EED distribution functions at the tails, it at least errors on the conservative side, i.e., it will predict a lower safety than actually exists and a lower reliability than actually exists.

The two-level test and analysis, then, is one technique which we have used to make, with limited samples, estimates of extreme probability of firing points. We could certainly devise more elaborate and sophisticated variations, but we wonder if those more skilled than we in statistical theory might not be able to recommend alternate procedures which can do the job better. More specifically, we have wondered about, and have planned to work on, the application of non-parametric statistical methods to the problem. The clinic's opinion and advice on this matter could be beneficial since, at the time of this writing (June 1964), we are only in the preliminary thinking stage.

Finally, we have been hopeful that some combination might be made of statistics and the underlying physics of the mechanism by which wire bridge EED's function, to put bounds on the degree of extrapolation needed in making our estimates. In this regard our work has shown that the heating of a wire bridge EED can be represented by the mathematical equation:

$$C_p \frac{d\theta}{dt} + \gamma\theta = p(t)$$

where C_p = heat capacity of bridge plug explosive

θ = temperature elevation above ambient

t = time

γ = heat loss factor, and $p(t)$ = power input.

The combination of this equation[6, 7] with Bowden's hot spot theory of explosions[8] has led to fairly accurate representation of EED firing characteristics over a limited range of input times (i. e., average powers). Since equipment is available for making independent measurements of C_p, γ, and C_p/γ, the cooling time constant, it appears possible to measure, on individual EED's, parameters which should be directly related to their individual firing characteristics.

REFERENCES

1. L. D. Hampton, J. N. Ayres, "Characterization of Squib Mk 1 Mod 0: Determination of the Statistical Model", NavWeps Report 7347, 30 January 1961.

2. Statistical Research Group, Princeton University, "Statistical Analysis for a New Procedure in Sensitivity Experiments", AMP Report 101.1R, SRG-P No. 40, (OSRD Report 4040) July 1944.

3. J. N. Ayres, L. D. Hampton, I. Kabik, "The Prediction of Very Low EED Firing Probabilities", NOLTR 63-133, 4 September 1963.

4. L. D. Hampton, J. N. Ayres, I. Kabik, "Estimation of High and Low Probability EED Functioning Levels", NOLTR 63-266, 3 February 1964.

5. Jour. of American Statistical Association, Vol. 48, p. 565-599, 1953, Joseph Berkson, "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response Based on the Logistic Function".

6. I. Kabik, L. A. Rosenthal, A. D. Solem, "The Response of Electro-Explosive Devices to Transient Electrical Pulses", 4th Navy Science Symposium, Pasadena, Calif., March 1960.

7. I. Kabik, L. Rosenthal, A. Solem, "The Response of Electro-Explosive Devices to Transient Electrical Pulses", NOLTR 61-20, 17 April 1961.

8. Initiation and Growth of Explosion, Bowden & Yoffe, Cambridge at the University Press, 1952.

APPENDIX A

ILLUSTRATIVE EXAMPLE

The units used for X are in terms of the transformed variable.

The twenty trial Bruceton gave a mean of 20.314 and standard deviation of 0.589.

The two test levels are then

$m+0.4s = 20.55$
$m+1.3s = 21.08.$

The results at these levels were

Near level 35/50 = 70%
Remote level 113/130 = 86.92%.

The upper 95% confidence limit at the near level is 78.68%. The lower 95% confidence limit at the remote level is 81.94%.

A straight line through the observed points is

$$Y = 1.13019\,X - 22.7014$$

(Y in Normits).

This gives estimates as follows:

95% point	21.542
99% point	22.144
99.99% point	23.347.

The equation for the lower 95% confidence band assuming the heterogeneity factor to be unity is

$$Y = 1.13019\,X - 22.7014 - 1.645\sqrt{0.014909 + 0.213434(X - 20.857)^2}$$

This gives estimates as follows:

95% point	22.9
99% point	24.8
99.99% point	29.0 .

The straight line through the binomial limits on the observed points has the equation

$$Y = 0.2226\,X - 3.7784\ .$$

This gives the following estimates

95% point	24.37
99% point	27.43
99.99% point	33.69 .

Using the same data with the logistic assumption, we have the following analysis

at the near level 35/50 = 70%

$$L = \ln \frac{35}{15} = 0.8473$$

at the remote level 113/130 = 86.92%

$$L = \ln \frac{113}{17} = 1.8942\ .$$

The straight line through these points is

$$L = 1.975\,X - 39.7447\ .$$

This gives the following estimates

	L	X
95%	2.9444	21.6
99%	4.5951	22.4
99.99%	9.2102	24.8 .

The binomial confidence limits as before are

near level 1.306; remote level 1.512 .

The straight line through these points is

$$L = 0.389\,X - 6.688$$

which gives

95% point	24.8
99% point	29.0
99.99% point	40.9 .

The hyperbola for the lower 95% confidence band has the equation

$$L = 1.9753\,X - 39.7447 - 1.645\sqrt{0.039557+0.579926(X-21.86)^2}$$

which has an asymptote

$$L = 0.723\,X - 13.6224\ .$$

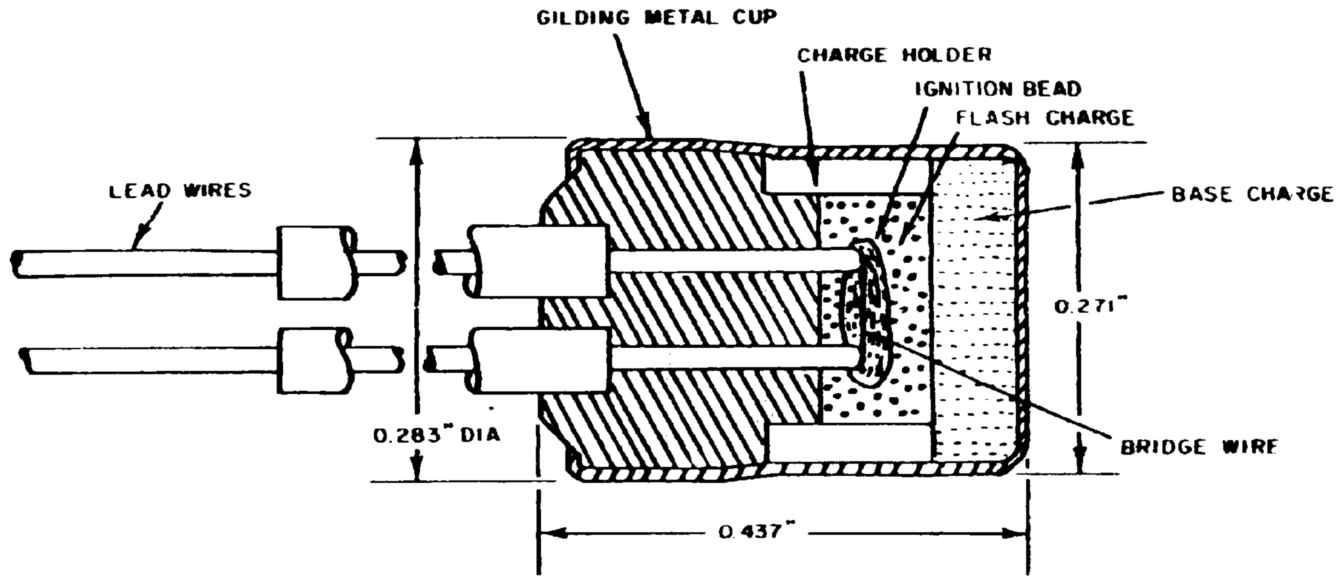
Estimates are

95% point	22.9
99% point	25.2
99.99% point	31.6

Summary of these calculations results

	Normal			Logistic		
	Straight Line	95% conf. band	Binomial	Straight line	Conf. Band	Binomial
95% point	21.54	22.9	24.4	21.6	22.9	24.8
99% point	22.14	24.8	27.4	22.4	25.2	29.0
99.99% point	23.35	29.0	33.7	24.8	31.6	40.9

Comparison of these values shows the more conservative nature of the logistic distribution. The difference is not marked at the 95% point but does show up at the more extreme points.



NOTES:

1. IGNITION BEAD - APPROX. 5 MG DDNP/$KClO_3$
2. FLASH CHARGE - APPROX. 45 MG BLACK POWDER
3. BASE CHARGE - APPROX. 45 MG BLACK POWDER
4. BRIDGE WIRE - 0.001" PLATINUM-IRIDIUM 0.060" LONG

FIG. I SQUIB MK I MOD 0

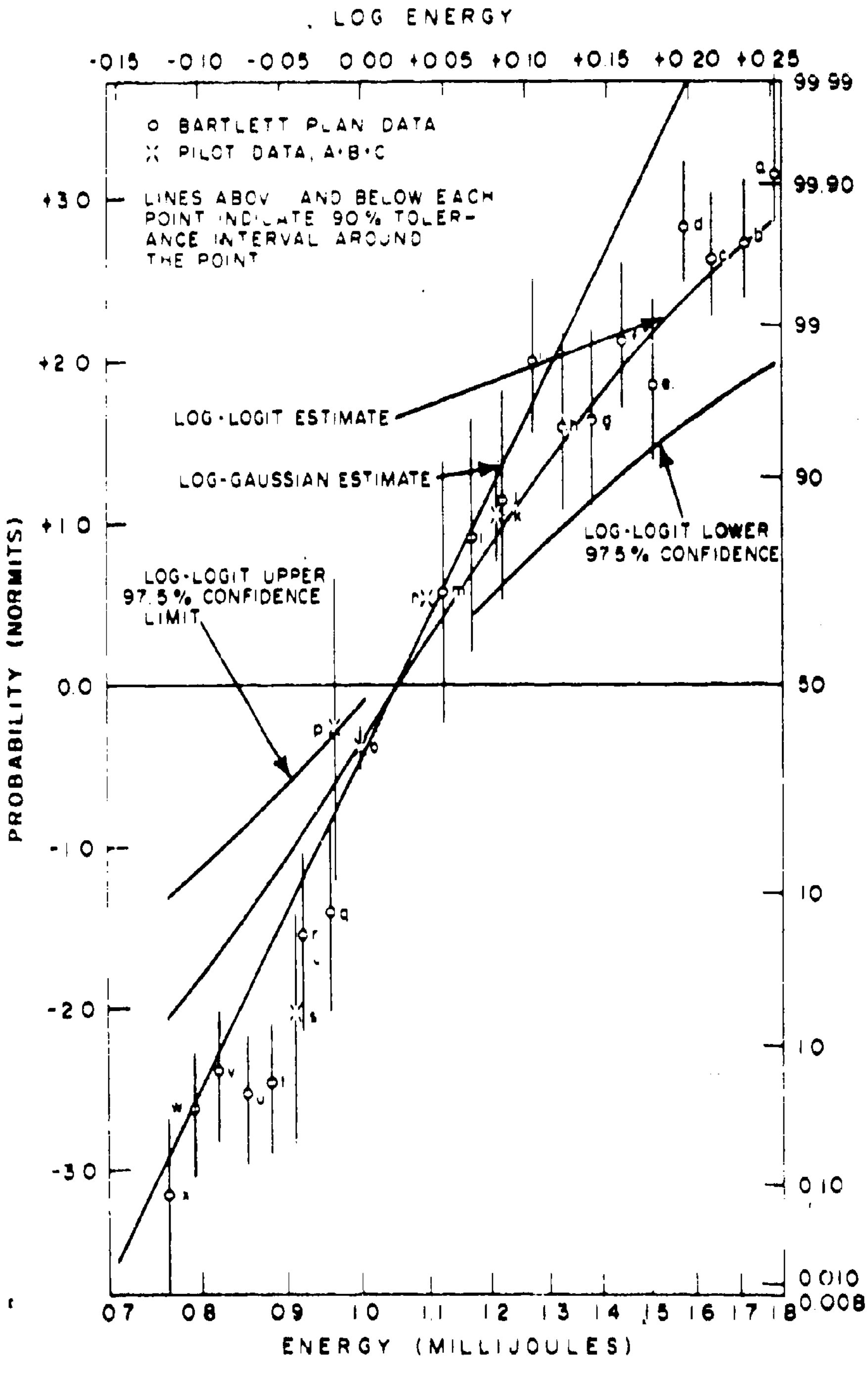


FIG. 2 LOG-GAUSSIAN, LOG-LOGISTIC FITS OF FIRING DATA

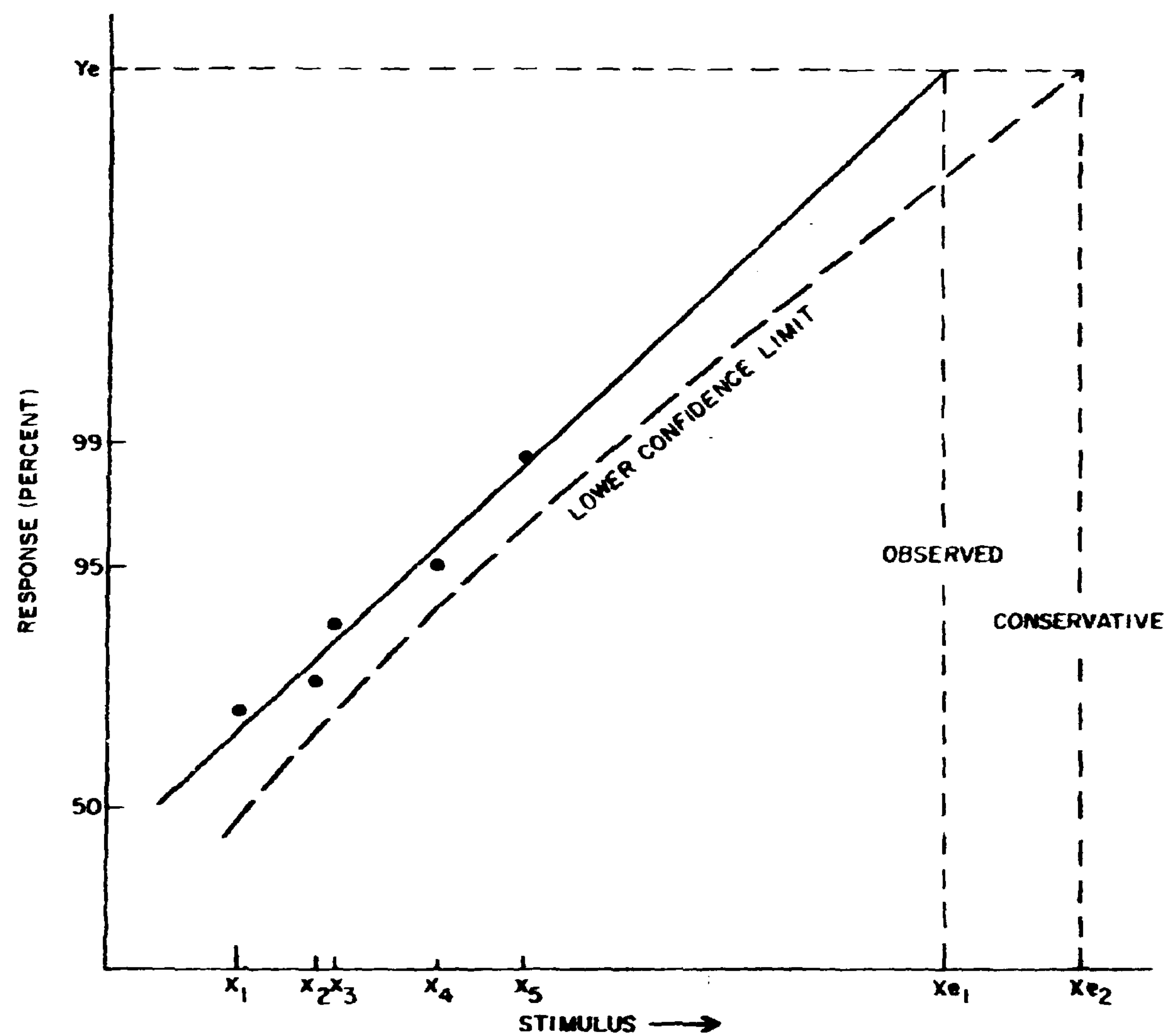


FIG. 3 MULTIPOINT-DATA ESTIMATE

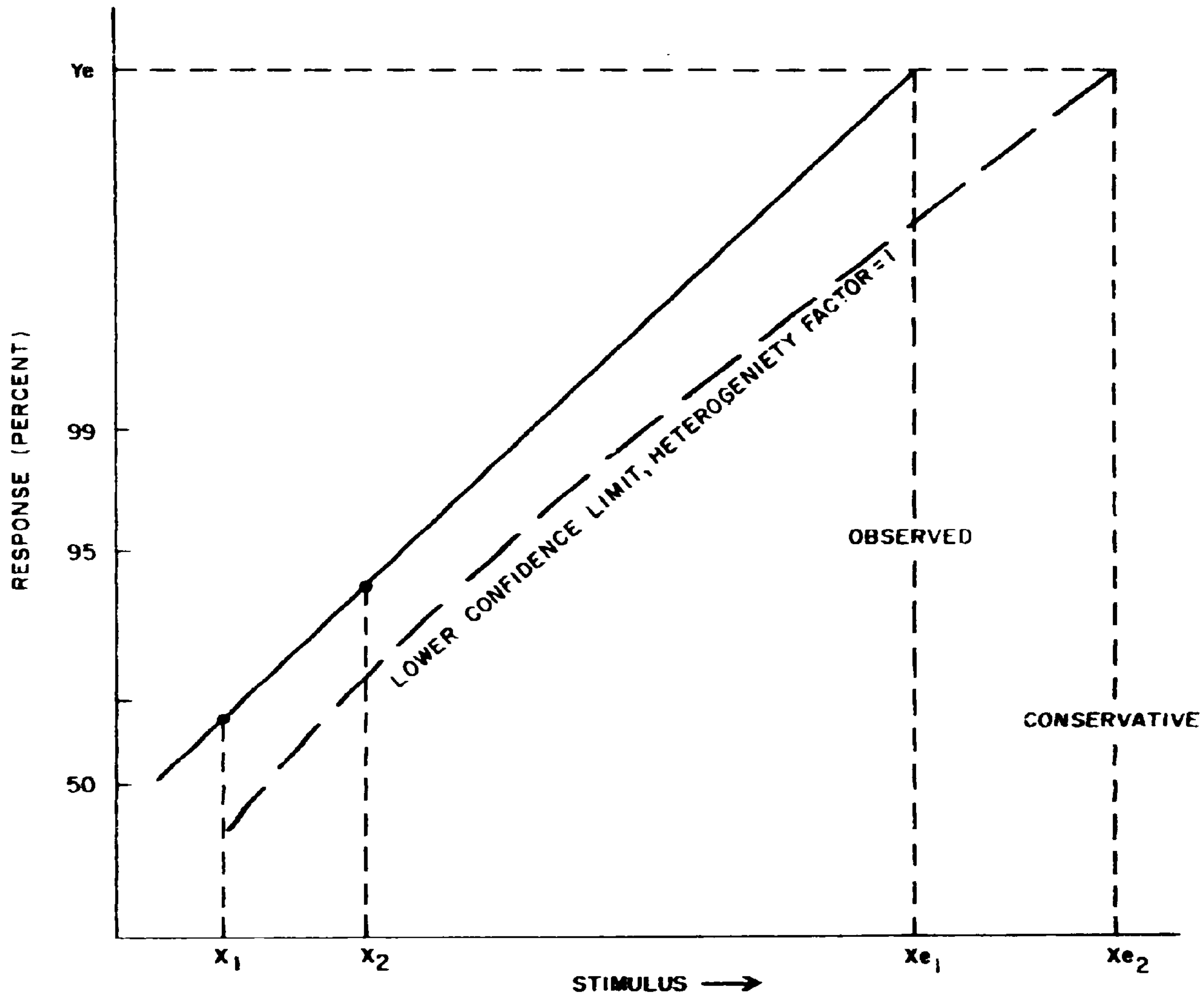


FIG.4 2-POINT ESTIMATE WITH CONFIDENCE LIMIT

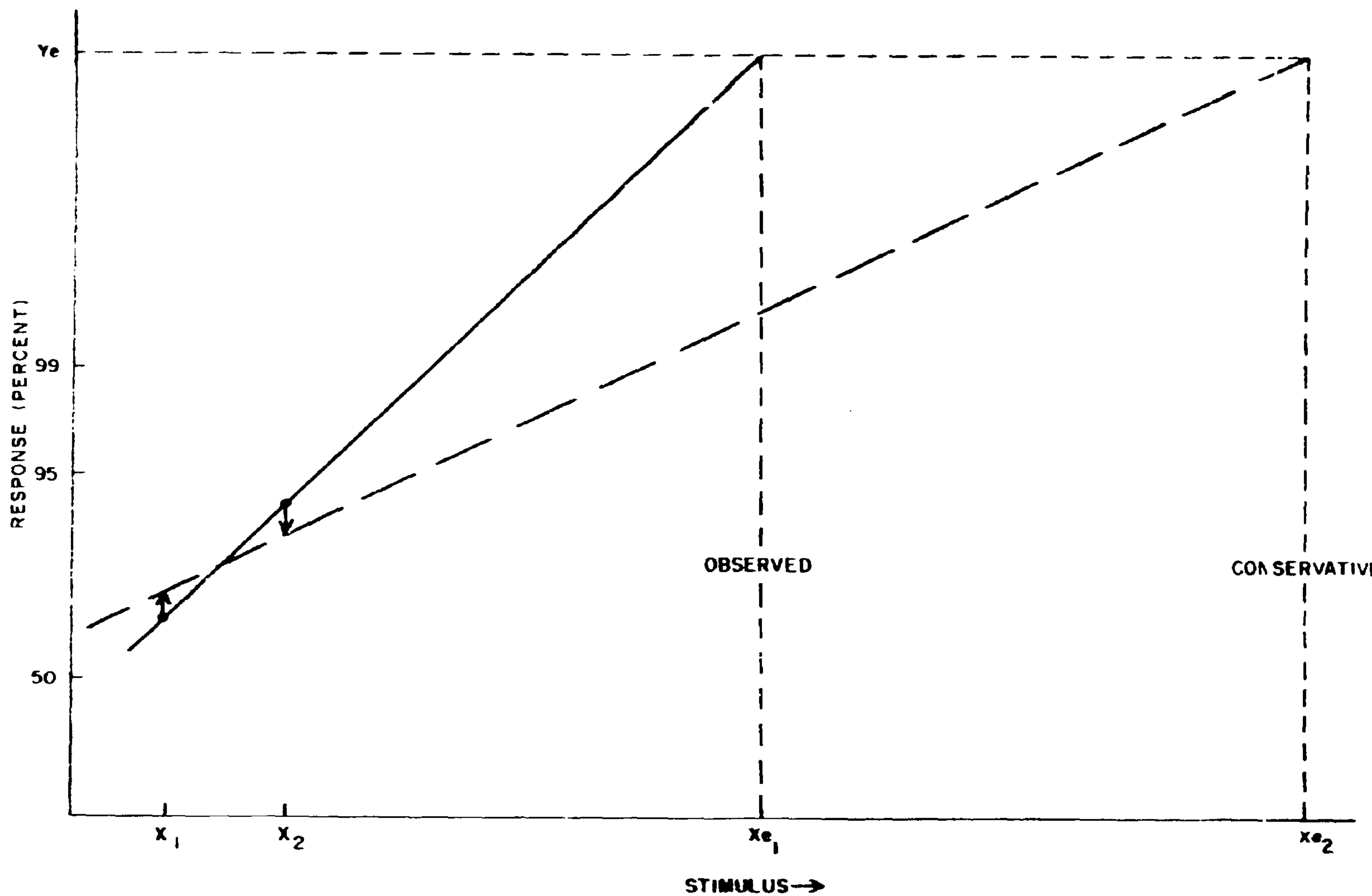


FIG. 5 2-POINT ESTIMATE WITH BINOMIAL LIMITS

CYCLIC DESIGNS*

H. A. David and F. W. Wolock
University of North Carolina at Chapel Hill and Boston College

1. INTRODUCTION. Cyclic designs are incomplete block designs consisting in the simplest case of a set of blocks obtained by cyclic development of an initial block. More generally, a cyclic design consists of combinations of such sets and will be said to be of size (n, k, r), where n is the number of treatments, k the block size, and r the number of replications.

It is well known (e.g. Bose and Nair [2]) that cyclic development of a suitably chosen initial block is one of the methods of generating designs with a high degree of balance in the arrangement of the treatments such as balanced incomplete block (BIB) designs and partially balanced incomplete block designs with two associate classes (PBIB(2) designs). Again, the cyclic type is a rather junior partner among the five types into which Bose and Shimamoto [3] classify PBIB (2) designs. The emphasis in these and many related papers has been understandably on the number of associate classes, the cyclic aspect being incidental. In the present article we proceed in opposite fashion putting the cyclic property first. It will be shown how cyclic designs may be systematically generated and how the non-isomorphic designs of given size may be enumerated and constructed. All such designs are PBIB designs but may have up to $\frac{1}{2}n$ associate classes. For $n \leq 15$ and k = 3, 4, 5, tables of the most efficient cyclic designs are presented and comparisons with BIB and PBIB (2) designs are made.

Points which make cyclic designs attractive are:

(i) Flexibility. A cyclic design of size (n, k, ik) exists for all positive integers n, k, i. If n and k have a common divisor d then a "fractional set" of size (n, k, k/d) exists corresponding to each d. Fractional sets may be combined with designs of size (n, k, ik) to form fresh designs, or used by themselves especially if n is large. Thus there are cyclic designs for many sizes (n, k, r) for which no PBIB (2) design is available, but the reverse may also happen.

---

*Research supported by the Army Research Office, Durham, and the National Institutes of Health. This paper has been submitted for publication in the "Annals of Mathematical Statistics."

(ii) Ease of representation. No plan of the experimental layout is needed since the initial block or blocks suffice.

(iii) Youden type. In view of their method of generation cyclic sets with r = k, and hence combinations of such sets, provide automatic elimination of heterogeneity in two directions.

(iv) Analysis. For cyclic designs the coefficient matrix of the normal equations is a circulix. The inverse matrix may therefore be obtained explicitly (as another circulix), thus making possible a general method of analysis. Questions of analysis will not be considered further here since methods given in a special case by Kempthorne [9] continue to apply with minor modifications. However, details and aids to analysis are presented in [12].

Cyclic designs as a class in their own right were introduced for k = 2 by Kempthorne [9] and Zoellner and Kempthorne [13]. Design aspects for the case k = 2, which has some special features, were considered in [6] and [7], and will not be treated in this paper. For general k cyclic designs are closely related to the circular designs of Das [5]. See also the survey of non-orthogonal designs by Pearce [11] who calls cyclic designs a "little publicized class." PBIB designs have been studied from an algebraic point of view in a series of papers by Masuyama. In some of these (e.g. [10]) reference is made to cyclic designs but no detailed results are obtained.

2. CYCLIC SETS. Label the treatments 0, 1, 2, ..., n-1. To fix ideas consider the arrangement of n = 7 treatments in blocks of size k = 3. The complete design of $\binom{7}{3}$ = 35 distinct blocks may be set out as follows:

	{012}	:	012	123	234	345	456	560	601
	{013}	:	013	124	235	346	450	561	602
(1)	{014}	:	014	125	236	340	451	562	603
	{015}	:	015	126	230	341	452	563	604
	{024}	:	024	135	246	350	461	502	613

From any block the others in the same row may be obtained by increasing each object label in turn by 1, 2, 3, 4, 5, 6, and reducing modulo 7. The rows have been arranged to start with the block of lowest numerical value and are designated by the initial block placed in braces. We call each row a <u>cyclic set</u>.

A block may also be conveniently represented by identical beads spaced regularly on a circular necklace. Fig. 1 shows the blocks 012 and 123.

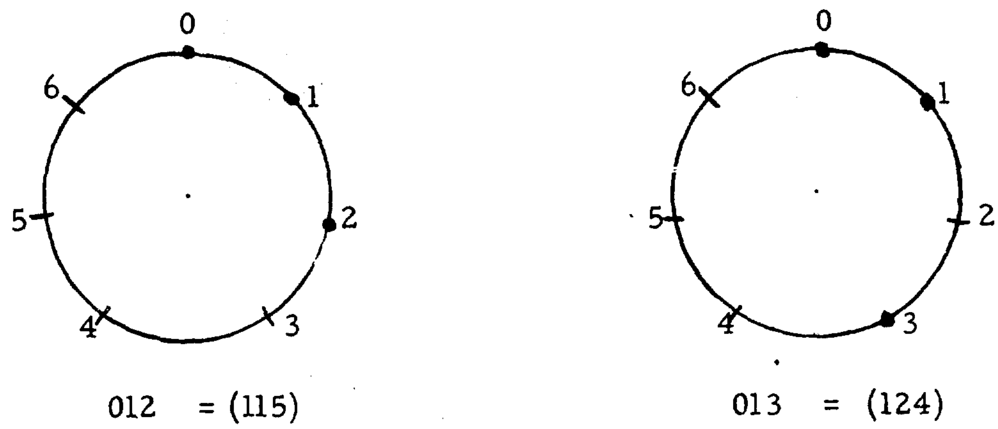


Figure 1

The set {012} is then generated by successive unit rotations.

It is not difficult to show that each cycle set forms a partially balanced incomplete block (PBIB) design with b (no. of blocks) = n and r (no. of replications) = k. If objects i and j are α-th associates so are i and n-j. Thus the number m of associate classes is at most $\frac{1}{2}(n-1)$ for n odd and $\frac{1}{2}n$ for n even, but may be less, with m = 1 for a balanced (BIB) design. An additional feature of a cyclic set is that each object occurs once in each position within a block. Order effects are therefore automatically balanced out and the sets are Youden Type designs, balanced (m = 1) or partially balanced (m > 1).

The same procedure can be used for any n and k except that when n and k are not relative primes fractional sets arise consisting of n/d blocks, where d is any common divisor of n and k. In terms of Fig. 1 such sets correspond to arrangements of beads which can be reproduced in fewer than n rotations of the necklace.

For the purpose of systematically enumerating all cyclic sets it is convenient to characterize each set by a circular partition of n. Thus we may replace $\{0x_1\ x_2\ x_3 \ldots x_{k-2}\ x_{k-1}\}$ by $(x_1,\ x_2-x_1,\ x_3-x_2, \ldots,$

$x_{k-1}-x_{k-2},\ n-x_{k-1})$.

Example 1. For n = 8, k = 4 the set {0123} becomes (1115). The cyclic sets may now be written down in increasing order of the numerical value of the corresponding partition: (1115), (1124), (1133), (1142), (1214), (1223), (1232), (1313), (1322), (2222). After (1142) we omit (1151) this being identical with (1115), etc. As the repetition of digits indicates the set (1313) consists of the 4 blocks

0145 1256 2367 3470 (r = 2)

and (2222) of the 2 (disconnected) blocks 0246, 1357 (r = 1). These are still PBIB designs but, of course, no longer of the Youden Type. We shall say that the corresponding arrangements of beads on a necklace have periods 4 and 2, respectively. As a check note that all $\binom{8}{4}$ blocks are accounted for since 8x8+4+2 = 70.

For any n and k, the total number of sets, being equal to the number of distinct arrangements of k white beads and n-k black beads on a necklace of n beads (which may not be turned over) is given by (Jablonski [8])

$$(2) \qquad N(k,\ n-k) = \frac{1}{n} \ \Sigma \ \phi(d) \frac{(n/d)\,!}{(k/d)\,!\ [(n-k)/d]\,!},$$

where the summation is over all integers d (including unity) which are divisors of both k and n-k, and $\phi(x)$ is Euler's function, the number of integers less than and prime to x. Thus

$$N(4,\ 4) = \frac{1}{8}\ \left(\frac{8!}{4!\ 4!} + \frac{4!}{2!\ 2!} + 2.\ \frac{2!}{1!\ 1!}\right) = 10.$$

The number of cyclic sets of various sizes making up this total is tabulated in [7] for n ≤ 15.

If a design of size $n = b = 7$ and $k = r = 3$ is required a look at the association schemes of the 5 sets in (1) leads to $\{013\}$ or $\{015\}$, both being BIB designs. For most sizes there will be no balanced set and the choice is less clear but might be based on the usual efficiency factor. Combinations of sets provide larger designs and again the question of optimal selection of sets arises. This presents a formidable task for all but small designs. Our principal aim is to show that this task can be greatly simplified if certain isomorphisms between cyclic sets are recognized. A systematic approach for the construction of optimal cyclic designs is then developed.

3. <u>EQUIVALENCE CLASSES</u>. Let us now apply to $\{012\}$ of equation (1) the re-numbering or permutation

$$R(7, 3) = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 3 & 6 & 2 & 5 & 1 & 4 \end{pmatrix}$$

obtained by multiplying each of the 7 labels by 3 (mod 7). Then $\{012\}$ becomes

036 362 625 251 514 140 403,

a Youden Type design which is merely a re-arrangement of $\{014\}$. We write $\{012\} \xrightarrow{3} \{014\}$. Thus $\{012\}$ and $\{014\}$ are isomorphic. Two further applications of R(7, 3) give $\{024\}$ and the original $\{012\}$. We have therefore established the equivalence class $\{012\} \sim \{014\} \sim \{024\}$. No blocks need be written in the process if partition notation is used: $\{012\} \xrightarrow{3} \{036\} = (331) = (133) = \{014\} \xrightarrow{3} \{035\} = (322) = (223) = \{024\}$. Likewise $\{013\} \xrightarrow{3} \{032\} = \{023\} = (214) = (142) = \{015\}$, so that $\{013\}$, $\{015\}$ form a second equivalence class.

The same procedure can be used for any prime n and any k. To see this note that the permutations R(n, 1) (the identity permutation), R(n, 2), . . . , R(n, n-1), form a group under "multiplication" * defined by

$$(2) \qquad R(n, i) * R(n, j) = R(n, ij \bmod n)$$

which is isomorphic with the multiplicative group of residues mod n. Hence all elements R(n, i) are generated by powers of R(n, g), where g is a primitive root of n (i.e., $g^x \not\equiv 1 \bmod n$ for $x = 1, 2, \ldots, n-2$ but

$g^{n-1} \equiv 1 \bmod n$). But a permutation σ which changes one cyclic set into another must be of the form R(n, i) if we assume without loss of generality that σ leaves 0 unchanged; for if a, b, c, d, are elements of the residue set with a and b = a+d two elements in the same block we require that

$$\sigma(b) - \sigma(a) = \sigma(d) \qquad \text{all } a, b, d$$

or

$$\sigma(a) + \sigma(d) = \sigma(a+d),$$

showing that σ is multiplicative: $\sigma(a) = ca$. Thus all possible isomorphisms between cyclic sets can be established conveniently by repeated application of R(n, g).

When n is not prime the R(n, i) continue to form a group under * of (2) provided i and j are restricted to be integersrelatively prime to n. The group is now of order $\phi(n)$ and is clearly isomorphic with the multiplicative group of the reduced set of residues. g is said to be a primitive root of n if $\phi(n)$ is the smallest power making $g^{\phi(n)} \equiv 1 \bmod n$. Primitive roots exist only if n equals 2, 4, p^n, or $2p^n$, where p is any prime > 2 and n any integer. For values of n admitting a primitive root we proceed as before; otherwise, multiplication by each member of the reduced set of residues will establish most isomorphisms.

Example 1 (cont'd.) Since 8 does not have a primitive root we begin by applying R(8, 3) to the sets of Example 1 and find

$$(1115) \xrightarrow{3} (1232), \quad (1124) \xrightarrow{3} (1223), \quad (1142) \xrightarrow{3} (1322).$$

The other sets are unchanged by the transformation. Likewise R(8, 5) gives

$$(1115) \xrightarrow{5} (1232), \quad (1124) \xrightarrow{5} (1322), \quad (1142) \xrightarrow{5} (1223).$$

R(8, 7) produces "mirror images" obtained by reading a circular partition anti-clockwise rather than clockwise. E.g. $(1124) \xrightarrow{7} (4211) = (1142)$. This isomorphism had already been established by R(8, 3) and R(8. 5) because $5 \equiv -3$. However, an additional isomorphism can be obtained by the permutation

$$\binom{0}{0} \quad \binom{1}{1} \quad \binom{2\ 6}{6\ 2} \quad \binom{3\ 7}{7\ 3} \quad \binom{4}{4} \quad \binom{5}{5}$$

which takes (1133) into (1214). This is the only instance we have come across where the equivalence of two cyclic sets cannot be demonstrated by a multiplicative permutation.

A listing of all equivalence classes for cyclic sets in experiments with $n \leq 15$ and $k = 3, 4, 5$, is given in [12]. The efficiences of these sets regarded as designs have also been tabulated. When $n = 8$, $k = 4$ we find

Design			E	E_1	E_2	E_3	E_4
{0123}	=	(1115)	.812	.922	.834	.760	.712
{0124}	=	(1124)	.851	.867	.873	.810	.868
{0125}	=	(1133)	.851	.867	.809	.867	.877
{0134}	=	(1214)	.836	.863	.810	.869	.807
{0145}	=	(1313)	.779	.802	.803	.668	.800 (r = 2).

Here E is the overall efficiency and E_j ($j = 1, 2, 3, 4$) is the efficiency factor relating to the comparison of j-th associates. On the basis of E the choice of optimal design for $r = 4$ among the five sets (the fifth duplicated) lies between {0124} and {0125}, with the latter preferable in having only 3 associate classes. It should be noted that except for fully balanced designs the highest value of E does not necessarily correspond to the design with the smallest number of associate classes. Other optimality criteria might be used but the choice of cyclic design is in any case reduced to one of the non-isomorphic sets. Moreover, it is only combinations of these sets (and possible disconnected sets) which need to be considered in the construction of larger cyclic designs. In Table 1 we list the most efficient cyclic sets for $n \leq 15$ and $k = 3, 4, 5$.

Cyclic sets with two associate classes. For purposes of comparison we have made a corresponding compilation in Table 2 of two-associate PBIB designs of all types as given by Bose et al. [1] and (with asterisks) by Clatworthy [4]. The BIB designs in this range are also included. It will be noted that Table 2 has gaps for several (n, k) combinations although the symmetrical case is favorable to the existence of designs with a high degree of balance. The table also shows that a cyclic design with more than two associate classes may be more efficient than any two-associate PBIB.

It is of some interest that every regular (R) group divisible PBIB of Table 2 may be laid out as a cyclic design; this is already done in [1] in some cases and may be effected for the remaining designs by suitable relabeling. We find the following isomorphisms:

$n = 6$:	$R1 \sim \{013\}$,	$R2 \sim \{0124\}$;	
$n = 8$:	$R5 \sim \{013\}$,	$R108^* \sim \{01235\}$,	$R109^* \sim \{01246\}$;
$n = 9$:	$R8 \sim \{0136\}$,	$R112^* \sim \{01346\}$;	
$n = 10$:	$R114^* \sim \{01257\}$;		
$n = 12$:	$R15 \sim \{0137\}$,	$R116^* \sim \{01356\}$,	
	$R117^* \sim \{01249\}$,	$R118^* \sim \{0147\underline{10}\}$;	
$n = 14$:	$R24 \sim \{0146\}$;		
$n = 15$:	$R27 \sim \{0137\}$.		

There are only two other cyclic designs with two associate classes in the range under consideration. For $n = 13$ we have $C1 \sim 014$; for $n = 12$ the design $\{01247\}$ has the same association scheme as $R116^*$ but is not isomorphic with it.

4. <u>COMBINATIONS OF CYCLIC SETS</u>. Cyclic sets for given n may be combined to produce a wide variety of cyclic designs, still of PBIB form. This can always be done if the number of replications r is a multiple of k but will also be possible for certain other values of r if fractional sets exist. We shall say that the combined design is of size (n, k, r). Equivalence classes may again be established. However, the most efficient cyclic design of given size is not necessarily one made up of the most efficient cyclic sets.

Example 2. For $n = 9$, $k = 3$ we have the equivalence classes

A : (117) , (225) , (144) ;

B : (126) , (243) , (153) , (162) , (234) , (135) ;

C : (333) ($r = 1$) .

The order within a class has been arranged so that successive sets are obtained by the application of $R(9, 2)$, the primitive root of 9 being 2. There are clearly two non-isomorphic designs of size (9, 3, 4) obtained by combining (333) with any member of class A or class B. Of these the latter, which may be written as {013, 036}, is the more efficient, with $E = 0.713$ and 4 associate classes.

To get designs with $r = 6$ we can take two sets from A, two from B, or one from each. Call the sets A_1, A_2, A_3, and B_1, $B_2, \ldots, B_6$. We then have the following seven equivalence classes;

A_1A_2 , A_2A_3 , A_3A_1 ;

B_1B_2 , B_2B_3 , B_3B_4 , B_4B_5 , B_5B_6 , B_6B_1 ;

B_1B_3 , B_2B_4 , B_3B_5 , B_4B_6 , B_5B_1 , B_6B_2 ;

B_1B_4 , B_2B_5 , B_3B_6 ;

A_1B_1 , A_2B_2 , A_3B_3 ; A_1B_4 , A_2B_5 , A_3B_6 ;

A_1B_2 , A_2B_3 , A_3B_4 , A_1B_5 , A_2B_6 , A_3B_1 ;

A_1B_3 , A_2B_4 , A_3B_5 , A_1B_6 , A_2B_1, A_3B_2 .

Calculations show that the most efficient cyclic design is A_1A_2 with $E = 0.731$ and 4 associate classes.

The present example has been chosen to bring out the enumeration procedure required when the original cyclic sets fall into several equivalence classes.

Actually, for $r = 6$ as many as four PBIB(2) designs are available, viz. SR13, R10, LS3, and LS9*, of which LS3 is the most efficient having $E = 0.741$. When $r = 4$ the only tabulated PBIB(2) design is LS6, with the relatively low efficiency $E = 0.667$. For $r \leq 10$ Table 3 lists a selection of cyclic designs in cases where no such PBIB(2) designs are known to exist or are all of more than trivially inferior efficiency.

It is of interest to note that the number of non-isomorphic designs made up of s sets all chosen from the same class of S sets is just N (s, S-s), where N is defined by (2). This is so because we can now regard the beads of Fig. 1 as representing sets rather than blocks. The operation R(n, g), where g is a primitive root, produces a unit turn. The enumeration of non-isomorphic designs when sets are from more than one class proceeds exactly as described in [7] for $k = 2$.

5. FRACTIONAL SETS. The number nk of observations required for a cyclic set of size (n, k) will often be greater than desired, especially when n is large. In this situation fractional sets are very useful. As pointed out in Example 1 such sets are characterized by a repetitive pattern in their partition representation. No such design is possible if n is prime. For n composite fractional sets exist corresponding to every divisor d ($1 < d < n$) of n since there must be at least one partition of n consisting of d repetitions. Clearly, k must be a multiple of d, and $r = k/d$; (however, $r = 1$ gives a disconnected set). From a cyclic set with parameters $(n/d, k/d)$ a fractional set with parameters $(n, k, r = k/d)$ can always be obtained.

Example 3. For $n = 30$ connected fractional sets exist for $k = 4, 6, 8, 9, 10, \ldots$. Suppose we require a design with $k = 6$. The non-isomorphic connected cyclic sets of size (15, 3) are (1113), (1212), (1311), (1410), and (159). Of these (1212) leads to the most efficient design of size (30, 6, 3), viz. (12121212) or {01315 16 18} with $E = 0.762$.

In [12] a selection of the most efficient fractional sets of given size is tabulated for $n \leq 100$.

6. ACKNOWLEDGMENT. Much of this work was done while the authors were at the Virginia Polytechnic Institute. We are grateful to Dr. Dale M. Mesner for several helpful comments.

REFERENCES

[1] Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S. (1954). Tables of partially balanced designs with two associate classes. North Carolina Agric. Expt. Sta., Tech. Bull. No. 107.

[2] Bose, R. C. and Nair, K. R. (1939). Partially balanced incomplete block designs. Sankhyā, 4 337-372.

[3] Bose, R. C. and Shimamoto, T. (1952). Classification and analysis of partially balanced designs with two associate classes. J. Amer. Statist. Assoc. 47 151-190.

[4] Clatworthy, W. H. (1956). Contributions on partially balanced incomplete block designs with two associate classes. Nat. Bur. Stand., Applied Mathematics Series No. 47.

[5] Das, M. N. (1960). Circular designs. J. Indian Soc. Agric. Statist. 12 46-56.

[6] David, H. A. (1963). The structure of cyclic paired-comparison designs. J. Austral. Math. Soc. 3 117-127.

[7] David, H. A. (1965). Enumeration of cyclic paried-comparison designs. Amer. Math. Monthly, 72 (In press).

[8] Jablonski, E. (1892). Théorie des permutations et des arrangements circulaires complets. J. Math. Pure Appl., 4th series, 8 331-349.

[9] Kempthorne, O. (1953). A class of experimental designs using blocks of two plots. Ann. Math. Statist. 24 76-84.

[10] Masuyama, M. (1962). On the classes of blocks. Rep. Stat. Appl. Res., JUSE 9 83-87.

[11] Pearce, S. C. (1963). The use and classification of non-orthogonal designs. J. Roy. Statist. Soc. A 126 353-369.

REFERENCES (cont'd.)

[12] Wolock, F. W. (1964). Cyclic designs. Ph.D. dissertation, Virginia Polytechnic Institute.

[13] Zoellner, J. A. and Kempthorne, O. (1954). Incomplete block designs with blocks of two plots. Iowa Agric. Expt. Sta., Res. Bull. No. 418.

Table 1. Most efficient symmetric cyclic PBIB design D for n treatments and block size k, and its efficiency E.

	k=3		k=4		k=5	
n	D	E	D	E	D	E
6	{013} [2]	·784	{0123}	·895	{01234} [1]	·961
7	{013} [1]	·778	{0124} [1]	·876	{01234}	·932
8	{013} [2]	·748	{0125}	·851	{01235} [2]	·914
9	{013}	·722	{0134}	·836	{01235}	·898
10	{013}	·700	{0125}	·823	{01245}	·888
11	{013}	·676	{0125}	·817	{01247} [1]	·880
12	{014}	·673	{0137} [2]	·813	{01247} [2]	·870
13	{014} [2]	·667	{0139} [1]	·812	{01269}	·863
14	{014}	·670	{0146} [2]	·805	{01358}	·859
15	{015}	·641	{0137} [2]	·795	{012410}	·853

N. B. Superscripts [1, 2] denote respectively BIB and PBIB(2) designs.

Table 2. Balanced (BIB) and two-associate PBIB designs with codes and efficiencies from Bose et al. [1] and Clatworthy* [4].

	k=3		k=4		k=5	
n	D	E	D	E	D	E
6	R1	.78	S2, R2	.88, .89	BIB	.96
7	BIB	.78	BIB	.88		
8	R5	.75	SR7	.84	R108*, R109*	.91, .90
9	SR12	.73	R8, LS1	.80, .83	LS10, R112*	.90, .89
10	T6	.70	S17, T2	.79, .79	R114*	.88
11			T12	.82	BIB	.88
12			R15	.81	R116*, R117* R118*	.87, .87 .81
13	C1	.67	BIB	.81		
14			R24	.80		
15	T28	.66	R27	.80		

Table 3. Selected cyclic designs with r > k, corresponding optimal two-associate PBIB designs, and efficiencies E.

Size (n, k, r)	Cyclic design	E	PBIB(2) design	
8, 3, 6	{013, 014}	·756	R50*	·747
8, 4, 5	{0134, 0246}	·850	-	
9, 3, 4,	{013, 036}	·713	LS6	·667
10, 4, 6	{0147, 0156}	·825	T3	·789
10, 4, 8	{0126, 0148}	·830	R14	·823
11, 3, 6	{013, 026}	·727	-	
11, 3, 9	{013, 014, 027}	·730	-	
11, 4, 8	{0134, 0248}	·823	-	
13, 4, 8	{0125, 0159}	·807	C2	·797
13, 5, 10	{01247, 01258}	·865	-	
14, 3, 9	{014, 0211, 019}	·709	-	
14, 5, 10	{012410, 01710 12}	·862	-	
15, 3, 4	{015, 0510}	·682	T23	·673
15, 5, 6	{01257, 036912}	·856	T38*	·808

SOME RESULTS ON THE FOUNDATIONS OF STATISTICAL DECISION THEORY

Bernard Harris, J. D. Church, and F. V. Atkinson
Mathematics Research Center, U. S. Army
The University of Wisconsin

INTRODUCTION. A fundamental problem in statistical decision theory is concerned with establishing criteria for selecting a single decision procedure from the set of available decision procedures. In this paper, some criteria for optimality of statistical decision procedures are proposed and the consequences of these criteria are discussed. It is shown that these optimality criteria exclude a very general class of decision criteria, which contain as members, the minimax and minimax regret criteria. Finally, we note that these optimality conditions are consistent, in that there exists a decision procedure which satisfies all conditions, and a constructive procedure is given for determining such a decision procedure.

THE GENERAL STATISTICAL DECISION PROBLEM. A statistical decision problem is characterized by a set of states of nature S, whose elements will be denoted by s, and a set of pure (non-randomized) decisions D, whose elements will be denoted by d. The statistician selects an element d from D, and if nature is in state s, a loss $L(d, s)$ is incurred. An experiment is conducted and random variables $X_1, X_2, \ldots, X_N$ are observed where $X_1, X_2, \ldots, X_N$ has the probability distribution $P(x_1, x_2, \ldots, x_N \mid s)$. We require that the distributions $P(x_1, x_2, \ldots, x_N \mid s)$ be distinct for every $s \in S$. Then, since the decision is to be made following the experiment, the decision procedure is a function δ from the sample space to the space of decisions D. Let Δ be the set of such functions and note that d is then a random variable, i.e. $d = \delta(X_1, X_2, \ldots, X_N)$. This risk function $\rho(\delta, s)$ is then defined by

$$E[L(d, s)] = \rho(\delta, s).$$

The statistician's objective is to choose δ, so that $\rho(\delta, s)$ is small in some appropriate sense. It will frequently be desirable (in the sense of reducing risk) to augment the set of decisions to include the randomized decisions; and equivalently to augment the set of decision procedures Δ to Φ the set of randomized decision procedures, whose elements will be denoted by ϕ. Φ is the set of all probability mixtures of elements of Δ.

The fundamental problem of statistical decision theory is to decide how to choose an element $\phi \epsilon \Phi$. We can interpret this as consisting of two sub-problems.

1. What conditions should be imposed on a randomized strategy $\phi \epsilon \Phi$, so that we can regard strategies having those properties as being optimal?

2. Having decided which conditions are appropriate, how do we determine which elements $\phi \epsilon \Phi$ satisfy those conditions? Note that for some sets of possible conditions which one may wish to consider, it may happen that there are no strategies in Φ which satisfy them.

We will make the formal assumption that, in advance of the experiment, the statistician is in "complete ignorance" of which element s of S has been selected by nature. That is, that there is no a priori information available concerning the mechanism by which nature will select an element $s \epsilon S$.

The results stated in the succeeding sections have been established under the following hypotheses.

1. S and D are finite sets, i.e. $S = (s_1, s_2, \ldots, s_n)$, $D = (d_1, d_2, \ldots, d_r)$

2. With probability one, the random vector $(X_1, X_2, \ldots, X_N)$ assumes only a finite number of values.

As a consequence of the two hypotheses stated above, Δ is a finite set, and we can label its elements as $\delta_1, \delta_2, \ldots, \delta_m$.

Despite the restrictive nature of these assumptions, there are a substantial number of statistical problems to which they are applicable, and in addition, many problems may be approximate by problems satisfying the above hypotheses. As an example of a problem which satisfies the above restrictions, consider the following illustration.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random variables with

$$P\{X_i = 1\} = p_j, \quad P\{X_i = 0\} = 1 - p_j, \quad 0 < p_j < 1$$

for $i = 1, 2, \ldots, N$; $j = 1, 2$; and $S = \{1, 2\}$. Then, the sample space has 2^n elements. If we let $D = \{1, 2\}$, then Δ consists of all functions from the sample space to D, and hence Δ has 2^{2^n} elements. Hence, for this problem the above assumptions are all satisfied.

We can make this illustration more concrete by noting that the above is essentially the problem of testing whether a coin is fair ($p_1 = \frac{1}{2}$) or has probability ($p_2 = \frac{3}{4}$) of landing heads. We can interpret the two elements of D as being 1: Accept the hypothesis that $s = 1$, i.e. $p = \frac{1}{2}$; 2: Accept the hypothesis that $s = 2$, i.e. $p = \frac{3}{4}$. Thus, the illustration given is an "abstraction" of a test of a simple hypothesis against a simple alternative in a coin tossing problem.

It is well-known, that as a consequence of the above two assumptions we can identify the selection of a decision procedure ϕ with the selection of a point in a convex polyhedron C in Euclidean n-space, where C is generated as the convex hull of the points $(\rho(\delta_1 s_1), \rho(\delta_1 s_2), \ldots, \rho(\delta_1 s_n))$, $\delta \in \Delta$. If we define the matrix A, whose elements are a_{ij}, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ by $\rho(\delta_i, s_j) = a_{ij}$, then $C = C(A)$, the convex hull of the row vectors of A. Thus, we can use the natural relationship between the matrix A and the polyhedron $C(A)$, and freely characterise all relevant aspects of the problem in terms of either the matrix or the associated polyhedron. The reader is referred to the book by D. Blackwell and M. A. Girshick [1] for the relevant details.

We now turn to the characterization of desirable properties for decision procedures.

THE CHOICE OF A DECISION PROCEDURE. It is convenient at this time to introduce some definitions which will be needed in order to specify those properties of a decision procedure which will be considered desirable.

<u>Definition 1.</u> Two decision procedures ϕ_1, ϕ_2, in Φ will be said to be <u>equivalent</u> if

$$\rho(\phi_1, s_j) = \rho(\phi_2, s_j) \quad \text{for } j = 1, 2, \ldots, n$$

<u>Definition 2.</u> ϕ_1 is said to be dominated by ϕ_2 if

$$\rho(\phi_2, s_j) \leq \rho(\phi_1, s_j), \quad j = 1, 2, \ldots, n$$

with strict inequality holding for at least one j.

Note that if ϕ_1 is dominated by ϕ_2, then regardless of which state of nature s_j has been selected by nature, the risk using ϕ_1 is always at least as large as that using ϕ_2, and hence ϕ_2 is always to be preferred over ϕ_1.

<u>Definition 3.</u> A decision procedure ϕ_0 is said to be admissible if it is not dominated by any element $\phi \in \Phi$.

Since we have previously noted that dominated strategies are not desirable, then clearly the selection of a strategy should be made from among those that are admissible.

<u>Definition 4.</u> A decision procedure ϕ_0 is <u>essential</u> if it is admissible and if for every pair of decision procedures ϕ_1, $\phi_2 \in \Phi$, with ϕ_1 not equivalent to ϕ_0, and for every real number λ, $0 < \lambda < 1$,

$$\rho(\phi_0, s_j) \neq \lambda\rho(\phi_1, s_j) + (1-\lambda)\rho(\phi_2, s_j)$$

for at least one index j, $1 \leq j \leq n$.

The essential decision procedures are those which are admissible and in addition are also extreme points of the convex polyhedron $C(A)$. These decisions can then be used to generate all strategies which one may wish to consider.

The characterization of optimal decision procedures is equivalent to partitioning Φ into two sets, K - the set of decision procedures which are considered optimal, and Φ - K, those which are non-optimal. Equivalenty, we can characterize $Q(A) \subset C(A)$, the set of optimal vectors in Euclidean n-space.

We now propose eight properties which we believe will characterize a satisfactory decision procedure.

1. For every matrix A, Q(A) is a non-empty subset of C(A).

Clearly this condition is essential, since if Q(A) is empty, we have no decision procedures available for use.

2. If A' can be obtained from A by a permutation of the rows and columns of A, then Q(A') can be obtained from Q(A) by applying the permutation on the columns of A to the coordinates of vectors in Q(A).

Condition 2 says that the relabeling of the states of nature, and the (pure) decision procedures available to the statistician should not affect the decision procedure employed.

3. Every decision procedure with a risk vector in Q(A) is admissible.

This condition is just the observation that the only decision procedures that should be considered are the admissible decision procedures.

4. Q(A) is convex.

The motivation for this property is the following. If ϕ and ϕ' are both optimal, i.e. have their risk vectors in Q(A), then every probability mixture of ϕ and ϕ' will also be optimal.

5. If

$$A_1 = \lambda A_0 + \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_n \\ \cdots & \cdots & \cdots & \cdots \\ c_1 & c_2 & \cdots & c_n \end{bmatrix}$$

where λ is a positive real number and the vector $(c_1, c_2, \ldots c_n)$ is an arbitrary real vector, then

$$Q(A_1) = \{\lambda\tilde{x} + \tilde{c}\ ,\ \tilde{x} \in Q(A_0)\ ,\ \tilde{c} = (c_1, c_2, \ldots, c_n)\}\quad .$$

This requirement, includes, for example, invariance under the change of units of the loss function. In particular, if $\lambda = 1$ and $C_j = -\min_{1 \le i \le m} \rho(\delta_i, s_j)$, the matrix A_0 is reduced to its regret matrix.

6. If $C(A_1^T) = C(A_2^T)$, where A^T is the transpose of A, and in addition A_1 can be obtained from A_2 by deleting j columns from A_2, then $Q(A_1)$ can be obtained by deleting the corresponding coordinates from every vector in $Q(A_2)$.

Property 6 includes the column duplication property required by other writers, such as J. Milnor [4]. The point of this property, is that under complete ignorance, the decision problem for the statistician is essentially the same in both cases.

7. Let E_A be the submatrix of A corresponding to essential decision procedures in A. Then, if A_1 and A_2 are two matrices with $C(E_{A_1}) = C(E_{A_2})$, we require that $Q(A_1) = Q(A_2)$.

This says that the set of optimal decision procedures should depend only on those pure strategies which are candidates for good strategies. We might note that a risk vector $\tilde{x} \in C(A)$ is an essential strategy if and only if it uniquely minimizes the risk for some a priori distribution on the states of nature.

8. If $\{A_j\}$ is a sequence of matrices with $\lim_{j \to \infty} A_j = A_0$, and $\tilde{x}_j \in Q(A_j)$ for every $j \ge 1$, then every limit point of $\{x_j\}$ is an element of $Q(A_0)$.

This last condition is a continuity requirement. The reader's intuition may be aided by noting, that if one statistical decision problem may be approximated by another statistical decision problem, then this property requires that optimal decision procedures for the first problem are also approximated by the optimal decision procedures for the second problem.

R. D. Luce and H. Raiffa [3] give an extensive discussion of similar systems of optimal properties. The reader's attention is also specifically directed to papers by H. Chernoff [2] and J. Milnor [4], which deal with this problem.

CONSEQUENCES OF THIS CHOICE OF DESIRABLE PROPERTIES. Let $v_j = \min_{1 \le i \le m} \rho(\delta_i, s_j)$ and define $\tilde{v} = (v_1, v_2, \ldots, v_n)$. Define $\|\tilde{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$, $1 \le p \le \infty$, where $\|\tilde{x}\|_\infty$ is interpreted as $\sup_{1 \le i \le n} |x_i|$. Then, let the class of decision procedures $\Delta_p (1 \le p \le \infty)$ specify as optimal all $\tilde{x} \in C(A)$ which are admissible and satisfy

$$\|\tilde{v} - \tilde{x}\|_p \le \|\tilde{v} - \tilde{y}\|_p$$

for all $y \in C(A)$. Then, the following theorem can be established.

THEOREM. For $1 \le p < \infty$, Δ_p satisfies every property with the exception of property 6. Δ_∞ satisfies every property except property 8.

The reader should note that Δ_1 is Laplace's criterion and that Δ_∞ is the minimax-regret criterion restricted to admissible decision procedures. The failure of the minimax regret criterion to satisfy the above list of properties also establishes that the minimax criterion does not always satisfy the list of requirements given above.

Finally, we have the following theorem.

THEOREM. There is at least one decision procedure satisfying all of the above properties.

The proof of this last statement is accomplished by exhibiting a constructive process, which we now sketch.

Let $\{\epsilon_j\}$, $j = 1, 2, \ldots$ be a monotone non-increasing sequence of positive real numbers, with $\lim_{j \to \infty} \epsilon_j = 0$. Let $d(\tilde{x}, \tilde{y}) = \sup_{1 \le i \le n} |x_i - y_i|$ and let $Q_1 = C(A)$. Define $v_i^{(1)} = \min_{\tilde{x} \in Q_1} x_i$ and $\tilde{v} = (v_1^{(1)}, v_2^{(1)}, \ldots, v_n^{(1)})$. Then let $z_1 = \min_{x \in Q_1} d(\tilde{v}_1, \tilde{x})$. We now proceed inductively. For $h \ge 1$, define

$Q_{h+1} = \{\tilde{x} \in Q_h : d(\tilde{v}_h, \tilde{x}) \le z_h + \epsilon_h z_1\}$ where $v_i^{(h)} = \min_{\tilde{x} \in Q_h} x_i$ and

$\tilde{v}_h = (v_1^{(h)}, v_2^{(h)}, \ldots, v_n^{(h)})$ and $z_h = \min_{\tilde{x} \in Q_h} d(\tilde{v}_h, \tilde{x})$. Then, it can be shown

that $Q(A) = \bigcap_{h=1}^{\infty} Q_h$ satisfies all of the requirements.

One of the consequences of the above construction is that $Q(A)$ is a single point $\tilde{s}$. However, the specific single point obtained may depend on the choice of the sequence $\{\epsilon_j\}$ employed.

The reader's intuition concerning the above construction may be aided by considering the process as a limit of a sequence of minimax-regret procedures, as follows:

z_1 is the minimax regret decision procedure (more properly, it is the distance of the risk vector associated with the minimax regret decision procedure from $\tilde{v} = \tilde{v}_1$). Then a new convex polyhedron Q_2, is constructed, containing the risk vector for the minimax regret decision procedure, and the minimax regret decision procedure for Q_2 is determined. The process is repeated and converges to a single point $\tilde{s} = Q(A)$.

REFERENCES

[1] Blackwell, D. and Girshick, M. A. (1954). The Theory of Games and Statistical Decisions. Wiley, New York.

[2] Chernoff, H. (1954). Rational Selection of Decision Functions. Econometrica, 22, 422-443.

[3] Luce, R. D. and Raiffa, H. (1957). Games and Decisions. Wiley, New York.

[4] Milnor, J. W. (1954). Games Against Nrture. In Thrall, Coombs, and Davis, Decision Processes, Wiley, New York, 49-60.

PATHOPHYSIOLOGY OF INDIAN COBRA VENOM

James A. Vick, Henry P. Ciuchta, and James H. Manthei
Directorate of Medical Research
US Army Edgewood Arsenal
Chemical Research and Development Laboratories
Edgewood Arsenal, Maryland

INTRODUCTION. It has been reported that the venom of the hooded cobra, *Naja naja*, has a detrimental effect on the respiratory system of animals and man [1-3]. Several workers have attempted to fractionate the crude venom into its various toxic fractions [4, 5], they being: (a) neurotoxic fraction, (b) a cardiotoxic fraction and (c) a non-specific hemolytic fraction. Our study is concerned with the effect of crude cobra venom on cortical electrical activity (EEG), respiration and the cardiovascular system.

MATERIALS AND METHODS. In this study a total of 54 dogs and 5 monkeys of the Cynapthecoid group (sooty mangabey) were used. Of the above total, 44 adult mongrel dogs, anesthetized with sodium pentobarbital, 30 mg/kg, were used to study the effect of venom on the respiratory and cardiovascular systems. Femoral arterial pressure was monitored using a Statham strain gauge and a Grass polygraph recorder. The phrenic nerve was isolated at the level of the 5th cervical vertebra. The nerve was carefully dissected free of connective tissue and sectioned. Silver wire electrodes were connected to the central end of the phrenic from which nerve impulses were then monitored and amplified on a Tektronix oscilloscope. Permanent recordings were obtained photographically. Both EKG and respiratory rate were recorded in some of the animals using a Grass polygraph recorder. All of the above animals were administered (0.5 mg/kg) lyophilized crude cobra venom, which was reconstituted with normal saline and injected directly into the femoral vein.

The 44 animals were divided into the following groups: Group I was comprised of six animals used to study the overall effect of the venom on blood pressure and respiration. Group II - Eight animals ventilated with a Starling pump at respiratory arrest but prior to cardiovascular failure. The resultant effect on survival time was noted. Group III - The remaining thirty animals were used to study specific effects of the of the venom on the respiratory system, including the phrenic nerve and diaphragm. Nerve impulses over the central end of the cut phrenic nerve were continuously observed. The peripheral end of the cut phrenic nerve and the diaphragm were stimulated at intervals using a Grass model 54 stimulator. Diaphragmatic muscle contractions were recorded with a Grass Force Displacement Transducer. The effect of venom, artificial respiration and changes in pO_2 and pCO_2 tension on nerve activity were observed. Group IV comprised the remaining ten dogs and five monkeys which were used to monitor the effect of crude cobra venom (0.5 mg/kg) on cortical electrical activity. Blood pressure and respiratory effects were also recorded. The cortical electrical activity was recorded using bipolar silver electrodes which were surgically implanted directly on the dura of each hemisphere of the brain. Continuous electroencephalograms were recorded prior to and for up to 10 hours after the intravenous administration of the venom.

RESULTS.

Group I. The effect of venom on respiratory rate and arterial blood pressure is shown in Figure 1. Within 1-5 minutes post-injection there is an increase in respiratory rate as well as a sharp drop in blood pressure. This is followed by a progressive decrease in respiratory rate and volume to complete arrest at 90-120 minutes. During this time blood pressure makes a partial recovery remaining stable until respiratory failure, at which time cardiovascular collapse results. The average survival time of this group was 105 minutes.

Group II. The effect of venom on the artifically ventilated animal is shown in Figure 2. These animals were placed on a positive pressure respirator at time of respiratory cessation, with a resultant increase in survival time of from 4-6 hours. However, all animals untimately developed arrhythmias and progressive hypotension which led to death Figure 3. The average survival time for this group of animals was 7.5 hours post-venom.

Group III. Changes in phrenic nerve action potentials induced by cobra venom are shown in Figure 4. Action potentials prior to venom are synchronous corresponding to the inspiratory phase of respiration. Increase in both rate and amplitude are noted within 1-5 minutes after administration of venom. The central component of the nerve continues to discharge for from 5-10 minutes after complete cessation of respiration. During this period phasic discharges over the phrenic nerve become sporadic and irregular. These central impulses are eliminated by placing the animal on the artificial respirator. At any time prior to death impulse traffic can again be re-established by discontinuing artificial respiration, even though the animals do not breathe spontaneously. Phrenic impulses, as seen on the oscilloscope, continue with increasing frequency and amplitude until the animal either expires or is again ventilated.

The administration of 5 percent CO_2 to artifically ventilated animals initiates discharges over the phrenic nerve. This is quickly eliminated by removal of the stimulus. Phasic phrenic discharges can also be elicited in ventilated animals by the reduction of their tidal volume. Where such activity is noted the administration of 100 percent oxygen does not eliminate or appreciably alter their frequency or amplitude.

The terminal effect of venom on impulse traffic over the phrenic nerve is characterized by abnormal appearing bursts probably due to a combination of hypotension and central nervous system ischemia.

Spontaneous contractions of the diaphragm show a gradual decrease in force of contraction after venom ultimately leading to complete cessation of movement Figure 5 [6].

Group IV. The effect of crude cobra venom (0.5 mg/kg) on the EEG of the dog and monkey can be seen in Figure 6. Within 30-60 seconds following the administration of the venom there was complete loss of EEG, as well as corneal reflexes. There also occurred a sharp drop in arterial blood pressure shortly after cessation of all EEG activity. This hypotension was followed by a partial recovery. The effect of the venom on EEG was irreversible. As seen in Table I all animals expired, with an average survival time of 1.4 hours in the dog and 2.0 hours in the monkey.

DISCUSSION. This study has characterized the effects of crude cobra venom (0.5 mg/kg) on the peripheral respiratory mechanism, cardiovascular system and cortical electrical activity (EEG) of the dog and monkey. The respiratory effect is apparently due to a blockage of nerve impulses at the neuromuscular junction of the diaphragm. This is supported by the fact that the respiratory center remains functional after venom. There are continued phrenic dischagres, although somewhat modified following the venom. The muscle of the diaphragm remains in tact in that it retains its response to stimuli. This same stimulation when applied to the phrenic nerve produces no response in the diaphragm. It appears, therefore, that transmission of impulses is interferred with at the level of the neuromuscular junction. The character of this block is unknown.

The primary lethal effect of cobra venom, respiratory arrest, was shown to be alleviated with the application of artificial ventilation. This, however, was a temporary phenomena in that all animals eventually developed cardiovascular failure. The etiology of this phenomenon has not been studied but may be related to the action of venom on motor end plates [7]. The effect of venom on cardiovascular hemodynamics may also be due in part to its strong hemolytic effect, producing a high serum potassium which may result in cardiac failure [6].

The cortical electrical activity of the brain of the dog and monkey has been shown to be severely depressed by the action of cobra venom. The exact action of venom is not clear but may also, in some way, be related to its blocking effect on neuromuscular transmission [8].

SUMMARY. This study has dealt with the effects of cobra venom, Naja naja, on the respiratory system cardiovascular system and the cortical electrical activity of the dog and monkey. Results have indicated that death is primarily due to respiratory failure, which appears due to peripheral neuromuscular blockade. The character of this block is unknown. The respiratory center, phrenic nerve and diaphragmatic muscle fibers appear to be relatively unaffected by the venom. Survival time was increased several hours with artificial ventilation, however, all eventually developed cardiovascular difficulties terminating in death. This effect may be due to the extended action of venom on the areas of the body.

In addition, venom has been shown to have a severe depressant effect on the cortical electrical activity of the dog and monkey. The exact mechanism by which this effect is produced has not as yet been defined.

REFERENCES

1. Ganguly, S.N. and Malkana, M.T.: Indian F. med. Res. 24: 281, 1936.

2. Sarkar, B.B., Mitra, S.R and Ghosh, B.N.: Indian F. med. Res. 30: 453, 1942.

3. Gautrelet, J., Halpern, N. and Cortiggiani, E.: Arch. int. Physiol. 38: 293, 1934.

4. Master, R.W.P. and Rao, S.S.: F. biol. Chem. 236: 1986, 1961.

5. Sarkar, N.K.: Ann. Biochem. exper. Med. 8: 11, 1948.

6. Vick, J.A. Ciuchta, H.P. and Polley, E.H.: Arch. int. Pharmacodyn. Vol. 153, No. 2: 424-429, 1965.

7. Chang, C.C. and Lee, C.Y.: Arch. int. Pharmacodyn. 144: 214, 1963.

8. Vick, J.A., Ciuchta, H.P. and Polley, E.H.: Nature. Vol. 203, No. 4952: 1387-1388, Sept. 26, 1964.

LEGENDS FOR ILLUSTRATIONS

Figure 1. The effect of cobra venom on arterial blood pressure and respiratory rate.

Figure 2. Modification of venom effect by use of artificial respirator.

Figure 3. The effect of cobra venom on cardiovascular function after respiratory arrest and subsequent artifical ventilation.

Figure 4. Changes in phasic phrenic discharges produced by cobra venom. Effects of artificial respiration and administration of 5 percent CO_2 after cessation of spontaneous respiration are shown.

Figure 5. Effect of cobra venom on blood pressure, phrenic nerve discharges and diaphragmatic contractions. Note: Loss of diaphragmatic response to direct phrenic stimulation (PS). Diaphragmatic responses to direct stimulation (DS) are retained.

Figure 6. The effect of cobra venom on EEG and blood pressure.

TABLE I

Effect of Cobra Venom on Cortical Electrical Activity

	No. of animals	EEG change	Average survival time (h)
Dogs	10	10/10	1.4 (0.3-2.2)
Monkeys	5	5/5	2.0 (1.1-3.1)

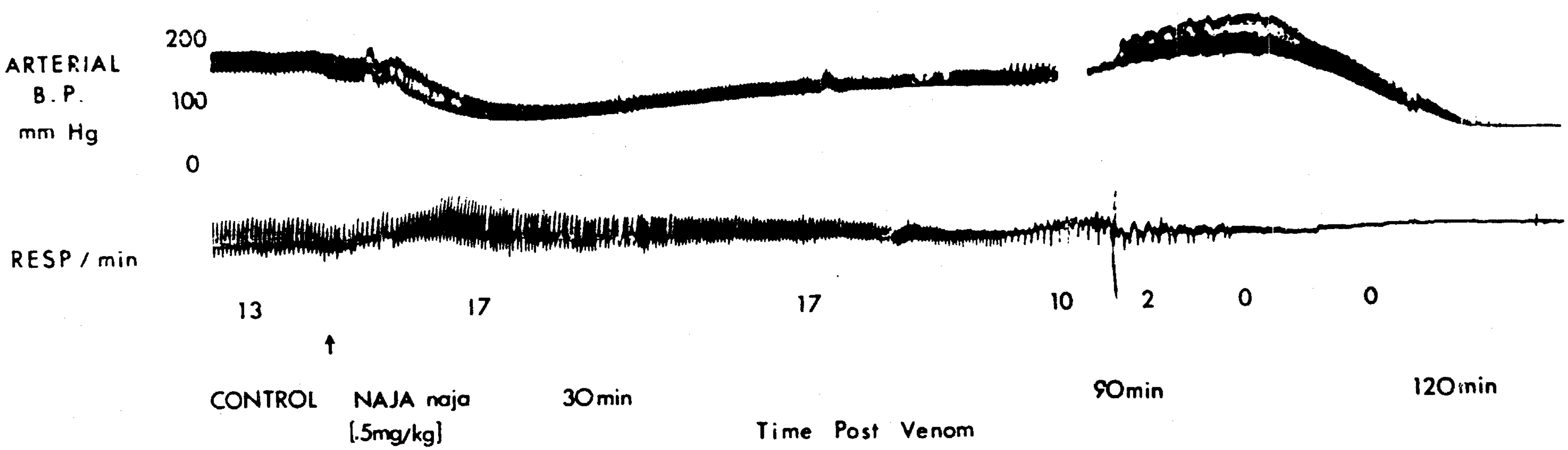

ARTERIAL
B.P.
mm Hg
200
100
0
RESP / min
13
17
17
10
2
0
0
CONTROL
NAJA naja
[.5mg/kg]
30min
Time Post Venom
90min
120min

Figure 1

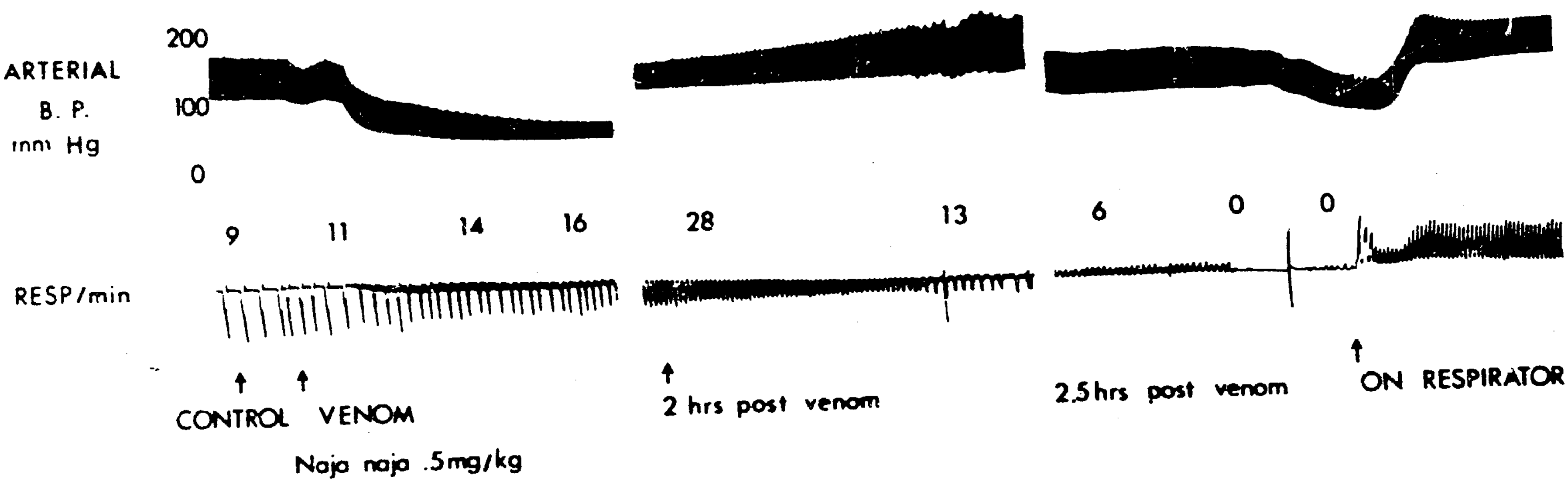

ARTERIAL
B. P.
mm Hg
200
100
0
9
11
14
16
28
13
6
0
0
RESP/min
CONTROL
VENOM
Naja naja .5mg/kg
2 hrs post venom
2.5hrs post venom
ON RESPIRATOR

Figure 2

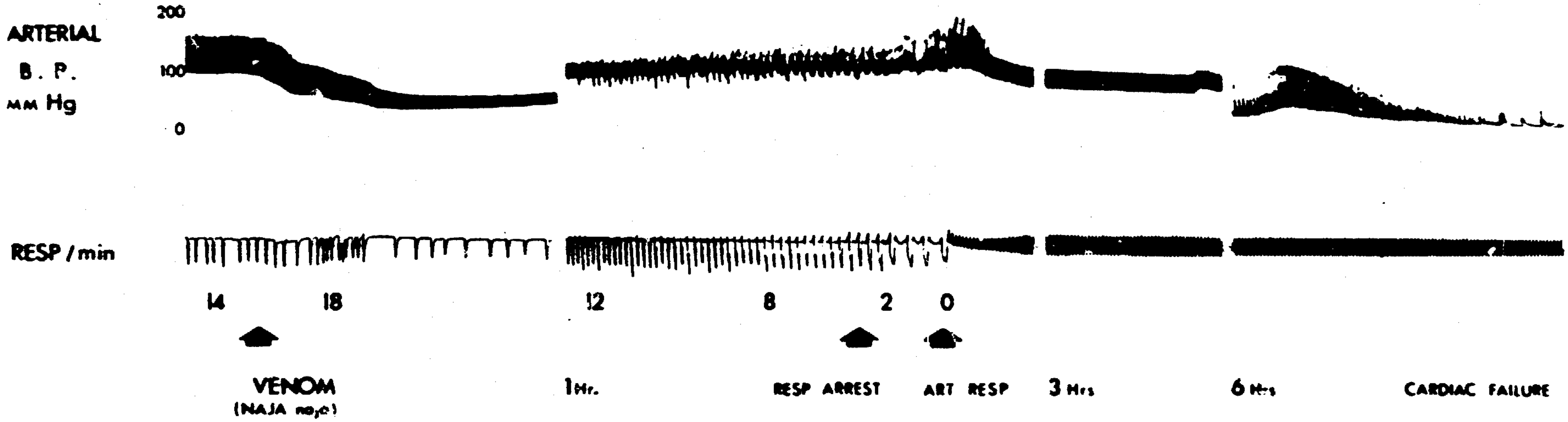

ARTERIAL
B. P.
MM Hg
200
100
0
RESP / min
14
18
12
8
2
0
VENOM
(NAJA naja)
1Hr.
RESP ARREST
ART RESP
3 Hrs
6 Hrs
CARDIAC FAILURE

Figure 3

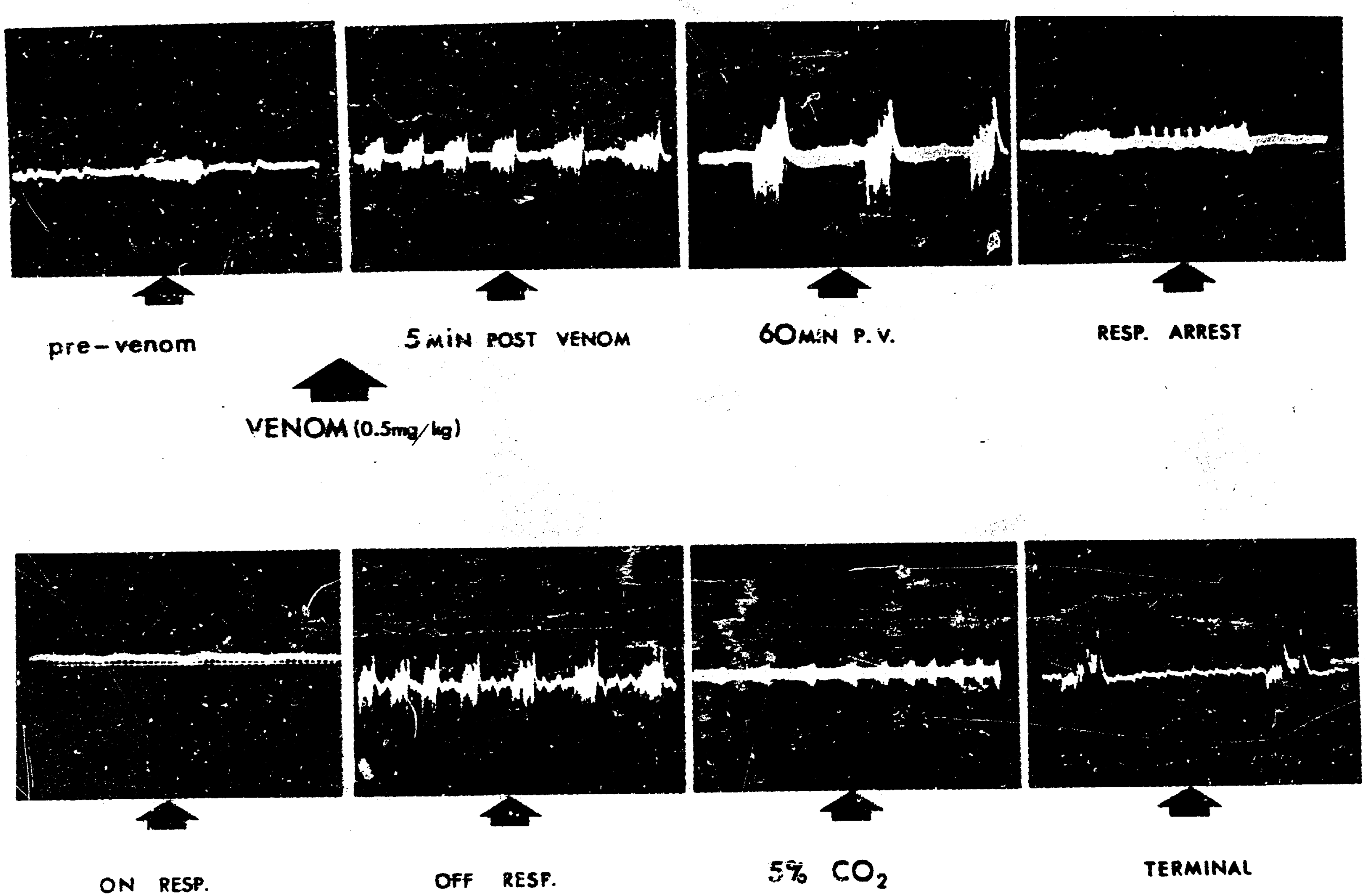

pre-venom
5 MIN POST VENOM
60 MIN P.V.
RESP. ARREST
VENOM (0.5mg/kg)
ON RESP.
OFF RESP.
5% CO2
TERMINAL

Figure 4

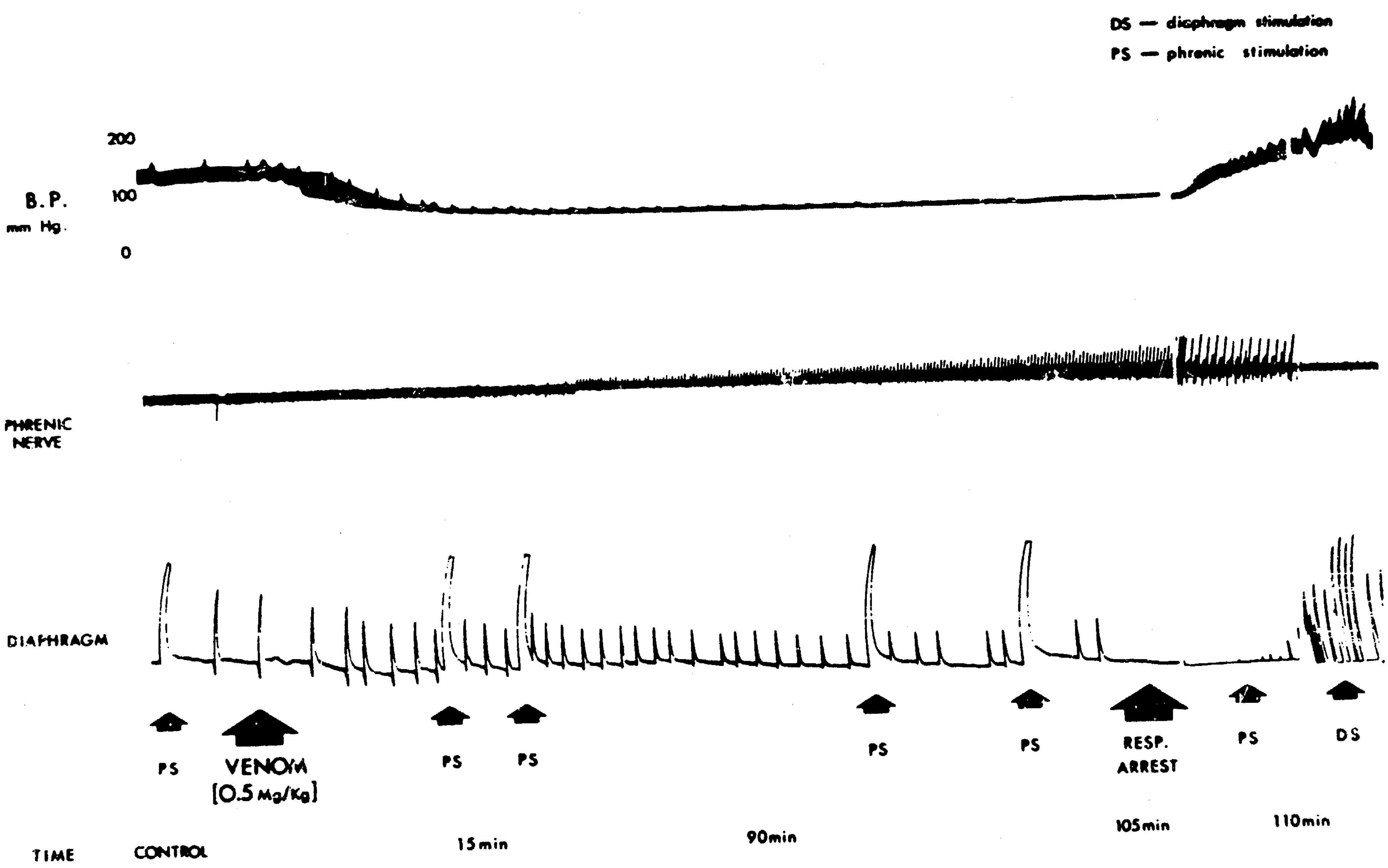

DS — diaphragm stimulation
PS — phrenic stimulation
200
100
0
B.P.
mm Hg.
PHRENIC
NERVE
DIAPHRAGM
PS
VENOM
[0.5 Mg/Kg]
PS
PS
PS
PS
RESP.
ARREST
PS
DS
TIME
CONTROL
15 min
90min
105min
110min

Figure 5

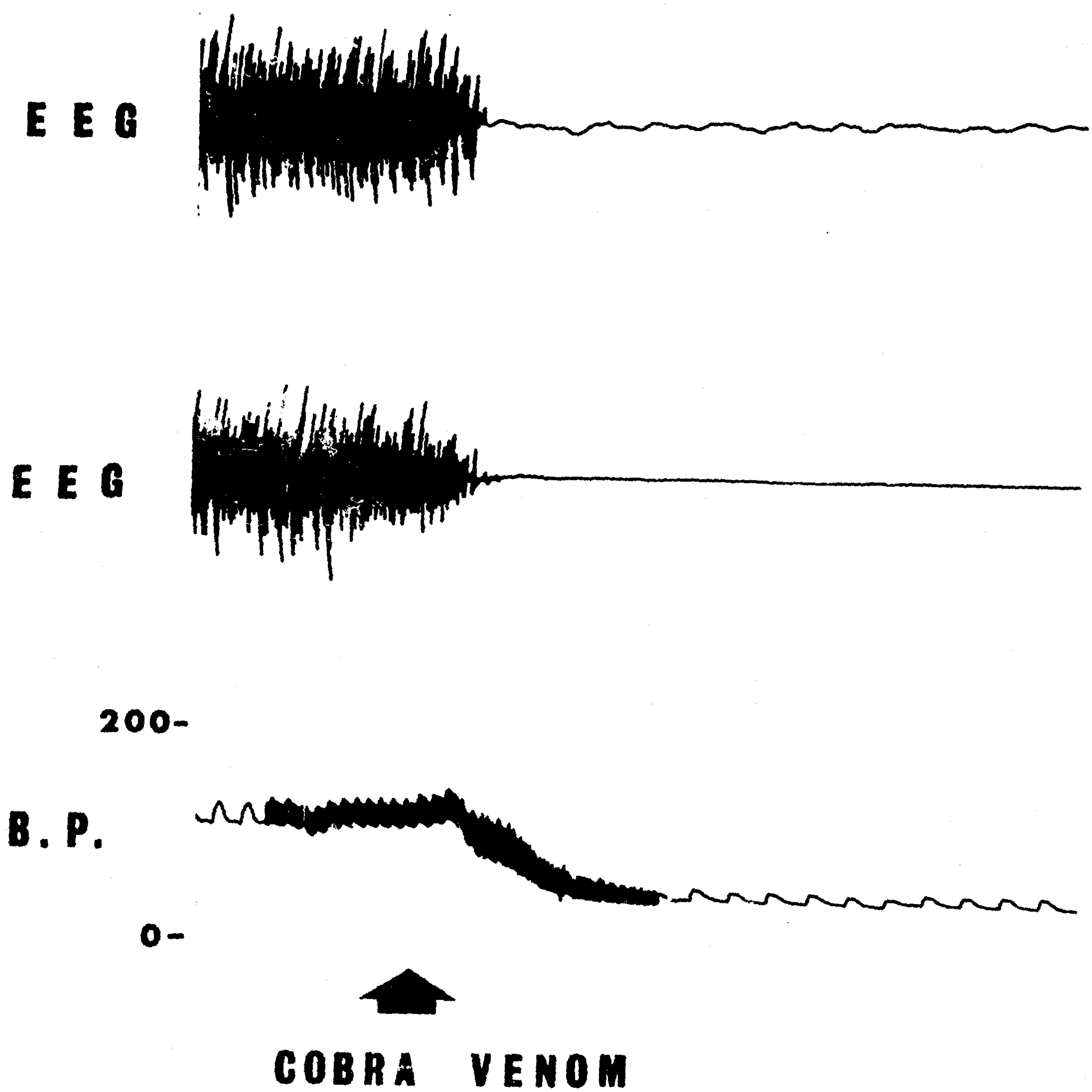


Figure 6

COMPUTER ANALYSIS OF VISUAL DISCRIMINATION DATA

John C. Atkinson
Directorate of Medical Research,
Chemical Research & Development Laboratories
Edgewood Arsenal, Maryland

One of the methods used by the Directorate of Medical Research, Chemical Research and Development Laboratories in evaluating the effect of various drugs on an animal's performance is a visual discrimination test. This is a conditioned visual discrimination between a triangle and a square in which monkeys are trained to avoid or escape an electric shock by pressing a lever under the correct symbol, the triangle. Thus, successful performance involves sensory perception (vision) decision making and motor activity (pressing the lever).

If a drug interfers with any of these activities the result will be a slowed or inaccurate performance. An obvious correlation can be seen between this test and many tasks performed by a soldier during combat.

In our operation Rhesus monkeys are used as test subjects. Each monkey is placed in a sound attenuated booth which is enclosed to prevent visual as well as audio distraction. The monkey is restrained by a Wilinski harness*[1]. By this means the monkey is kept in front of a panel on which there are two screens at an equal level. At the beginning of a trial a triangle appears on one screen and a square on the other. If the monkey presses the lever under the triangle, the symbols disappear from the screen and the trial is over. If he presses the lever under the square he receives a punishment in the form of a mild electrical shock for twenty (20) milliseconds. This is called an incorrect response. If the monkey does not press the designated lever in an interval of five (5) seconds he receives a negative reinforcement in the form of a mild electrical shock. This shock continues for five (5) seconds unless sooner shut off by pressing the lever under the triangle. Pressing the correct lever before the shock is considered an avoidance response. Pressing the correct lever after the shock has started is considered an escape response. Never

---

*Patent Pending

[1] Frank T. Wilinski - Effects of Atropine Sulfate on Trained Monkeys. Manuscript in progress.

pressing the correct lever is a no response. The time interval between the trial start and a correct response is considered response latency. Pressing either lever when their is no figure on the screen is called an intertrial response.

The electrical equipment associated with trial presentation and the paper tape punch are rack mounted behind each booth. Two (2) loops of punched mylar tape on each rack control the presentation of the trials. The shorter loop initiates trial starts and is punched at random intervals in order that no discernible trial start pattern will be presented to the monkey. Circuity in the rack presents the triangle on the right or left screen in a random order with the restriction that the long term expectation of the number of presentations on the two sides be equal. The longer tape starts and stops the trial presentation tape. The monkeys are given five (5) sessions per day, 55 minutes each, with a five (5) minute break between sessions. No presentations are made for the remaining 19 hours. The monkeys live in the test booths for several days during testing. Food and water are provided adlibitum and the cage provides enough room for the monkey to lie down. A tray of absorbent material underneath the woven wire floor is provided for excretions.

A record of the experiment is made on punched paper tape containing six (6) information codes. When a trial is initiated the punch emits a "start of trial punch" and continues to run at 10 characters per second, emitting a code associated with latency until a correct response is made or the trial is automatically terminated. Separate codes are made for latencies following a right or left screen presentation. At present no distinction is made between right or left latencies upon computer analysis. A code is associated with the avoidance response, with the escape response and with an incorrect response. A separate code for right or left presentations is provided for an intertrial response. Since it is quite possible that two or three of the above things could happen in a single 1/10 second period, the code is designed for this. Forty-one separate codes may appear. Since the tape has 6 information channels it is possible to represent up to 64 codes thus 41 presents no coding problem. At the end of a session an end of session code is automatically punched.

At the end of a day the punched paper tapes are removed from the take-up roll on each equipment rack and sent to the computer group for analysis. During the 55 minute sessions an average of 104 trial presentations are made. To obtain frequent measurements of the progress of the monkey, the data are considered as 4 subgroups by the computer. These subgroups are termed segments. The first 3 segments contain exactly 26 trials while the last contains the number of trials remaining.

For each segment and for the session the geometric mean of the trial latencies is computed. The computer determines each trial latency by counting the number of tape frames between the start of trial punch and an avoidance or escape punch. If neither occur in 100 tape frames this is considered a no response, and the latency is taken as 10 seconds. The standard error is computed in terms of log latencies for each segment and each session. The 95% fiducial limits are computed for the geometric mean latency for each segment and for each session, and the mean and its limits are printed. Analysis in terms of log latencies is done to minimize the skewness of the latencies which results from the physical inability of the animal to react in less than 2 or 3 tenths of a second. This would truncate the deviations on the minus side. Deviations on the positive side are only truncated after the cut off time of 10 seconds. Since the mean response time is generally from 1/2 to 1 second the positive deviations can be many times as large as the negative ones. This causes skewness. Conversion of the latencies to their logarithms minimizes this skewness.

Session one of each day is considered a control run and any drug is administered between session one and two. A "t" test for significance is made between the mean latency in terms of logarithms of each of the 4 subsequent sessions and the control run. The statement "not significant, or significant at 95%, or significant at 99%, or significant at 99.9%" is printed after each of the sessions 2, 3, 4 and 5. The sum of all latencies for a session is printed at the end of each session. In addition to the analysis of the latencies the computer counts and prints for each segment the number of occurrences of each of the following: avoidance responses, escape responses, incorrect responses done with the right hand, incorrect responses done with the left hand, intertrial responses done with the right hand, intertrial responses done with the left hand and the no responses. No analysis is made of these figures at the present time.

A typical computer printed output is presented as Figure I. The numeric portion of this output is simultaneously punched into paper tape. This tape is to be converted into Holorith cards for storage and will allow future manipulation of the test results.

FIGURE I

COMPUTER OUTPUT FROM VISUAL DISCRIMINATION ANALYSIS

G. MEAN	L. LIMIT	U. LIMIT	A.R.	E.R.	R. INC.	L. INC.	R. INT.	L. INT.	N.R.
.6879	.6033	.7843	26.0000	.0000	.0000	.0000	7.0000	8.0000	.0000
.7576	.6954	.8254	26.0000	.0000	.0000	.0000	5.0000	2.0000	.0000
.7480	.6839	.8181	26.0000	.0000	.0000	.0000	8.0000	.0000	.0000
.7471	.6886	.8106	25.0000	.0000	.0000	.0000	4.0000	2.0000	.0000
.7345	.7001	.7706							80.2998
.8709	.7947	.9544	26.0000	.0000	.0000	.0000	5.0000	4.0000	.0000
.8873	.7980	.9866	26.0000	.0000	.0000	.0000	3.0000	2.0000	.0000
.8404	.7905	.8934	26.0000	.0000	.0000	.0000	4.0000	1.0000	.0000
.9255	.8774	.9763	24.0000	.0000	.0000	.0000	3.0000	2.0000	.0000
.8796	.8455	.9152							93.3999
SIGNIFICANT 99 PER CENT									
.9500	.8715	1.0356	26.0000	.0000	.0000	.0000	4.0000	2.0000	.0000
1.0081	.8948	1.1357	26.0000	.0000	.0000	.0000	12.0000	6.0000	.0000
1.0831	.9571	1.2258	26.0000	.0000	.0000	.0000	7.0000	1.0000	.0000
1.0840	.9932	1.1831	21.0000	.0000	.0000	.0000	9.0000	.0000	.0000
1.0271	.9744	1.0826							107.2998
SIGNIFICANT 99.9 PER CENT									

G. MEAN = GEOMETRIC MEAN
U. LIMIT = UPPER LIMIT (95 PER CENT CONFIDENCE)
L. LIMIT = LOWER LIMIT (95 PER CENT CONFIDENCE)
E.R. = ESCAPE RESPONSES
A.R. = AVOIDANCE RESPONSES
L. INC. = LEFT HAND INCORRECTS
R. INC. = RIGHT HAND INCORRECTS
L. INT. = LEFT HAND INTERTRIALS
R. INT. = RIGHT HAND INTERTRIALS
N.R. = NO RESPONSES
NUMBER AT END OF EACH SESSION UNDER NO RESPONSE COLUMN IS TOTAL LATENCY FOR SESSION

FATIGUE-LIMIT ANALYSES AND DESIGN OF FATIGUE EXPERIMENTS

A. H. Soni and R. E. Little
Oklahoma State University
Stillwater, Oklahoma

INTRODUCTION. It is generally accepted that there is as much, if not more scatter associated with fatigue than with any other mode of failure. Consequently, fatigue presents a challenging problem to both the engineer and the statistician.

The purpose of fatigue analyses is to adduce information about the probability of relatively rare events, not to describe the mean or modal event. Accordingly, the statistical problem in fatigue is to establish the alternating stress amplitude that corresponds to the optimum economic level of tolerable failures.

A brief resume of the nature of fatigue is presented here before discussing existing data and the design of future experiments.

NATURE OF METAL FATIGUE. Fatigue is caused by continued cycle stressing. A fatigue failure can be recognized by fitting the two broken pieces back together and observing the original geometry. As indicated in Figure 1, there is no evidence of gross plastic deformation prior to failure by fatigue.

Fatigue cracks are the cumulative result of micro-inelastic behavior occuring within the substructure of the metal. Electron microscopy and X-ray diffraction studies have shown:

(1) the physical mechanisms associated with fatigue are of a 10^{-3} to 10^{-7} cm observation level, and

(2) these physical mechanisms are intimately related to actual defects (dislocations) in the theoretical atomic arrangement.

The statistical nature of fatigue is intuitively apparent when fatigue is viewed as being caused by these minute substructural defects.

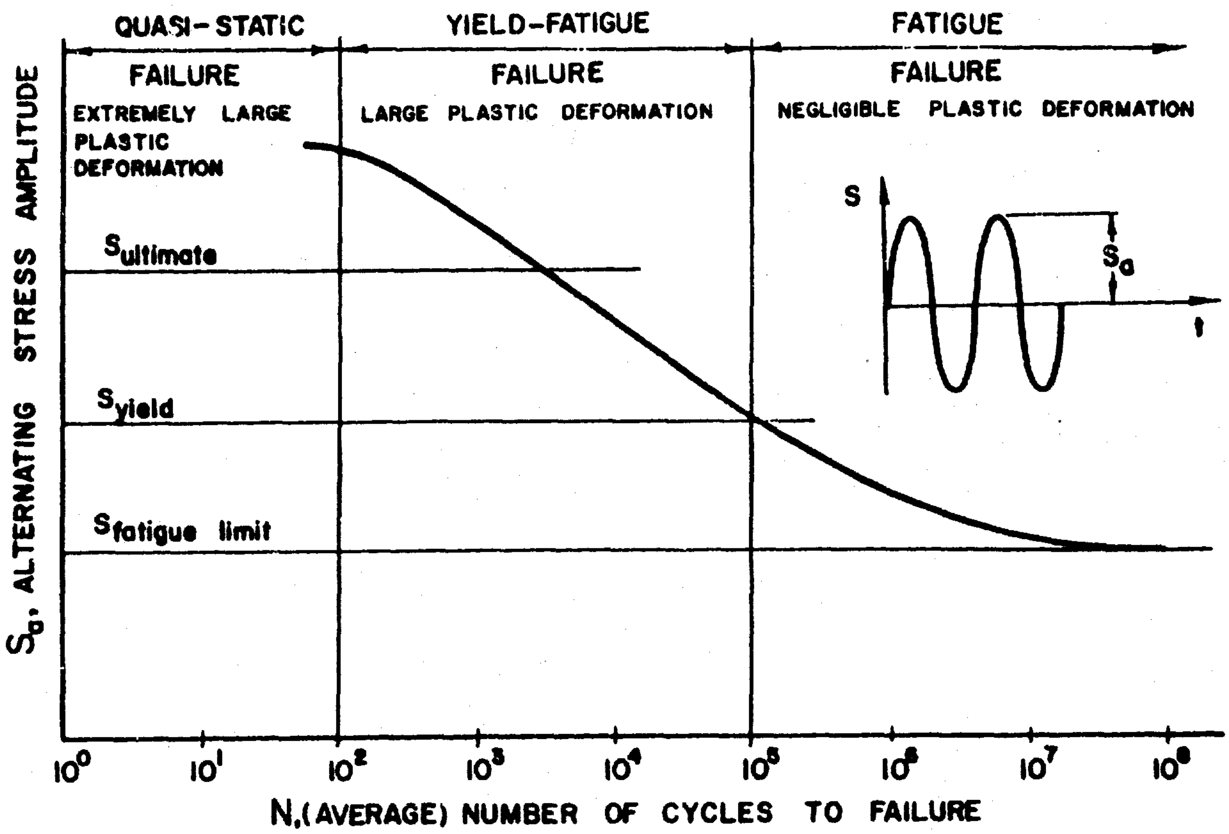


Figure 1 S-N Curve
The lower the alternating stress amplitude
the greater the over-all fatigue life, N_f

This intuitive view can be enhanced by considering an idealized material model. First, recall that metals are aggregate structures of randomly oriented crystallites (grains), and that individual crystallites are anisotropic (exhibit differenct properties and strengths in different directions). Now consider the static yield strength of the metallic tensile specimen shown in Figure 2. It theoretically has a unique yield strength only if all crystallites are perfect and have the same orientation. But, since the crystallites of commercial metals have defects and are randomly oriented, the crystallites within this specimen must exhibit a strength distribution.

Observe in Figure 2 that only a few crystallites experience yielding at low stress levels. But, under alternating stressing (Figure 1), these few crystallites yield first in tension, then in compression, then again in tension, and so forth. This localized reversed slip deformation will eventually lead to a fatigue crack in crystallites where the slip magnitude (fatigue intensity) is high. Thus, the number of crystallites that serve as potential fatigue crack initiation sites as well as the fatigue intensity at these sites are directly related to the crystallite strength distribution. Accordingly, fatigue is a statistical problem.

Fatigue failure theories are in their infancy---theory lags experimental work. The present criterion for the relative evaluation of various statistical functions is simply their goodness of fit with regard to data. Figure 3 shows the two types of fatigue data considered, namely:

(1) data stated in terms of a life distribution.

(2) data stated in terms of a strength distribution.

In turn, the over-all objective of all statistical analyses of fatigue data is to develop the P-S-N surface shown in Figure 4.

Existing data indicates that the P-S-N surface is warped and cannot be described in its entirety by a simply mathematical function. This paper treats a small but significant portion of this surface---the statistical analyses of fatigue-limits in terms of a strength distribution.

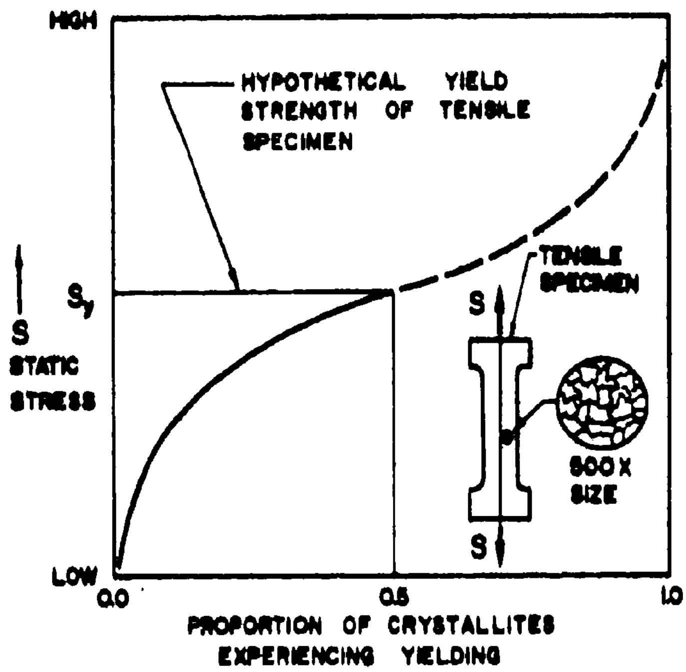


Figure 2 Strength Distribution of Tensile Specimen Crystallites.

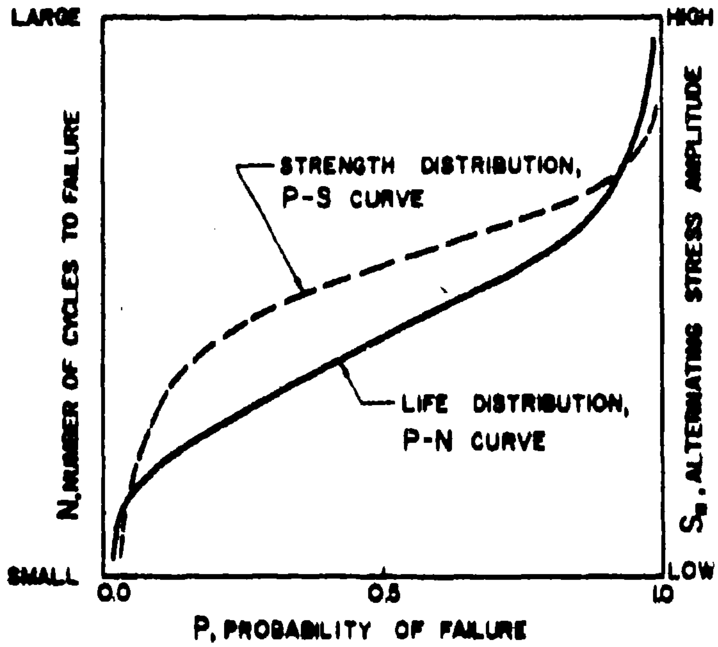


Figure 3 Fatigue Strength and Fatigue Life Distributions
The life distribution is markedly skewed to the right.

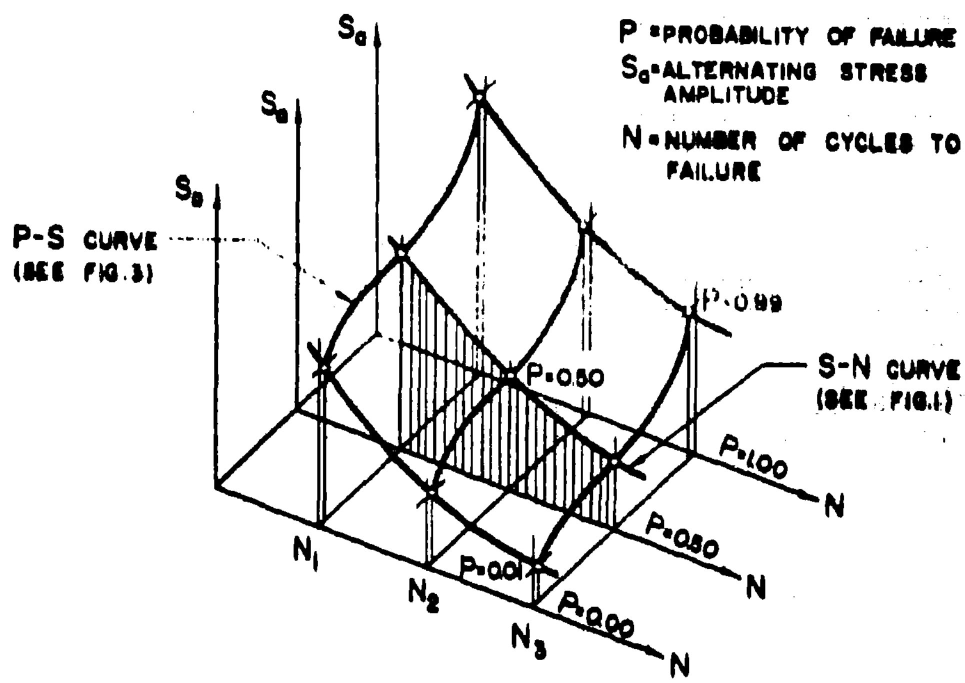


Figure 4 P-S-N Surface
This surface completely defines fatigue failure.

PART I - ANALYSES OF EXISTING FATIGUE-LIMIT DATA

COMMON DISTRIBUTIONS. The three common statistical functions applied herein to fatigue-limit data are listed rows 1, 2, and 3 of Table 1.

Typical fatigue-limit data appears in Table 2. Observe that the statistics recorded are simply the alternating stress amplitudes and the corresponding proportion of specimens failed prior to the given fatigue life.

These common functions are fitted to the observed statistics by using a minimum residual χ^2 approach. For example, the logistic function is fitted by minimizing the logit $\chi^2 = \Sigma\ Npq(\ell - \hat{\ell})^2$, where $\hat{\ell} = \hat{\alpha} + \hat{\beta}\ s$. Taking the partial derivative of the logit χ^2 with respect to $\hat{\alpha}$ and $\hat{\beta}$ and then setting these expressions equal zero; simultaneous solution of the two resulting equations yields the expressions for the estimates listed in rows 4 and 5 of Table 1.

OTHER DISTRIBUTIONS. The goodness of fit of the common (two-parameter) functions can be evaluated by examining the goodness of fit for three-parameter functions, i.e., determining whether the third parameter is really required to describe the data.

Table 3 lists these three-parameter functions. The estimates listed in rows 4, 5, and 6 are established by taking the partial derivative of χ_3^2 with respect to $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, respectively; setting these expressions equal to zero, and then solving these three equations simultaneously.

The significance of the third parameter, γ, can be now determined from the magnitude of

$$F \approx \frac{\chi^2 - \chi_3^2}{(\chi_3^2 / K-3)}$$

where F has one and (K-3) degrees of freedom. (At least five datum points are desirable for comparative residual χ^2 analyses.)

TABLE 1. COMMON DISTRIBUTIONS

	Normal	Logistic	Extreme Value - Type I
Probability Function	$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{t^2}{2}} dt$	$P = \frac{1}{1 + e^{-(\alpha + \beta s)}}$	$P = 1 - e^{-e^{-(\alpha + \beta s)}}$
Linear Transform	$Y = \alpha + \beta s$	$\ell = \ln \frac{P}{Q} = \alpha + \beta s$	$g = \ln \ln \frac{1}{1 - P} = -(\alpha + \beta s)$
Parameters	α, β	α, β	α, β
Estimate of β	$\hat{\beta} = \frac{\Sigma w \Sigma w Y s - \Sigma w Y \Sigma w s}{\Sigma w \Sigma w s^2 - (\Sigma w s)^2}$ Weights, w, are listed in Ref. [4].	$\hat{\beta} = \frac{\Sigma w \Sigma w \ell s - \Sigma w \ell \Sigma w s}{\Sigma w \Sigma w s^2 - (\Sigma w s)^2}$ Where $w = Npq$	$\hat{\beta} = -\frac{K \Sigma g s - \Sigma g \Sigma s}{K \Sigma s^2 - (\Sigma s)^2}$ Where K = No. of Stress Amplitudes
Estimate of α	$\hat{\alpha} = \frac{\Sigma w Y - \hat{\beta} \Sigma w s}{\Sigma w}$	$\hat{\alpha} = \frac{\Sigma w \ell - \hat{\beta} \Sigma w s}{\Sigma w}$	$\hat{\alpha} = -\frac{\Sigma g - \hat{\beta} \Sigma s}{K}$

TABLE 2. RESULTS OF ROTATING BENDING FATIGUE TESTS ON SAE 4340 STEEL. ($N = 10^7$ cycles)

$S_u = 190$ ksi, $K_t = 2.6$.

(Data by Cummings, Stulen, and Schulte)

Test	Stress Level s, ksi	Number Tested	Number Failed	Proportion Failed p
1	32	110	1	0.0091
2	35	60	3	0.0500
3	38	30	6	0.2000
4	41	20	14	0.7000
5	42	20	16	0.8000

TABLE 3. THREE-PARAMETER DISTRIBUTIONS

	Normal	Logistic	Extreme Value--Type I	Weibull[(a)]
Probability Function	$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{t^2}{2}} dt$	$P = \frac{1}{1 + e^{-(\alpha + \beta s + \gamma s^2)}}$	$P = 1 - e^{-e^{-(\alpha + \beta s + \gamma s^2)}}$	$P = 1 - e^{-\frac{(s-\gamma)^{\beta}}{\alpha'}}$
Modified Transform	$Y = \alpha + \beta s_1 + \gamma s_2$ where $s_2 = s_1^2 = s^2$	$\ell = \alpha + \beta s_1 + \gamma s_2$ where $\ell = \ln \frac{P}{1-P}$; $s_2 = s_1^2 = s^2$	$g = -(\alpha + \beta s_1 + \gamma s_2)$ where $g = \ln\ln \frac{1}{1-P}$; $s_2 = s_1^2 = s^2$	$g = \alpha + \beta x$ where $\alpha = \ln \alpha'$ and $x = \ln(s - \gamma)$
Parameters	$\alpha,\ \beta,\ \gamma$	$\alpha,\ \beta,\ \gamma$	$\alpha,\ \beta,\ \gamma$	$\alpha,\ \beta,\ \gamma$
Estimate of β[(b)]	$\hat{\beta} = \frac{\Sigma wS_1S_2\Sigma wYS_2 - \Sigma wYS_1\Sigma wYS_2^2}{(\Sigma wS_1S_2)^2 - \Sigma wS_1^2\,\Sigma wS_2^2}$ Weights, w, are given in ref. [4].	$\hat{\beta} = \frac{\Sigma wS_1S_2\Sigma wLS_2 - \Sigma wLS_1\Sigma wS_2^2}{(\Sigma wS_1S_2)^2 - \Sigma wS_1^2\,\Sigma wS_2^2}$ where $w = Npq$	$\hat{\beta} = -\frac{\Sigma S_1S_2\Sigma GS_2 - \Sigma GS_1\Sigma S_2^2}{(\Sigma S_1S_2)^2 - \Sigma S_1^2\,\Sigma S_2^2}$	$\hat{\beta} = \frac{K\Sigma gx - \Sigma g\Sigma x}{K\Sigma x^2 - (\Sigma x)^2}$
Estimate of γ[(b)]	$\hat{\gamma} = \frac{\Sigma wS_1S_2\Sigma wYS_1 - \Sigma wS_1^2\,\Sigma wYS_2}{(\Sigma wS_1S_2)^2 - \Sigma wS_1^2\,\Sigma wS_2^2}$	$\hat{\gamma} = \frac{\Sigma wS_1S_2\Sigma wLS_1 - \Sigma wS_1^2\,\Sigma wLS_2}{(\Sigma wS_1S_2)^2 - \Sigma wS_1^2\,\Sigma wS_2^2}$	$\hat{\gamma} = -\frac{\Sigma S_1S_2\Sigma GS_2 - \Sigma S_1^2\,\Sigma GS_2}{(\Sigma wS_1S_2)^2 - \Sigma S_1^2\,\Sigma S_2^2}$	γ is estimated by an iterative procedure -- see footnote (a).
Estimate of α	$\hat{\alpha} = \frac{\Sigma wy - \hat{\beta}\Sigma ws_1 - \hat{\gamma}\Sigma ws_2}{\Sigma w}$	$\hat{\alpha} = \frac{\Sigma w\ell - \hat{\beta}\Sigma ws_1 - \hat{\gamma}\Sigma ws_2}{\Sigma w}$	$\hat{\alpha} = -\frac{\Sigma g - \hat{\beta}\Sigma s_1 - \hat{\gamma}\Sigma ws_2}{K}$	$\hat{\alpha} = \frac{\Sigma g - \hat{\beta}\Sigma x}{K}$

(a) Weibull's function is fitted using an iterative procedure, i.e., the correlation between g and x is maximized by iterating with regard to γ. This correlation coefficient is given by

$$\rho = \Sigma GX \ / \ \sqrt{G^2 X^2} .$$

(b) β and γ are estimated by adjusting the various variables about their means, i.e.,

$Y = y - \Sigma wy/\Sigma w$; $\quad L = \ell - \Sigma w\ell/\Sigma w$; $\quad G = g - \Sigma g/K$;

$S_1 = s_1 - \Sigma ws_1/\Sigma w$; $\quad S_2 = s_2 - \Sigma ws_2/\Sigma w$; $\quad X = x - \Sigma x/K$.

EXISTING DATA [1]. The mean-square error associated with fitting the two- and three-parameter functions appears in Table 4. Although the two-parameter functions are similar, the logistic and the extreme value functions fit the data slightly better than the integrated normal curve. See Figure 5. In turn, the three-parameter functions fit the data somewhat better than the two-parameter functions as shown in Figure 6. However the third parameter is required for only about one-half the data.

Table 5 emphasizes the similarities in the descriptive abilities of these functions. The respective (calculated) 10, 50, and 90 per cent responses are identical for practical purposes. These functions differ only at their tails as indicated by the extrapolated 0.1 per cent response. (These 0.1 per cent responses are computed only for illustrative purposes, and are not intended for use in design.)

Clearly, these data are not adequate to discern which function, if any, precisely describes the nature of the fatigue-limit. Consequently, further experimental study is required. The second part of this paper deals with the design of these tests.

PART II - DESIGN OF FUTURE FATIGUE-LIMIT EXPERIMENTS

EXPERIMENT DESIGN. The design of fatigue-limit experiments must overtly reflect efficiency in terms of over-all cost. Thus it is imperative to exploit fatigue testing. In turn, two considerations are basic to exploitation of fatigue testing:

(1) the minimum number of specimens required (for testing at a given alternating stress amplitude) to attribute a prescribed level of confidence in the position of the datum point, and

(2) preselected spacing of the different alternating stress amplitudes (datum points) to describe the distribution in an efficient manner.

The following discussion shows how simple statistical concepts can be used to design more efficient fatigue tests.

TABLE 4. RESPONSE MEASURED IN TERMS OF MEAN SQUARE ERROR

Test Series No.	Material	Ultimate Strength S_u (ksi)	Two-Parameter Distributions			Three-Parameter Distributions			
			Integrated Normal	Logistic	Extreme Value--Type I	Integrated Normal	Logistic	Extreme Value--Type I	Weibull
(1)	SAE 4330 Unnotched	130	1.145	0.872	0.794	0.571*	0.586	0.822	0.983
(2)	SAE 4340 Unnotched	239	0.511	0.484	1.472	0.766	0.708	0.699	0.789
(3)	SAE 4350 Unnotched	300	2.107	1.308	0.930	0.022***	0.003***	0.491	0.971
(4)	SAE 4350 Notched (1)	300	4.13	3.114	1.525	0.351**	0.307**	0.0001***	3.241
(5)	SAE 4340 Notched	190	0.533	0.132	0.127	0.119*	0.084	0.133	0.136

(1) $K_t = 2.6$

* Level of Significance for Y Pr < 0.05

** " " " " " Pr < 0.01

*** " " " " " Pr < 0.001

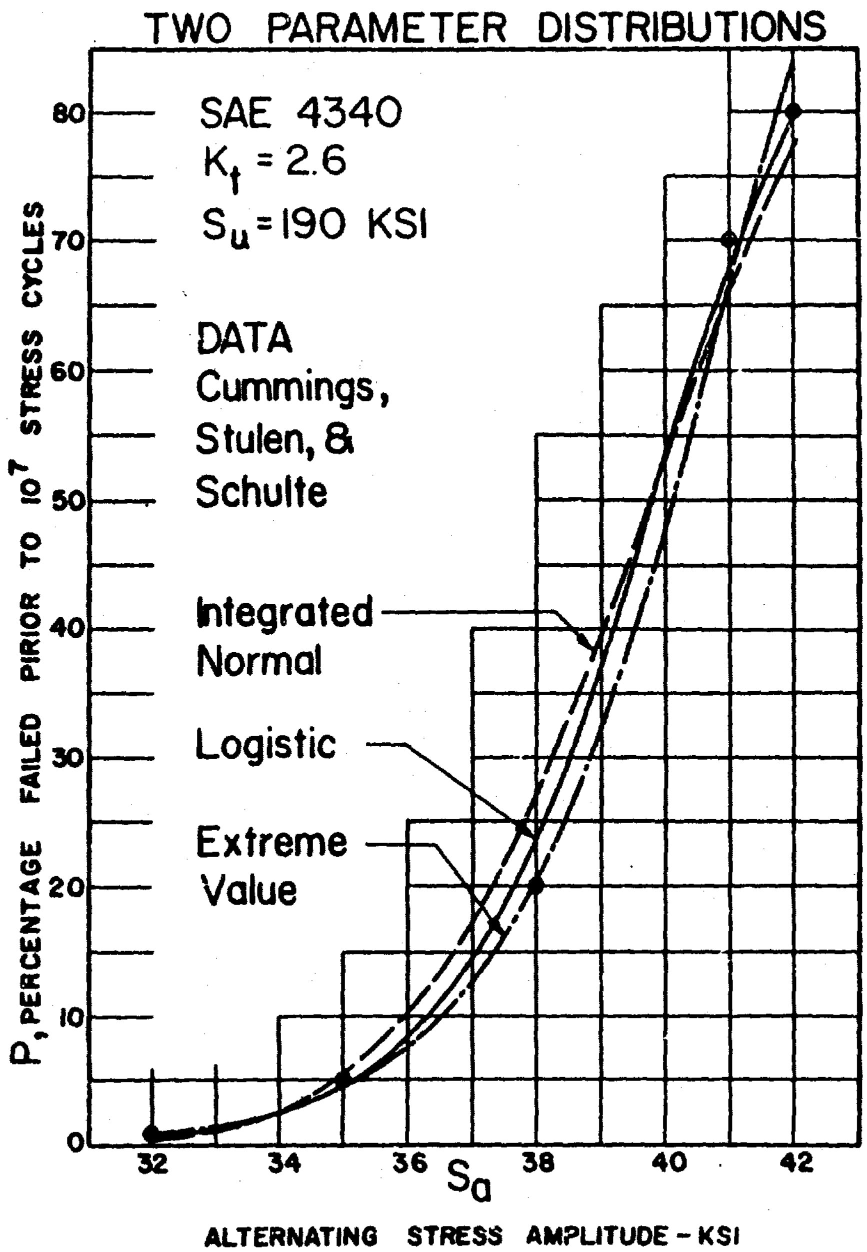


Figure 5 Typical Performance of the Two-Parameter Functions

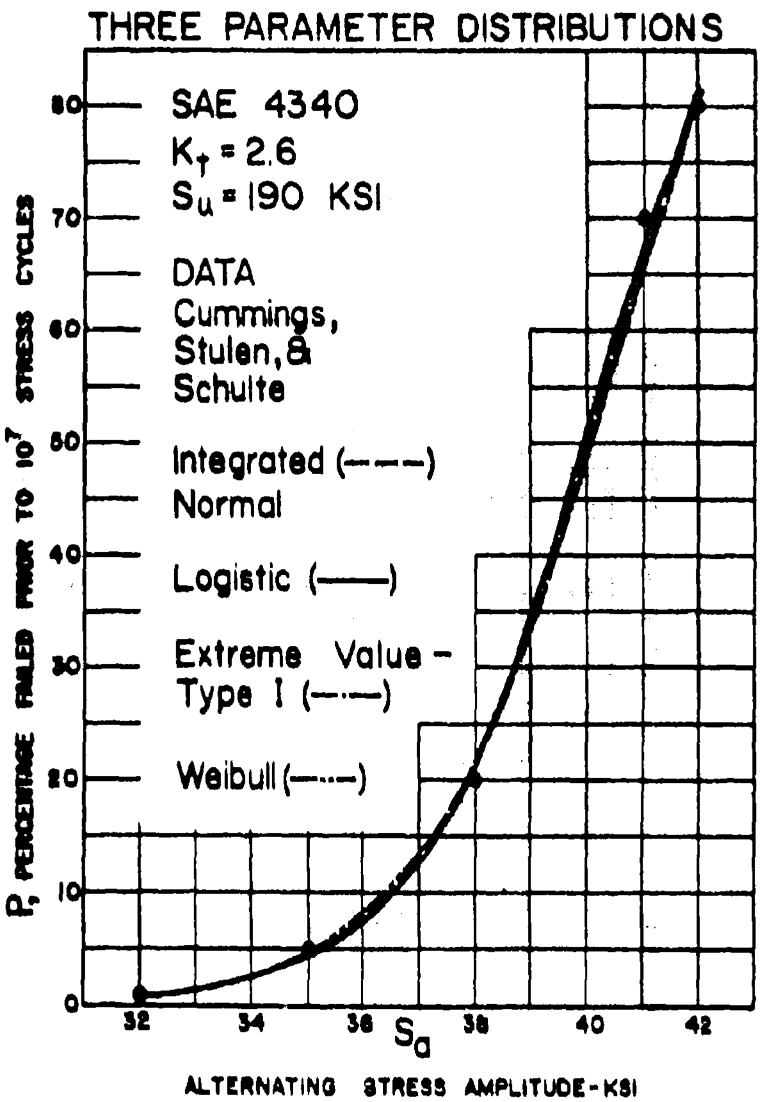


Figure 6 Typical Performance of the Three-Parameter Functions

TABLE 5(a). COMPARISON OF STRESS AMPLITUDES PREDICTED BY VARIOUS DISTRIBUTIONS FOR 10, 50, AND 90 % FAILURES.

The Integrated Normal, Logistic, and Extreme Value--Type I responses pertain to the respective two-parameter functions. The response predicted by the Weibull function is typical of the responses predicted by the other three-parameter functions. (The Weibull γ parameter has no physical meaning.)

Material (S_u)	10 % Response				50 % Response				90 % Response			
	Integrated Normal	Logistic	Extreme Value--Type I	Weibull	Integrated Normal	Logistic	Extreme Value--Type I	Weibull (γ)	Integrated Normal	Logistic	Extreme Value--Type I	Weibull
SAE 4330 (130)	62.61	62.85	62.96	62.72	69.11	69.09	69.76	69.20 (-150)	75.61	75.34	74.10	73.63
SAE 4340 (239)	82.05	82.14	81.73	81.99	89.51	89.53	90.95	89.70 (72.2)	96.98	96.93	96.82	97.02
SAE 4350 (300)	91.87	92.49	92.36	92.07	99.77	99.93	100.49	100.26 (-300)	107.68	107.37	105.67	105.54
SAE 4350* (300)	54.15	54.53	53.86	53.24	60.78	61.03	60.99	61.04 (-300)	67.42	67.53	65.53	65.04
SAE 4340* (190)	35.95	36.34	36.55	36.36	39.80	39.85	40.15	40.09 (0.0)	43.64	43.36	42.45	42.67

* Notched K_t = 2.6

TABLE 5(b). COMPARISON OF THE (Extrapolated) STRESS AMPLITUDES PREDICTED BY THE VARIOUS DISTRIBUTIONS FOR 0.1 % FAILURE.

Test Series	Material	Two-Parameter Distributions			Three-Parameter Distributions
		Integrated Normal	Logistic	Extreme Value--Type I	Weibull*
1	SAE 4330	53.44	49.47	46.13	46.90
2	SAE 4340	71.52	66.28	58.93	74.54
3	SAE 4350	80.73	76.53	72.25	72.65
4	SAE 4350	44.78	40.61	36.24	37.08
5	SAE 4340	30.53	28.81	27.63	28.56

* The other three-parameter distributions are not listed here because: (a) the third parameter is not needed, or when it is needed, (b) the 0.1 % values correspond to "imaginary roots" of the quadratic expression.

The minimum number of fatigue specimens required for testing at a given alternating stress amplitude may be deduced by considering the possible variation in the observed quantal response. For simplicity, assume that the specimen response is described by a binomial distribution that has parameters P and $\sigma_p^2 = \frac{PQ}{N}$ and a coefficient of variation of $C.\ V. = \sqrt{\frac{Q}{NP}}$. Reliable estimates of P require a small C. V.--on the order of 0.2. Thus, approximately 225 specimens should be tested to estimate P = 0.1. No such experimental results are available. Moreover it is likely that none will be forthcoming in the immediate future because this test alone could cost up to $10,000. (A single fatigue machine running at 10,000 RPM night and day would take eight years to complete such a test if the desired fatigue life is 5×10^8 cycles).

Clearly, statistical efficiency must be sacrificed in fatigue-limit tests. A coefficient of variation on the order of 0.5 is probably the best that can be expected. Even then, approximately 400 specimens are required to estimate P = 0.01. Thus, it appears that the coefficient of variation approach to deducing the number of fatigue specimens required in testing will satisfy neither the statistician nor the materials analyst.

It is possible to mitigate this problem somewhat by estimating the number of specimens required by a different approach, viz., selecting N such that the parameters have a negligible bias. The logistic function is selected here to illustrate this approach. (This selection is made on the basis of ease of computation. . . . there is relatively little difference in the descriptive abilities of any of the functions considered here within the probability ranges of existing fatigue-limit data.)

The linear transform of the logistic function is given by

(1) $$\ell = \alpha + \beta s + \epsilon$$

where ϵ is the (random) error associated with ℓ. This transform is used to fit the logistic function to the data. However, to accomodate subsequent graphical solution of β, this transform is temporarily redefined as [2, 3]

(2) $$\ell' = \ln\left[\frac{P + \frac{1}{2N}}{Q + \frac{1}{2N}}\right] = \alpha + \beta s + \epsilon.$$

The error and variance of ℓ' are given by

(3) $$\pm E(\ell' - \alpha - \beta s) = e^{-Npq}\left\{Npq\ln 3 + \frac{1}{2!}(Npq)^2\ln 5 + \frac{1}{3!}(Npq)^3\ln 7 + \dots\right\} - \ln(2Npq)$$

(4) $$V(\ell') = e^{-Npq}\left\{Npq(\ln 3)^2 + \frac{1}{2!}(Npq)^2(\ln 5)^2 + \dots\right\} - e^{-2Npq}\left\{Npq\ln 3 + \frac{1}{2!}(Npq)^2\ln 5 + \dots\right\}^2$$

and the asymtotic mean and variance are:

(5) $$E(\ell') = \alpha + \beta s$$

(6) $$V(\ell') = \frac{1}{Npq}.$$

Thus, it is clear that bias is a function of Npq. This relationship is shown in Figure 7, where it can be seen that a value of Npq of two or larger affords unbiased estimates of the populations parameters. Accordingly, the minimum number of specimens required at a given alternating stress amplitude can be read from Figure 8.

The spacing of the different alternating stress amplitudes should be sufficiently wide to attain an efficient estimate of β. Considering the logistic function:

(7) $$\hat{\beta} = \frac{\ell_1' - \ell_2'}{s_1 - s_2}$$

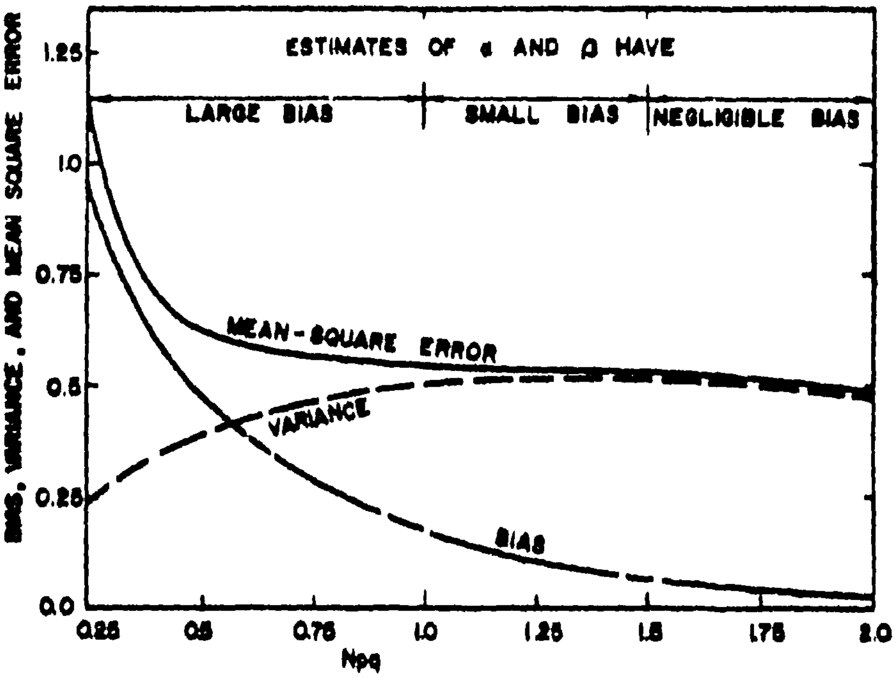


Figure 7 Bias, Variance, and Mean Square Error for $\hat{\alpha}'$

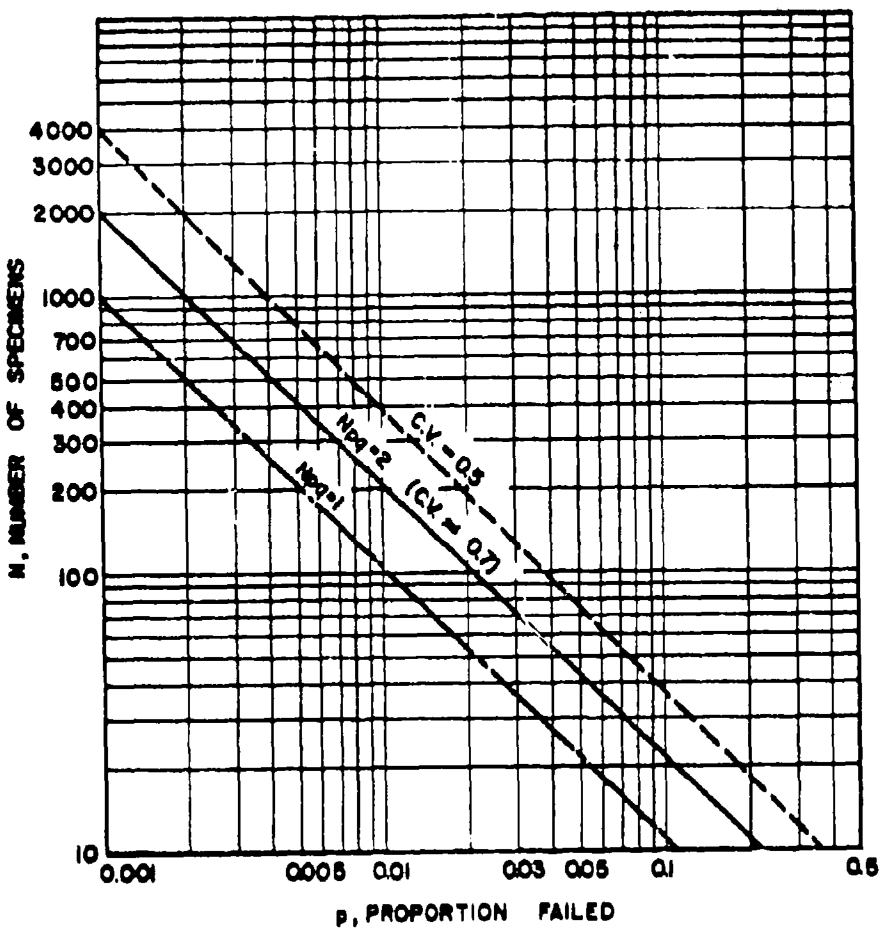


Figure 8 Minimum Number of Specimens Required to Estimate the Population Parameters Efficiently. (C.V. is coefficient of Variation)

where $\ell_1' > \ell_2'$; $s_1 > s_2$; and $(s_1 - s_2) = d$, the spacing.

Equation 7, restated in terms of d, becomes

$$(8) \qquad d = \frac{1}{\hat{\beta}} \left[\ell_1' - \ell_2' \right].$$

Now, selecting p_2 such that

$$(9) \qquad p_1 - p_2 > t \sqrt{\left(\frac{pq}{N}\right)_1 + \left(\frac{pq}{N}\right)_2}$$

this inequality can be rewritten as

$$(10) \qquad p_2 < p_1 - t \sqrt{\left(\frac{pq}{N}\right)_1 + \left(\frac{pq}{N}\right)_2} \; .$$

Finally, substitution of Equation (10) into Equation (8) gives the desired spacing

$$(11) \qquad d_{min} = \frac{1}{\beta} \left[\ln \left\{ \frac{p_1 + \frac{1}{2N_1}}{q_1 + \frac{1}{2N_1}} \right\} - \ln \left\{ \frac{p_1 + \frac{1}{2N_2} - t\sqrt{2\frac{N_1^2 + N_2^2}{N_1^2 N_2^2}}}{q_1 + \frac{1}{2N_2} - t\sqrt{2\frac{N_1^2 + N_2^2}{N_1^2 N_2^2}}} \right\} \right]$$

when $Npq = 2$. This spacing is shown in Figure 9. Note that the spacing can be qualitatively deduced from Equation 9 which indicates that $p_1 - p_2$ can approach zero as N becomes very large.

HYPOTHETICAL FATIGUE TEST. Suppose that the materials analyst has only 100 AISI-1020 annealed steel specimens (Ultimate Strength = 70 ksi), but wishes to obtain the most information concerning the strength distribution. Figure 10 suggests a trial value of the alternating stress amplitude

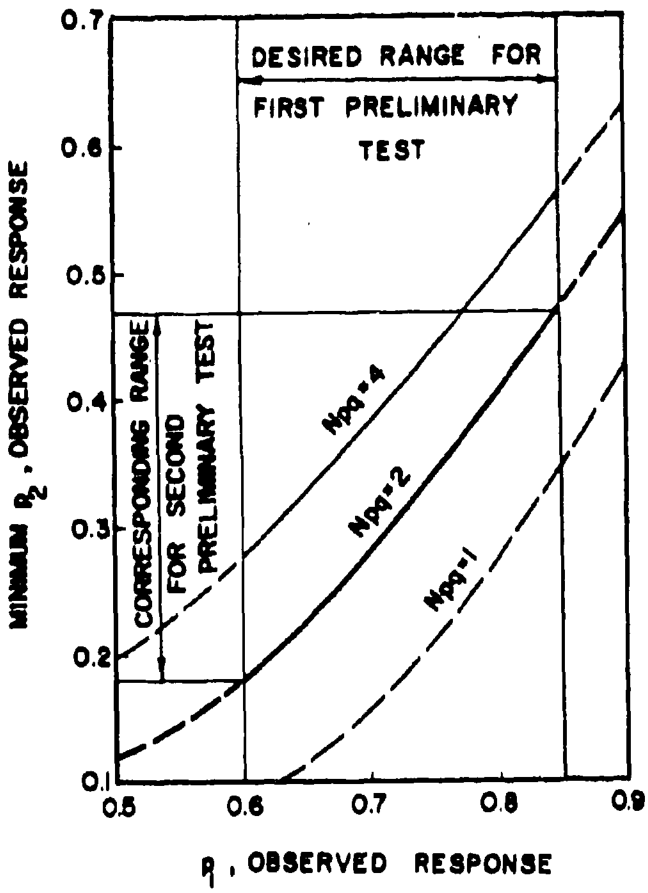


Figure 9a Relationship Between p_1 and p_2 for Efficient Estimation of β.

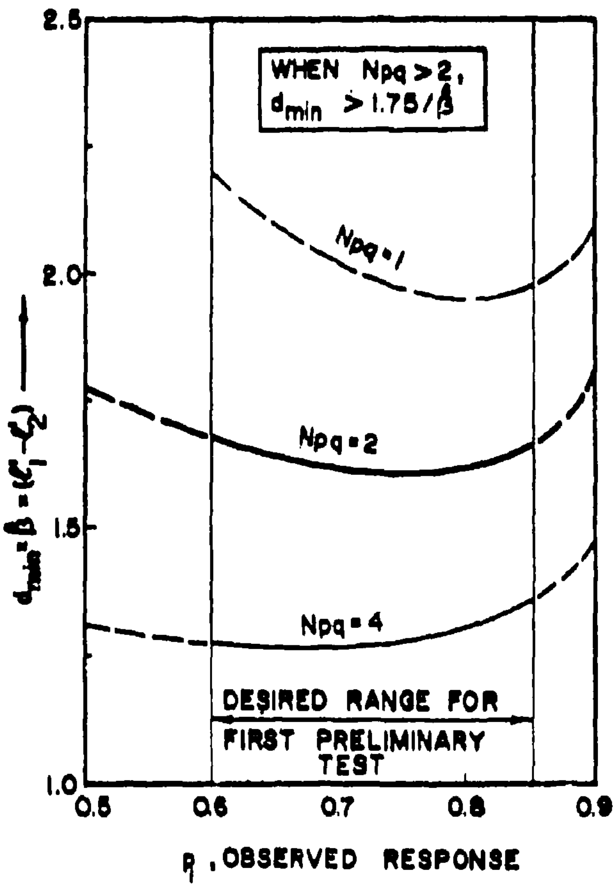


Figure 9b Relationship Between ℓ_1' and ℓ_2' for Efficient Estimation of β. The values of ℓ_1' and ℓ_2' correspond to p_1 and p_2, respectively. Observe that $d_{min} > 1.75/\hat{\beta}$.

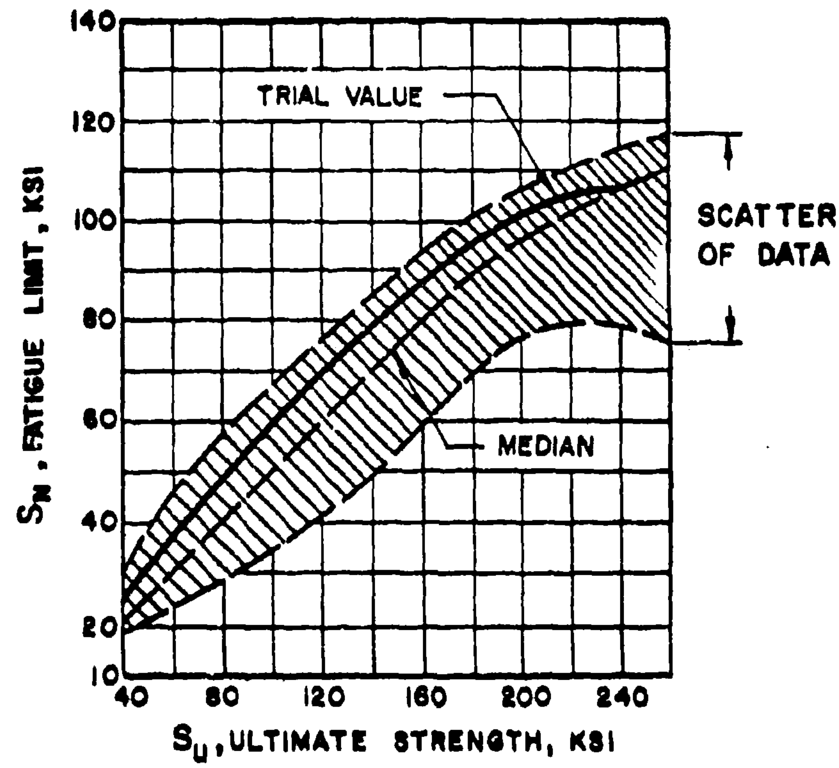


Figure 10 Trial Value for First Preliminary Test Diagram taken from S_N-S_U relationship developed for steel by Bullens [5].

that corresponds to a P of roughly 0.50 to 0.75. Using this trial value, 10 specimens are tested at S_a = 42 ksi and it is observed that 7 specimens fail prior to 10^7 cycles. Setting d = 5 ksi (based on Figure 10), the second test is conducted at S_2 = 37 ksi. In this second test, only 4 of the 10 specimens tested fail. The required spacing in subsequent tests can now be determined by estimating β (using Equation 7). In this hypothetical test

$$\hat{\beta} = \frac{\ell_1' - \ell_2'}{S_1 - S_2} = 0.225 \quad .$$

Thus, d is taken as 7 or 8 and the number of specimens is established as indicated in the following table:

Test	Alternating Stress Amplitude (d=7)	p estimated by graphical solution ($\hat{\beta}$= .225)	Corresponding N for Npq=2	Adjusted N ($Npq \approx 1.75$)
Fourth	23	0.04	52	45
Third	30	0.15	16	15
(Second)a	(37)	(0.40)	(10)	(10)
(First)a	(42)	(0.70)	(10)	(10)
Fifth	49	0.90	22	20
			trial total = 110	adjusted total = 100

(a) preliminary tests

Note that Npq is greater than two for the preliminary tests. Actually 10 specimens are not required in either case. The weights can be calculated as these preliminary tests progress and the next test can be started when the respective Npq approaches two. Then, the "saved" specimens can be tested at the most appropriate stress amplitude at the conclusion of the over-all test.

The over-all test data are then listed in tabular form (Table 2) and fitted as outlined in Part I (Tables 1 and 3.)

SUMMARY. Fatigue data will always be somewhat limited because fatigue tests are expensive. Thus, it is necessary to design fatigue tests to be statistically more efficient. This means that care must be given to the preselection of the number of specimens tested and to the spacing of the respective alternating stress amplitudes considered.

Present analyses can only compare the relative performance of different functions with regard to goodness of fit of limited ranges of data.

REFERENCES

1. H. N. Cummings, F. B. Stulen, and W. C. Schulte, "Investigation of Materials Fatigue Problems," Wright Air Development Center, WADC Technical Report 56-611, (March - 1957).

2. F. J. Anscombe, "On Estimating Bionomial Response Relations," Biometrika, 43, 461-4, (1956).

3. S. E. Hitchcock, "A Note on the Estimation of the Parameters of the Logistic Function, Using the Minimum Logit χ^2 Method," Biometrika, 49, 250-252, (1962).

4. D. J. Finney, "Probit Analysis," Cambridge University Press, London, (1952).

5. D. K. Bullens, Steel and Its Heat Treatment, Vol. 1, P. 157, Fifth Edition, J. Wiley & Sons, New York, (1948).

PERTINENT REFERENCES

6. "A Tentative Guide for Fatigue Testing and the Statistical Analysis of Fatigue Data," American Society for Testing Materials, Special Technical Publication No. 91-A. (Supplement to Manual on Fatigue Testing, ASTM, STP, No. 91 (1958).

7. E. J. Gumbel, "Statistics of Extremes," Columbia University Press, New York, (1958).

8. A. M. Freudenthal, and E. J. Gumbel, "On Statistical Interpretation of Fatigue Tests," Proc. Roy. Soc., A, 216, 309, (1953).

9. W. Weibull, "A Statistical Representation of Fatigue Failures in Solids," Trans. Roy. Inst. Tech. Stockholm No. 27 (1949).

10. A. H. Soni and R. E. Litt' , "Statistical Analysis of Fatigue Limits Using the Logistic Function" The University of Michigan, Industry Program of the College of Engineering, IP-634, (Oct. 1963) Also see, "Materials Research and Standards," Vol. 4, No. 9, pp. 471-473, September 1964.

GETTING REGRESSION ANALYSIS IMPLEMENTED*

W. H. Ammann
U. S. Army Aviation Materiel Command
St. Louis, Missouri

INTRODUCTION. The idea for this presentation came as a result of unsuccessful attempts to solve an analytical problem which was complicated by restraints placed on the collection of data for analysis. Figure 1. This situation is not an isolated one but generally occurs when much data are already being gathered and they are not sufficient for the analysis desired. Alteration of the existing data collection system just to satisfy the needs of a supposedly isolated and parochial study effort is generally not feasible. So, it is necessary to consider the existing data limitations as part of the problem to be solved.

In this case, the success of the analytical effort depends on the relationship which is established between the kind and amount of information which is needed to define the problem and the kind and amount of information available for solving the problem as defined. When this problem-defining and solving effort does not provide meaningful results (Figure 2), three questions are appropriate: has the problem been inadequately defined because of ignorance about the nature of the operation being considered?; are the data collected not sufficient in kind and/or quantity to establish the desired relationships?; and are the data being inadequately analyzed because of the ignorance of the analysts? It is generally necessary to assume that data collected for analysis are not erroneous to the extent that they would be the principal cause for the lack of meaningful analytic results because it is seldom feasible to double check the correctness of the data.

ONE EXAMPLE. To illustrate the foregoing remarks, a recent study will now be described. To appreciate the need for this study, it is necessary to point out that AVCOM's supply effort relates to keeping Army aircraft from being too often deadlined due to a lack of parts (commonly referred to as an Equipment Deadlined for Parts or briefly an EDP situation) while incurring no more than the least costs necessary to obtain such results. It was recognized that this effort might be made more

*The views contained herein have not been approved by the Department of the Army, and represent only the views of the author.

effective and/or efficient if it could be analytically demonstrated how the rate, at which aircraft are EDP, varies with various supply actions and ultimately with the costs associated with each action.

A study to obtain the desired analytical results was developed.

a. Concept: It was recognized that the total time that aircraft are EDP is affected by how often an EDP situation occurs and how long it takes to satisfy each EDP situation. Therefore, the study was naturally subdivided into a study of the frequency of occurrence of EDP situations and a study of the time required to satisfy EDP situations. Immediately, obstacles were encountered.

(1) Frequency of Occurrence: How often each aircraft is EDP during the month is not reported. However, when an aircraft is EDP for a part which is supposed to be furnished by AVCOM action and that part cannot be obtained below depot level, an EDP requisition is sent to AVCOM. Therefore, there are at least as many instances of aircraft EDP as there are valid EDP requisitions received at AVOCM and it is operationally known that there are more such occurrences since aircraft are EDP for parts which are supplied without AVCOM action. The term valid EDP requisitions is needed because those for common hardware parts which could not render an aircraft operationally deadlined were excluded as being invalid.

Fortunately, the total time that each aircraft is EDP is reported to AVCOM. So, it was hoped that an estimate of the amount of change which might be achieved in the days aircraft are EDP by a supply action which might reduce the rate at which EDP requisitions occur at AVCOM by a particular amount, might be obtained by regression analyses by aircraft type. The results of these analyses will be indicated later.

(2) Time to Satisfy: Meanwhile, attempts to relate the time to satisfy an aircraft EDP situation encountered similar data constraints. When an aircraft EDP situation is satisfied without an AVCOM action in response to an EDP requisition, the time required to obtain such satisfaction is unknown at AVCOM. So, it was necessary to assume that such instances have a random effect on the total time aircraft are EDP in a month. Then, a meaningful correlation might be discovered between the time aircraft are EDP and the time required to satisfy an EDP requisition at AVCOM.

Also, the complete time required to satisfy an EDP requisition at AVCOM could not be easily obtained. The time that was obtained is the time between the date an EDP requisition is initiated and the date on which materiel release at the appropriate depot is confirmed.

In other words, the time consumed after a material release confirmation is sent to AVCOM and until the part arrives at the site of the particular EDP aircraft was not readily measurable and had to be left out of the study. Again, it was necessary to assume that the effect of this time on the total time aircraft are EDP is random and that a meaningful correlation might exist between the time aircraft are EDP and the principal portion of the time to satisfy an EDP requisition measured in this study.

On the other hand, since an EDP requisition does not identify the specific aircraft which is awaiting the part, it is possible that an EDP aircraft has been made serviceable by using a part obtained from some other source such as off of a crash damaged aircraft and yet the pertinent EDP requisition is not satisfied. It was hopefully assumed that such instances might compensate for some of the excluded shipping time.

b. Sample Selection: By now, hopes to obtain fruitful analytical results were waning and yet the worse was yet to come. Since it was desired to obtain some useful results as soon as possible and the information about aircraft days EDP is available only on a monthly basis, six months data or six data points were chosen for analysis. After the data were gathered, there was reason to believe that data concerning all EDP requisitions received by AVCOM during the first three months observed had not been obtained. Further, it could not be determined whether the sample EDP requisitions could be validly claimed to be a representative sample. Therefore, only the latter three month's data were used for regression analysis. At this point, the problem being described can be summarized as shown. Figure 3.

c. Results Obtained: Approximately nine months elapsed before efforts to obtain the preferred analytical results were exhausted. A total of 14 aircraft types were considered. Needless to say, the results obtained were disheartening even though not unexpected.

(1) Frequency of Occurrence: Table 1 contains estimates of the relationship between the days aircraft are EDP and the quantity of EDP requisitions received at AVCOM.

(2) Time to Satisfy: Table 2 contains estimates of the relationship between the days aircraft are EDP and the major portion of the time to satisfy EDP requisitions at AVCOM.

(3) It is recognized that three data points are not enough to preclude apparently conflicting results from occurring because of sampling variations but there were no more reliable data points which could be used to reduce this likelihood. However, the occurrence of both positive and negative correlation coefficients is disconcerting. In the case of negative ones, it is implied that a reduction in aircraft days EDP can be obtained by increasing the frequency of occurrence of EDP instances or by taking more time to satisfy EDP requisitions. Both of these implications are unreasonable. With the hope that the three questionable data points might be good ones, regression analyses using six points were made but no more reasonable results were obtained.

(4) To preclude some wrong implications, it must be pointed out that this nine month study effort did not consume much more than one analyst's time. The study time had to take six months to obtain six months of data. Additional time was required to allow EDP requisitions received near the end of the sixth month to be satisfied. In addition, several by-product analyses were made with the data collected. In other words, it would be unfair to conclude that this analytical effort was not worthwhile. Also, it seems that it could be concluded that the desired results were not obtained for at least the first two of the reasons listed on Figure 2; namely, inadequate representation of the problem and insufficient data collected both in type and quantity.

d. Question: However, the question still remains: What can be done to increase the effectiveness of the analytical effort being expended in the manner just described?

ANOTHER EXAMPLE. Before attempting to present any subjective answers to the question just stated, another analytical problem area can be used to suggest that there is a related question that also needs answering. This analytical problem is suggested by a review of budgeting and funding practices.

It is not necessary to know the exact budgeting and funding procedures to appreciate the features which are useful here. Figure 4.

a. The preparation of a budget must be in accordance with guidance furnished by higher headquarters. This guidance has usually been different from year to year. This situation implies that a generally sound budgeting procedure has not yet been determined.

b. In addition, forecasted budget requirements are never completely honored. Somewhere up the line, limitations are set below the accumulated forecasted requirements and these limitations are somehow partitioned and passed down to each organization involved.

c. Further, each organization's general objective is to make commitments nearly equal the limitations appropriate at the time of each within year review. In other words, if there is only a mid-year review, commitments should be nearly equal to one half of the annual limitation otherwise it might be concluded that even less funds will suffice and limitations will be decreased accordingly. As a result of these within year reviews, particular fund limitations for the remainder of the year are revised; sometimes upward and sometimes downward.

d. At this point, it is well to hypothesize the logic which supports this budgeting and funding practice. It is initially assumed that no one can forecast an organization's budgetary requirements more accurately than the organization itself. Therefore, forecasted requirements are made by each organization and these are the starting point for the budgeting cycle. Since fund limitations have always been set less than forecasted budget requirements, organizations find it expedient to compensate for such reduction by somehow inflating estimates of requirements. It seems reasonable to assume that the extent of this inflation cannot be accurately determined by the people who set limitations otherwise budgetary guidance could preclude such inflation and forecasted requirements could be honored. Also, since the practice of setting limitations less than forecasted requirements has never been considered responsible for serious operational shortages, the practice has been continued without fear.

It seems that this strategic exercise must persist until it has been definitely learned that the allotment of different quantities of funds leads to the achievement of recognizably different accomplishments. Only then can superiors choose the desired amount of accomplishments and fund accordingly. Thus, the question arises:

How can the regression analysis effort, necessary to establish a sound relationship between money allotted and results achieved thereby, be obtained?

CONCLUDING REMARKS. In review, it seems that the two situations just described indicate a need for a way; of improving the effectiveness and efficiency of the analytical effort trying to do regression analysis in a subordinate command such as AVCOM; and of getting regression analysis implemented in a higher headquarters in the budgeting-funding subject area.

a. In the first case, it is possible to take the viewpoint that certain analytical efforts should be dropped when data collection restraints are too restrictive or that it is worthwhile to expend some effort to remove as many of those restraints as necessary. However, the potential value of certain analytical results is sufficient to preclude their being dropped until it has been indisputably demonstrated that they cannot be obtained in spite of existing constraints. Also, the removal of data collection restraints to satisfy local analytical needs is practically impossible since existing data collection and reporting requirements have been entrenced by tradition and austerity measures in the manpower area preclude the collection and reporting of additional data for local analyses that have not been specifically required by higher headquarters. Therefore, it seems that some outside, authoritative intervention is needed if the situation confronting local, investigative analyses is to be improved.

b. In the second case, since budgeting guidance is furnished by higher headquarters and must be adhered to by subordinate commands, it seems that regression analysis must be attempted and found successful at the top before the official authorization to do such analysis at the bottom can be expected and before the cooperation necessary to have a reasonable chance at being successful with this effort will be forthcoming.

In other words, it seems that it is not enough to hire analysts at all levels in the Department of the Army and then allow organizational tradition to render such analysts ineffective and inefficient. The situation could be significantly improved if (Figure 5) the Office of the Chief of Research and Development (OCRD) would form a Survey Team of renowned analysts who would visit selected Army headquarters to determine the extent and kind of analytical program that seems appropriate for each organizational level and would then prepare a recommended Department of the Army program. Then, OCRD could coordinate this program as appropriate and direct that the coordinated program be done. This type of positive action seems a bit extreme and probably impossible to obtain and so a solicitation for a less extreme improvement action and one more within the authority of a subordinate organization is hereby extended.

Table 1

A/C Type	A/C Days EDP vs Qty of EDP Rqns	Correlation Coefficients
OH-13	$y = 1921 + 0.777x$	0.927
UH-19	$y = 397 + 2.610x$	0.903
CH-21	$y = -419 + 21.177x$	0.933
OH-23	$y = 2701 - 0.890x$	-0.120
CH-34	$y = 399 + 5.472x$	0.621
CH-37	$y = 588 - 5.447x$	-0.683
UH-1	$y = 1572 + 1.679x$	0.415
O-1	$y = 519 + 20.107x$	0.783
U-6	$y = 908 - 1.543x$	-0.347
U-8	$y = 533 - 2.403x$	-0.713
U-1	$y = 391 - 1.582x$	-0.281
OV-1	$y = 468 + 2.610x$	0.976
CV-2	$y = 362 - 0.522x$	-0.649
CH-47	$y = 885 - 5.567x$	-0.997

y = is in terms of aircraft days EDP

x = is in terms of quantity of EDP requisitions received at AVCOM

Table 2

A/C Type	A/C Days EDP vs Qty of EDP Rqns	Correlation Coefficients
OH-13	y = 1923 + 0.126x	0.998
UH-19	y = 567 + 0.103x	0.394
CH-21	y = 509 - 0.046x	-0.088
OH-23	y = 3194 - 0.457x	-0.662
CH-34	y = 482 + 0.382x	0.777
CH-37	y = 577 - 0.220x	-0.883
UH-1	y = 6913 - 1.551x	-0.388
O-1	y = 3151 - 6.119x	-0.498
U-6	y = 950 - .184x	-0.465
U-8	y = 56 + 0.530x	0.999
U-1	y = 302 + .116x	0.681
OV-1	y = 431 + 0.132x	0.619
CV-2	y = 359 - 0.023x	-0.728
CH-47	y = 690 - 0.092x	-0.566

y = is in terms of aircraft days EDP

x = is in terms of the principal quantity of days required to satisfy EDP requisitions received at AVCOM

Problem To Be Solved

Analytical
Problem

Plus

Data Collection
Restraints

FIGURE 1

No Useful Solution Obtained

Inadequate Representation of Problem?

Insufficient Data Collected?

Inadequate Analysis?

FIGURE 2

Sample Problem

Effect of Supply Actions on
Acft EDP Rate

Plus

Inexact EDP Frequency

Unknown Extent of EDP Change
Due to Non-AVCOM Action

Unknown Shipping Times

Only Three Reliable Data Points

FIGURE 3

Budgeting & Funding

Guidance Changes Annually

Limitations Less Than
Forecasted Requirements

Commit Full Limitations

Within Year Reviews

Revised Limitations

FIGURE 4

For Consideration

PROFESSIONAL TEAM:

Conduct Survey
&
Describe Analytical Program

CHIEF, RESEARCH & DEVELOPMENT:

Require Program Be Done

FIGURE 5

ASSESSMENT AND CORRECTION OF DEFICIENCIES IN PERT

H. O. Hartley and A. W. Wortham
Institute of Statistics
Texas A&M University
College Station, Texas

1. INTRODUCTION. As is well known, the technique known under the name of PERT (Program Evaluation Review Technique) is concerned with a 'project' comprising a large number of successive 'activities' which are arranged in a complex 'network' (see e.g. Figure 2). Each activity 'commences' at a particular 'point' of the network but not until all activities 'terminating' at that point are completed. Specifically, PERT is concerned with computing the expected time required to complete all activities of the project; -Assuming that the time taken to complete a particular activity follows a specified distribution of completion times, the total time needed to complete the project the so called 'critical time' is a statistical variable and is given by the total of completion times along the 'critical path', i.e. along that sequence of activities in the network which for a given sample of completion times takes longest to reach every point along its path. The expected value of this critical time is the expected time to complete the project.

Now it is well known that PERT does not compute the correct critical time as defined above but instead uses for each activity the average completion time and then determines a unique and fixed critical path as the sequence of activities for which the sum of the expected completion times is at a maximum. Critical path determination by this method may be badly misleading and may result in a serious underestimate of the expected time to complete the project. Moreover, it may also lead to erroneous information on the identification of 'critical activities', i.e., activities which are crucially responsible for the delay in completion of the project.

Whilst this shortcoming of PERT has been known from its initiation and the above method is deliberately used as an approximate short-cut, we do not think that the magnitude of the bias in this short-cut method is fully appreciated. Indeed it can be shown (see e.g. section 8) that under certain circumstances PERT may underestimate the correct expected completion time by 50% or more. Moreover, for a general network, PERT provides the correct answer only under the (completely unrealistic) assumption that there is essentially no variability in the completion times for each activity.

One of the objectives of this paper is therefore to eliminate this bias from PERT; in fact, we shall provide a method of computing the probability distribution of criticaltimes and thereby supply not only the correct value of its expectation but likewise of its variance and percentage points.

It may rightly be argued that our exact method of critical path analysis is based on the assumed distribution of completion times for each activity, and that there is usually a notorious lack of information on such timings. This point is well taken. However, we feel that the deplorable lack of input data should not excuse us from using a method accurately utilizing at least all the available information. Moreover and more positively our method enables the analyst who is uncertain about the completion times of (say) a particular activity in the network to evaluate the effect of altering his assumptions about that activity on the critical time and path. We consider the provision of such a 'sensitivity analysis of PERT' as an important contribution to planning a project 'under uncertainty'.

Mathematically our method utilizes the following devices: -

(a) A classification of networks into different types depending on their degree of involvement and complexity.

(b) An operational calculus by which the distribution of critical times will be derived by numerical analysis, notably numerical integration. This method will provide the solution to our problem for the basic types of networks.

(c) A Monte Carlo procedure providing an approximate solution for the more involved networks.

(d) Analytic solutions for particularly simple networks and particularly simple distributions of completion times. These are mainly used for illustration purposes.

2. GENERAL DEFINITIONS AND 'UNCROSSED NETWORKS'. In order to provide a mathematically rigorous theory of PERT analysis for networks, it is necessary to introduce certain definitions and concepts. We therefore give the following definitions and explanations: -

2.1. An activity is represented by one or two line segments in the network (see Figure 1). It 'commences' at one of its ringed end points and 'terminates' at the other ringed end point, the 'direction of the time flow being indicated by the arrow. The numbering of the activities is explained in 2. 3.

2.2. A Network Point: - These are represented by ringed points in Figure 1. A network point represents any stage in the network occurring at the beginning and/or end of an (or several) activity(ies) (e. g., event 5 in Figure 1 is a network point since activity (7; 2, 5) terminates and activities (10; 5, 8) and (11; 5, 8) commence at that stage of the network.

2.3. Codes: - 'Network points' carry a serial number (ringed in Figure 1) identifying them. The order of the numbering is immaterial at this stage. An activity also carries a 'serial number' (preceding the ;) but also the number of the network point at which it commences followed by the network point number at which it terminates. Thus (7; 2, 5) denotes activity No. 7 commencing at point No. 2 and terminating at point No. 5.

2.4. Two consecutive activities are defined as activities numbered (t; i, j) and (s; j, k) i. e., the first terminates at point j whilst the second commences at point j.

2.5. A Path from i to j is a 'sequence of consecutive activities' starting at point i and finishing at point j (e. g., activities (2; 0, 2), (7; 2, 5) and (10; 5, 8) starting at point (0) and terminating at point (8)).

2.6. A complete path - A path starting at the beginning and finishing at the end of the project (e. g., the path formed by (1; 0, 1), (5; 1, 4), (9; 4, 7) and (15; 7, 10)).

2.7. A Universal Point - A network point through which all complete paths pass (the only universal points in Figure 1 are at 0 and 10).

2.8. Consecutive Points - Point j is consecutive to Point i if both j and i are universal and if all paths starting from i pass through j before passing through any other universal point (if any).

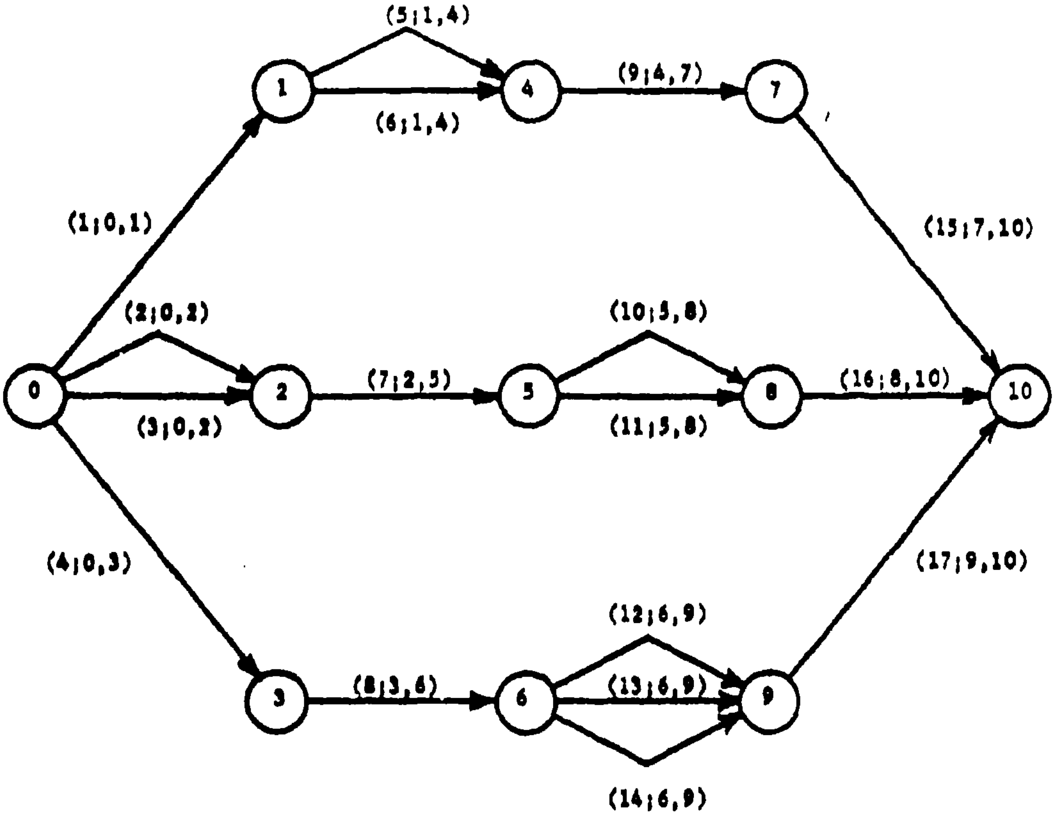


FIGURE 1
NETWORK NOTATION

2.9. Sets of first order branches - Consider the set of all paths commencing at a universal point i and terminating at a universal point j consecutive to i. Subdivide the set of these paths into exhaustive subsets such that any two paths in different subsets have only points i and j in common but any two paths in the same subset have at least one more point in common. (This is always possible since we may place, if necessary, all paths in the same subset.) These mutually exclusive subsets are called '1st order branches.' (e.g., in Figure 1 the paths formed by connecting points 0,1,4,7,10 form the first 1st order branch, the paths formed by connecting points 0,2,5,8,10 the second 1st order branch and the paths formed by 0,3,6,9,10 the third 1st order branch.) If there are only two consecutive points in the network (i.e., the start and the end) and there is only <u>one</u> set of paths as described above, we shall term it a zero order branch. For example, Figure 3 would constitute a single zero order branch, so would a single activity network.

2.10. Sets of 2nd order branches - Consider a particular 1st order branch starting at a universal point i and ending at a universal point j consecutive to i and regard it as a separate network. Apply definitions 2.1 to 2.9 to this network, then any 1st order branches of this first order branch are called second order branches, but any zero order branch of a first order branch still be called a 1st order branch. (e.g., in Figure 1 activities (2; 0, 2), and (3; 0, 2) connecting points (0) and (2) are two second order branches belonging to the second first order branch. Likewise (7; 2, 5) is a second order branch belonging to this first order branch.

2.11. The uncrossed network - If by the repeated application of definitions 2.1 to 2.10 all individual activities in the network can be identified as different k-th order branches (for some $k \geq 0$), the network is said to be "uncrossed." (e.g., the network in Figure 1 is uncrossed and all activities are recognized as different 2nd order branches. The network in Figure 2 is likewise uncrossed with some of the individual activities being 2nd order branches and some 3rd order branches. However, the network in Figure 3 is crossed - there being only one (0 order) branch comprising all activities.

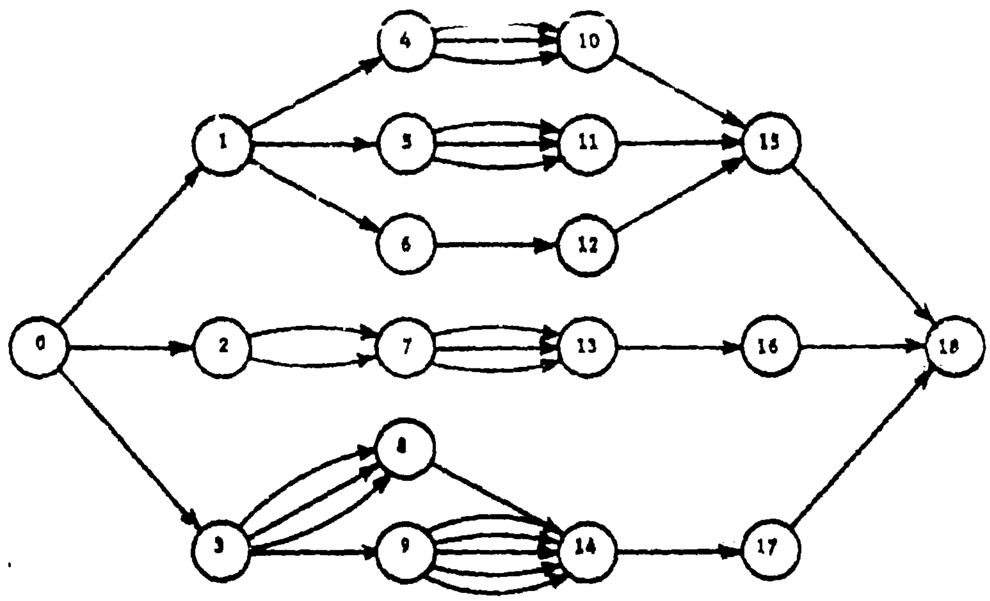


FIGURE 2
AN EXAMPLE OF AN "UNCROSSED" NETWORK

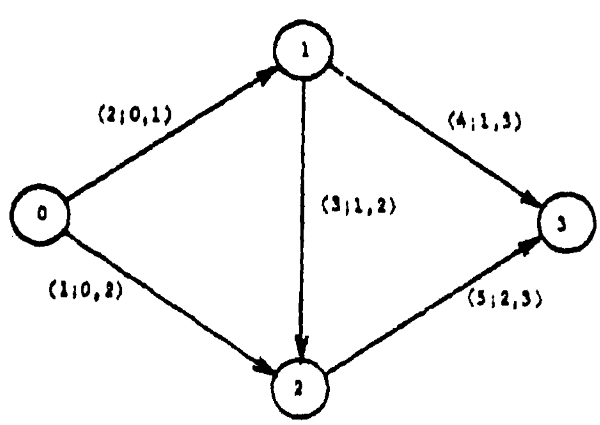


FIGURE 3
AN EXAMPLE OF A "CROSSED" NETWORK
"WHEATSTONE BRIDGE"

3. CROSSED AND MULTIPLE-CROSSED NETWORKS. The arrangement shown in Figure 3, called the 'Wheatstone bridge', has been quoted in the previous section as an example of a crossed network. It consists of the five activities (1; 0, 2), (2; 0, 1), (3; 1, 2), (4; 1, 3) and (5; 2, 3). If now each of these five single activities is replaced by an uncrossed network, as defined in Section 2, we shall reach a network called a '1st order crossed network.' More specifically we define a 0-order crossed network as an uncrossed network in which at least one of the 'activities' is replaced by a Wheatstone bridge (see Figure 3). With the help of this network we define a t^{th}-order crossed network (for $t \geq 1$) as a 0-order crossed network in which any 'activity' may be replaced by a k^{th}-order crossed network with $0 \leq k \leq t-1$, but at least one activity is replaced by a $(t-1)^{st}$-order crossed network.

Although most practical situations of activity networks will be recognized as a t^{th} order crossed network for some order t. There are clearly quite small networks which do not belong to this category, as for example the network shown in Figure 4:

4. OPERATORS FOR EXACT SOLUTION BY NUMERICAL ANALYSIS. Consider first the case of an uncrossed network as defined in section 2. It is easy to show (see e. g. Section 5) that an uncrossed network can be built up from individual activities by two basic operations which can be briefly described as follows: -

Operation π: - Placing activities in parallel

Operation S: - Placing activities in series

These basic operations, well known concepts in electric circuit theory, are illustrated in Figures 5 and 6.

Corresponding to these two basic networks we now develop the simple equation for the c. d. f. (cumulative distribution function) of the 'critical time' in the two basic networks.

a. Parallel activities: -

Denote the serial number of the k activities in parallel by s so that s = 1, 2, . . . , k (k = 5 in Figure 5) and denote the time required to

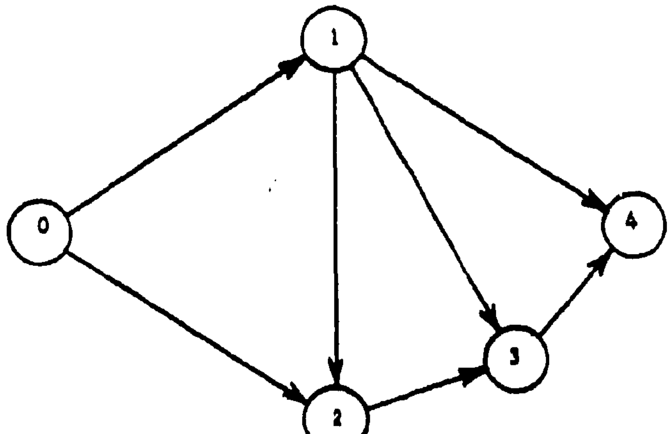


FIGURE 4
EXAMPLE OF A NETWORK NOT IDENTIFIABLE
AS t^{th} ORDER CROSSED NETWORK

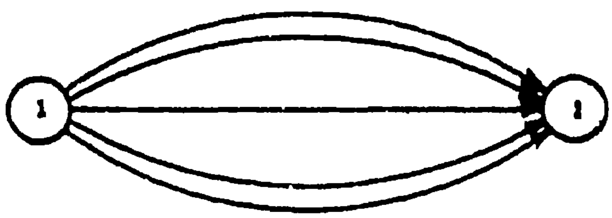


FIGURE 5
ACTIVITIES IN PARALLEL

FIGURE 6
TWO ACTIVITIES IN SERIES

complete the s-th activity by t_s. If the c.d.f. of t_s is denoted by $F_s(t_s)$ then the critical time t for this simple network is clearly given by $t = \max_s t_s$ so that the c.d.f. of t is obtained as

$$(1) \qquad F(t) = \Pr\left\{\max t_s \leq t\right\} = \prod_{s=1}^{k} F_s(t)$$

b. Two activities in series.

Denote the times required to complete the two activities by t_1 and t_2 respectively and their c.d.f.'s by $F_1(t_1)$ and $F_2(t_2)$. Then the critical time for this simple network is clearly given by $t = t_1 + t_2$ so that the c.d.f. of t is obtained as

$$(2) \qquad F(t) = \int_0^F F_1(t-t_2)dF_2 \text{ , where } F = F_2(t) \text{ and } t_2 = F_2^{-1}(F_2).$$

It should be noted that equations (1) and (2) yield the c.d.f. F(t) for the basic network from the c.d.f.'s of the individual activities. Therefore, these basic networks can subsequently be regarded as 'individual activities' and entered as $F_s(t_s)$ in subsequent operations of the type (1) and (2). It is obvious therefore that by repeated application of (1) and (2) the c.d.f. of an uncrossed network such as in Figure 1 and Figure 2 can be obtained. The operational logic for this is given in section 5.

Next we deal with 1st order crossed networks and to this end must evaluate the c.d.f. of the critical time t for the Wheatstone bridge (figure 3). Denoting by $t_1, \ldots, t_5$ the completion times for the five activities $s=1,2,\ldots,5$ as arranged in Figure 3 and by $F_i(t_i)$ their respective c.d.f.'s we obtain by elementary probability calculus the c.d.f. of the critical time t as a sum of three integrals as shown in (3) below: -

$$F(t) = \int^{a} dF_2 \int^{b} dF_3 \int^{c} dF_5 \; F_1(t_3+t_2)\, F_4(t_3+t_5)$$

$$(3) \qquad + \int^{a} dF_2 \int^{d} dF_4 \int^{f} dF_5 \; F_1(t_2+t_4-t_5)\, F_3(t_4-t_5)$$

$$+ \int^{g} dF_1 \int^{h} dF_5 \int^{j} dF_2 \; F_3(t_1-t_2)\, F_4(t_1+t_5-t_2)$$

where $t_i = F_i^{-1}(F_i)$ are the inverse functions of $F_i(t_i)$, all variables of integration are the F_i with integrations starting at $F_i = 0$ and ending at points 'a' to 'j' given by

$$a = F_2(t),\; b = F_3(t-t_2),\; c = F_5(t-t_2-t_3)$$

$$(4) \qquad d = F_4(t-t_2),\; f = F_5(t_4),\; g = F_1(t)$$

$$h = F_5(t-t_1),\; j = F_2(t_1).$$

It should be noted that the three terms in (3) correspond to the three mutually exclusive and exhaustive situations (a), (b), (c) shown below

(a) Critical path $t = t_2 + t_3 + t_5$

(b) Critical path $t = t_2 + t_4$

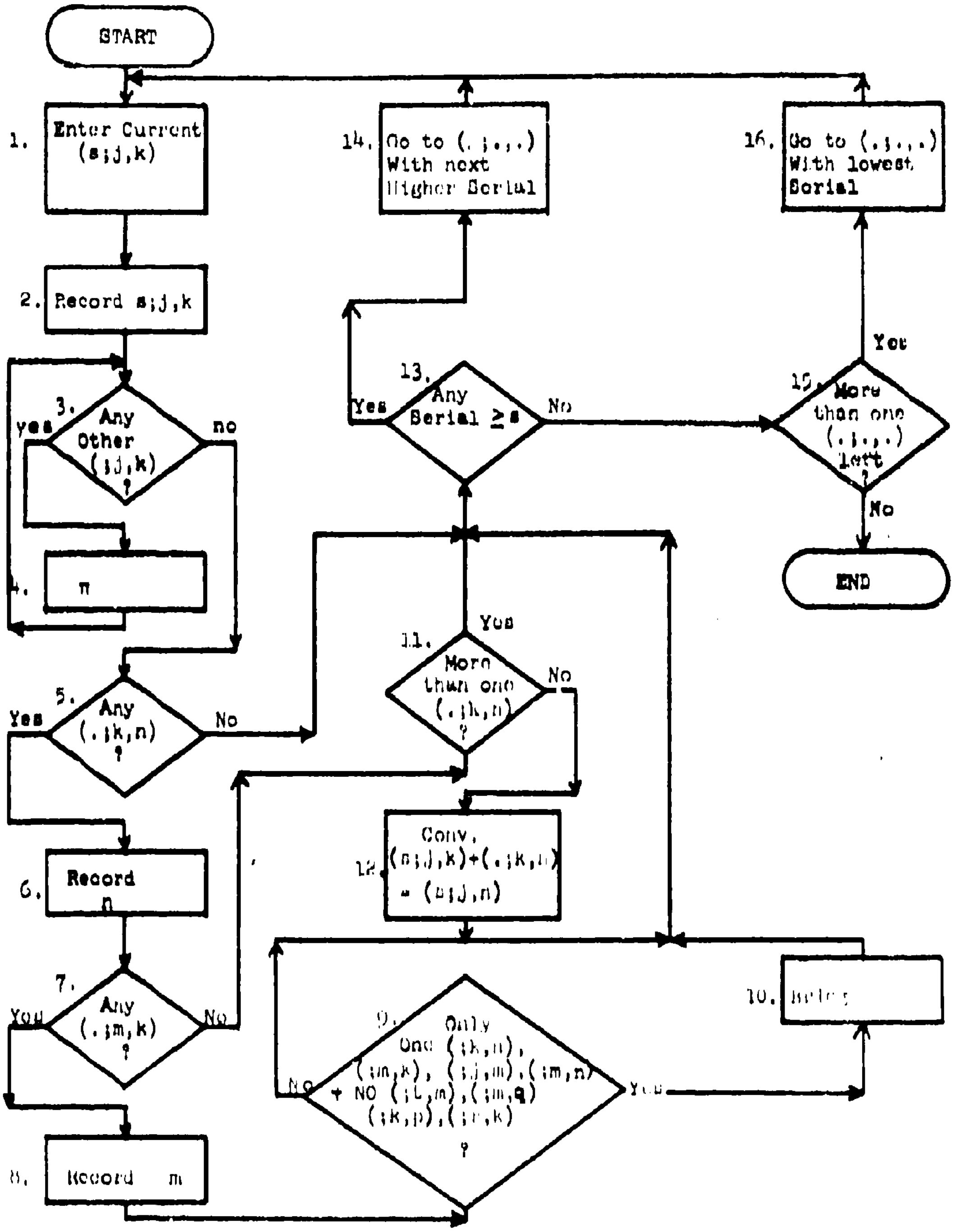
(c) Critical path $t = t_1 + t_5$.

The general case of a t-th order crossed network is finally covered by repeated application of the above operators as shown in section 5.

5. <u>THE COMPUTATIONAL LOGIC FOR t-th ORDER CROSSED NETWORKS.</u> The computer logic shown in Figure 7 will compute the c.d.f. of the critical time in a t-th order crossed network from c.d.f.'s of the completion times of the individual activities.

FIGURE 7

FLOW DIAGRAM FOR t^{th} ORDER CROSSED NETWORKS



The initialization of the computation consists of loading the code numbers of all activities $(s; j, k)$ (see section 2.3) as well as readying the tape giving all their c.d.f. functions. If the serial number of the activity is immaterial we shall use the symbol $(.; j, k)$. In the course of the operations certain code numbers will be deleted and the retained code number activities have their c.d.f.'s modified. We should give the following explanations of some of the operations involved: -

Box 1; 2 An activity $(s; j, k)$ with the current serial number s and starting at j and ending at k (see 2.3) is processed i.e. s; j k are recorded and the associated c.d.f. $F_1(t)$ loaded.

Box 3 A test is made as to whether there is a 2nd activity starting at j and ending at k

Box 4 If the 2nd activity starting at j and ending at k has a code $(u; j, k)$ and a c.d.f. of $F_2(t)$, replace $F_1(t)$ by $F_1(t)\ F_2(t)$ and delete $(u; j, k)$ from the list of code numbers and $F_2(t)$ from the tape of c.d.f. functions.

Box 12 If the c.d.f. functions of activities $(s; j, k)$ and $(\cdot; k, n)$ are denoted by $F_1(t)$ and $F_2(t)$ respectively we replace $F_1(t)$ by $\int_0^F F_1(t-t_2)\, dF_2$ with $F = F_2(t)$, and $t_2 = F_2^{-1}(F_2)$ replace the code $(s; j, k)$ by $(s; j, n)$ and delete code $(\cdot; k, n)$ and $F_2(t)$.

Box 9 A test is made as to whether the current activity $(s; j, k)$ and associated activities $(\cdot; j, m)$, $(\cdot; m, k)$, $(\cdot; m, n)$ and $(\cdot; k, n)$ can be identified with the activities $(1; 0, 2)$, $(2; 0, 1)$, $(3; 1, 2)$, $(4; 1, 3)$ and $(5; 2, 3)$ of the Wheatstone bridge of Figure 3.

Box 10 The five c.d.f. functions involved on the Wheatstone bridge operation are combined in accordance with equation (3). The resulting $F(t)$ replaces $F_1(t)$, the code $(s; j, n)$ replaces $(s; j, k)$ and all other codes and c.d.f. are deleted.

The proof that the logic of the flow diagram in Figure 7 does indeed result in the computation of the c.d.f. of the critical time for any multiple-crossed network is given in the Appendix.

6. MONTE CARLO SOLUTIONS FOR THE MORE COMPLEX NETWORKS. As is well known and as was mentioned in section 1 the currently used PERT algorithm determines that path in the network for which the total of average completion times is a maximum. Now imagine that we apply the same algorithm to a random sample of completion times, each drawn from the distribution relevant to its activity. The 'critical time' so computed will be a single random variable from the distribution of critical times defined in section 1 and discussed in section 5. A large number of repetitions of this computation will therefore yield a Monte Carlo solution of the distribution of critical times. Such a solution will therefore be available for any network (and not just for multiple crossed networks).

Suppose now we are faced with a complex network (not necessarily multiple crossed). If we apply the algorithm of section 5 to such a network we would in general reduce the number activity - codes by the operations 'π', 'Conv' and 'Bridge'. However, if the network is not multiple crossed we shall not be able to reduce the network to a single activity. As soon as we find therefore that no reduction of codes has occurred on too consecutive cycles we would output the reduced network activities and associated c.d.f.'s so that it can be solved by Monte Carlo as indicated above. The operational calculus of section 5 will considerably reduce the complexity and extent of the network so that the subsequent Monte Carlo calculations are much simplified.

An IBM 709 computer program performing the above Monte Carlo computations of the distribution of critical times was prepared by L. L. McGowan (1964), in his M. Sc. thesis at the Institute of Statistics at Texas A&M University.

7. SENSITIVITY ANALYSIS AND GUIDE TO MANAGEMENT. The previous sections have been concerned primarily with the establishment of the mathematical, statistical, and logical aspects of determining the distribution of completion times for a project. The methods developed have further applications in analysing the effects of making specified changes in the original network and thereby providing guides for management actions. Basically, the analyses most readily recognized in this area are concerned with (1) assessing the impact of modifying the distribution of specified activities (e.g., a change in their average completion times); (2) assessing the impact of modifying blocks of

activities; (3) comparing two or more networks to establish the organization of the project for minimum time, minimum cost, or some other optimum; and (4) assessing progress or remaining time for the completion of the project.

All of the above assessments are permissible under the method developed in this paper. In fact once the logic is established on a computer, all four assessments are possible with the same computer programs. It is only necessary to vary the input and certain problem parameters according to the assessment required.

It should be pointed out that the assessments gained via this logic will be more comprehensive than a similar PERT assessment. With the present logic the impact on the c.d.f. of project completion times will be observable. This means that our sensitivity analysis provides estimates of the impact of production schedule changes on the expected completion time but also of the impact on its variance, percentiles, confidence intervals and other statistical parameters.

8. SPECIAL CASES OF BIAS DEMONSTRATION. As noted earlier bias enters the solution of a network problem due to inadequate treatment of the statistical considerations and approximate logic. In order to demonstrate this bias a few examples will be worthwhile for illustrative purposes. The following examples will also demonstrate the dependence of the solution on the distribution form and network composition.

EXAMPLE 1. Consider the case where k activities are in parallel as is illustrated in Figure 5. Assume further that each t_i is a random variable with exponential c.d.f.

$$F_i(t_i) = 1 - e^{-\lambda_i t_i}, \; i = 1, 2, \ldots, k, \; t \geq 0 \; .$$

The c.d.f. of the maximum time t is then given by

$$F(t) = \prod_{i=1}^{k} (1 - e^{-\lambda_i t}) \qquad (5)$$

If $\lambda_i = \lambda$ for all i, the mean of $F(t)$ is given by

(6) $$\mu = \frac{1}{\lambda} \sum_{i=1}^{k} \frac{1}{i} .$$

Clearly, since all $\lambda_i = \lambda$ and hence all $\mu_i = \mu^* = 1/\lambda$, the conventional PERT solution under this condition is $\mu^* = 1/\lambda$. The bias is then given by

(7) $$\mu - \mu^* = \mu^* \sum_{i=2}^{k} \frac{1}{i} .$$

Thus if there are only k=4 activities in parallel the bias will be $\mu - \mu^* = \mu^*(\frac{13}{12})$ or more than 100 percent of the PERT solution, whilst with k=8 activities in parallel the bias is 1.718 or 172%. It should of course be remembered that the above bias applies to the particular network in Figure 5 which, in general would only constitute a small section of the large network. Therefore, the % bias in the PERT - computed expected completion times will not, in general, be as large as the above example would indicate. However, PERT will always make underestimates of the critical time intervals (see e.g., Fulkerson (1962), p. 808) so that the biases from individual network sections will cumulate.

EXAMPLE 2: Consider the same network as above but with the density functions given by $f(t_i) = \frac{1}{c}$; $0 \leq t_i \leq c$.

In this case

(8) $$F(t) = (t/c)^k , \quad 0 \leq t \leq c .$$

The mean value of F(t) is then

(9) $$\mu = \frac{kc}{k+1} .$$

The PERT solution would be the mean value of t_i which is $\mu^* = \frac{c}{2}$. The bias is found to be

$$\mu - \mu^* = \frac{k-1}{k+1}\,\mu^* \tag{10}$$

In this example the bias is at least bounded in that it cannot exceed 100% of the PERT solution. It does increase very rapidly however, with the number of activities in parallel. If k=4 as in the first example the bias is 60% of μ^*, when k=8 it is 78%.

EXAMPLE 3. To illustrate the dependence of the solution upon the form of the densities involved consider the following network.

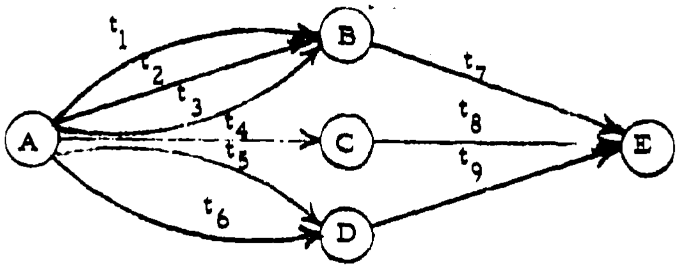


FIGURE 8
SHIFT OF CRITICAL PATH WITH FORM OF DISTRIBUTION

In this case suppose that the activities represented by the t_i have expected times as follows: -

Activity	Expected Time
t_1, t_2, t_3	9
t_4	11
t_5, t_6	10
t_7	3
t_8	6
t_9	4

If conventional PERT is applied, path ACE will be critical with a sum of expected times of 17 units. On the other hand, if the densities of the t_i are exponential and the operational logic of this paper is applied the expected time for ABE is 19 1/2 units, for ACE 17 units, and ADE 19 units, thus making ABE critical. This distribution dependence is further emphasized if the t_i are rectangularly distributed. In such a case the expected time for ABE is 16 1/2 units, and for ADE 17 1/3 units, thus making ADE critical.

The above examples, though somewhat elementary and academic, demonstrate the consequences of inadequate statistical treatment and approximate logic. The impact can be even more pronounced and the consequences more significant in a realistically large program plan.

9. RELATION TO THE EXISTING LITERATURE ON PERT. Most of the published work on PERT is concerned with computations based on the mean values of the completion times and deliberately ignores the bias discussed in this paper. There are undoubtedly situations when this bias is not serious notably in networks when

(a) There is a low degree of parallelism in the activities of the network and most operations are sequential and/or

(b) When some activities are carried out in parallel but one of them has a considerably longer expected completion time than the others parallel to it.

It will be agreed that the above conditions are not usually satisfied. In view of the very extensive, detailed and costly computations involved in the currently practiced PERT analysis it is surprising that so little attention has been paid to the bias affecting them.

We believe that whilst the possibility of a statistical approach (such as is here presented) has sometimes been considered (see e.g., Department of the Navy (1958), Appendix A, and Fulkerson, D. R. (1962)) it has apparently been regarded as leading to unsolvable or unmanageable mathematics. Indeed, Fulkerson (1962) who fully recognises the existence of the bias (see page 308) and offers an interesting approximate method to correct it, states (page 309): - "Since a typical PERT network may involve hundreds and thousands of arcs, the precise calculation of expected critical path lengths would, of course, be out of the question." Now it must of course be remembered that the method of numerical analysis here offered gives the solution only for the special case of multiple-crossed networks as here defined. We do not claim that the networks encountered in practice will usually belong to this category. However, if the algorithm described in section 5 is applied to a general network it will reduce it considerably so that the distribution of the critical time for the reduced network can be obtained by the Monte Carlo procedure described in section 6. Moreover, we could enlarge the scope of the numerical method of section 5 by adding (to the Wheatstone bridge operation for the network in Figure 3) similar basic crossed networks (such as that of Figure 4) and incorporate a calculation of the critical time (similar to that given by equation (3)) for such configurations. The feasibility and economy of such additions is under investigation.

Since we only give a hand full of references in spite of the vast literature on the subject, we should perhaps include the extensive Bibliography (Bolling Air Force Base (1963)) in our list.

REFERENCES

Bolling Air Force Base (1963). PERT Orientation and Training Center. Bibliography, PERT and Other Management Systems and Techniques.

Fulkerson, D. R. (1962). "Expected Critical Path Lengths in PERT Networks." Rand Corporation, RM 2075, P. R., Santa Monica, California.

McGowan, L. L. (1964). Monte Carlo Techniques Applied to PERT Networks. M. Sc. Thesis in Statistics, Texas A&M University.

Navy Special Projects Office (1958). PERT Summary Report, Phase I. (Gov't. Printing Office. Catalog No. D217.2, p. 94/958.