Eighth DOE Conference (1962) – NOT Available from DTIC

ARO-D Report 63-2

PROCEEDINGS OF THE EIGHTH 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. 63-2
December 1963

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

Sponsored by the Army Mathematics Steering Committee

conducted at

Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D. C.
24-26 October 1962

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

TABLE OF CONTENTS

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

** This paper was not presented at the Conference. It does not appear in this technical manual.

TABLE OF CONTENTS (contd.) Page

---

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

TABLE OF CONTENTS (contd) Page

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

This page intentionally left blank.

FOREWORD

The Walter Reed Army Institute of Research, Walter Reed Army Medical Center served as the host for the Eighth Conference on the Design of Experiments in Army Research, Development and Testing. This Washington, D. C. Conference was held 24-26 October 1962. Colonel Conn L. Milburn, Jr., Director of WRAIR, issued the following letter to those attending this scientific meeting:

"The staff and faculty of the Institute feel privileged and honored that Walter Reed Army Institute of Research has been selected as the place for this Eighth Conference on the Design of Experiments in Army Research, Development and Testing. To each of you we extend a most cordial welcome.

It is a pleasure to have you with us and we sincerely hope that your stay here will be both enjoyable and professionally rewarding."

The Army Mathematics Steering Committee takes this opportunity to thank Colonel Milburn for his welcoming remarks and for the use of the excellent facilities under his command. The thanks of the Committee are also due to the two Local Chairmen, Lt. Col. Stefano Vivona and Major Paul J. Wentworth, appointed by him, for their very efficient handling of all local arrangements.

The following information about the host installation was extracted, with minor alterations, from the Medical Annals of the District of Columbia, Vol XXX, No. 11, November 1961.

"The Walter Reed Army Institute of Research (WRAIR) was founded in 1893 as the Army Medical School, the first school of preventive medicine in the United States. The school was established by Surgeon General George M. Sternberg, one of the great military men of science, thus bringing to fruition an idea first proposed by Surgeon General William A. Hammond in 1862. On General Sternberg's recommendation the War Department issued General Orders No. 51, dated 24 June 1893, founding the school.

Captain (later Major) Walter Reed, in whose honor the Army Medical Center is named, was a member of the first faculty. Captain Reed was Professor of Clinical and Sanitary Microscopy and Director of the Pathological Laboratory. He was also the first Secretary of the Faculty.

The Army Medical School marked its fifthieth anniversary on 18 December 1943 with graduation exercises for the class completing courses in military and tropical medicine. The class of 124 included medical officers from the armies of the United States, Canada, and Peru, and officials of state health departments. An honored guest at the exercises was Colonel Deane C. Howard, retired, who held the highest rating in the first class graduated from the Army Medical School in 1894.

Since World War II the School has changed its name several times. The name, Army Medical Department Research and Graduate School, was adopted in 1947. This was changed to The Army Medical Service Graduate School in 1950 and finally, in 1955, the present name, The Walter Reed Army Institute of Research, was adopted.

The basic mission of the Walter Reed Army Institute of Research is 'To provide the medical research and professional graduate training required by the Army to fulfill its role in the national defense.' Actually, this broad mission is divided into 4 parts: (1) serving as a research and development center, (2) providing education and training to officers of the Army Medical Service (also to some officers from other branches of the Armed Forces and from other countries and some civilians), (3) serving as a central reference laboratory for the Army Medical Service and, in a number of fields, for the Navy, Air Force and Veterans Administration, and (4) manufacturing more than 100 biological products, not obtainable through commercial sources, which are supplied to all the Armed Forces and the Veterans Administration.

The first class began on 1 November 1893 with 5 students officially enrolled. From this small beginning the original Army Medical School has grown into the present great Institute of Research whose activities cover all parts of the world."

At the Eight Conference, invited addresses were delivered by Drs. Robert P. Abelson, Herbert C. Batson, Herman Chernoff, Egon S. Pearson, and Marvin A. Schneiderman. Decisions under uncertainty, bio-assay, optimal designs, assessing optimal performance of weapons, screening theory were, respectively, the areas discussed by these five specialists. Dr. Harold F. Dorn served as Chairman for the Panel Discussion on Diet and Heart Disease. Dr. George V. Mann initiated the Panel Discussion by presenting data and known facts about diet and heart diseases. Mr. Jerome Cornfield commented on the material presented by Mann; then he and the Chairman led the discission and answered questions by members of the audience. Miss Beatrice Orleans showed and discussed a film on the Design of Experiments. Two other important parts of the conference were the three papers discussed in the Clinical Sessions, and the twenty-six papers delivered in the Technical Sessions.

The present volume of the Proceedings contains twenty-eight papers which were presented at this meeting. The Army Mathematics Steering Committee, the sponsor of this series of conferences, has asked that these articles on modern statistical principles in the Design of Experiments, as well as certain applications of these ideas, be made available in the form of these Proceedings.

The Eighth Conference was attended by registrants and participants from over 80 different organizations. Speakers and panelists came from Booz-Allen Applied Research, Inc.; Bureau of Ships; C-E-I-R, Inc.; Ford Motor Co.; Iowa State University of Science and Technology; National Bureau of Standards; National Cancer Institute, NIH; National Heart Institute, NIH; North Carolina State College; Operations Research, Inc.; Princeton University; Research Analysis Corporation; Research Triangle Institute; Stanford University; University College, London; University of Illinois, College of Medicine; University of Michigan Institute of Science and Technology; University of North Carolina; University of Wisconsin, U. S. Army Mathematics Research Center; Vanderbilt University; Virginia Polytechnic Institute; Yale University, and thirteen Army Facilities.

Before closing, the Chairman wishes to express his sincere thanks to his Advisory Committee: F. G. Dressel (Secretary), Fred Frishman, B. G. Greenberg, Frank E. Grubbs, Boyd Harshbarger, H. L Lucas, Jr., Clifford J. Maloney, Lt. Col. Stephano Vivona, and Marvin Zelen for their suggestions and assistance in selecting the invited speakers and formalizing the plans for this conference. The Chairman is especially grateful to Dr. Dressel for coordinating the Conference program and seeing these Proceedings through publication.

S. S. Wilks
Professor of Mathematics
Princeton University

This page intentionally left blank.

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

24-26 October 1962

Walter Reed Army Institute of Research

Wednesday, 24 October

0900-0930 REGISTRATION
Lobby of Sternberg Auditorium (WRAIR)

0930-1215 GENERAL SESSION I
Sternberg Auditorium

Calling of Conference to Order
Lt. Colonel Stefano Vivona, Local Chairman

Welcome to Walter Reed Army Medical Center
Major General A. L. Tynes, Commanding

Welcome to Walter Reed Army Institute of Research
Colonel Conn L. Milburn, Jr., Commanding

Announcements
Major Paul J. Wentworth, Chairman on Local Arrangements

Chairman: Dr. Churchill Eisenhart, Statistical Engineering Laboratory, National Bureau of Standards

A Statistician's Place in Assessing the Likely Operational Performance of Army Weapons and Equipment
Professor Egon S. Pearson, University College, London

A General Survey of Screening Theory
Dr. Marvin A. Schneiderman, National Cancer Institute, National Institutes of Health

1220-1320 LUNCH - Ballroom, Officers' Open Mess, WRAMC

Technical Sessions I and II as well as Clinical Session A will start at 1330. After the coffee break Technical Session III and IV will convene and run from 1510-1700. Starting at 1700 there will be a Social Hour. This will provide an opportunity for old friends to get together and for new friends to get acquainted. We hope that everyone at the conference will be able to stay for this phase of the meeting.

1330-1440 TECHNICAL SESSION I - Sternberg Auditorium

Chairman: Sidney Sobelman, U.S. Army Munitions Command, Picatinny Arsenal

Estimation of Service Life from Fatigue Testing Results on Full Scale Specimens
John P. Purtell, Research and Engineering Division, Watervliet Arsenal.

The Simulated versus National Environment in Military Testing and Operations
C. Bruce Lee, Systems Analysis Section, Advanced Design Branch, Research & Engineering Directorate, Ordnance Tank-Automotive Command, Center Line, Michigan

1330-1440 TECHNICAL SESSION II - Room 358

Chairman: A.C. Cohn, Jr., The University of Georgia

Application of 2^{8-4} Fractional Factorials in Screening of Variables Affecting the Performance of Dry Process Zinc Battery Electrodes
Nicholas T. Wilburn, U.S. Army Electronics R & D Laboratory, Fort Monmouth, New Jersey

Applications of the Calculus for Factorial Arrangements
B. Kurkjian, Diamond Ordnance Fuze Laboratories
M. Zelen, U.S. Army Mathematics Research Center, University of Wisconsin

1330-1440 CLINICAL SESSION A - Room 341

Chairman: Frank E. Grubbs, Ballistics Research Laboratories

Panelists: W.T. Federer, U.S. Army Mathematics Research Center, The University of Wisconsin
B.G. Greenberg, The University of North Carolina
H.O. Hartley, Iowa State University of Science & Technology
H.L. Lucas, North Carolina State College
M.A. Schneiderman, National Institutes of Health

Statistical Procedures for the Evaluation of Thrust Curves
Paul C. Cox, White Sand Missile Range, New Mexico

1440-1510 COFFEE BREAK
Lobby Sternberg Auditorium

1510-1700 TECHNICAL SESSION III - Room 341

Chairman: Gerhard J. Isaac, U.S. Army Medical Research and Nutrition Laboratory, Denver, Colorado

The Independent Action Theory of Mortality as Tested at Fort Detrick
Francis M. Wadley, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Trial and Station Variability in P^{32} for Tripartite Collaborations
Walter D. Foster, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Design and Analysis of Entomological Field Experiments
William A. Brown, Dugway Proving Ground, Dugway, Utah
Scott A. Krane, C-E-I-R, Inc., Dugway Field Office, Dugway, Utah

1510-1700 TECHNICAL SESSION IV - Sternberg Auditorium

Chairman: Vaughn LeMaster, Ammunition Procurement and Supply Agency

Comparison of Two Approaches to Obtaining a Transformation Matrix Effecting a Fit to a Factor Solution Obtained in a Different Sample
Cecil D. Johnson, U.S. Army Personnel Research Office

Some Least-Squares Transformations of Regression Estimators of Orthogonal Factors
Emil F. Heermann, U.S. Army Personnel Research Office

A Reliability Test Method for "One-Shot" Items
H. J. Langlie, Aeronutronics Division, Ford Motor Company

1700-1900 SOCIAL - Ballroom, Officers' Open Mess, WRAMC

Thursday, 25 October

Technical Session V and VI will be held from 0800-0920. Clinical Session B and Technical Session VII will be in order from 0950-1120. The first paper in Clinical Session B carries a classification of SECRET and the second paper a classification of CONFIDENTIAL. General Session 2 will start at 1130. After lunch Technical Session VII and IX will convene at 1330 and end at 1440. The Panel Discussion on Diet and Heart Disease is scheduled to start at 1510. A film entitled "The Design of Experiments" will be shown following the panel discussion.

0800-0920 TECHNICAL SESSION V - Sternberg Auditorium

Chairman: Ralph Brown, U.S. Army Materiel Command
Frankford Arsenal

Investigation in Temperature Control of Hydraulic Systems in Rough Terrain Fork Trucks
Irving Tarlow, Quartermaster Research and Engineering Command, Natick, Massachusetts

Precision of Simultaneous Measurement Procedures
W.A. Thompson, Jr., University of Delaware and National Bureau of Standards

0800-0920 TECHNICAL SESSION VI - Room 341

Chairman: B.G. Greenberg, The University of North Carolina

The Ultraviolet Microscopy of Tissues
George I. Lavin, Ballistic Research Larobatories, Aberdeen Proving Ground, Maryland

Redundancies in Human Biomechanics and their Application in Assessing Military Man-Task Disability Performance Resulting from Ballistic Agents
William H. Kirby, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

0920-0950 COFFEE BREAK - Lobby Sternberg Auditorium

0950-1120 TECHNICAL SESSION VII - Room 358

Chairman: Paul C. Cox, White Sands Missile Range

Half-Normal Plots for Multi-Level Factorial Experiments
Scott A. Krane, C-E-I-R, Inc., Dugway Field Office, Dugway Proving Ground, Dugway, Utah

Proportional Frequency Designs
Sidney Addelmann, Statistics Research Division, Research Triangle Institute

0950-1120 CLINICAL SESSION B - Room 341

Chairman: Clifford J. Maloney, National Institutes of Health

Panelists: O.P. Bruno, Ballistics Research Laboratories
Boyd Harshbarger, Virginia Polytechnic Institute
S.S. Wilks, Princeton University
Marvin Zelen, U.S. Army Mathematics Research Center, The University of Wisconsin

Security Classification: SECRET

Design of an Experiment to Evaluate Aerosol and Storage Characteristics of a Viral Slurry
Samuel N. Metcalfe, U.S. Army Biological Laboratories, Fort Detrick, Maryland
Bertram W. Haines, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Security Classification: CONFIDENTIAL

Sensitivity Analysis of Outputs from a Computer Simulation Model of a Chemical Weapons System
Reynold Greenstone, Operations Research, Inc., Silver Spring, Maryland
Ira A. DeArmon, U.S. Army CBR Agency, Operations Research Group, Army Chemical Center, Maryland

1130-1230 GENERAL SESSION 2 - Sternberg Auditorium

Chairman: Fred Frishman, Army Research Office, Office, Chief of R & D

Optimal Design of Experiments
Professor Herman Chernoff, Stanford University

1230-1330 LUNCH - Ballroom, Officers' Open Mess, WRAMC

1330-1440 TECHNICAL SESSION VIII - Room 341

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

Vibration Experiments
F. Pradko, Dynamic Simulations Laboratory, U.S. Army Ordnance Tank-Automotive Command

Size Effects in the Measurement of Soil Strength Parameters
B. Hanamoto, Land Locomotion Lab, Research Division, Research and Engineering Directorate, ATAC
Emil H. Jebe, Institute of Science & Technology, The University of Michigan

1330-1440 TECHNICAL SESSION IX - Sternberg Auditorium

Chairman: Edwin Cox, Agriculture Research Service, Beltsville, Maryland

Effectiveness of Certain Experimental Plans Utilized in Sensory Evaluations
J. Wayne Hamman and Jan Eindhoven, Quartermaster Food and Container Institute for the Armed Forces, Chicago, Illinois

An Evaluation of Radiation-Processed Foods for the Military Rations
Donald M. Boyd, Research Analysis Corporation

1440-1510 COFFEE BREAK - Lobby Sternberg Auditorium

1510-1700 GENERAL SESSION 3 - Sternberg Auditorium

Panel Discussion on Diet and Heart Disease

Chairman: Dr. Harold F. Dorn, Biometrics Research Branch, National Heart Institute

Panelists: Dr. George V. Mann, School of Medicine, Vanderbilt University
Mr. Jerome Cornfield, Biometrics Research Branch, National Heart Institute

1700-1730 SPECIAL SESSION - Sternberg Auditorium

Chairman: Beatrice Orleans, Bureau of Ships

Film on the Design of Experiments

Friday, 26 October

Technical Sessions X and XI run from 0800-1000. General Session 4 will start at 1030 and end at 1230. After the lunch hour your host has planned for you an interesting and informative tour.

0800-1000 TECHNICAL SESSION X - Room 341

Chairman: Earl Atwood, Walter Reed Army Institute of Research

Some Consequences of Some Assumptions with Respect to the Physical Decay of a Chamber Aerosol Cloud
Theodore W. Horner, Booz-Allen Applied Research, Inc., Bethesda, Maryland

The Role of Intuition in the Scientific Method
Nicholas M. Smith, Research Analysis Corporation

How to Design War Games to Answer Research Questions
W. L. Pierce, Research Analysis Corporation

0800-1000 TECHNICAL SESSION XI - Sternberg Auditorium

Chairman: H. L. Lucas, North Carolina State College

Evaluation of Performance Reliability
Seymour K. Einbinder, Picatinny Arsenal
Ingram Olkin, Stanford University

Evaluation of Various Laboratory Methods for Determining Reliability
A. Bulfinch, Picatinny Arsenal

Confidence Intervals for Systems Reliability
Jasper Dowling, Picatinny Arsenal

1000-1030 COFFEE BREAK - Lobby Sternberg Auditorium

1030-1230 GENERAL SESSION 4 - Sternberg Auditorium

Chairman: Dr. S. S. Wilks, Princeton University

An Experimental Design for Decisions under Uncertainty
Dr. Robert P. Abelson, Yale University

Bio-Assay
Dr. Herbert C. Batson, University of Illinois, College of Medicine

1230-1330 LUNCH - Ballroom, Officers' Open Mess, WRAMC

1330-1500 TOUR - Assemble in Sternberg Auditorium

This page intentionally left blank.

A STATISTICIAN'S PLACE IN ASSESSING THE LIKELY OPERATIONAL PERFORMANCE OF ARMY WEAPONS AND EQUIPMENT

E. S. Pearson
University College, London, England

THE BACKGROUND OF THIS PAPER. It has been a special honour to receive an invitation from the organising committee of this Conference to make the journey from England and to address you today. In thinking how I could best repay the compliment, it seemed to me that I should look for a subject in illustrating which I could draw on my own particular experiences, gained in working for the British armed services both during and since the second world war. From 1939 to 1946 I was attached with a number of members of the University College, London Statistics Department, to the British Ordnance Board. This is an organisation of some historic interest for I believe its foundation can be traced back to an appointment made in 1414, the year before the Battle of Agincourt! It is now concerned with certain aspects of the development and acceptance of weapons for both the Army, the Navy and the Air Force. Then, for some years after the war, I was a member of the Ordnance Board Anti-aircraft Lethality Committee and very recently I have been pulled back to be chairman of an advisory committee concerned with the general problem of assessment in connection with army weapons and equipment.

My main experience was with the subject which has been described as terminal ballistics and in particular with the lethal effectiveness of anti-aircraft fire. We were concerned also with field artillery fire and with the medium and small bombs of those days, in so far as fragmentation of the casing rather than blast played an important part in their effectiveness. It is of course true that the weapons and the army requirements of 15-20 years ago have been to a large extent out-dated, but if I make my main topic today a piece of historical recording, it is because I believe that a number of general principles and lessons emerge from such a study which are still relevant to the practice of experimentation and analysis in Army Research today.

It seemed to me that there were two advantages in taking illustrations from World War II experience. In the first place I could speak of matters about which I had the 'feel' from first hand knowledge and so perhaps could be more interesting as well as convincing in any arguments put forward. Secondly, it was easier to be factual without running into the danger of using classified material. What I shall try to do, therefore, is to give you first some account of the difficulties with which we were faced in the years 1939-45 in constructing a model which could be used to help determine how to improve the effectiveness of anti-aircraft fire. In describing this problem, it should be possible to indicate a number of lessons which are still relevant in a much wider field. There are also many points of difference which it will be instructive to emphasise.

THE STATISTICIAN'S PLACE. I should perhaps confess straight away that I shall say very little about statistics or about what is commonly thought of as the design of experiments. To this extent you may think that the leading phrase in the title of this paper is misleading, unless you interpret the words in the personal sense as referring to the statistician who is giving this address! But there is, I think, a point here which I should like to make. At the fourth of this series of Conferences, held in 1958, Dr. A. W. Kimball read a paper with the title: "Errors of the 3rd kind in statistical consulting"; in this he discussed and illustrated the fault of giving a perfectly sound statistical answer to a problem which is not the real one needing solution.

Many of us are I think conscious of what might perhaps be called an error of a 4th kind; that which the statistician makes when he allows his interest in the statistical elements of a problem and its potential for statistical elegance and sophistication to obscure what should be his prime objective, the solution of the real matter at issue. The fault is not so much that wrong statistical methods are used (Kimball's 3rd kind of error) but that the situation does not justify the use of any refined statistical methods at all until the outstanding problem has been solved of obtaining data which are both relevant and reliable. The statistician, indeed, is called upon to be a scientist in the fullest sense of that term--to apply scientific method, not merely statistical techniques, to the job on hand.

When he has completed some piece of mathematical or arithmetical analysis, he needs to ask himself searchingly: does this answer make sense? I can recall, as no doubt some of you can too, war-time reports which appeared both in my country and in yours, containing a pretty piece of algebraic development or some standard analysis of variance, the conclusions from which obviously did not make sense. Perhaps such reports from youthful enthusiasts would never have appeared but for the inevitable shortage of experienced and critical supervision in rapidly expanding organisations. They are likely, however, to discourage the idea that mathematics or statistics were of value in problems of weapon development and testing, because the experienced non-statistical layman, the military or naval technical officer who had the feel of the problems, could see at once that the data would not bear the confident interpretation which was often placed on them.

Certainly in my own experience at the Ordnance Board it was the physical difficulty in securing meaningful experimental data which had always to be faced. There was very little opportunity for design as it is understood in agricultural or biological trials. There was no paramount function for the application of advanced statistics--we used to say that the only statistical tools which were needed were the normal distribution in 1, 2 and 3 dimensions, the Poisson and the bionomial. But it is true to say that the statistician's training, with the understanding which should follow of the meaning of variation and correlation, of randomness and probability, with its emphasis on the importance of adopting a critical outlook on assumptions--all this is likely to provide an excellent preparation for the kind of work we are discussing, but on one essential condition--that the training has been carried out in conjunction with practical application to data analysis. The trend in the teaching of mathematical statistics at our universities today is often increasingly away from any real application to data.

There is another point which I think is worth emphasising. One of the surest ways to cure the statistician from any tendency to over-sophistication is to arrange that he is present at experiments or trials, the data from which he is to use. In this respect we were lucky in England; we attended firing trials on the Shoeburyness Ranges, we were hot on the scene after bombs had been dropped on parked aircraft, trucks and wooden dummies in slit trenches on a special bombing range in the New Forest, and--as a wartime experience--we might happen to be present at a gun-site when German aircraft were the target. Under such conditions it is easier to come to grips with the meaning and limitations of data.

THE ANTI-AIRCRAFT PROBLEM. First let me try to put this problem into its setting of 20 or more years ago. As far as the Ordnance Board group was concerned, we had not to consider the problems of the deployment of guns, of the acquisition of targets, of the handling of mass attacks or other important tactical matters. These were questions for the Anti-aircraft Command and its Operational Research Section which was formed in the summer of 1940. Our work was closely related to the question of design, to understand more clearly the individual relationship between predictor, gun, shell, fuze and enemy target in order to advise what improvements were possible and likely to be worthwhile.

In this field of research where the terminal action in which one is interested may be taking place several thousand feet above ground, no overall experiment bringing in all the factors concerned is conceivable; the reasons for this are so obvious that I do not need to list them. As a consequence, it is absolutely essential to construct a mathematical model of the terminal engagement, and then to consider how the parameters of this model may best be estimated. As in so many other problems of military science, the model even if necessarily simplified, serves as an essential means of defining the relationships of the situation, showing how research investigation can be broken into separate pieces and emphasising at what points our lack of sure information is greatest and most hampering.

Let me now outline the problem and its solution in some detail, first describing the mathematical model and then discussing the three main headings under which gaps in knowledge had to be filled, namely:

(i) positioning errors (until the introduction of the proximity fuze in 1943-44 it was easy to combine the error of the time fuze with the predictor, gun-laying and ballistic errors);

(ii) fragmentation characteristics of the shell;

(iii) target vulnerability.

The difficulties which had to be overcome, largely through ignorance of physical properties in this hitherto unexplored field, are I think sufficiently instructive to be worth including as part of the story. Much the same problems were I know faced later on (building perhaps on our experience) in Section T of the Applied Physics Laboratory at Silver Spring and the associated Proving Ground near Albuquerque, where research and trial work was carried out for the U. S. Navy. I did not myself have any direct contact with U. S. Army investigations.

THE MATHEMATICAL MODEL. The first simplified model which was used involved:

(a) A three-dimensional normal distribution of positioning errors about the target, with a major axis along the shell trajectory and the standard errors in directions perpendicular to this axis equal, i. e., the density contours were taken to be ellipsoids with circular cross sections in planes perpendicular to the principal axis.

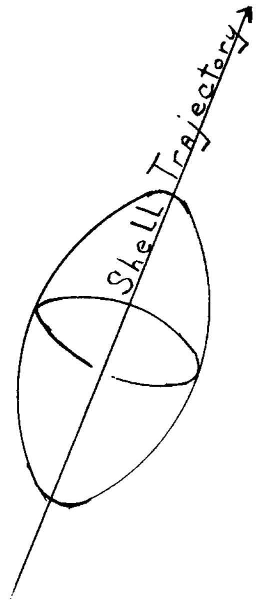


Figure 1

(b) A main fragment zone lying between two cones whose axis was that of the shell axis and the trajectory at time of burst, and a small sudsidiary nose cone.

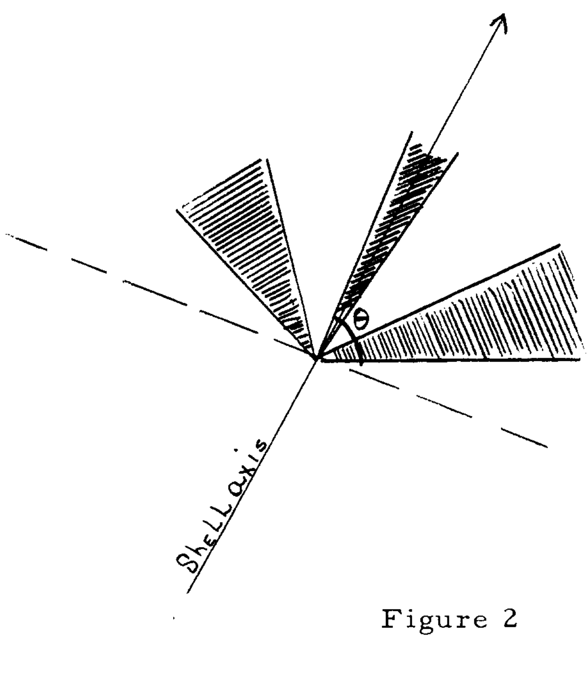


Figure 2

The density of fragments within the main zone was not of course uniform, though it might be treated as such for a first approximation. For any zone within which the average density of fragments of a given penetrating power could be regarded as constant, the probability distribution of strikes was taken as Poisson.

(c) For the aircraft, we first used what was termed an 'equivalent vulnerable target' represented by a sphere of a few feet in radius such that its 'perforation' by at least one 'lethal' shell fragment would result in a kill. Later, this representation had to be treated in more detail.

This simple model based on the trivariate normal and the Poisson distributions, with bounding surfaces consisting of ellipsoids, cones and spheres was amenable to computation, provided that meaningful numerical values for the various parameters could be estimated. But the task of filling in these unknown elements was immense and for a time the more we learnt, the more we realised our ignorance. Consider then some of the gaps to be filled.

THE POSITIONING ERRORS. The original data were collected from Practice Camp firings at towed 'sleeves', using kine-theodolites to measure the relative position of shell bursts and target. This was much too slow a target and the Practice Camp computational analysis was not very accurate. Later, in April 1940, a special trial of predictor accuracy was staged, following a free flying aircraft, and using camera recordings of the predictor output dials synchronised with kine-theodolites tracking the target. However, when German aircraft began to come over England later in 1940, it was at once clear that the aiming errors under operational conditions were much greater than those estimated from trials. We were up against the problem of increased operator inaccuracy under stress.

I remember P. M. S. Blackett (who was then in charge of the newly formed A. A. Command, Operational Research Group), wondering after watching the shell bursts in the night sky and a searchlight-held enemy aircraft, whether it would be possible to determine roughly an operational error distribution with appropriate photo-positioning equipment. I think that we later gave up all hope of estimating the actual aiming errors under operational conditions and made our calculations for a variety of different error combinations, which was often all that was needed in reaching conclusions about the relative merits of different types of shell, etc. It was only towards the end of the war when we were faced with that ideal straight-line-flying target, the V1 flying bomb, and when using proximity fuzes that a rough operational check on the overall adequacy of the model could be made.

THE FRAGMENTATION PROBLEM. Before the war, the standard trials for determining the fragmentation characteristics of shell were:

(a) Fragmentation in a sand-bag 'beehive', the shell fragments being recovered, passed successively through various sizes of sieve and (above a certain minimum size) counted and weighed.

(b) Trials to measure the dispersion and penetrating power of fragments by detonating the shell some 5 ft. above ground, in a surround of 2-inch-thick wooden targets, placed in a semicircle of, say, 30, 60, 90 or 120 ft. radius. The detonation was either at rest or obtained by firing the shell with appropriate remaining velocities against a light bursting screen.

With the war-time allocation of additional scientific effort onto weapon lethality problems, the number of questions which were posed for answering was greatly increased. The shell and bomb fragment attack on many targets besides aircraft had to be considered. On the one side it was necessary to have means of projecting individual fragments of various sizes at known velocities, against a variety of targets. On the other it was important to know more about the size-velocity-directional pattern as well as the retardation of the fragments projected by a complete shell burst in flight.

As soon as forward planning is attempted it becomes necessary to generalise the characteristics of a weapon; in the case of A. A. shell the ultimate objective was to be able to predict the characteristics of the fragment distribution from

(i) the drawing board design,

(ii) a knowledge of the particular explosive filling to be used, and

(iii) for any desired forward velocity of the shell.

It became clear that the old form of trials mentioned in the first paragraph of this section was inadequate. When shells were burst in flight in a wood target surround the resulting pattern of perforations could not be accurately related to the pattern from a static burst, merely by adding the component forward velocity of the shell. Nor was it easy to link the distribution of fragment sizes from the sand-bag collection with the number of perforations in the wood, using any simple assumptions about velocities and retardations. The essential need was for more basic physical experimentation; without this we could not generalise.

Here we were lucky in getting help from a very skilled scientific team at our Safety in Mines Research Establishment at Buxton, who initiated a programme of research which gradually succeeded in disentangling the picture. Shells on which small letters were engraved in successive rings round the circumference were fired at rest, within a surround of strawboard, against which a large number of small velocity measuring screens were placed. Fragments subsequently collected and weighed could be identified with a particular zone of the shell, and velocities estimated either by direct measurement or more crudely from depth of penetration into the strawboard.

It then became clear that the initial velocity of fragments varied very considerably with the part of the casing from which they came and similarly, that size or weight also varied with position. To some extent this initial velocity could be related to the charge/weight ratio of the section of the shell (perpendicular to its axis) from which the fragments originated. With this information, we began at last to get a surer picture of how fragments would be projected from different designs of shell detonated at any given velocity in free air.

It should be noted that the angle of the fragment zone, in particular the rather sharply defined 'cut-off angle' or semi-vertical angle θ of the backward bounding cone of my Figure 2 became particularly important with the introduction of proximity fuzes. If the pattern of fuze functioning was not co-ordinated with that of fragmentation the shell might generally burst in positions relative to the target such that fragments were bound to miss the more vulnerable parts of the aircraft.

<u>AIRCRAFT VULNERABILITY.</u> In the earliest trials carried out shortly before the war, an aircraft and an arc of large 2-inch thick vertical wooden screens were placed beyond and on opposite sides of a small burster screen at which the shell (with percussion fuze) was fired at a prescribed velocity.

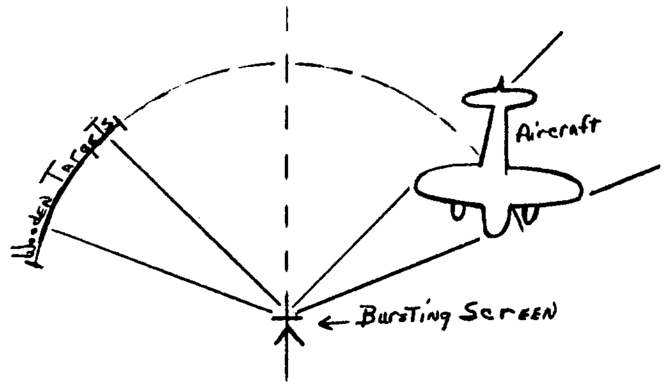


Figure 3

It was in this way possible to correlate the damage done to the aircraft with the density of fragments which perforated two inches of wood in a second, similarly constituted fragment stream. By noting and painting round the fragment holes after each round was fired, the same target could be used a large number of times, varying the aspect of attack and distance of detonation as desired.

It was from the observed correlation of density of 'throughs' (fragment capable of perforating 2 ins. of wood) and damage that it was possible to introduce into the model calculations a simplified 'equivalent vulnerable target'. This was the first method of attack. At a later stage after experimental techniques had become more refined and the Royal Aircraft Establishment assessors more experienced, it became possible to dissect the problem still further. The overall vulnerability picture was then built up from information gained by firing from high velocity barrels individual fragments of predetermined sizes, housed in specially designed cups, at a variety of aircraft components, which were screened where necessary by aluminium plates representing wing surfaces or fuselage.

The information so obtained could of course be used directly both in trying to draw conclusions about optimum fragment sizes and velocities and in considering ways of improving the protection of our own aircraft. Viewed in this way the problem may not appear to be statistical at all, but it did assume a statistical character as soon as one had to try and make use of this information in the 'model', with its shells bursting in a probability distribution around an aircraft and each projecting a composite stream of fragments, whose frequency distribution of strikes on equal areas of an intervening target would be roughly of Poisson form.

SOME CONCLUSIONS DRAWN FROM THIS SURVEY. Looking back now after a number of years, it seems to me that by 1944 we had really broken the back of the problem. It became possible to make recommendations with some confidence of a number of matters: on the optimum design characteristics of time fuzed and of proximity fuzed shell; on the relative importance of case thickness and explosive filling; on what might be achieved by using methods to control the size of fragments; on the relative gains to be won by improvement in fire control and in design of shell. Few such questions could have been answered with any confidence in 1939.

It is of course a truism that much of the fundamental research bearing on military problems is only rounded off when it is becoming too late to be of use in the war which provided the stimulus for the effort; and by the next war, the whole conditions of warfare are changed. This seems particularly true in regard to the ground-to-air weapons. But I think that the work I have been describing brought to the front a number of general principles, a sample of which I will bring to your attention in concluding this account.

The ease with which important factors may be overlooked. A common experience when the human mind starts to investigate the unknown is the way in which important considerations which seem so obvious afterwards are only realised through a process of slow and perhaps painful discovery.

(a) We did not for long appreciate the effect of ground ricochet in our firing trials. The influence of ricochet and other factors arising from proximity to the ground on the directional distribution of fragments would be natural operational effects in the case of field artillery or dropped bombs, but were very confusing when we were seeking information about the character of shell-bursts thousands of feet above ground. I know that the American experimenters appreciated this effect before we did and were the first to introduce ricochet traps into A. A. shell trials. Perhaps the most convincing demonstration of its existence which I recall occurred when we burst a 500 lb. bomb statically, with axis inclined at 30° to the vertical. The target screens showed a striking pattern of holes; a tilted belt like the forward-arm of a V from direct hits and another, like the other arm, from the ground ricochets. As long as bombs or shell were burst with their axes horizontal (or vertical), the effect remained unnoticed.

(b) Again, when studying the size distribution of fragments, the amount of secondary break-up on striking the collecting medium after detonation, was only realised when strawboard was used in place of sand

and the paths of these pieces, broken on first strike, could be traced through the successive layers of board.

(c) Another point not fully appreciated was the effect of emotional stress on the human element under battle conditions. The assessment of its magnitude, especially under circumstances and conditions which cannot be precisely foretold, is one of the hardest problems of the moment.

The place of basic research. In many instances it may not be too difficult to carry out a realistic trial of a particular weapon, against a given target under specified environmental conditions. But a more fundamental knowledge is necessary to assess the performance of weapons, perhaps still on the drawing board, under a wide variety of conditions. It was in this connection that the detailed experimental work on fragmentation performed to laboratory standards was essential, even if the laws of initial velocity, of size distribution and of retardation which resulted were to some extent empirical.

The value of having something up your sleeve. Observation of the amount of the metal casing which appeared to be broken up into dust or very small fragments*, on detonation, suggested that the destructive power of the anti-aircraft shell might be considerably increased by 'controlling' the size of fragments. It was over this matter that the help of the Safety in Mines Research Establishment was first called on, and by the end of the war this research group had developed a variety of techniques, relatively easily applied, by which it was possible to control the size and shape of shell and bomb fragments to a remarkable degree. These techniques were never used** but they were available to put into operation should any new target have had to be faced, e.g. a tough one against which only large fragments could be effective.

* It was realised later that some of this effect was due to secondary break-up of the large fragments on striking the collecting medium.

** It was found later that the Germans had applied a system of external grooving to some of their A. A. shell, apparently to increase the fragment size.

These are some of the still relevant points which I have noted in again coming into contact with problems of weapon research and development after a gap of several years. I am sure there are other lessons to be drawn from these World War II investigations, and without doubt those scientists who have carried on continuously in government service will have quietly absorbed them, so that they form part of their whole attitude of approach to the problems of today.

THE POSITION TODAY. There are, of course, many obvious differences between:

(a) The war-time problem, which was essentially that of trying to establish an understanding of a weapon system in service, in order to determine how its effectiveness could be improved, under conditions which were not expected to change radically from those known to exist; and

(b) The problem of today, which is greatly concerned with predictive assessments of the operational performance of future systems, taking many years to develop and to be used against an opponent whose future equipment, weapons and tactics must be to a large extent a matter of guesswork.

In the course of war, even when action has to be taken to meet a new situation, this can be done by working on the basis of information which possesses some element of reality. A good example of this occurred in 1944 with the launching of the V.1 flying bombs against London. Within a few days a complete bomb which had been shot down without exploding was recovered, and immediate steps could be taken to estimate its vulnerability to shell-fire and fighter attack.

As far as I can recall, priority trials were undertaken to determine (a) the burst pattern of a proximity fuze around such a target, and (b) the nature and extent of its vulnerability to A. A. shell fragments. How quickly we went as far as inserting these new parameters into our probability model, I cannot remember; but it must have been soon evident that the V.1 was a target which could be successfully engaged by 3.7 inch anti-aircraft guns with existing shell, provided they were supplied with proximity fuzes. The large-scale delivery of American fuzes and the appropriate re-deployment of guns, when achieved after some weeks when the fighter aircraft had been forced to take the leading defence roll, played a very large part in countering the menace.

The scientific effort, when it became accepted as of value by the armed services, was quite naturally first directed to the study of the performance of individual weapons or pieces of equipment; the radar set, the proximity fuze, the terminal ballistics of a shell or of a variety of anti-tank weapons. Today there is a special demand for scientific aid in the intractable job of peering into the future. The lead for this activity was of course provided by the Operational Research Sections which were closely associated with various operational commands during the war. In this very difficult field of prediction in which the last war's operational experience becomes less and less relevant, the scientific line of attack must consist in welding together a great number of elements.

The following scheme of relationships illustrates what I mean by the many-sided approach:

$$\left.\begin{array}{l}\text{Mathematical models}\\ \text{Paper study}\end{array}\right\} \begin{array}{c}\longrightarrow\\ \longleftarrow\end{array} \left\{\begin{array}{l}\text{Laboratory experiments}\\ \text{Simulator trials}\\ \text{Range trials}\end{array}\right.$$

$$\left.\begin{array}{l}\text{War games}\\ \text{Computer games}\end{array}\right\} \begin{array}{c}\longrightarrow\\ \longleftarrow\end{array} \left\{\begin{array}{l}\text{Field experiments}\\ \text{Army exercises}\end{array}\right.$$

The overall inferences to be drawn from the whole build-up are not of course matters of statistics; but the use of the theory of probability and of stochastic processes is implicit in the studies on the left-hand column, while statistical planning plays its part in the laboratory experiments and the range trials--even to some extend in the field trials.

I have already tried to illustrate the great value of a mathematical model in forming the structure against which an evaluation problem may be broken up into parts for separate study. In so doing attention is drawn to the links in the construction where essential information in quantitative form is most needed and perhaps most lacking. Again, and this is important, by permitting a good deal of elasticity in the mechanism and allowing for the introduction of factors which might conceivably operate in a future situation, the model may be used to extrapolate beyond the envelope of engagement conditions tested during field trials or even accepted as likely under present combat conditions.

The application of the model approach to the problem of ground-to-air missile evaluation is the natural successor to the war-time investigations which I have described. The break-up of the problem for study under four headings still remains as before.

(a) Engagement geometry, (b) fuze performance,

(c) warhead effectiveness, (d) target vulnerability.

But problem (a) has taken a much more complex shape, involving perhaps the use of both analogue and digital computers. The war game has an essential part to play as a research tool in the combined attack on the problem of developing weapons, equipment and tactics for the future. Its main function is perhaps to aid thought and analysis rather than to obtain directo results. By injecting the human decision process into the study, it provides an insight into the complex nature of land battle which it would be hard to get in any other way. In this form of study, as elsewhere, the essential need to formulate rules, focuses attention on the limiting conditions which have to be accepted by whatever route we try to make predictions of the performance of future systems.

As a final illustration of where we now stand, let me refer to a problem of considerable present interest in whose solution a number of the techniques tabled above might be called in. This is the problem of comparing the merits of the free flight gun and the guided missile in the ground attack on armour. Both types of weapon depend, though in very different degree, on the human operator:

The free flight gun. Here we have a system, fairly well understood which has been studied for years and for which a reasonable idea of performance under operational conditions is available. The operator has only to concentrate while laying the gun and, after firing, plays no further part in the fate of that particular round. The greatest element of uncertainty lies in the vulnerability of his own gun, and to assess this requires rather extensive study of visibility and audibility in a variety of environments.

The guided weapon. The advantage of this weapon is that its firing position can be concealed behind the crest of a hill. However, the human controller who must see the target, has to concentrate for a considerable time (depending on the range) in guiding the weapon onto the target. That he can do this with fair success has been demonstrated on a simulator and with live weapons used under trial conditions. The open question here is whether he can maintain this performance in an operational setting, when subject to the fears and emotions to which he would be exposed in battle.

A sound basis for any policy decision on these alternative systems must depend on a comparative quantitative assessment; this cannot be completed without these missing pieces of information--the vulnerability of the gun to enemy counter action, the fall-off in human performance in a battle setting, and now adding to the puzzle, the observational power of the helicopter. Success in solution depends not only in not overlooking these considerations but, in persuading authority to provide the means of proceeding to the answers. How often one wonders have important decisions on weapon policy had to be taken in the past when the basic information for a real comparison was not available, although with greater foresight, perhaps, it might have been obtained in time.

Finally, it may again be asked: what of the statistician? Have I pushed him out of the picture: I think not. You must remember that I have been concentrating on a particular aspect of this matter of research, development and testing--the assessment of operational performance of weapons. In this peculairly difficult field, the statistician becomes the scientist who must merge his statistical identity into that of a group of men trained in several disciplines, but prepared to give no undue weight to any one of them in searching for answers to the problems in hand. That at any rate has been my personal experience.

This page intentionally left blank.

APPLICATION OF 2^{8-4} FRACTIONAL FACTORIALS IN SCREENING OF VARIABLES AFFECTING THE PERFORMANCE OF DRY PROCESS ZINC BATTERY ELECTRODES

Nicholas T. Wilburn
U. S. Army Electronics Research and Development Laboratory
Fort Monmouth, New Jersey

Among its research and development activities on zinc-silver oxide batteries for special applications, USAELRDL is investigating the preparation of zinc electrodes by dry processes. These involve the application of dendritic zinc powders under pressure to the grids. The interlocking properties of the dendritic zinc particles make it possible to form the electrodes with moderate pressures such that the porosity and related high surface area of the electrode is not destroyed. It is expected that dry process zinc electrodes will have many advantages over conventional electrodeposited sponge zinc electrodes, including higher discharge efficiency, greater uniformity of performance and better adaptability to mechanized production with resultant economics.

Due to the large number of variables affecting the discharge performance of the electrodes, it was decided to design and conduct a fractional factorial experiment to isolate the significant variables. These variables, and any controlling interactions between them, could then be studied further to arrive at the optimum conditions for the production of electrodes of maximum discharge efficiency. The fractional factorial experiment was thus intended for the preliminary screening of all major variables acting simultaneously. As such, it was recognized that it is the most efficient and economical process known for accomplishing this, in addition to providing valuable data on interactions between the variables which cannot be obtained from the more widely used one or two at a time variable investigations.

There were two categories of variables in the electrode investigation, those related to the electrolytic formation of the dendritic zinc powders and those related to the electrode preparation itself. Although several more variables were considered, it was decided to limit the number of variables to eight, keeping all other factors constant. The eight selected variables are shown in Figure 1. High and low levels as shown were assigned to each variable. Although considerable thought was given to determination of the levels, it is seen in retrospect that wider ranges might have been assigned in some instances. Based on available literature, these were, however, considered sufficiently wide ranges.

---

*Figures can be found at the end of this article.

Variable G, the electrolyte temperature during the plating operation, turned out to be impossible to control with the plating equipment which was prepared for the experiment and with the selected plating current densities. The experiment thus became one involving seven variables, each at two levels. Rather than to eliminate G, however, it was felt that this paper would serve a broader purpose if it gave the fractional factorial procedures for the study of eight as well as seven variables. Variable G should, therefore, be considered both in and out of the figures and analysis.

Having established the variables and the high and low levels, the fractional replicate design was established, as shown in Figure 2. This is the first design, a second having been run later for reasons to be discussed. The design involves eight variables (or seven) each at two levels to be studied with a total of sixteen different electrodes. A full 2^8 factorial experiment, eight variables each at two levels, would involve 2^8 or 256 trials (128 trials for a 2^7 factorial). Therefore, this design represents a one sixteenth fraction of the 2^8 factorial (one eighth of the 2^7), expressed as a 2^{8-4} design (or 27-3 for seven factors). The design is based on extension of a basic 2^4 or 16 trials. Thus the first four variables are arranged in standard order for the 2^4 factorial. The other variables are then introduced by making a basic assumption that three factor interactions between the first four variables are negligible. Therefore E is introduced by equating it to the interaction between variables A, B, and C. Regarding the high level as plus and the low level as minus, the level of E for the first box is $- \times - \times - = -$, $+ \times - \times - = +$ for the second box, etc. Similarly, variable F is introduced to equating it to the BCD interaction, variable G to the ABD interaction and variable H to the ACD interaction. Normally in a seven variable design, the seventh variable would be equated to ABD instead of ACD as was done here since variable G was dropped out. However, this does not affect the 2^{7-3} experiment in any way.

Having thus established the fractional design, the sixteen electrodes were prepared in accordance with the high and low level criteria for each variable. The electrodes were then tested one at a time under standard and carefully controlled conditions to give the sixteen yields or responses as shown. The response is the discharge efficiency of the electrode expressed in percent as the ratio of the output capacity (at a discharge current density of 1.1 amp/square inch to a 0.3 volt change for the electrode) to the theoretical capacity of the zinc active material. For comparison the conventional sponge zinc electrodes give an average efficiency of about 20% under comparable discharge conditions. The average of the sixteen responses is 31.2%. Before proceeding with the analysis of the responses, it should be noted how the design can be used to relate the sixteen mean effects obtained from the response analysis to the variables and their two factor interactions. Consider the high level boxes in the first four columns. The first mean effect, no high levels, will be for twice the average. The second will be for variable A, the third for B, the fourth for the AB interaction, the fifth for C and so on. The eighth (A, B and C at the high level) will be for variable E which was originally equated to ABC. Similarly the twelfth for G, the fourteenth for H and the fifteenth for F. Since this is a one sixteenth fraction, each of these principal effects will be confused with fifteen other effects. However, these for each of the eight variables will be three factor or higher order interactions which are considered negligible, the basis on which the design was established. Each two factor interaction, AB for example, will be confused with fifteen other effects of which three are other two factor interactions. The sixteenth row of the First Design will represent such a combination of four 2 factor interactions.

The analysis of the sixteen responses is shown in Figure 3. The technique used here is the Yates' Algorithm which is a rapid method for obtaining the same mean effects that would be obtained from a formal and lengthy analysis of variance. The Yates' Algorithm is applicable to any factorial experiment. Its advantages become more apparent the larger the experiment.

The mechanics of the Yates' calculations are very simple. The first figure in Column (1) is the sum of responses 1 and 2; the second, the sum of responses 3 and 4, etc. The ninth figure is the sum of responses 1 and 2 with the sign of response 1 reversed. Column (2) is derived from Column (1) in the same manner. Additional columns are introduced until a column is completed the first figure of which is equal to the sum of the responses. The arithmetic in each column is checked before proceeding to the next column. The sum of Column (1) is equal to twice the sum of the even numbered responses; the sum of Column (2) is equal to four times the sum of every second even numbered response; the sum of Column (3) is equal to eight times the sum of every fourth even numbered response; and the sum of Column (4) is equal to sixteen times the sixteenth response. The mean effects are obtained by dividing each figure of Column (4) by eight. The sum of the mean effects is also checked. The 62.363 figure, twice the average, is not used in the subsequent analysis. The effects measured by the mean effects are given in the last column. The effects A, B, AB, C, AC, etc. are those as previously read off the first design chart. The three factor and higher order interactions which they are confused with are of no importance since the design is based on their being assumed negligible. The other two factor interactions are important and must be known. They are obtained from the effects A, B, AB, etc. and the defining contrasts of the design. These defining contrasts are obtained from the equating that had been done in establishing the fractional design; E = ABC, F = BCD, G = ABD and H = ACD. Sixteen defining contrasts are required for the eight variable design. The first defining contrast is always I, the next four are ABCE, BCDF, ABDG and ACDH. The remaining eleven are found by exhaustively multiplying these contrasts using the rule that like factors are cancelled out.

(1) I
(2) $ABC \times E = ABCE$
(3) $BCD \times F = BCDF$
(4) $ABD \times G = ABDG$
(5) $ACD \times H = ACDH$
(6) $A\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}DF = ADEF$
(7) $\cancel{A}\cancel{B}CE \times \cancel{A}\cancel{B}DG = CDEG$
(8) $\cancel{A}B\cancel{C}E \times \cancel{A}\cancel{C}DH = BDEH$
(9) $\cancel{B}C\cancel{D}F \times A\cancel{B}\cancel{D}G = ACFG$
(10) $B\cancel{C}\cancel{D}F \times A\cancel{C}\cancel{D}H = ABFH$
(11) $\cancel{A}B\cancel{D}G \times \cancel{A}C\cancel{D}H = BCGH$
(12) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}B\cancel{D}G = BEFG$
(13) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}C\cancel{D}H = CEFH$
(14) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{A}\cancel{B}\cancel{D}G \times A\cancel{C}\cancel{D}H = AEGH$
(15) $\cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}\cancel{B}\cancel{D}G \times \cancel{A}\cancel{C}DH = DFGH$
(16) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}B\cancel{D}G \times ACDH = ABCDEFGH$

(These sixteen defining contrasts may also be obtained by simply reading off the low level boxes for each of the sixteen trials in the First Design chart. However this procedure will not apply in all cases, e.g. in the 2^{7-3} it would give sixteen defining contrasts when only eight are required.)

The principal effects are found by multiplying the first effect, A, B, AB, etc., by the sixteen defining contrasts. The two factor interactions confused with AB are found for example to be CE, DG and FH.

$$AB \times I = AB$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}CE = CE$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}DG = DG$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}FH = FH$$

The other twelve effects confused with AB, all three factor or higher order interactions, could be found with the other twelve defining contrasts if desired. The total 256 effects found by multiplying each starting effect by the sixteen defining contrasts would give the full 256 effects, from I to the interaction ABCDEFGH, which would be obtained in conducting a full 2^8 factorial experiment.

The defining contrasts for the 2^{7-3} experiment are I, ABCE, BCDF, ACDH, AEDF, BDEH, ABFH, and CEFH. The effects measured are found in the same way as for the eight variable design. The only difference is that, in this case, all effects involving G drop out. The twelfth set of effects measured, identified by the asterisk, then becomes eight interactions, all three factor or higher order.

The fifteen mean effects with their identifying effects are then ordered by arranging them in order of magnitude without regard to sign. The thus ordered set of mean effects is then arranged in a half-normal plot as shown in Figure 4 to interpret the relative significance of the effects. The ordinate is the order number of the fifteen effects from smallest to largest. The abscissa gives the mean effect magnitudes. The fifteen points which are plotted are identified by the proper major effects and two factor interactions. The plot is given for both the full eight variables, and the actual seven variable experiment. In the latter case, the G factor and the G interactions drop out. The asterisk again denotes high order interactions. In the plot an error best straight line has been drawn through the lowest seven points. High magnitude effects falling significantly off the line are judged to be distinct from error and therefore controlling factors in the process being investigated. In a plot of this type, it is considered unusual, however, to have a well defined error line with as many as eight points falling clearly off it. To gain insight into this unusual behavior as well as to gain more precision in the estimation of all of the effects, it was decided to conduct a second phase of the experiment, another group of sixteen electrodes differing from the first. Since an interaction between A (the highest magnitude effect) and B appeared reasonable, and since F and H were both high magnitude effects which might interact with each other, it was decided to establish the second design in such a way as to separate the AB and FH interactions. Although this can be done in several equally effective ways, it was decided to do it by reversing the levels for factors A and D, giving the design as shown in Figure 5.

This design is identical to the first except for the reversal of levels of variables A and D. Sixteen electrodes were prepared in accordance with the indicated variable levels. The electrodes were then tested to give the listed responses. The next step in the experiment was to determine the mean effects generated by these responses and to combine them with the mean effects resulting from the first design. This may be done in two ways. The first is to combine the two sets of sixteen responses into an overall group of thirty-two and then to conduct a Yates' computation to arrive at the thirty-two mean effects. The second way is to conduct a Yates' computation on the second design responses and then combine the mean effects with those of the first design. The first method is less time consuming and therefore preferable. The second method gives a clearer picture of the separation of effects and is therefore now given. (The first method is given in Appendix A-1)

Figure 6 shows the Yates' computations on the responses from the second design. The operations are identical to those as described for the first design. It is noted that the signs of all elements involving A and D in the Effects Measured column are now minus since the levels of A and D had been reversed. This reversal of signs permits the separation of effects as shown in Figure 7.

The mean effects derived from the first design are given in the column marked X. Those derived from the second design are given in the column marked Y. An example of the computation to separate the effects is as follows:

The second mean effect in the X column is for variable A. The similar mean effect in the Y column is for minus variable A with the difference in the absolute magnitudes of the mean effects being due to experimental error. Reversing the sign of the column Y mean effect and averaging it with that of the column X mean effect will give the -4.81 mean effect for variable A and certain high order interactions confused with it. $1/2(X+Y)$, -1.36, then gives the mean effect for the remaining high order interactions originally confused with A. Similar calculations are performed to separate all of the other effects, those containing A or D from each of the others. The thirty-one statistics thus obtained, not including the 61.22 (twice the average) figure, are then arranged in order of magnitude without respect to sign and plotted in a thirty-one factor half-normal plot as shown in Figure 8.

It is seen that the error best straight line is now established by twenty-three of the thirty-one points reflecting the greater precision achieved by doubling the experiment and thereby reducing the variance of the estimated effects by one-half. Of the eight points which are clearly off the line, two of them, denoted by asterisks, are combinations of high order interactions. Their relative significance cannot be interpreted within the limits of the experiment. As will be seen, equating them to zero will not affect the results. The BF interaction is not far off the line and its significance may be questioned. Variables E and A are both clearly controlling factors in the efficiency of the zinc electrodes. The signs of their mean effects are both minus, indicating that higher efficiencies can be obtained at the lower levels of the ranges studied, in other words at the lower pressing temperature, 80° F, and with the smaller weight of zinc per plate. The interpretation of variables H, the formation current density, and F, the presence or absence of zinc oxide in the formation electrolyte, is more complex. This results from the probable significance of FH, the interaction between them (CE is very unlikely to be significant due to the low magnitude of C). In general when an interaction is large, as in this case, the corresponding mean effects cease to have much meaning. The effect of F is clearly dependent upon the level of H and vice versa. The three effects F, H and their interaction FH may best be interpreted as a single highly significant effect. Further experimental work at intermediate levels for the two variables is definitely indicated.

The final step in the analysis was to determine if the conclusion was correct that only the effects E, A and the combined effects of H, F and FH were significant, i.e., distinct from experimental error. Part of the purpose of this final step was to determine if the entire experiment was valid, in other words, that there were no large errors made in the actual responses which could have seriously altered the mean effects. The procedure used was to determine the standard error of the individual observed responses by analyzing the thirty-one mean effects. A second standard error, for the differences between observed and predicted responses, was then obtained with the predicted responses based on the assumption that all mean effects other than those for E, A, H, FH and F were indeed zero. If the two standard errors would then be equivalent, both the total experiment and the conclusions derived from it would be proved valid.

Since the error straight line on the final half-normal plot was established by the twenty-three lowest magnitude points, a half-normal plot for these points was prepared. The standard error for the individual observed responses as derived from this plot was 3.0. (See A-5). All mean effects other than the twice the average effect, E, A, H, FH and F were then equated to zero and a reverse Yates' computation was conducted to obtain the thirty-two predicated responses (See A-2). These responses were compared with the observed responses and a list of the thirty-two differences between the observed and predicted responses was prepared (See A-3). The magnitudes of the differences were plotted on normal probability paper. A standard error was obtained from this plot for the difference between individual observed responses and individual predicted responses. The value of this standard error was 3.2 (See A-4). This was in excellent agreement with the standard error of the individual observed responses, thus proving the validity of the experiment and the conclusions derived from it.

In conclusion, a fractional factorial designed experiment, involving two 2^{7-3} fractional designs, has been conducted to determine the significance of seven major variables, and two factor interactions between them, on the discharge efficiency of the dry process zinc battery electrodes. A total of thirty-two electrodes was prepared and tested. Analysis of their responses has indicated the controlling influence of two of the variables, pressing temperature and the amount of zinc per plate, and of the interaction between two other variables relating to the plating conditions under which the zinc material was prepared. The other three variables have been shown to be unimportant in comparison, within the range of levels selected. The experiment has fulfilled its basic purpose, narrowing down the range of variables to permit extensive investigation of the truly important variables in order to arrive at the optimum electrode preparation procedures in the most expeditious manner.

Figure 1

Experiment Variables

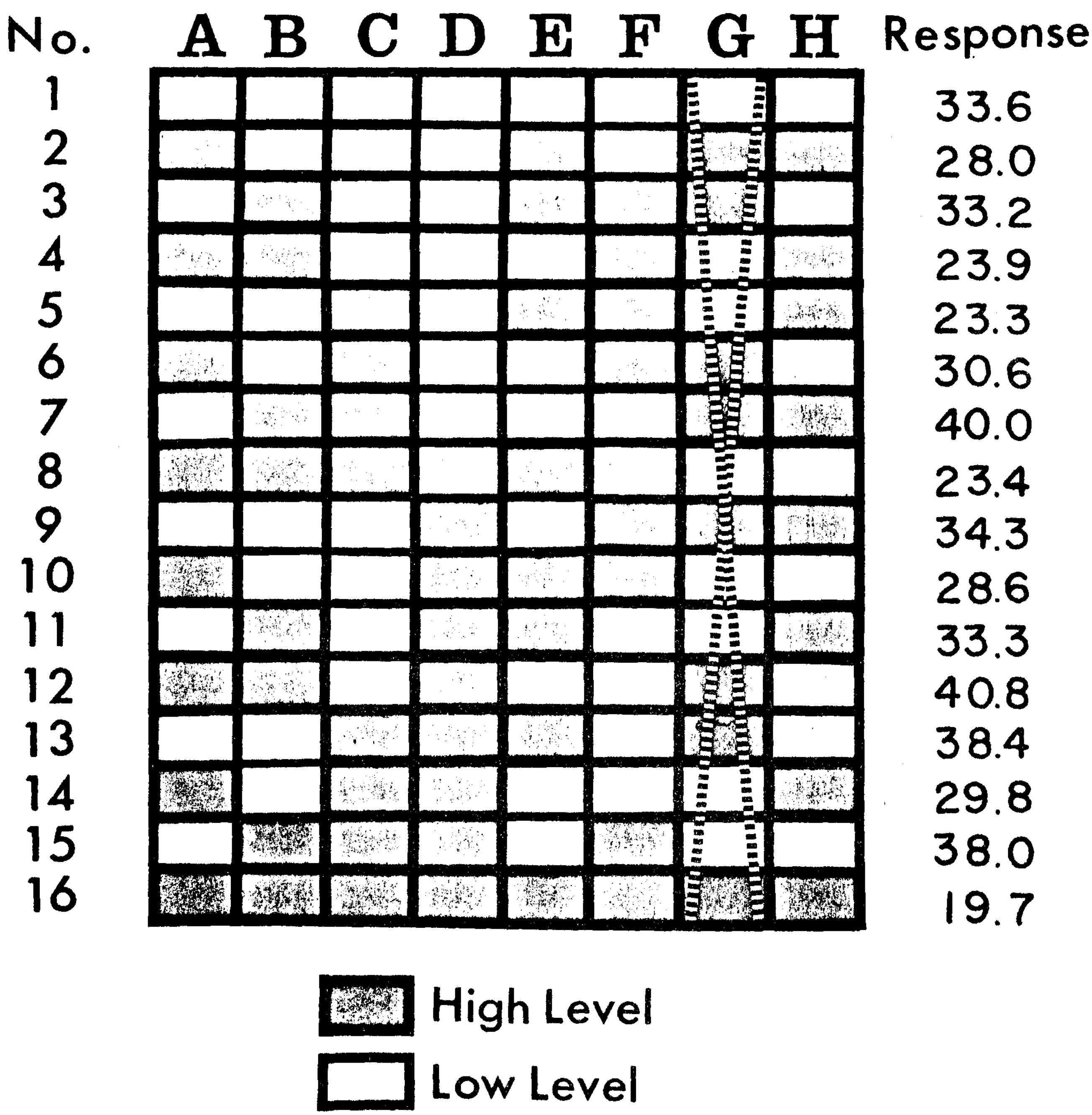
	Pressing Variables	Units	High	Low
A	Zinc weight	grams	5.23	2.62
B	Pressure	psi	1,840	1,230
C	Particle size	sieve mesh	100	200
D	Pressure time	minutes	15	1
E	Pressure temp.	°F	300	80
	Formation Variables			
F	ZnO in electrolyte	gm./liter	20	0
~~G~~	~~Electrolyte temp.~~	~~°F~~	~~100~~	~~80~~
H	Current density	amp./sq. in.	1.0	0.75

This page intentionally left blank.

Figure 2

FIRST DESIGN

Variables



This page intentionally left blank.

FIRST DESIGN ANALYSIS, YATES' ALGORITHM

NO.	RESPONSE	(1)	(2)	(3)	(4)	MEAN EFFECTS (4)/8	EFFECTS MEASURED
1	33.6	61.6	118.7	236.0	498.9	62.363	-
2	28.0	57.1	117.3	262.9	-49.3	-6.163	A
3	33.2	53.9	137.0	-24.2	5.7	0.713	B
4	23.9	63.4	125.9	-25.1	-24.1	-3.013	AB + CE + ~~DG~~ + FH
5	23.3	62.9	-14.9	5.0	-12.5	-1.563	C
6	30.6	74.1	-9.3	0.7	-23.1	-2.888	AC + BE + DH + ~~FG~~
7	40.0	68.2	1.8	-27.6	-7.7	-0.963	AE + BC + DF + ~~GH~~
8	23.4	57.7	-26.9	3.5	-43.1	-5.388	E
9	34.3	-5.6	-4.5	-1.4	26.9	3.363	D
10	28.6	-9.3	9.5	-11.1	-0.9	-0.113	AD + ~~BG~~ + CH + EF
11	33.3	7.3	11.2	5.6	-4.3	-0.538	~~AG~~ + BD + CF + EH
12	40.8	-16.6	-10.5	-28.7	31.1	3.888	~~G~~*
13	38.4	-5.7	-3.7	14.0	-9.7	-1.213	AH + BF + CD + ~~EG~~
14	29.8	7.5	-23.9	-21.7	-34.3	-4.288	H
15	38.0	-8.6	13.2	-20.2	-35.7	-4.463	F
16	19.7	-18.3	-9.7	-22.9	-2.7	-0.338	AF + BH + ~~CG~~ + DE
CHECKS	498.9	449.6	431.2	344.8	315.2	39.396	

449.6
431.2
344.8
315.2
39.400

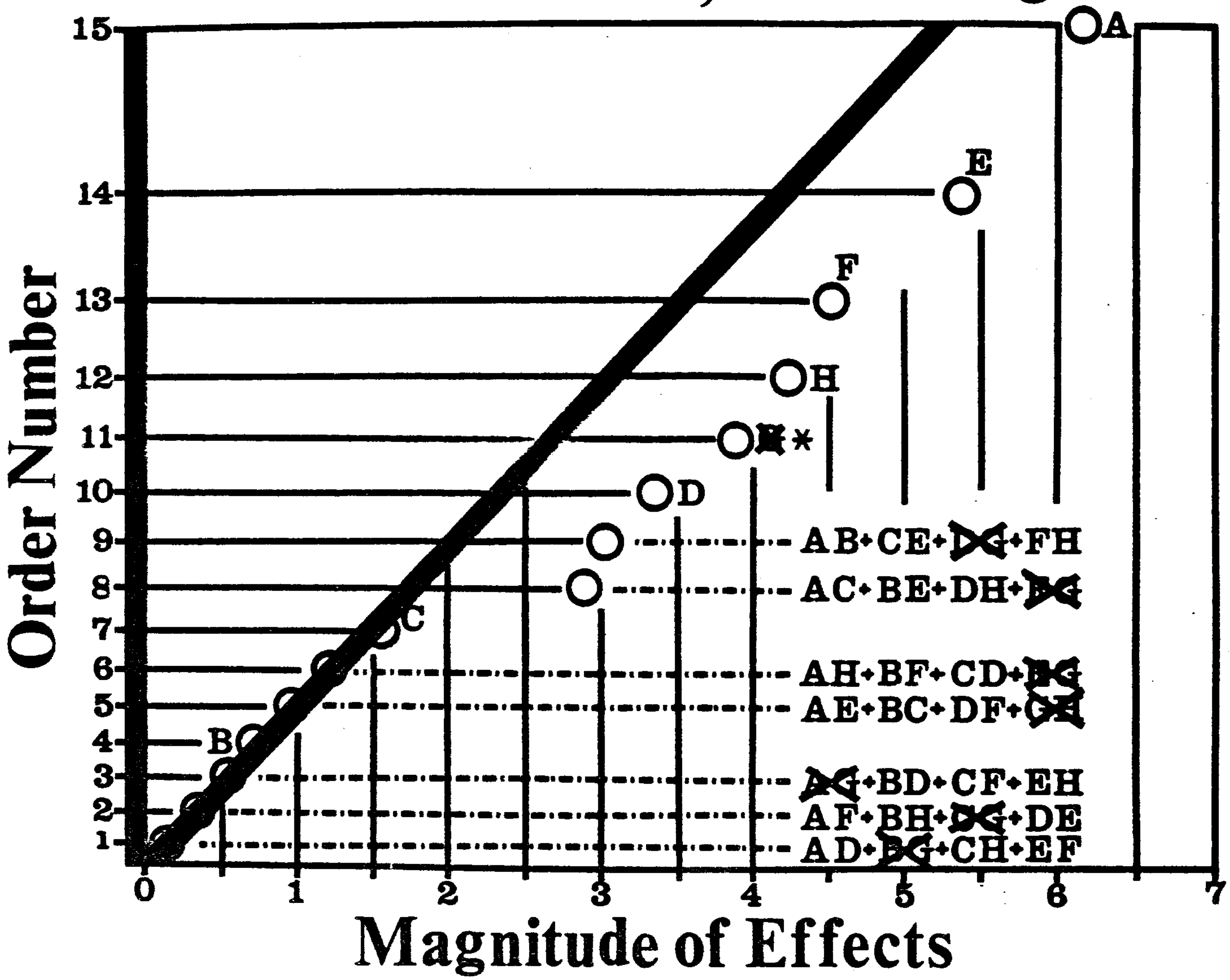
*3 factor and higher order interactions

Figure 3

This page intentionally left blank.

Figure 4

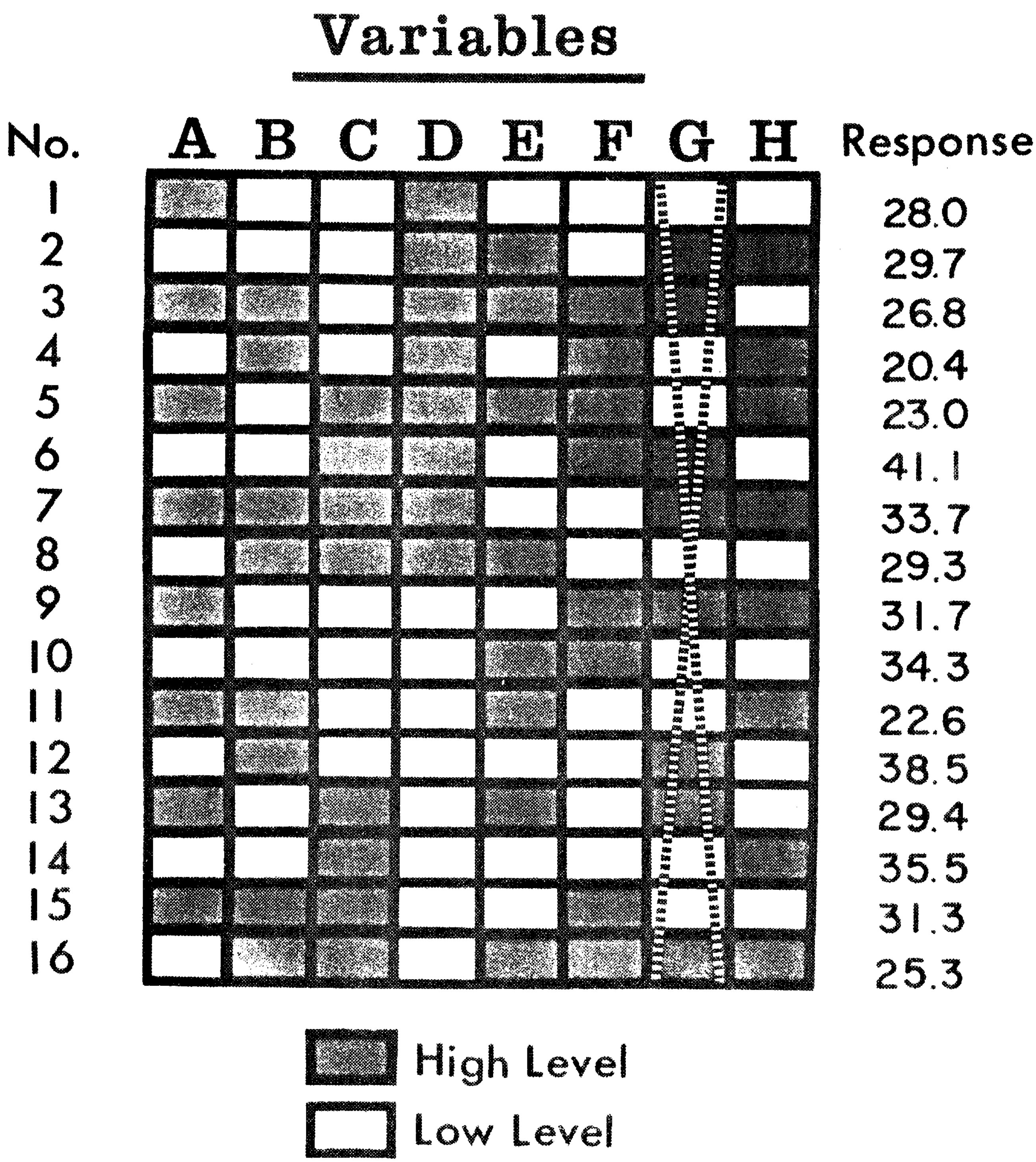
Half-Normal Plot, First Design



This page intentionally left blank.

Figure 5

SECOND DESIGN



This page intentionally left blank.

SECOND DESIGN ANALYSIS, YATES' ALGORITHM

NO.	RESPONSE	(1)	(2)	(3)	(4)	MEAN EFFECTS (4)/8	EFFECTS MEASURED
1	28.0	57.7	104.9	232.0	480.6	60.075	-
2	29.7	47.2	127.1	248.6	27.6	3.450	-A
3	26.8	64.1	127.1	9.0	-24.8	-3.100	B
4	20.4	63.0	121.5	18.6	-29.4	-3.675	-AB + CE - ~~DG~~ + FH
5	23.0	66.0	-4.7	-11.6	16.6	2.075	C
6	41.1	61.1	13.7	-13.2	0.0	0.000	-AC + BE - DH + ~~FG~~
7	33.7	64.9	18.5	-30.6	6.0	0.750	-AE + BC - DF + ~~GH~~
8	29.3	56.6	0.1	1.2	-39.8	-4.975	E
9	31.7	1.7	-10.5	22.2	16.6	2.075	-D
10	34.3	-6.4	-1.1	-5.6	9.6	1.200	AD + ~~BG~~ + CH + EF
11	22.6	18.1	-4.9	18.4	-1.6	-0.200	-~~AG~~ - BD + CF + EH
12	38.5	-4.4	-8.3	-18.4	31.8	3.975	~~G~~*
13	29.4	2.6	-8.1	9.4	-27.8	-3.475	-AH + BF - CD + ~~EG~~
14	35.5	15.9	-22.5	-3.4	-36.8	-4.600	H
15	31.3	6.1	13.3	-14.4	-12.8	-1.600	F
16	25.3	-6.0	-12.1	-25.4	-11.0	-1.375	-AF + BH + ~~CG~~ - DE
CHECKS	480.6	508.2	454.0	436.8	404.8	50.600	

508.2
454.0
436.8
404.8
50.600

*3 factor and higher order interactions

Figure 6

This page intentionally left blank.

DETERMINATION OF EFFECTS, COMBINED DESIGNS

No.	X Mean Effects First Design	Y Mean Effects Second Design	½(X + Y)	Effects Measured	½(X - Y)	Effects Measured
1	62.363	60.075	61.22	-	1.14	Blocks
2	-6.163	3.450	-1.36	*	-4.81	A
3	0.713	-3.100	-1.19	B	1.91	*
4	-3.013	-3.675	-3.34	CE + FH	0.33	AB + ~~DG~~
5	-1.563	2.075	0.26	C	-1.82	*
6	-2.888	0.000	-1.44	BE + ~~FG~~	-1.44	AC + DH
7	-0.963	0.750	-0.11	BC + ~~GH~~	-0.86	AE + DF
8	-5.388	-4.975	-5.18	E	-0.21	*
9	3.363	2.075	2.72	*	0.64	D
10	-0.113	1.200	0.54	AD + ~~BG~~ + CH + EF	-0.66	*
11	-0.538	-0.200	-0.37	CF + EH	-0.17	~~AG~~ + BD
12	3.888	3.975	3.93	~~G~~*	-0.04	*
13	-1.213	-3.475	-2.34	BF + ~~BG~~	1.13	AH + CD
14	-4.288	-4.600	-4.44	H	0.16	*
15	-4.463	-1.600	-3.03	F	-1.43	*
16	-0.338	-1.375	-0.86	BH + ~~CG~~	0.52	AF + DE
			45.01		-5.61	
			-5.61			
			34.90			
			x 8			
			315.20	(Check)		

* 3 factor and higher order interactions

Figure 7

This page intentionally left blank.

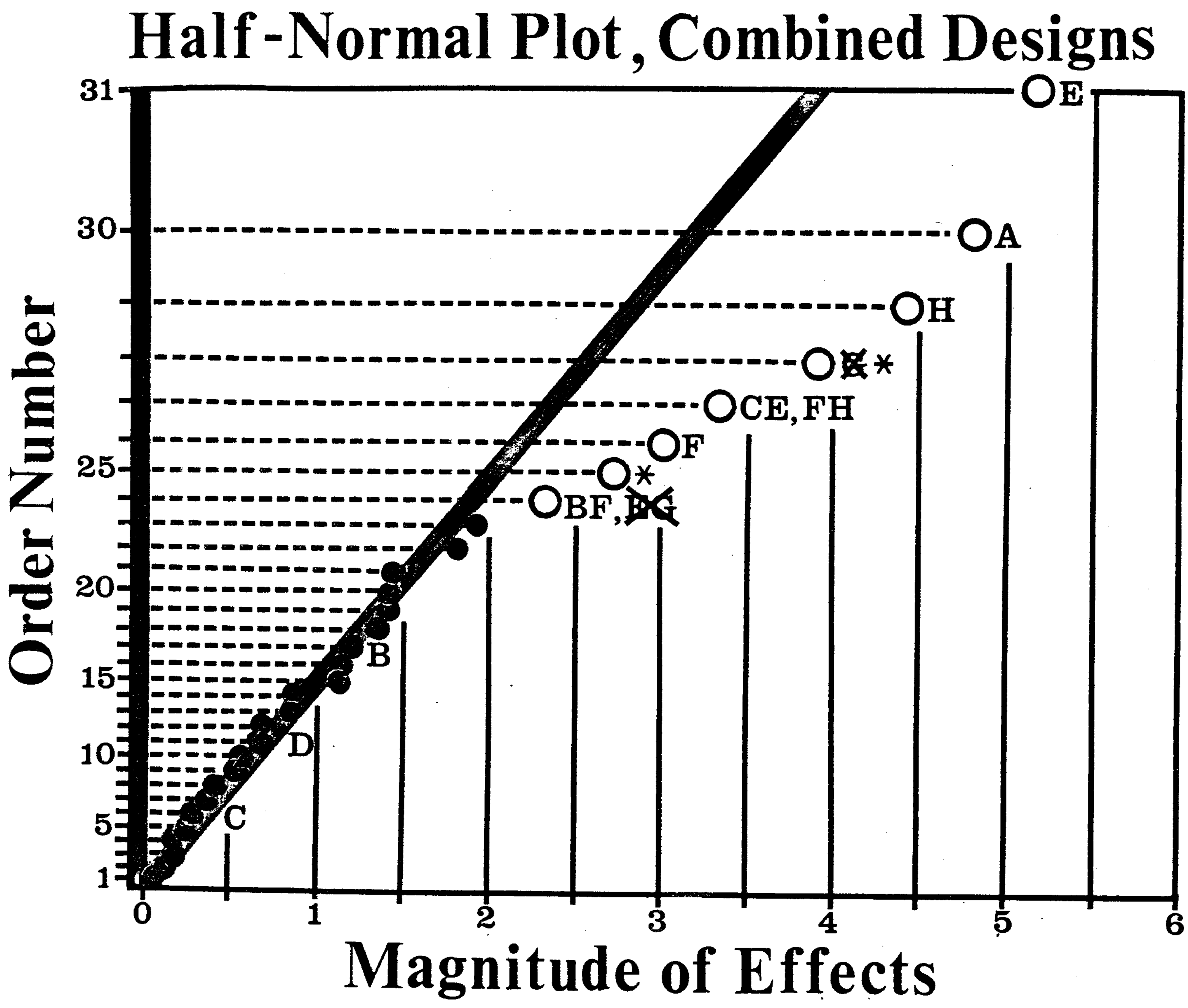

Half-Normal Plot, Combined Designs
Order Number
Magnitude of Effects
31
30
25
20
15
10
5
1
0
1
2
3
4
5
6
E
A
H
*
CE, FH
F
*
BF,
B
D
C

Figure 8

This page intentionally left blank.

APPENDIX

This appendix contains supplemental data as follows:

A-1. Yates' Algorithm computation for thirty-two responses, combination of Designs 1 and 2.

A-2. Reverse Yates' Algorithm computation for thirty-two effects, assuming all effects are zero except those for average, E, A, H, FH and F.

A-3. Comparison of observed and predicted responses.

A-4. Probability plot to obtain standard error of individual differences between observed and predicted responses.

A-5. Half-normal plot of twenty-three mean effects to obtain standard error of individual observed responses.

A-1. Yates' Algorithm computation for thirty-two responses, combination of Designs 1 and 2.

No.	Responses	(1)	(2)	(3)	(4)	(5)	(5)/16	Effects Measured
1	33.6	61.6	118.7	236.0	498.9	979.5	61.22	-
2	28.0	57.1	117.3	262.9	480.6	-21.7	-1.36	*
3	33.2	53.9	137.0	232.0	-49.3	-19.1	-1.19	B
4	23.9	63.4	125.9	248.6	27.6	-53.5	-3.34	CE+FH
5	23.3	62.9	104.9	-24.2	5.7	4.1	0.26	C
6	30.6	74.1	127.1	-25.1	-24.8	-23.1	-1.44	BE+FG
7	40.0	68.2	127.1	9.0	-24.1	-1.7	-0.11	BC+GH
8	23.4	57.7	121.5	18.6	-29.4	-82.9	-5.18	E
9	34.3	57.7	-14.9	5.0	-12.5	43.5	2.72	*
10	28.6	47.2	-9.3	0.7	16.6	8.7	0.54	AD+BG+CH+EF
11	33.3	64.1	1.8	-11.6	-23.1	-5.9	-0.37	CF+EH
12	40.8	63.0	-26.9	-13.2	0	62.9	3.93	G
13	38.4	66.0	-4.7	-27.6	-7.7	-37.5	-2.34	BF+EG
14	29.8	61.1	13.7	3.5	6.0	-71.1	-4.44	H
15	38.0	64.9	18.5	-30.6	-43.1	-48.5	-3.03	F
16	19.7	56.5	0.1	1.2	-39.8	-13.7	-0.86	BH+CG
17	28.0	-5.6	-4.5	-1.4	26.9	-18.3	-1.14	-Blocks
18	29.7	-9.3	9.5	-11.1	16.6	76.9	4.81	-A
19	26.8	7.3	11.2	22.2	-0.9	-30.5	-1.91	-*
20	20.4	-16.6	-10.5	-5.6	9.6	-5.3	-0.33	-AB-DG
21	23.0	-5.7	-10.5	5.6	-4.3	29.1	1.82	-*
22	41.1	7.5	-1.1	-28.7	-1.6	23.1	1.44	-AC-DH
23	33.7	-8.6	-4.9	18.4	31.1	13.7	0.86	-AE-DF
24	29.3	-18.3	-8.3	-18.4	31.8	3.3	0.21	-*
25	31.7	1.7	-3.7	14.0	-9.7	-10.3	-0.64	-D
26	34.3	-6.4	-23.9	-21.7	-27.8	10.5	0.66	-*
27	22.6	18.1	13.2	-9.4	34.3	2.7	0.17	-AG-BD
28	38.5	-4.4	-9.7	-3.4	-36.8	0.7	0.04	-*
29	29.4	2.6	-8.1	-20.2	-35.7	-18.1	-1.13	-AH-CD
30	35.5	15.9	-22.5	-22.9	-12.8	-2.5	-0.16	-*
31	31.3	6.1	13.3	-14.4	-2.7	22.9	1.43	-*
32	25.3	-6.0	-12.1	-25.4	-11.0	-8.3	-0.52	-AF-DE
	979.5	957.8	885.2	781.6	720.0	809.6	50.62	

Checks 957.8
885.2
781.6
720.0
809.6
50.60

A-2. Reverse Yates' Algorithm computation for thirty-two effects assuming all effects are zero except those for average, E, A, H, FH and F.

No.	Effects Measured	(5)	(5')	(6')	(7')	(8')	(9')	(10')	Predicted Response (10')/32
1	1	979.5	979.5	979.5	926.0	1,008.9	1,128.5	1,205.4	37.7
2	-	-21.7	0	53.5	-82.9	-119.6	-76.9	850.6	26.8
3	-	-19.1	0	0	0	-76.9	927.5	1,049.6	32.8
4	CE+FH	-53.5	-53.5	82.9	119.6	0	76.9	812.4	25.4
5	-	4.1	0	0	-76.9	950.1	972.7	800.4	25.0
6	-	-23.1	0	0	0	22.6	-76.9	1,061.6	33.2
7	-	-1.7	0	71.1	0	76.9	889.3	1,170.2	36.6
8	E	-82.9	-82.9	-48.5	0	0	76.9	885.8	27.7
9	-	43.5	0	-76.9	1,033.0	950.1	723.5	966.2	30.2
10	-	8.7	0	0	82.9	-22.6	-76.9	895.8	28.0
11	-	-5.9	0	0	0	-76.9	1,138.5	1,004.4	31.4
12	-	62.9	0	0	-22.6	0	76.9	1,051.6	32.9
13	-	-37.5	0	0	76.9	1,008.9	1,093.3	1,039.6	32.5
14	H	-71.1	-71.1	0	0	119.6	-76.9	1,016.4	31.8
15	F	-48.5	-48.5	0	0	76.9	962.7	1,215.4	38.0
16	-	-13.7	0	0	0	0	76.9	646.6	20.2
17	-	-18.3	0	979.5	1,033.0	843.1	889.3	1,051.6	32.9
18	-A	76.9	76.9	-53.5	82.9	119.6	-76.9	1,040.4	31.4
19	-	-30.5	0	0	0	-76.9	972.7	895.8	28.0
20	-	-5.3	0	-82.9	22.6	0	76.9	966.2	30.2
21	-	29.1	0	0	-76.9	1,115.9	972.5	646.6	20.2
22	-	23.1	0	0	0	-22.6	-76.9	1,215.4	38.0
23	-	13.7	0	-71.1	0	76.9	1,128.5	1,016.4	31.8
24	-	3.3	0	-48.5	0	0	76.9	1,039.6	32.5
25	-	-10.3	0	76.9	926.0	1,115.9	962.7	812.4	25.4
26	-	10.5	0	0	-82.9	22.6	-76.9	1,049.6	32.8
27	-	2.7	0	0	0	-76.9	1,093.3	850.6	26.6
28	-	0.7	0	0	-119.6	0	76.9	1,205.4	37.7
29	-	-18.1	0	0	76.9	843.1	1,138.5	885.8	27.7
30	-	-2.5	0	0	0	-119.6	-76.9	1,170.2	36.6
31	-	22.9	0	0	0	76.9	723.5	1,061.6	33.2
32	-	-8.3	0	0	0	0	76.9	800.4	25.0
			979.5	1,862	3,918	7,836	15,672	31,344	980.0
			-48.5		x2	x2	x2		(979.5)
			931	1,959	7,836	15,672	31,344		
			x2	x2					
			1,862	3,918					

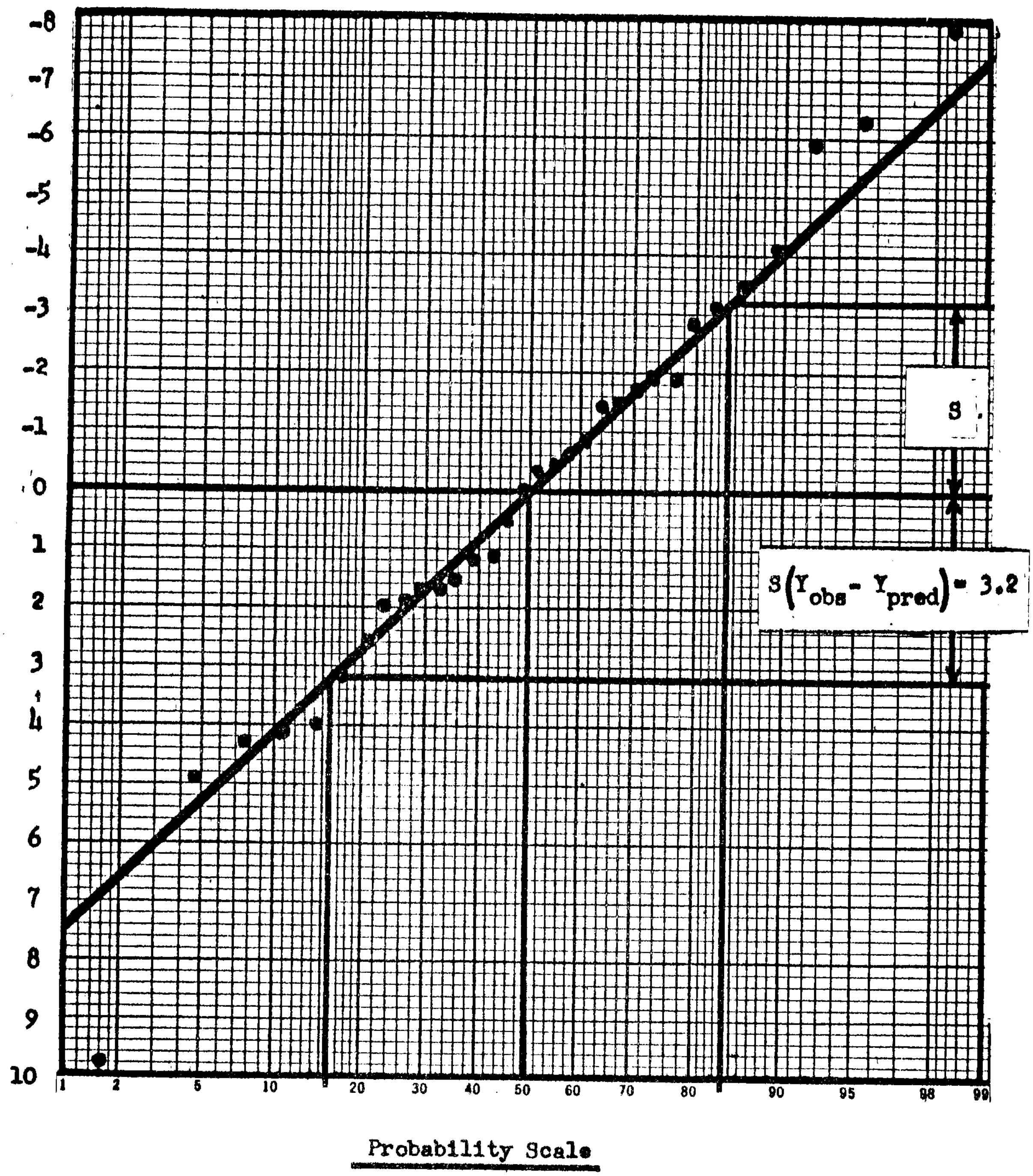
A-3. Comparison of observed and predicted responses.

No.	Observed Responses	Predicted Response	Obs. Resp.- Pred. Resp.	Ordered Series
1	33.6	37.7	-4.1	-9.8
2	28.0	26.6	1.4	-4.9
3	33.2	32.8	0.4	-4.3
4	23.9	25.4	-1.5	-4.1
5	23.3	25.0	-1.7	-4.0
6	30.6	33.2	-2.6	-3.2
7	40.0	36.6	3.4	-2.6
8	23.4	27.7	-4.3	-2.0
9	34.3	30.2	4.1	-1.9
10	28.6	28.0	0.6	-1.7
11	33.3	31.4	1.9	-1.7
12	40.8	32.9	7.9	-1.5
13	38.4	32.5	5.9	-1.2
14	29.8	31.8	-2.0	-1.1
15	38.0	38.0	0	-0.5
16	19.7	20.2	-0.5	0
17	28.0	32.9	-4.9	0.3
18	29.7	31.4	-1.7	0.4
19	26.8	28.0	-1.2	0.6
20	20.4	30.2	-9.8	0.8
21	23.0	20.2	2.8	1.4
22	41.1	38.0	3.1	1.5
23	33.7	31.8	1.9	1.7
24	29.3	32.5	-3.2	1.9
25	31.7	25.4	6.3	1.9
26	34.3	32.8	1.5	2.8
27	22.6	26.6	-4.0	3.1
28	38.5	37.7	0.8	3.4
29	29.4	27.7	1.7	4.1
30	35.5	36.6	-1.1	5.9
31	31.3	33.2	-1.9	6.3
32	25.3	25.0	0.3	7.9
	979.5	980.0	-44.5	
			+44.0	
			-0.5	

This page intentionally left blank.

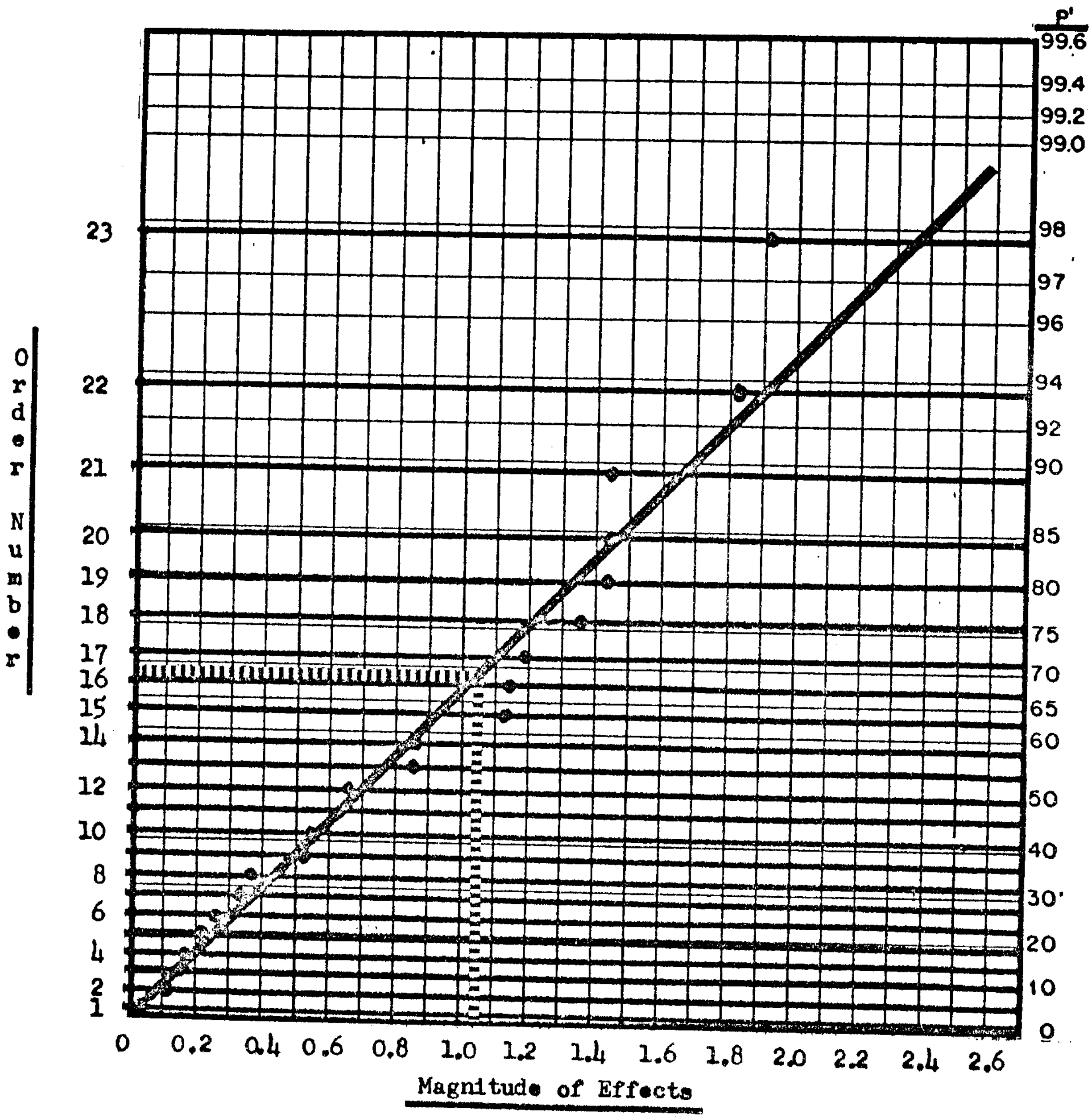
A - 4

Probability plot to obtain standard error of individual differences between observed and predicted responses



This page intentionally left blank.

A - 5 Half-normal plot of twenty-three mean effects to obtain standard error of individual observed responses.



$$S_Y = \frac{1.04}{\sqrt{\frac{1}{16} + \frac{1}{16}}} = 3.0$$

This page intentionally left blank.

Applications of the Calculus of Factorial Arrangements
I. Block and Direct Product Designs

Badrig Kurkjian and Marvin Zelen
Harry Diamond Laboratories
and
Mathematics Research Center, U. S. Army, University of Wisconsin

ABSTRACT

This paper deals with some applications of a general theory for the analysis of factorial experiments as reported by the authors in the June 1962 issue of the Annals of Mathematical Statistics.

General expressions are given for the usual quantities associated with the analysis of variance for the cases where simple treatments or factorial treatment-combinations are applied to Randomized Blocks, Balanced Incomplete Blocks, Group Divisible designs, and a wide class of Kronecker Product designs.

The main point of the new theory is that, for a wide class of the more practical designs, the complete analysis can be carried out almost by inspection of the normal equations, with no requirement for inverting the normal equations.

The complete version of this paper is published in BIOMETRIKA, Vol. 50, Parts 1 and 2, June 1963.

This page intentionally left blank.

STATISTICAL DESIGN OF EXPERIMENTS FOR CONTINUOUS DATA

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

I. INTRODUCTION. I wish to talk on the subject of how one would design an experiment and analyze the data when the results come in the form of a continuous curve, rather than just a single value. This is an area that would appear to have extensive application in science and engineering. For example: velocity data, trajectory data, meterological data, thrust data, etc. As I just mentioned, I am interested in the design of experiments, which means I am not concerned with the evaluation of a single curve, but many curves obtained as a result of testing under several sets of conditions, and very likely each set of conditions will have some replications.

To illustrate what I have in mind, I will use rocket motor thrust curves, although I could have used some other type of curve equally effectively. Now many results from rocket motor tests can easily be analyzed. For example: average exhaust velocity; effective (average) pressure; total impulse; specific impulse, etc. These are simple to analyze because the data for a given test usually comes in the form of one single number. However, if we want to estimate a typical or average thrust curve when a motor is tested under given conditions, this is quite a different problem.

To keep this report unclassified, the thrust data which will be discussed will be completely fictitious. The data is not, to my knowledge, appropriate for any existing rocket motor, but the general shape of the curve is similar to what may be expected for a number of motors currently in use.

The extensive information available in such areas as: Regression Analysis; Random Processes; Power Spectral Analysis; Time Series; Analysis of Covariance; Multivariate Analysis; and similar fields may easily cover the problem I am going to present. Therefore, my first question is, if the solution is readily available in the literature, I would (1)
like some references. My second question is, if it is not readily available but you know the answers, I will appreciate the information. (Please note (2)
that there are numbers along the margins of this report. These numbers refer to specific questions which may be found at the end of this report.)

I will conclude the introduction by stating that the five panel members, W. T. Federer, B. G. Greenberg, M. A. Schneiderman, H. L. Lucas, and H. O. Hartley have all prepared and forwarded their comments. These are included at the end of the report in the order in which they were received.

II. BASIC ASSUMPTIONS.

A. We will begin by assuming that the thrust curves to be considered in this study will look something like the one illustrated in Figure 1, although an actual curve will be somewhat more irregular than the example. The letters along the curve will represent the points which we will consider critical, and we will refer to these points many times in the future.

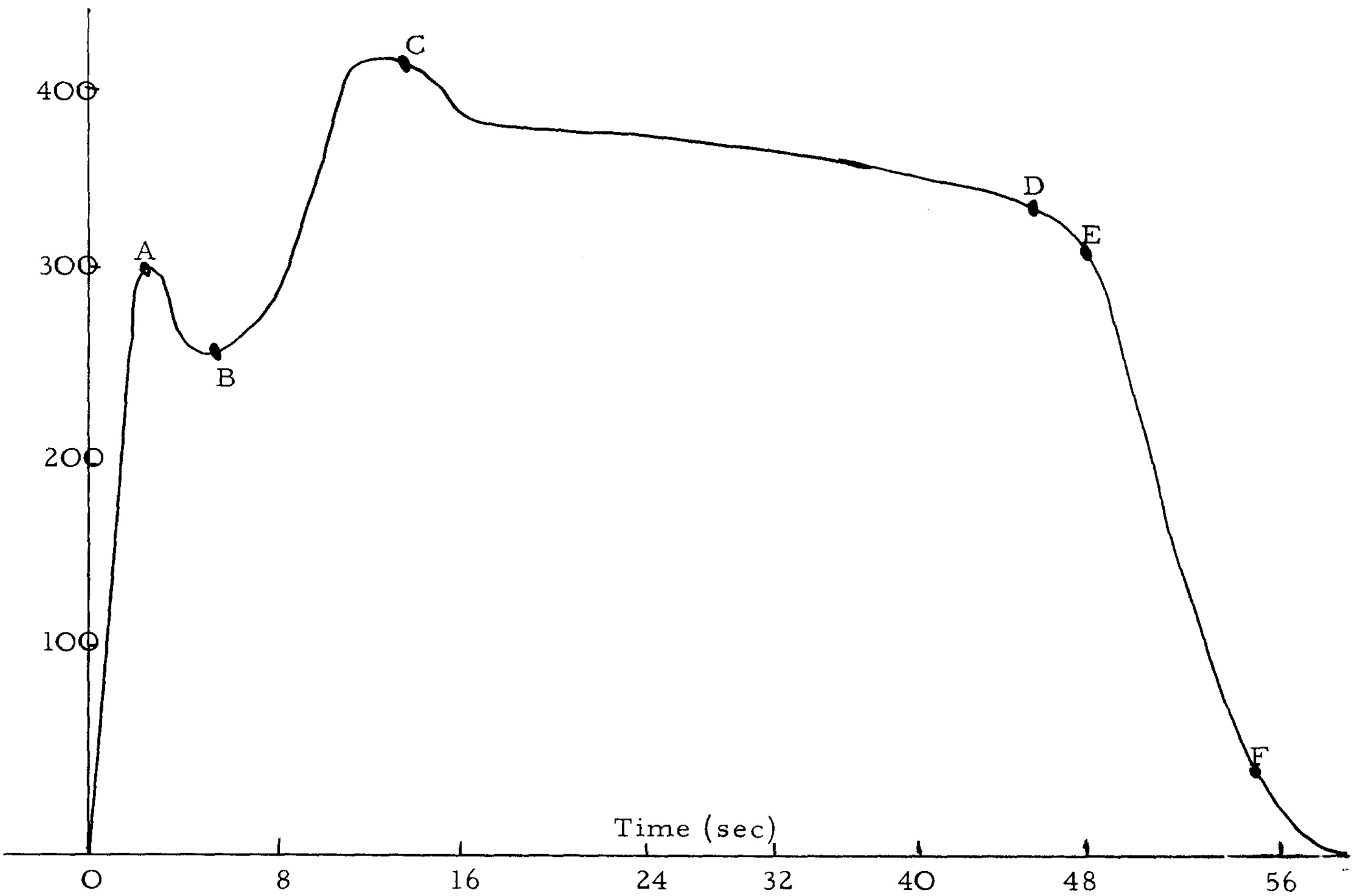


Figure 1

A Typical Thrust Curve for This Study

B. The Second assumption we will make is that the propellant mix from whence the motors are selected as well as the preconditioning temperature can influence the shape of the thrust curve. We will use motors selected from the three propellant mixes (A, B, and C), and conditioned at three temperatures (0^{o}, 50^{o}, and 100^{o} F). We will use 27 motors, nine randomly selected from each mix, and of the nine, three conditioned at each of the temperatures. This arrangement is illustrated in Figure 2.

Mix	Conditioning Temperature		
	0^{o}	50^{o}	100^{o}
A	3	3	3
B	3	3	3
C	3	3	3

Figure 2

Number of motors selected from each mix and conditioned at each temperature.

C. The third assumption is that temperature conditioning will have an effect similar to that illustrated in Figure 3. These three curves actually represent the average of the curves selected from Mix A and conditioned according to the specified temperature.

D. The fourth and final assumption is that the Test Engineer will want the following information:

1. Does Mix difference or temperature conditioning have a significant effect upon the shape of a thrust curve?

2. For a given propellant mix or temperature what is the mean or expected thrust curve?

3. In addition to the mean thrust curve, confidence and tolerance bounds are desired.

III. <u>ANALYSIS OF VARIANCE.</u>

A. If you refer again to Figure 1, you will observe a slightly declining plateau between points C and D. I resolved, first of all, to compare

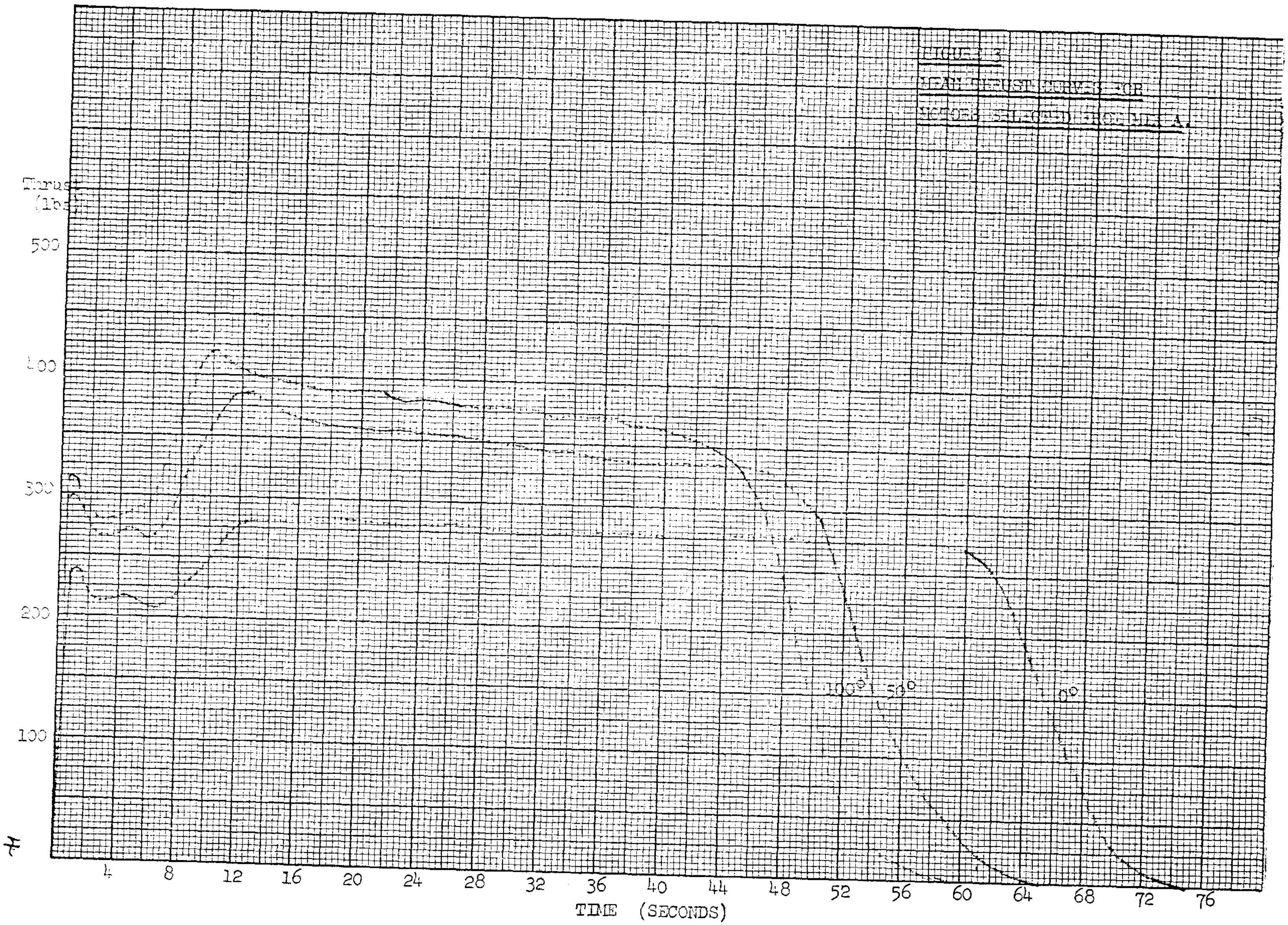


FIGURE 13
MEAN THRUST CURVES FOR
MOTORS SELECTED FROM MIX A.

performance to this plateau area. One problem is evident from Figure 2, namely the plateau areas are of different length at different temperatures. I therefore decided to study only the region from 15 to 42 seconds inclusive. For all 27 rounds, I read the values at 3 second intervals, that is to say (15, 18, ... 42 seconds). The variances appeared to be homogeneous in the region, so we performed the analysis of variance illustrated in Fig. 3.

Source of Variance	d/f	SS	MS	F
Mix	2	43217.71	21608..89	82.72**
Temperature	2	488387.22	244193.61	934.78**
Time	9	12325.61	1369.50	5.24**
(Mix)(Temperature)	4	19794.61	4948.65	18.94**
(Mix)(Time)	18	513.70	28.52	0.11
(Time)(Temperature)	18	2954.70	164.15	0.63
(Mix)(Time)(Temperature)	36	24459.88	679.44	2.60**
Error	180	47022.04	261.23 $=(16.16)^2$	
Total	269	638674.97		

*Significant at .05 level (Single and double asterisk used in later tables
**Significant at .01 level have the same meaning indicated here.)

Figure 4

Analysis of Variance of Thrust Values between 15 and 42 seconds for 27 motors

From Figure 4 it appears that for these curves, temperature has a highly significant effect, mix has an important effect, and there is a significant, downward trend during this time interval. While the mix-temperature interaction is significant, its F value is relatively small; so we will assume it is not really critical. It may also be observed that the pooled estimate of variance is 261.23, and we will therefore assume a standard error of 16 lbs. with 180 degrees of freedom. I then proceeded to construct 9 graphs, one for each combination of temperature and mix. Figure 5 illustrates the graph for Mix B and 50° temperature. I took the sample of three and plotted the mean values for the interval from 15 to 42 seconds. Next, I used the pooled estimate of variance and plotted 95% confidence bounds on this curve, and then on the outside of this, I plotted tolerance bounds, $\gamma = 95\%$, $P = 90\%$.

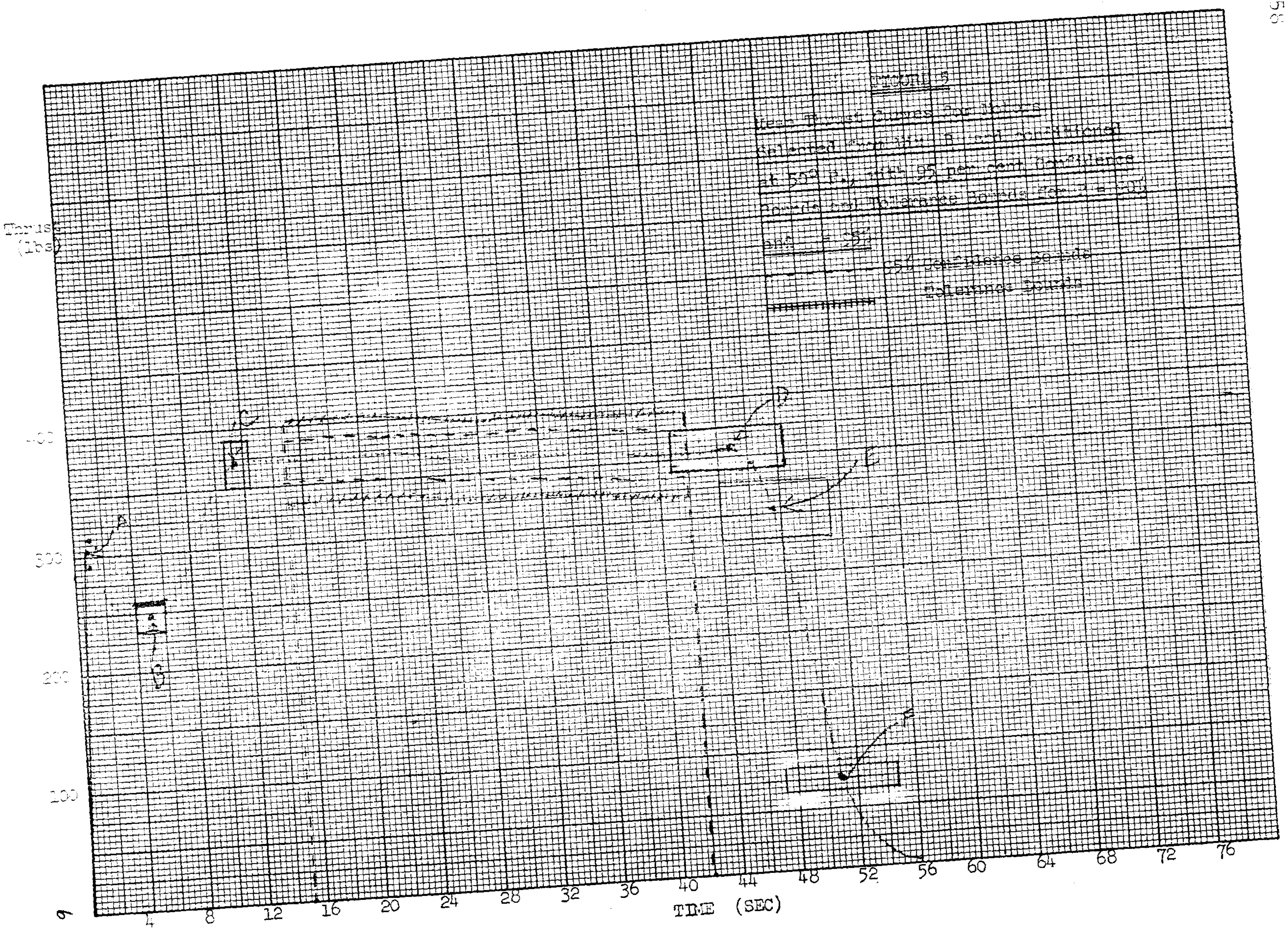


FIGURE 5
Mean Thrust Curves for Motors
Selected from Lot B and Conditioned
at 50° F., with 95 per cent Confidence
Bounds and Tolerance Bounds for P = .90
and γ = .95
95% Confidence Bounds
Tolerance Bounds
Thrust (lbs)
400
300
200
100
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
TIME (SEC)
A
B
C
D
E
F

Clearly, the question at this point is that of the propriety of arbitrarily taking a time curve, observing the values at stated intervals (every three seconds in this case) and considering time, along with mix and temperature,(3) as one of the treatments in the analysis variance. It is interesting that the error term has 180 degrees of freedom. Had we arbitrarily chosen 2 second intervals, for example, instead of 3 second intervals for taking our readings the degrees of freedom for error would have increased to 270. There is clearly something illogical at this point.

B. Referring again to Figure 1, you will note that I have arbitrarily selected six critical points. I then proceeded by performing an analysis of variance for both the X and Y component for each of these critical components. The Analysis of Variance for the Y component of A is given in Fig. 6 and the X component in Figure 7.

Sources of Var	SS	d/f	MS	F
Mix	198	2	99	0.74
Temp	27,746	2	13,873	104.30
Interaction	200	4	50	0.38
Error	2,398	18	133	
Total	30,542	26		

Figure 6

Analysis of Variance for the Y (Thrust in lbs) Component at Point A.

Sources of Var	SS	d/f	MS	F
Mix	.1267	2	.0634	1.80
Temp	.8339	2	.4170	11.88**
Interaction	.3129	4	.1564	4.48*
Error	.6317	18	.0351	
Totals	1.9052	26		

Figure 7

Analysis of Variance for the X (Time in seconds) Component for Point A

From this it may be observed that temperature had a significant effect but mix did not. Returning to Figure 5, the mean value was located for Point A and a confidence rectangle was drawn about it. This procedure was also followed for the other five critical points and these points were then connected. For the time component, temperature had a significant effect for all six critical points, mix at points D, E, and F. For the thrust Component, temperature had a significant effect at all points except F and mix had a significant effect at points B, D, and E.

One will obviously be concerned at this point by the fact that the six critical points were arbitrarily chosen and are not precisely defined. This could easily result in considerable inaccuracy in collecting data for these points. However, this fact will not necessarily be emphasized since it is not really relevant to the basic purpose of this paper.

However, the matter of performing separate analysis of the time and thrust components of each critical point is highly questionable, and I am certain that a procedure applying the bivariate normal distribution would be in order.

Referring either to Figures 1 or 5, it would appear reasonable that if an analysis of variance is appropriate for C-D, then it would probably be equally appropriate for A-C and possibly for E-F. In fact, it would appear more sensible than attempting to locate and evaluate critical points.

IV. REGRESSION ANALYSIS. A second approach, and one which I feel offers more promise is in the area of regression analysis and polynomial fitting. I will discuss a few ideas along this line at the present time.

If you will refer to Figure 1 again, I arbitrarily broke the graph up into four distinct segments. These are: O-A; A-C; C-D; and E-F. Then, for all 27 motors, I fitted the most appropriate polynomial, that is to say, I fitted a cubic to A-C and straight lines to the other three segments.

I will discuss the procedures I followed in analyzing segment C-D and state little more than that analyses were performed on the other segments, and upon completion all segments were plotted until they intersected. Using the values from 15 to 45 seconds and recording the data at 3 second intervals, a straight line was fitted for all 27 sets of data and an analysis of variance was performed for a, b, and r in the equation $Y = a + b.\ (X-30)$. Figure 8 gives the analysis of variance and mean values for a, while Figure 9 gives the analysis of variance and means values for b, (note that mix had no significant effect upon b, so the mean values reflect only temperature). (6) Neither mix nor temperature had any significant effect upon r (the correlation coefficient), but the mean value of the 27 correlation coefficients was 82%.

Sources of Variation	SS	d/f	MS	F
Mix	4411	2	2206	6.28**
Temperature	45959	2	22980	65.42**
Interaction	1959	4	488	1.39
Error	6323	18	351	
Total	58644	26		

Mix	Temperature		
	0°	50°	100°
A	280	348	375
B	292	365	385
C	264	309	376
Ave	278	340	379

Figure 8

The Analysis of Variance and the Mean Values for a, when Fitting the Equation, $Y = a + b \cdot (X-30)$ $15 \text{ sec} \leq x \leq 45 \text{ sec}$ (all means computed from a sample of 3, y = thrust in lbs., x = time in seconds).

Sources of Variation	SS	d/f	MS	F
Mix	0.618	2	0.319	1.040
Temperature	3.689	2	1.844	6.209**
Interaction	0.967	4	0.242	.815
Error	5.340	18	0.297	
Total	10.644	26		

Temperature	0°	50°	100°
Mean Value b	-.410	-.881	-1.315

Figure 9

The Analysis of Variance and Mean Values for b, obtained from fitting the equation $Y = a + b \cdot (X-30)$, 15 sec. $\leq x \leq$ 45 sec. (all means computed from a sample of 9)

The data in Figure 9 indicates that "b" increases almost linearily with temperature. In fact, the formula $b = -.410 - (.00905)$ temp, might serve as a guide for selecting "b" in the region $0^{\circ} \leq \text{temp} \leq 100^{\circ}$. If one desires a formula for estimating "a" in the region for $0^{\circ} \leq \text{temp} \leq 100^{\circ}$, he might try the formula: $a = 278 + 1.4 \cdot (\text{temp}) - .0041(\text{temp})^2$ which averages out the mix effect.

Again referring to Figures 8 and 9, one may observe that the standard error for "a" is 18.7 lbs. with 18 degrees of freedom, and the standard error for "b" is 0.545, also with 18 degrees of freedom.

In addition to this, each time a line is fitted by least squares, it is possible to obtain a standard error of estimates and standard errors for "a" and "b". For the 27 curves, I pooled these standard errors and obtain the following results:

Pooled Standard Errors of estimates:
3.40 lbs. with 243 d/f.

Pooled Standard Error for "b":
0.324 with 243 d/f.

Pooled Standard Error for "a":
1.025 with 243 d/f.

As may be expected, the estimates for the standard error for both "a" and "b" are larger in Figures 8 and 9 than the pooled estimates listed above. This is reasonable since the estimates in Figures 8 and 9 include the dispersions that exist among curves from the same lot and conditioned at the same temperature, while the pooled estimates reflect the variation within only a single curve.

Inasmuch as I am attempting to classify any curves which come from a given mix and a given temperature, it would seem more appropriate to use the estimates of variability in Figures 8 and 9. One other argument for this lies in the fact that when the standard error of estimates was computed for each of the 27 curves, it was computed from 11 points, selected from the thrust curves at 3 second intervals. 15 sec $\leq$ time $\leq$ 45 sec. Again the question arises concerning the arbitrariness in choosing 3 second intervals instead of some other intervals.

Now confidence bounds for a regression line at a point x_i may be computed from the formula

$$\bar{Y}_i - t \cdot S(\bar{Y}_i) \leq E(Y_i) \leq \bar{Y}_i + t \cdot S(\bar{Y}_i)$$

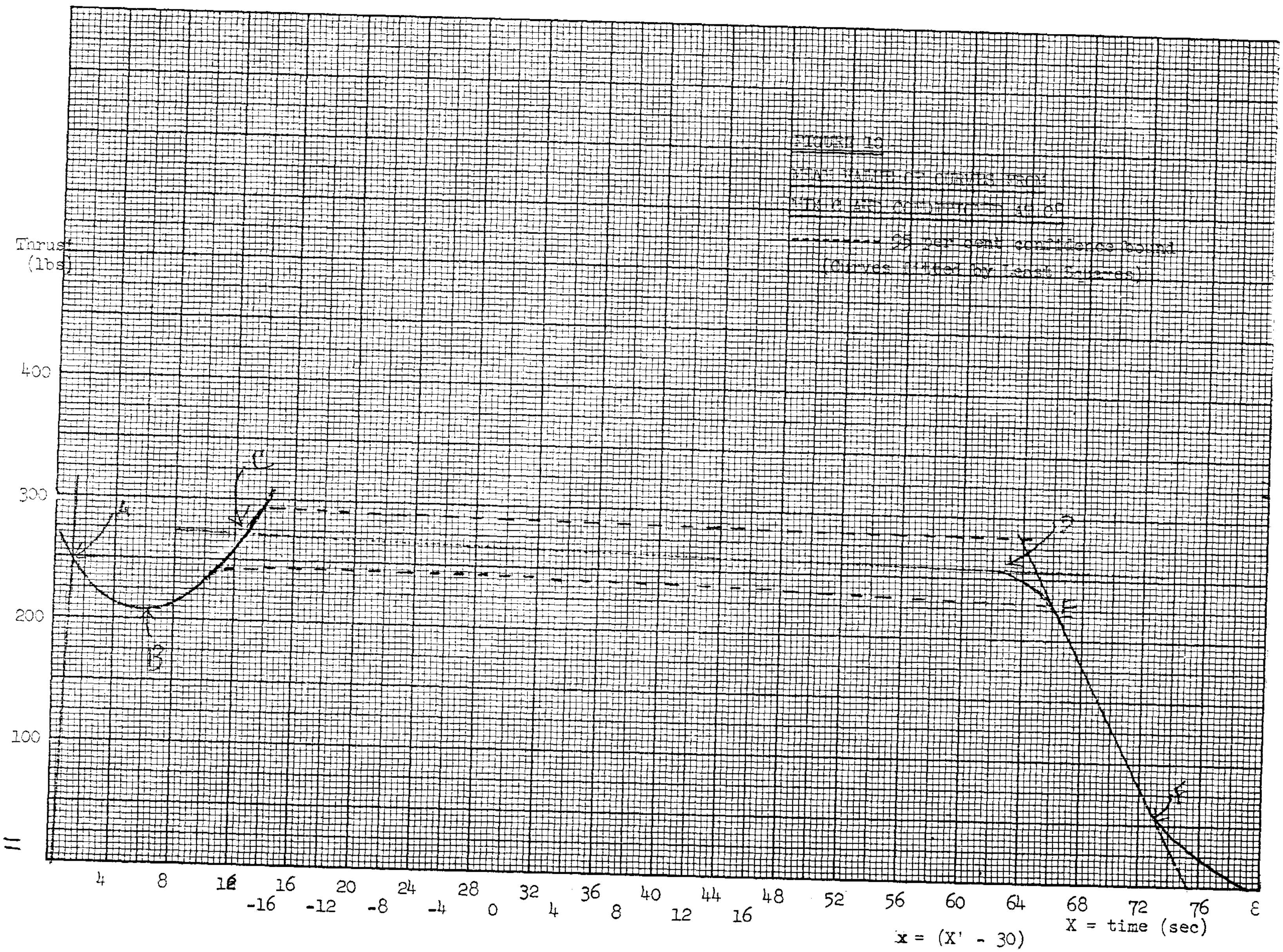
where $\bar{Y}_i = a + b\,x_i$

and $S^2(\bar{Y}_i) = S^2(a) + x_i^2 \cdot S^2(b);\ \bar{x}_i = 0.$

Since each "a" represents an average of 3 numbers and each "b" represents an average of 9 numbers, we have:

$$S^2(\bar{Y}_i) = \frac{351}{3} + x_i^2 \cdot \frac{.297}{9} = 117 + .033\,x_i^2\ .$$

These ideas are illustrated in Figure 10, in which we consider Mix C and a temperature of 0°, and fit the curve from 15 sec $\leq X \leq$ 45 sec. (Segment C-D). This curve is given by $\bar{Y} = 264 - .410\,X$. or $\bar{Y} = 264 - .410 \cdot (X-30)$,


per cent confidence bound
Thrust
(lbs)
400
300
200
100
A
B
C
D
E
F
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
-16
-12
-8
-4
0
4
8
12
16
x = (X' - 30)
X = time (sec)

and the variance $S^2(\bar{Y}) = 117 + .033\,X_i^2$ was used to compute a confidence bound about the curve. Incidentally, the confidence bound in Figure 10 is very close to the one illustrated in Figure 5.

The procedures for fitting the segments (O-A) and (E-F) could be quite similar to that of fitting the segment (C-D). In fact the segment (O-A) should be even simpler. To fit the segment (A-C), it is suggested that a cubic equation be fitted, using data points at one second intervals. It is further suggested that orthogonal polynomials be used when fitting a cubic or higher degree equation to simplify the process of obtaining the variance(7) and confidence bounds.

CONCLUSIONS. Frequently it is desired to design an experiment when the results of the test are a continuous curve rather than a single quantitative value. Scientists and Engineers frequently want to know whether certain levels of a given treatment will have a significant effect upon the curve obtained, and what will an average or expected curve be for a given set of conditions.

I have made a few suggestions based largely upon analysis of variance or regression analysis. I will greatly appreciate comments on the proposed solutions, but more important, I would like suggestions for better approaches to the problem.

QUESTIONS.

1. Has the problem of designing an experiment when the results come as a continuous curve rather than a single value ever been solved? If so, are useful references available?

2. Do you have any ideas of additional approaches beside those suggested in the paper?

3. When studying a section of the curve such as C-D, is there any justification in arbitrarily selecting a set of times between C and D, computing the thrust at each of these times for all available curvex, and performing an analysis of variance similar to that given by Figure 4? Perhaps this would be in order with certain changes in procedure.

4. What procedures would you suggest when attempting to locate a point in terms of both its X and Y components and then obtaining both a confidence and tolerance region about this point?

5. Is the analysis of variance for a, b, and r, as illustrated in Figures 8 and 9 appropriate?

6. Have you any suggestions regarding the validity of the techniques, using regression analysis, that were discussed in this section?

COMMENTS ON PRESENTATION BY PAUL COX

Walter T. Federer
Mathematics Research Center, United States Army
University of Wisconsin, Madison, Wisconsin

The paper presented by Mr. Cox is written in a somewhat provocative manner. I appreciate this style of presentation as it affords the Panel ample opportunity to illustrate several statistical points.

The first point I wish to make relates to the definition and use of terms in current statistical literature. There is a tendency in statistical literature for vague and imprecise usage of such terms as the design of experiments, analysis of variance, error rate, etc. It is instructive and useful to define and to use words or phrases in a specified manner. Any departure from specificity should be described. Personally, I would prefer to use definitions of the following form:

i) Experimental design (or experiment design) - The arrangement of the observations in the experimental area or space or the procedure for obtaining the observations in an experiment.

ii) Treatment design - The arrangement or selection of treatments for the experiment (e.g., the selection of levels and combinations of factors in factorial experiments, etc.)

iii) Determination of sample size - The number of observations necessary to achieve a prescribed objective. (Authors of some ranking procedures papers refer to the determination of numbers of observations as the design of experiment rather than as the determination of sample size.)

iv) Analysis of variance - The partitioning of the sum of squares into component parts. (One segment of statistical literature utilizes the term analysis of variance to be synonymous with an F test while another segment utilizes this term to refer to the estimation of variance components and so it goes.)

v) Analysis of experimental data - This term includes the last above but not vice versa. It refers to all statistical computations relevant to a set of experimental data. An analysis of experimental data refers to the reduction of data to summary form and is useful in, but does not replace, the interpretation of experimental results. The interpretation of statistical results must be made in light of the objectives, conditions, and related circumstances of the experimental results.

vi) Significance level - Type I error = size of the test = α, have all been used to refer to the same thing but unfortunately nothing is said about the base for computing "α".

vii) Valid estimate of the error variance - Fisher has defined this term but unfortunately many statistical writers by-pass this important concept with the phrase "given that σ^2 is the error variance." In much of experimentation the definition of error variance cannot be so glibly by-passed, but requires a thorough knowledge of the experimental conditions.

We could go on with other terms but now let us return to Mr. Cox's paper. The title of the paper is "Statistical Design of Experiment for Continuous Data"; it deals only with the analysis of experimental results with no reference either to the experimental or treatment design as defined above. Mr. Hartley has discussed some considerations to be given to the treatment design for experiments with specified objectives. Mr. Lucas will, I hope, make some comments about the actual experimental design used in this study and illustrate where confounding has taken place. Mr. Cox's paper is concerned with what to do with a set of data and not with how to obtain the data. He has raised a number of questions but rather than address myself to the specific question I prefer to proceed in another manner which, I hope, will furnish answers to or illustrate the relevance of the questions.

As Messrs. Grubbs, Greenberg, Hartley, and Schneiderman have already stressed we must first set up a Mathematical Model for the data which will be consistent with the experimental and treatment designs and with the nature and objectives of the experiment. For example, let us suppose that thrust = y, may be characterized by the following:

$$y = f(\underline{\epsilon}, \underline{t}, \underline{\theta})$$

where the response variable y is a function of error components denoted by the vector $\underline{\epsilon}$, of time components denoted by the vector $\underline{t}$, and of a set of parameters denoted by the vector $\underline{\theta}$. Our first job then is to define to nature of the function. If we are totally ignorant of the response curve then we could use a form of polynomial regression as follows:

$$E(y) = \sum_{i=o}^{b} \beta_i t^i$$

where β_i is the i^{th} regression coefficient and t the time variable. After we are satisfied that a suitable mathematical formulation of the problem has been made, the parameters of the response curve are estimated. The analysis of the estimates may be made using the results of R. A. Fisher (Jour. Agric. Sci. 11:107, 1921 and Phil. Trans. Roy. Soc. B, 213:89, 1925) and others. Also, multivariate analysis procedures may be pursued for summarizing the results for many estimates of a set of parameters. For example, if it is desired to discriminate between response curves, then an a priori or an a posteriori (These terms are not reserved solely for use by Bayesians.) weighting of coeffients in the discriminate function may be utilized.

As a part of the characterization of the model and of the problem it should be determined if the total response curve segements of the total curve, or specified points (e.g. points of inflection) on the curve are of interest. After this has been specified then the statistician proceeds with the estimation problems. Haziness on form or type of response desired leads to a confusion of issues.

One specific question raised by Mr. Cox related to the sample size N for response curves for continuous data. Now if the data are truly continuous N = infinity, but we all know that the recording machine records an impulse over a measurable period of time, say one-tenth of a second. In any event N is very large. Several of the previous Panel speakers have discussed the non-independence of two successive impulses or recordings by a recording machine. However, I wonder about the relevance of this since we use, or should use, these values only to estimate the parameters in the response curve. This procedure is, or should be, repeated for many response curves and the variation among response curves treated alike forms a basis for the variances and covariances among the estimates of parameters where each response curve represents but one observation.

At this point I do not see the importance of obtaining a variance of a single response curve. However, if such is desired, then as an approximation I would suggest segmentation of the total curves into small segments of time where small is such that the estimates are relatively unaffected by smaller segmentation. Course groupings could affect the results considerably. Some account may need to be taken of the relationship among adjoining segments as described by Messrs. Greenberg and Hartley.

The response curves presented in the paper bother me somewhat. Frankly, I believe (i) that the curves in Figure 3 are not very fictitious, (ii) that the area under each curve is relatively constant from the conservation of mass theory, (iii) that a heart-to-heart talk with the physicists and engineers would do much to simplify the nature of the problem, and (iv) that maybe Mr. Cox should be considering acceleration = z instead of thrust = y.

Summed up this means that I would want some education in this area before any analyses would be performed on thrust or any other data. It may be possible to reparameterize the problem by using a function of the time variable instead of the time variable itself. Some simple function such as log t might suffice.

COMMENTS ON PRESENTATION BY PAUL COX

Bernard G. Greenberg
School of Public Health
The University of North Carolina
Chapel Hill, N. C.

As the first discussant, my main concern will be with what I consider the the most important aspect of this problem. It involves the question of the basic underlying model and what the purposes of the investigator were in designing this experiment. As our chairman has just pointed out, this is our primary consideration.

This example illustrates what I think is a truism in statistical design and experimentation, viz. that no amount of statistical knowledge and methodology, however great, is a sufficient substitute of substantive knowlege of the field of application and in being able to discuss with the investigator the important questions. For instance, Mr. Cox has indicated that his six points were arbitrarily selected and the analysis of variances relating thereto were not necessarily meaningful. This may be so in these six points but does the person who has the expert knowledge of thrust curves tend to associate important meanings to them, or any other set of critical points. It may very well be that point A has direct application to the understanding of the model underlying thrust curves and that the actual value of A, the time elapsed from the origin, or the rate of ascent from the origin to A are of primary concern to the rocket motor designer.

This is best illustrated from examples in my own field of biology and medicine. We have similar kinds of data which may not be as closely continuous as the thrust curve, but the points of observation in the time series are spaced close enough together that we treat them as such. Thus, we have tracings for electrocardiograms, growth curves, and epidemic curves. In the electrocardiogram, the distance and regularity between two waves is extremely important to the cardiologist and deviations from normalcy are based upon this pattern. In an epidemic curve, we are interested in the length of time between peaks such as in the periodicity of measles. We are also interested in the amplitude of the waves in order to know expected numbers of cases.

In other words, without discussing with the engineer of the rocket motor thrust curve what his problem is, it is difficult for me to know how to handle best these data. It may be the measurement of the critical points, the length of time until each maximum is attained, or even the integral summing up the total amount of thrust.

If one is convinced that he does want to fit a mathematical equation to this curve, or segments of it, a nasty problem arises because of the time series nature of the data. The observations in time are not independent and the residuals may be autocorrelated. This is not a new problem. The econometricians have been considering this problem for years in their analysis of time series.

I might suggest one reference to Mr. Cox in this connection which may help in fitting equations to segments of the curve and the analysis thereof. A paper by Elston and Grizzle (Biometrics, Vol. 18, No. 2, June 1962, pp. 148-159.) considers the various ways of estimating time-response curves and the ideas in that paper should prove helpful in this problem.

COMMENTS ON PRESENTATION BY PAUL COX

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

The panelists so far have covered just about everything that I had in mind. I certainly agree in the main with the comments they have made regarding just what particular points on the observed curve or what particular function of the observations may be of interest. Also, I agree with the comments regarding the desirability of fitting a "rational" model, which presumably can be supplied, at least in approximate form, by the engineers. I wish, however to expand on a very important point.

Many of the remarks of panelists about design and analysis have been engendered by the existence of "noise" along the curve for an individual motor and the probable lack of independence of successive observations. I wish to emphasize that there is another, and probably much more important, "noise" component involved. The latter arises from the fact that a group of motors which are constructed and treated alike, insofar as can be managed, will, nevertheless, have inherently somewhat different curves. That is, there is "between-motor" noise as well as "within-motor" (along-the-curve) noise. The existence of between-motor noise must be taken into account for proper experiment design and analysis.

It is instructive to formalize the situation in a way which encompases the two noise components. For the jth experimental unit (here the motor, but in other cases a machine or an animal, etc.) on the ith treatment, we can write the model,

$$y_{ij}(t) = \phi(t;\ \underline{\theta}_{ij}) + \epsilon_{ij}(t) \qquad (1)$$

where

$y_{ij}(t)$ = observed time curve for the unit

$\phi(t;\underline{\theta}_{ij})$ = "true" time curve for the unit

$\underline{\theta}_{ij}$ = vector of parameters for the unit

$\epsilon_{ij}(t)$ = "within-unit" noise, For the jth unit on the ith treatment, we next write

$$(2) \qquad \underline{\theta}_{ij} = \underline{\theta}_i^* + \underline{\delta}_{ij}$$

where

$\underline{\theta}_i^*$ = expected value of $\underline{\theta}_{ij}$ for units on the ith treatment

$\underline{\delta}_{ij}$ = "between-unit" noise.

Substituting (2) into (1) yields the model desired, namely,

$$(3) \qquad y_{ij}(t) = \phi\left[t; (\underline{\theta}_i^* + \underline{\delta}_{ij})\right] + \epsilon_{ij}(t),$$

Suppose we compute $\underline{\theta}_{ij}$, an estimate of $\underline{\theta}_{ij}$, for each unit. We see that

$$(4) \qquad \hat{\underline{\theta}}_{ij} = \underline{\theta}_{ij} + \underline{\eta}_{ij}$$

where

$\underline{\eta}_{ij} = \underline{\eta}\left[t; \underline{\theta}_{ij}; \epsilon_{ij}(t)\right]$, a vector of errors with which $\underline{\theta}_{ij}$ is estimated; these stem from "within-unit" noise.

We are interested, however, in estimating $\underline{\theta}_i^*$. The relation of $\hat{\underline{\theta}}_{ij}$ to $\underline{\theta}_i^*$ can be seen by substituting (2) into (4) to obtain

$$(5) \qquad \begin{aligned} \hat{\underline{\theta}}_{ij} &= \underline{\theta}_i^* + \underline{\delta}_{ij} + \underline{\eta}_{ij} \\ &= \underline{\theta}_i^* + \underline{\delta}_{ij}\,. \end{aligned}$$

Note that $\underline{\delta}_{ij}^* = \underline{\delta}_{ij} + \underline{\eta}_{ij}$ is the total noise or error in $\hat{\underline{\theta}}_{ij}$ and stems from both "between" and "within" noise.

In view of the development just completed, it is certainly reasonable first to estimate $\underline{\theta}_{ij}$ for each individual unit and then as a second step, to analyze the $\hat{\underline{\theta}}_{ij}$ according as the experimental design dictates. Since $\hat{\underline{\theta}}_{ij}$ is a vector, multivariate methods may be desired. Note that the procedure is a "robust" one.

Some papers in which the "robust" approach has been employed are [1], [3], [4], [5], [6].

In view of the remarks of some of the other panelists about choice of points along the time curve and about correlation between successive observations along the curve, the following comments seem in order. In my experience, the contribution of the "between" noise, $\underline{\delta}_{ij}$, to the variance of $\hat{\underline{\theta}}_{ij}$ as an estimate of $\underline{\theta}_i^*$ is dominant over the contribution of the "within" noise as summed up in $\underline{\eta}_{ij}$. In fact, in some instances, the "between" noise, $\underline{\delta}_{ij}$, is large relative to the "within" noise, $\epsilon_{ij}(t)$, itself; in this event, the contribution of $\underline{\eta}_{ij}$ is negligible. With $\underline{\delta}_{ij}$ dominant over $\underline{\eta}_{ij}$, it is clear that one need not worry much about the correlation between successive observations on the same unit, that any reasonable method of computing $\hat{\underline{\theta}}_{ij}$ will do, and that one needs use only the minimum number of points along the ij^{th} curve consistent with the complexity of ϕ and the obtaining of moderately efficient estimates of $\underline{\theta}_{ij}$.

This leads next to the design problem, a matter which has been discussed by the other panelists primarily from the standpoint of selecting points along the time curve. In view of my foregoing remarks, I cannot see that the pattern for selection of points along the time curve is the really critical matter, just as long as the pattern is a reasonable one. Instead, the important question is how to select an optimum set of treatment combinations.

To comment further about the design problem, it is again advantageous to be somewhat formal. We note that $\underline{\theta}_i^*$ is a function of the levels of the treatment variables (here, temperature and mixture); i.e.,

$$\underline{\theta}_i^* = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha}) \tag{6}$$

where

γ = a vector of functions of the vectors $\underline{x}_i$ and $\underline{\alpha}$

$\underline{x}_i$ = the vector of levels of the treatment variables characterizing the i^{th} treatment;

$\underline{\alpha}$ = a vector of parameters which depends on basic invariants and on the levels maintained for treatment-type factors not under study (i.e., factors held constant over all i).

Substituting (6) into (5) yields

$$\underline{\hat{\theta}}_{ij} = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha}) + \underline{\delta}_{ij}^* . \tag{7}$$

Now, if the functional forms represented by $\underline{\gamma}$ are known, the problem is to select a minimum optimal set of x-vectors such that all elements of $\underline{\alpha}$ can be estimated and that the estimate, $\underline{\hat{\alpha}}$, is "best" in a suitable sense. In general the optimum design depends on $\underline{\alpha}$, but, since $\underline{\alpha}$ is unknown, one must use previous estimates (or best guesses) about $\underline{\alpha}$ in order to arrive at a good design. Some ideas about this problem are given in [2]. If the forms of the functions, $\underline{\gamma}$, are subject to question, the design must have extra x-vectors so that tests about the assumed $\underline{\gamma}$ and insight about improvements can be obtained. The latter point is also discussed briefly in [2].

I have finished the main things I want to say. There are, however, a couple of other matters that come to mind.

The first has to do essentially with what function of ϕ and hence of $y_{ij}(t)$ is really of concern to the investigator. Although, in some instances, only a particular univariate function of ϕ may ever be of interest, my experience indicates that this is not generally true. I suggest, therefore, that ordinarily it will be best to study ϕ; i.e., to fit the parameters, $\underline{\theta}_i^*$, or more basically, $\underline{\alpha}$. Given such fits, anything desired can be ascertained.

Finally, in the first analysis Mr. Cox outlined, he failed to distinguish "between" and "within" noise. The variance sources for his analysis were

Treatment

Time

Time by treatment

Residual.

They should have been

Treatment

Motor within treatment (Error for treatment; corresponds to δ^*_{ij}

Time

Time by treatment

Time by motor within treatment (Error for time and time by treatment; corësponds to ϵ_{ij}).

In closing, I should note that Mr. Cox, in all but his first analysis, adopted the "robust" approach. I stress the approach, however, because it is important, and because judging from his first analysis, Mr. Cox appeared not to be very clear on the implications of the existence of both "between" and "within" noise.

REFERENCES

[1] Box, G.E.P. 1950. Problems in the Analysis of Growth and Wear Curves. Biometrics 6, No. 4, 362-389.

[2] Box, G.E.P. and H.L. Lucas. 1959. Design of experiments in non-linear situations. Biometrika 46: 77-90.

[3] Elston, R.C. and J.E. Grizzle. 1962. Estimation of Time-Response Curves and Their Confidence Bands. Biometrics 18, 2, 148-159.

[4] Fisher, R.A. 1920. Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain from Broadbalk. Jour. Agric. Sci. 11: 107-135.

[5] Fisher, R.A. 1924. The Influence of Rainfall on the Yield of Wheat at Rothamsted. Phil. Trans. Roy. Soc. B, 213. 89-

[6] Vanasse, N.A., I. D. Jones and H. L. Lucas. 1951. Specific gravity-dry matter relationship in potatoes. Am Potato Jour. 28: 781-791.

COMMENTS ON PRESENTATION BY PAUL COX

H. O. Hartley
Statistical Laboratory and Department of Statistics
Iowa State University
Ames, Iowa

As the last of the discussants to give my comments in writing I have the advantage of a preview of what my predecessors have said. I will attempt to summarize and amplify their competent comments.

Mr. Cox has certainly described a problem which has raised many questions of considerable interest. His problem is concerned with the design and analysis of an experiment in which the 'response' is measured in the form of a curve (thrust curve of a motor is his example). Actually Mr. Cox is almost exclusively concerned with analysis. I will later make a few remarks on the design aspect.

Let me start by saying that we cannot really talk about an 'appropriate analysis' of a set of experimental response curves without being clear about

(a) The purpose or the objectives of the experiment

and at least to some extent about

(b) The physical mechanism generating the experimental responses.

With regard to (a) Mr. Cox has described essentially two objectives, namely, the effect of 'Conditioning Temperature' and 'Mix' on the 'Shape of the Thrust Curve' and the 'Total Impulse'. The latter is a single response clearly defined as an integral of the response curve and obtainable (say) by numerical integration. The latter requires clearer definition in terms of thrust curves characteristic of real interest to the engineer and to be specified by him. We may speculate that one of these may be the initial rate at which the thrust increases from zero, or possible the time at which it reaches the stationary stage, etc. Mr. Cox has, however, pointed to an important feature, namely that in general a multiplicity of responses will have to be computed from the curve representing the relevant summaries of interest to the engineer. Following Dr. Lucas's notation and denoting by $y_{ij}(t)$ the thrust for the j^{th} unit of the i^{th} treatment group observed at time t, we would compute for each curve k summaries $S_r(y_{ij}(t))$; $r = 1, 2, \ldots, k$, which in this example, may well be computed from standard formulas of numerical integration and differentiation. To answer the purpose of the experiment we may in many cases apply the well established techniques of multivariate (k-variate) analysis of variance (see, e.g., Smith, H; Gnanadesikan, R. and Hughes, J. B. (1962)) to the $S_r(y_{ij}(t))$ which would be a 3×3 factorial (i) by temperature and mix with 3 replicate units (j) in each cell. Much of the information will often be obtainable from a standard single variate analysis of variance applied to each of the $S_r(y_{ij}(t))$ separately. This sort of analysis which uses a 'between unit error' is also recommended by Dr. Lucas and Dr. Federer, called 'robust' by the former, and is in essence identical with Mr. Cox's analysis of variance for the regression intercept, a, and slope, b, fitted to a 'straightlooking' section of the curve. I question, however, whether Mr. Cox's procedure of 'arbitrarily' breaking up the curve into sections and fitting polynomials separately to the sections really contributes to our appreciation of the engineering aspects Is it really of interest to the engineer that a cubic term in the first section goes up with temperature? To my mind it is of the greatest importance to communicate with the engineer on the selection of relevant summaries $S_r(y_{ij}(t))$.

This brings me to (b), namely, the importance of a physical theory leading to a mathematical model for the thrust $y_{ij}(t)$, stressed by all discussants. Dr. Lucas postulates a model of the form

$$(1) \qquad y_{ij}(t) = \phi(t;\ \theta_{ij}) + \epsilon_{ij}(t)$$

where θ_{ij} is a (say) m vector of parameters. Whilst in the present example it should be quite feasible to obtain such a model from (say) the differential equations governing the dynamics of the thrust phenomenon, the statistician may be called upon to analyze curves arising in a situation in which the setting up of a mathematical model is difficult. I would stress, therefore, that summaries $S_r(y_{ij}(t))$ answering the purpose of the experiment can often be decided upon <u>without</u> reference to a mathematical model, although the study of their statistical efficiency is facilitated by the mode. Where the latter is available one may proceed as Dr. Lucas suggests to estimate the 'treatment averages' θ_i^* of the θ_{ij} although it may be argued to be more appropriate to estimate the treatment averages of the relevant summaries $S_r(\phi$ (t; $\theta_{ij})$ the two being differentially equivalent. Whatever method is used I believe that <u>some</u> attention <u>should</u> be given to the estimation of the individual units', θ_{ij}, and this raises the question of the 'within curve' error or noise. I agree with Dr. Lucas that this will often be relatively unimportant. However, it is of the same degree of relevance as, for example, the estimation of the mean life for a time mortality curve or the L. D. 50 of a dosage mortality situation both of which are usually 'within curve' estimation problems and both have received considerable (possible exaggerated) attention by statisticians. I will, therefore, answer the question raised by Mr. Cox concerning the within curve error structure: --The 'degrees of freedom' that are to be attached to a set of risiduals $y_{ij}(t_s) - \phi(t_s, \theta_{ij})$ computed at some arbitrarily selected time points, t_s, appear to depend on the choice of the t_s. Because of the time series correlogram the sum of squares of residuals

$$\sum_{s=1}^{s} (y_{ij}(t_s) - \phi(t_s, \theta_{ij}))^2$$

is approximately distributed as $c\chi^2$ based on an 'effective number' of degrees of freedom (see, e.g., Bayley and Hammersley (1946)) and ideally this should be invariant with the choice of grid points, t_s. Without the knowledge of the correlogram one cannot judge how the degrees of freedom of Mr. Cox's Figure 4 are affected. However, the analysis given in this figure in any case does not take proper account of the distinction between 'within curve' and 'between curve errors' as was pointed out by Dr. Lucas.

Finally a few words on the question of the design of an experiment with curve responses. First let me say that the choice of t_s values is not a question of the design (as Dr. Lucas rightly stresses but as far as I recall nobody said so during the discussion). This is a computational question of analysis. The design is concerned with the choice of the levels of 'temperature' and the composition of the mixes or, indeed, with questions of what treatment combinations should be chosen. Since the work by Box and Draper (1959) and Kiefer (1959) and others is mainly concerned with a single experimental response or a single response surface, much needs to be done about designing experiments which in some sense are optimum for the 'assessment' of multiple response surfaces, particularly if (as is the case with Mr. Cox's example) some factors (mix) are qualitative. In the absence of a comprehensive theory one would perhaps single out one important response from the $S_r(y_{ij}(t))$ and optimize the design for it using some such theory as the above but optimizing subject to tolerances for the bias and precision with which reponse surfaces for the other response surfaces can be estimated.

REFERENCES

Bayley, C. V. and Hammersley, J. M. (1946). Suppl. J. Roy. Stat. Soc. 8, 184.

Box, G. E. P. and Draper (1959). J. Amer. Stat. Assoc. 54, 622.

Smith, A., Gnanadesikan, R. and Hughes, J. B. (1962). Biometrics, 18, 32.

Kiefer, J. (1959). J. Roy Stat. Soc., Series B, 21, 272.

COMMENTS ON PRESENTATION BY PAUL COX

Marvin A. Schneiderman
Biometry Branch, National Cancer Institute
National School of Health, Bethesda, Maryland

It is always a pleasure (and enlightening) to attend a statistical session in which Professor Greenberg and Professor Hartley participate. They usually make important remarks, succintly, and thus save one from the responsibility of adding much more than "amen." Since they've spoken before me, I wish now to add my "not much more", (as well as "amen").

Mr. Cox has presented us with a very interesting problem in engineering which has a direct counterpart in medical and biological research. I have in mind among other measures the interpretation of electrocardiogram tracings, electro-encephalogram tracings, and the flood of apparently continuous measures that the physiologists are now capable of making. From this analogy, I am led to take slight issue, however with one of Professor Hartley's remarks.

A single response measure, such as the LD_{50} may well be inappropriate. Of course, Mr. Cox's engineer has been rather vague about what he really wants to know, and I suspect that more extended discussion with the statisticians might lead the engineer to specify his problem more. I am guessing that an LD_{50} would be inappropriate. Assume that the statistician finds, though, that he is interested in the shape of the curve, in some sense. That is, the specific shape of the curve, or the presence or absence of some specific wiggle tells him something about the physics underlying the system. For example, in the biological counterpart, a straight line inactivation curve (log response vs. time, for example) as was first postulated for the Salk vaccine for polio implies a simple one-step chemical process, or a single manner of excretion. A concave upward curve may imply a two-step process, or two (or more) modes of excretion, (i.e. a sum of exponentials) or some other functional arrangement. The investigation of the kinetics of such systems are a whole sub-field. Professor Hartley has some good advice to offer in the fitting of these sums of exponentials. Thus, the curves themselves may be the items of most importance to the experimenter. This should not be lost.

On the specific suggestions made by Mr. Cox for the analysis of the data, the physical meanings of the various points on the curve, A, B, C, D, E, F, might help guide the statistician into more fruitful lines--perhaps even into a solution of the problem the engineer wants solved. As Professor Hartley pointed out, the points on the curve are correlated, making the analysis of variance inappropriate because of the non-independence. One earlier reference on the non-independent regression problem then the Bailey-Hammersley reference that may be helpful is one by John Mandel, in the Journal of the American Statistical Association "Fitting a straight line to certain types of cumulative data," Vol. 52, p. 552 (1957). My recollection is that Mandel shows that a least squares approach gives an unbiased estimate of the parameters, but gives the wrong (too small) variance. He gives other references to this problem, too.

On the "shape" problem, there is a paper by G.E.P. Box and W.A. Hay which may be of interest. It appeared in Biometrics, Vol. 9, p. 304 (1953) "A statistical design for the efficient removal of trends occurring in a comparative experiment, with an application in biological assay." A recent doctoral dissertation by Francis J. Wall, from the University of Minnesota considers an aspect of the nearly continuous data problem in biology and medicine. The title is "Biostatistical linear models in longitudinal medical research problems." The title shouldn't mislead engineers. It's much the same problem as we had here.

THE INDEPENDENT ACTION THEORY OF MORTALITY AS TESTED AT FORT DETRICK

Francis Marion Wadley
U. S. Army Biological Laboratories
Fort Detrick, Frederick, Maryland

The independent action theory is sometimes used as an approach to all-or none dosage-effect problems instead of the more usual dosage-effect methods such as probit analysis. With the probit and similar analyses, the basic assumption is of varying susceptibility among the subjects. With the independent action theory in its simplest form the assumption is that any toxic unit reaching the site of action will be effective. Each unit is believed to have a small but definite chance of hitting its mark; a higher percentage response to larger doses is produced by multiplication of this chance. This theory obviously does not assume varying susceptibility among subjects, and will logically lead to the same slope for all trials. If $\underline{a}$ is the chance of hitting the mark, the chance of escape is $(1-a)$ for one toxic unit; for 2 units, it is $(1-a)^2$; for $\underline{n}$ units $(1-a)^n$. Danger from bullets on a battlefield has been used as one illustration.

The independent action theory was apparently first developed by Neyman and associates, according to K. L. Calder of Fort Detrick. It has been used by Watson (Phil. Trans. Roy. Soc. London, 1936) in studying transmission of plant viruses by insect migrants. It is also used by some workers in dosage-effect studies where the dose is of biological agents (S. Peto, Biometrics 1953; L. J. Goldberg and associates, 5th, 8th and other reports, Navy Biol. Lab., 1951-52). A. W. Kimball (1953 lectures, Fort Detrick) has applied the theorv to radioactive particles. Peto presents detailed procedure for calculation, and mathematical methods are also presented by Andrews and Chernoff (Tech. Rept. 17, Applied Math Lab, Stanford, 1952), and by W. G. Cochran (1946 lectures, N. C. State). Goldberg (l.c.) has worked out special plotting paper for quick graphic estimation of LD-50 and its error. The extensive work on dosage theory assuming varying susceptibility is conveniently summarized by Finney. ("Probit Analysis", Cambridge U. Press, 1952).

Where such agents as bullets and radioactive particles are considered, there can be no question that the idea of independent action will apply better than the concept of dosage and varying susceptibility. With chemical toxicants which can be measured out accurately, the idea of dosage and varying susceptibility undoubtedly applies better than the independent action concept. Susceptibility is know to vary. With biological agents such as pathogenic bacteria, we are on a middle ground where either procedure may have its advocates. The basic question appears to be whether susceptibility really varies substantially among subjects. If some individuals can use their provisions for combating invading agents to throw off effects of a moderate dose of organisms, while weaker subjects will succomb; the ordinary dosage treatment should apply.

The exponential approach obviously simplifies mathematical treatment of data, and in its simpler forms will allow calculation of an LD-50 from only one concentration giving partial mortality. Allowance can be made for varying susceptibility, but in so doing, simplicity if forfeited and advantages over probit analysis seem dubious.

With data of Fort Detrick, Goldberg's approach has given LD-50 estimates very similar to those from probit analysis. The graphic error estimates of his early publications seem inadequate. Where several concentrations give partial mortality, Goldberg's graphic method will yield several LD-50 estimates for the same experiment. These sometimes very incongruously for agents with characteristically low slope.

Critical test comparing the two approaches are very difficult since for ordinary experiments with small numbers results are apt to be quite similar. One possible test involves the form of the untransformed dosage-percentage curve. With the typical probit curve we have an asymetric sigmoid with a weakly defined but real lowe bend. With the exponential we have a single-bend curve of decreasing steepness. Demonstration of a lower bend in the zone of low mortality would be evidence for the probit approach, but would require hundreds of animals. In general, critical test would be expensive and would impede the progress of needed practical tests. We are at present limited largely to gleaning evidence from practical tests.

A preliminary test was afforded in 1953 by Fort Detrick data originated by A. N. Gorelick, of a number of toxic bacterial injections into mice. (See S. B. Job No. 433, Fort Detrick). Some 43 points based on 435 animals were available. If proportion of survival (q) with dosage $\underline{n}$ is estimated as $(1-a)^n$, then

$$\log q = n \log (1-a)$$

and dosage should be linear in relation to log survival. Significant departure from linearity should suggest that the logical basis of the independent action theory is weak in this material.

On plotting log survival against dose, a gentle curve was suggested by the chart. On fitting, a simple parabola gave a significant gain over a straight line. Statistics are as follows:

Source of Variation	Degree of Freedom	Mean Square
Linear	1	2.44
Quadratic (Additional)	1	0.33
Residual	40	0.03

Another test, not of a definite dosage affect study, but of some assumptions related to those of the independent action theory, was afforded by some of W. C. Patrick's data at Fort Detrick. It involved an encephalomyelitis virus injected intra-cerabrally into mice. A large number of test, made routinely in development work, were available. The agent is very toxic to mice, when injected intracerebrally at 0.03 ml of high dilutions. The regularity of results has led to some thought that any single infective particle reaching the site of action may be fatal.

Following this theory, in the high dioutions allowing survival, the survival is thought to be due simply to the fact that the small sample taken for injection contains no particles. This would imply a Poisson distribution of particles among such samples, with a rather small mean. This would throw us back on the independent action or "one-shot" theory of toxicity.

Patrick's numerous records offered a chance to test this theory. If infective particles have a Poisson distribution among injection samples, and if survival indicates a blank sample, the average number <u>M</u> of units per sample could be estiamted from the proportion of survivors <u>q</u>:

$$q = e^{-m}; \; M = -\ln q.$$

These estimates are quite easily made. Then with two successive concentrations, giving partial mortality; the ratio of two estimates of m in one test should approximate the dilution ratio (in these cases O. 5 log). This would not be realized exactly in any one comaprison, but with a long series the relation should appear. Filure of the M ratios to agree with the dilution ratios is regarded as evidence against the theory.

For illustration a firaly typical assay of an encephalomyelitis preparation by intra-cerebral injection in mice is taken. Unlike Patrick's series, dilutions were a log apart rather than half a log.

Log dilution	Response	o/o	p	q	Estimated m
7.O	16/16	100.O	1.OOO	O.OOO	--
8.O	16/16	100.O	1.OOO	O.OOO	--
9.O	9/15	60.O	O.600	O.400	O.92
10.O	3/16	18.8	O.188	O.812	O.21

From the first dilution (log is 9.O) showing partial mortality, the value of q is O.O4OO. The theory being tested would say that O.4 proportion of the injection samples contained zero particles. Solving the equation $q = e^{-m}$ with q taken as O.4O, m comes out as -ln(q) or O.92. The second dilution similarly treated gives as an estimate of m, O.21. The ratio is O.92/O.21 or about 4.4. This is far from the dilution ratio of 1O, to be expected if the theory holds.

With the aid of Pvt. Isen a large number of such ratios from Patrick's 1955 and 1956 test were assembled. Logs of computed ratios, from tests where 2 estimates from partial mortality were possible, were assembled and compared with the theoretical O.5O.

Year	No. Tests Used	Mean Log Ratio Of Estimates of M	95 o/o Confidence Limits
1955	1O4	O.33	O.27 - O.39
1956	166	O.4O	O.33 - O.47

Results do not bear out the theory that a Poisson distribution of infective particles will explain mortality or survival.

To sum up, experience with the independent action model in all-or-none tests at Fort Detrick, has not been very encouraging. Limited tests of the theoretical basis have not sustained the basic theory.

ANALYSIS OF A FUNCTION IN COLLABORATIVE EXPERIMENTATION

Walter D. Foster
Biomathematics Division
Fort Detrick, Frederick, Maryland

I. INTRODUCTION. The usual objective in collaborative or referee experimentation is to make comparisons among the set of participants with the over-all criterion that stations be no more diverse than runs at a station. Thus, station means and variances are the values for interstation comparison. This paper is concerned with the response variable for a particular class of referee experimentation and its analysis.

In this collaborative experiment, each of five laboratories ran a series of aerosol tests in which *P. tularensis* tagged with radioactive phosphorous (P^{32}) was aerosolized in rotating drums and sampled at eight points in time over a 22-hour period. The five laboratories with identical equipment achieved the series of tests at approximately the same time, going to extreme lengths to achieve homogenous methodology. Three treatments were introduced consisting of three relative humidity conditions in the rotating drums of 20%, 50% and 80%. Two aerosols or runs were completed per humidity at each participating laboratory on a randomized basis. It is of interest to note that three separate nations were represented in these five stations.

It was the objective of this experimentation to

(1) Comparestation means,

(2) Compare station variances,

(3) To identify stations whose results did not conform to those of the others,

(4) To examine the station by treatment interaction, i.e., whether the differences between treatments were consistent from one station to another.

II. DEFINITION OF THE RESPONSE VARIABLE. When an aerosol is monitored over a period of time, the measurement usually taken is the concentration at a series of points in time. Thus, the definition of the response variable to be analyzed could be the concentration, given a particular set of sampling times. However, this concept is likely to ignore the design restriction that only runs are random, not sampling points in a run. A second and better response variable is the function describing concentration and its change in time. Such a function in aerobiology is called a decay function. Previous research has identified a reasonably simple expression which is excellent for summarizing the course of an aerosol in time:

$$C = C_o(t + 1)^{-b} e^{-kt}$$

The usual univariate approach to the analysis of a function such as the one given above would be to analyze separately the parameters of this function, C_o, b, k. However, not only are these parameters known to be correlated because of the design of the experiment but they are also known to be stochastically correlated from one aerosol run to another. Therefore, it is the purpose here to show how the entire decay function, identified as the response variable, can be analyzed and interpreted through the usual analysis of variance technique.

III. ANALYSIS OF VARIANCE OF THE DECAY FUNCTION. With the decay function as the response variable, the following analysis of variance has been accorded this response for the purpose of examining stations levels, variability, and station by treatment interaction. The complete analysis of variance is shown in Table I in detailed form where all of the objectives have been answered. Its construction is given in a separate section.

TABLE I.

A. V. OF DECAY FUNCTION FOR STATIONS AND TREATMENTS

Line	Source	df	MS
15	Mean	3	387.1591
16	Stations	12	.1737
17	A vs Rest	3	.5894
18	Among Rest	9	.0352
19	Treatments	6	.0409
20	S x T	24	.0151
21	Runs in S x T	45	.0195
22	Runs in 20%	15	.0315
23	Runs in 50%	15	.0129
24	Runs in 80%	15	.0118
25	Deviations	150	.0014
26	TOTAL		

The following brief interpretation is accorded the analysis of variance shown in Table I in order to provide specific answers to the objectives of this experiment. Reading from the bottom of the table, the runs have been pooled over stations per treatment affording a test of homogeneity of variance from one treatment to another in lines 22-24. This departure from the original objective is better achieved than the original for estimating station variability because of the limited number of runs per treatment. There is a suggestion that the runs were less homogeneous at the 20% humidity than at the other two. In line 20, it is clear that the station by humidity interaction, if not zero, was small. On the other hand in line 16, differences among stations were obviously large compared to runs in S x T, line 21. The contrast of A versus the remaining stations, line 17, accounted for a large proportion of the station variability, with the variation attributed to the remaining stations being scarcely larger than the variation among trials at a given station. The purpose of this partition in line 17 was to investigate whether the variation among the remaining stations has been reduced to magnitude of trial-to-trial variation. Further partition is in order so long as it could be helpful in identifying and possibly eliminating factors at stations causing station departures.

This brief interpretation was developed completely on the basis of the analysis of variance in Table I. It would be desirable to present a tables of means to accompany the variance analysis. This is the point at which multivariate techniques in general are at a disadvantage, for there is no plainly defined quantity which is easily tabled. Two suggestions are given here as a means by which the interpretation can be visualized; these are first by graphs and secondly by the coefficients of the decay function. The graphs for each station are given in Figure 1 where the values have been averaged over all three humidity conditions. The coefficients computed as estimates of the parameters of the decay function are given below.

Values of Decay Function Constants

	Stations				
	A	B	C	D	E
Log C_o	2.227	2.053	1.879	1.868	1.972
$b \times 10^1$	.868	.994	.053	.406	.839
$k \times 10^3$	.264	.193	.219	.205	.073

It is appreciated that neither of these means of visualizing station differences is perfect; nevertheless, they are suggested here as the best which are easily available.

A few remarks are necessary here before describing in the next section the technique for the analysis of variance of a decay function. The question of auto-correlation always seems to appear in problems in time series such as these. However, it is contended here that because of the function approach the question of auto-correlation of the successive C_i does not arise. Only the residuals are important, and when the decay function is found to provide an excellent summary of the change of concentration in time the residuals may be considered as mutually independent. A second remark has to do with the potential use of the results of this kind of referee experimentation. With the variation noted here and appraised to be acceptable, these data afford a basis for constructing a quality control approach to future aerosol runs in which aberrant points, runs, and even stations may be readily identified.

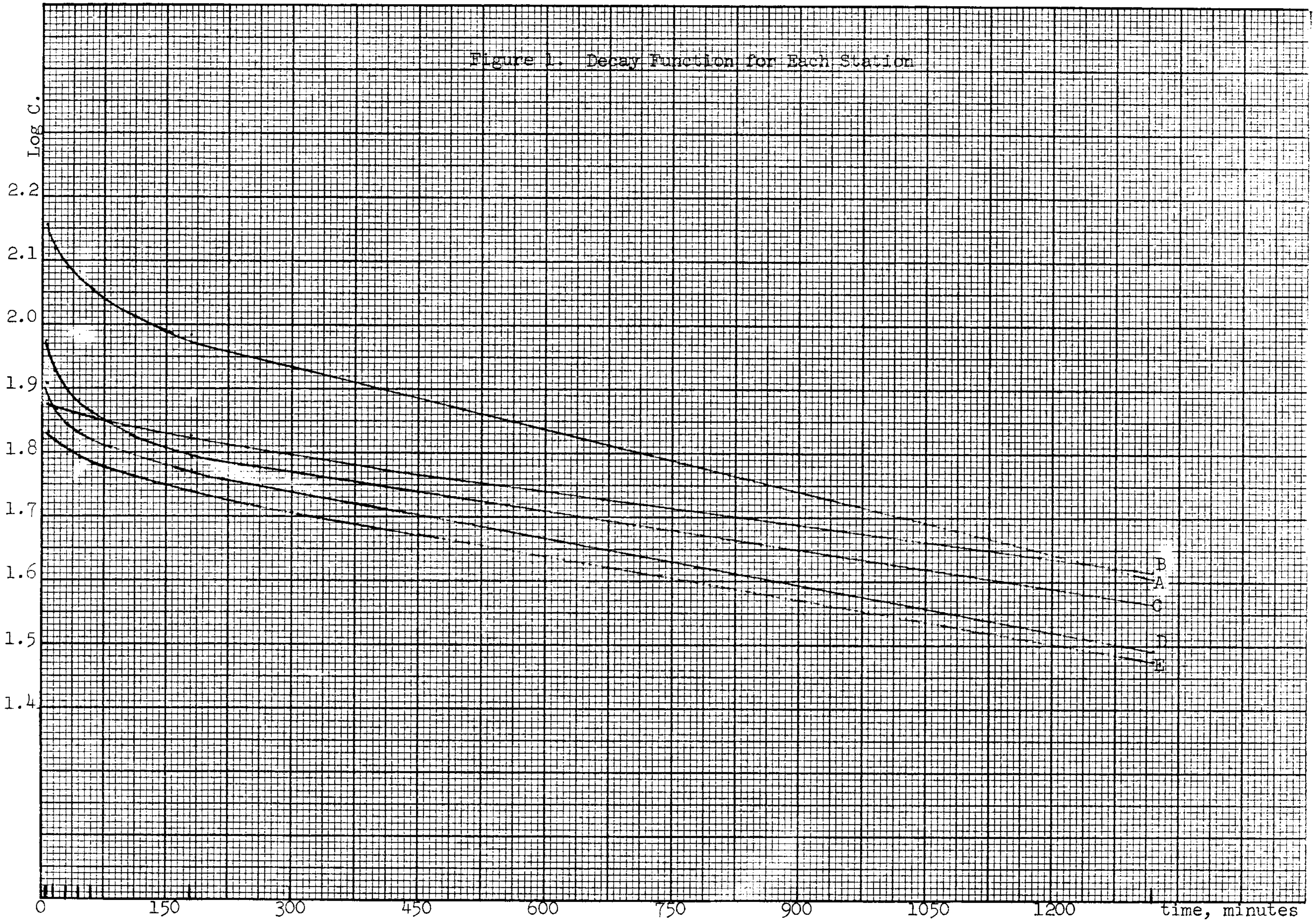


Figure 1. Decay Function for Each Station

IV. CONSTRUCTION OF THE ANALYSIS OF VARIANCE OF A DECAY FUNCTION. On an individual run basis, the familiar partition of variation is obtained as shown in Table II with the exception that no correction is shown separately for the mean -- the p parameters of the decay function are shown together.

TABLE II.

A.V. OF DECAY FUNCTION FOR A SINGLE RUN

Line	Source	df	e.g. df
1	Function	p	3
2	Deviations	n-p	5
3	TOTAL	n	8

The second step is to compute the analysis of variance for each station over the r runs for a given treatment as is shown in Table III. The sum of squares for line 4 are obtained as usual where the function is fitted to the entire set of values for the r runs and the computation is achieved on a per item (or per value) basis. The sum of squares for line 5 is obtained easily merely by summing the sum of squares for line 1 in Table II for the various runs and subtracting line 4. Similarly, line 6, deviations in runs, is obtained by summing the values in line 2 over all runs.

TABLE III.

A.V. OF DECAY FUNCTION FOR A STATION AND A TREATMENT

Line	Source	df	e.g. df
4	Mean	p	3
5	Among runs	p(r-1)	3
6	Deviations in runs	r(n-p)	10
7	TOTAL	rn	16

A small digression may be helpful at this point to explain the degrees of freedom shown thus far in the analysis of variance. The degrees of freedom in line 5 are shown to be the usual degrees of freedom for runs, r-1, multiplied by the number of parameters to be estimated in the decay function. Although these parameters are known not to be independent, they continue to be identified as restrictions in the least squares process for estimation and as such must be deducted as degrees of freedom. It is not likely that a further partition of these degrees of freedom could be achieved in a manner such as to show contrasts among the parameters themselves.

With the introduction of t treatments at a station, the analysis of variance as outlined in Table IV is appropriate for each station, where the partition is basically a nested one. As before, the function is fitted over all points in order to provide the sum of squares due to the function, line 8. The sum of squares for treatments is obtained through a two step procedure. First, the sums of squares shown in line 4 of Table III for each treatment are added. Then the sum of squares for the mean in line 8 is subtracted, the difference being specifically that due to variation among treatments and is entered in line 9. The sum of squares for runs in treatments, line 10, is obtained by summing the sums of squares for each trial separately for that particular treatment, i. e., the sum of lines 5 for that station. They can also be listed in partition as in lines 11 and 12 of Table IV. Similarly, the sum of squares for deviations are obtained by pooling for line 13.

TABLE IV.

A. V. OF DECAY FUNCTION AT STATION A WITH TREATMENTS

Line	Source	df	e. g. df
8	Mean	p	3
9	Treatments	p(t-1)	6
10	Runs in T	pt(r-1)	9
11	in T_1	p(r-1)	3
12	in T_2	p(r-1)	3
	etc	etc	3
13	Deviations	rt(n-p)	30
14	TOTAL	trn	48

The construction of the over-all analysis of variance as shown in Table I continues to be based upon the previous tables in a sort of a building block arrangement. The mean, line 15, is obtained by finding the sum of squares due to the function when fitted to all of the points in the combined collaborative experiment. Line 16 is obtained by a two step procedure: the sum of squares for line 8 in Table IV is summed over the s stations; from this sum of lines 8 the sum of squares in line 15 is subtracted. The difference then represents the sum of squares due to stations averaged over treatments.

The partition of the station sum of squares as initiated in line 17 depends upon which station appears to show the greatest departure from the other stations, following the philosophy given briefly in the interpretation of the example above. Assuming that this identification of the greatest departure can be made from a study of the graphs, line 17 then represents the contrast between the station with the maximum departure and the rest of the stations. This partition is accomplished in a three step procedure as follows. The sums of squares given in line 8 of Table IV are added for the four stations marked as "rest". This sum is entered as line "a" in the ancillary computation table below. The second step is to compute the sum of squares for the function when fitted to all the points represented by the four stations combined as "rest", having excluded the station with the maximum departure from the computation--line "b" below. The third step is to subtract the sum of squares in line "b" from the sum of squares in line "a", giving the "among rest" sum of squares as shown in line "c". Finally, the subtraction of line "c" sum of squares from line 16 is entered in line 17 and is identified as the contrast station A versus "rest". Further orthogonal partitioning for other "departures" can be computed in this fashion.

Ancillary Computation for Table I

Line	Source
a	Sum of line 8 for "rest" stations
b	Mean for "rest"
c	a-b = among "rest" stations

A new computation is required for line 19, the sum of squares due to treatments. This is accomplished by considering all points for the first treatment including those for the various stations and fitting the decay function. This is achieved for each treatment. These sum of squares are added over the various treatments. From this over-all sum, the value in line 15 is subtracted, giving the variation among treatments averaged over stations.

The interaction term, station by treatment, as shown in line 20, is obtained in the usual way. Briefly, it consists of summing line 4 over all stations and treatments. From this sum are subtracted lines 15, 16 and 19.

Line 21 is obtained easily by summing all lines in Table III. The partition of line 21 as shown in lines 22 and 23 is easily accomplished according to the purpose at hand merely by restricting the summing to the category desired.

Missing values will complicate this analysis and indeed will render the partition non-orthogonal if missing values are not restored to the analysis. Therefore, it is recommended that a simple procedure for estimating these missing values such as computing the value according to the function as estimated from the remainder of the points being inserted with one degree of freedom per missing value being subtracted from the degree of freedom assigned to deviations. Note that in the simpler analyses which are completely nested orthogonality does not depend upon equal numbers.

DESIGN AND ANALYSIS OF ENTOMOLOGICAL FIELD EXPERIMENTS

William A. Brown
Test Design and Analysis Office, Dugway Proving Ground

and

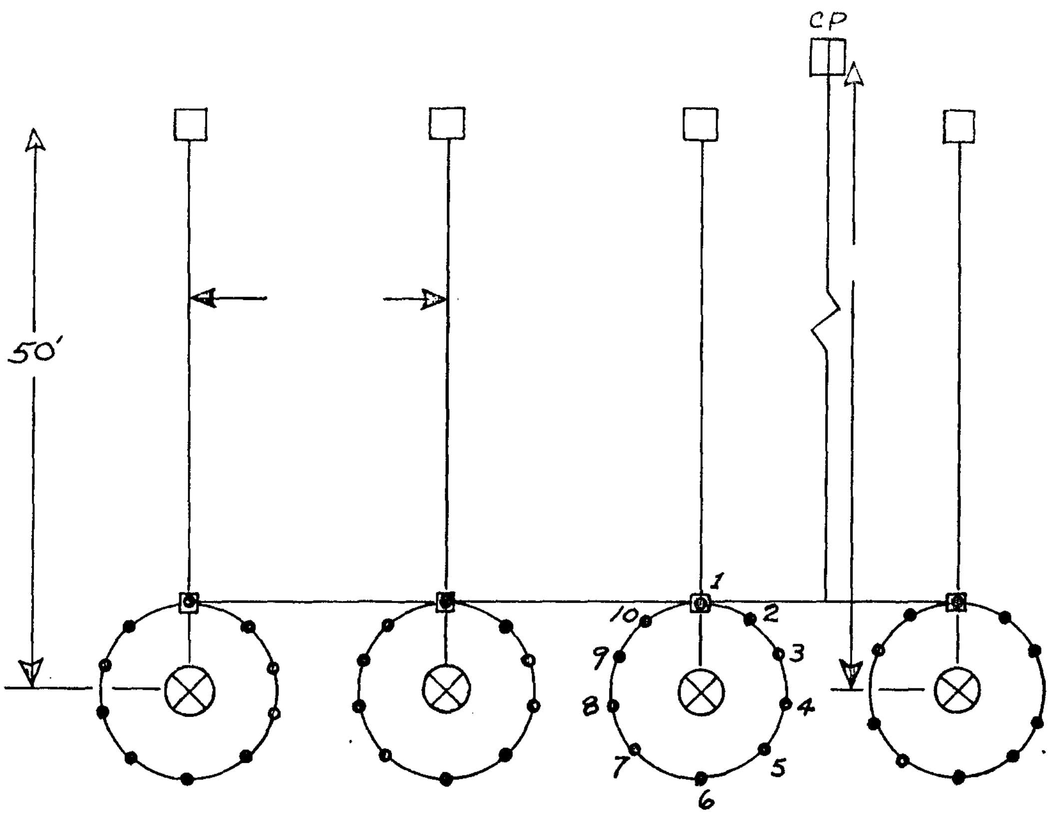
Scott A. Krane
Dugway Field Office, C-E-I-R, INC.

Recently two entomological field experiments were conducted at Dugway Proving Ground. The purpose of the first experiment was to compare the biting propensity of two strains of a species of insect. In each trial, four 15-foot radius circles were scribed, and 10 hosts, randomly selected, were positioned equidistantly along each circumference. The Number 1 position in each circle was oriented to true north. (See Figure 1.) At function time, 100 individuals of the appropriate strain were released at the center of each circle. In two of the circles, the A strain was used; in the other two circles, the B strain. The men were seated on the ground and remained relatively motionless throughout the trial. Sampling consisted of each man recording those bites actually received and entering the total number, for 5-minute intervals, on a data card. Sampling was conducted for 30 minutes following the release unless biting activity continued. In that event, sampling in all circles was extended for additional 5-minute periods until the biting activity had essentially ceased.

Comparisons between strains were thus subject to the variation found among circles. This variation was expected to be appreciably larger than the variation among men on a circle. By the nature of the experimental treatments (strains), however, it was necessary to separate the strains either in space or in time sufficiently that their ranges of biting activity did not overlap. Only in this manner could bites be accurately attributed to one strain or the other. The duplicate circles for each strain represented an effort to partially overcome this inherent insensitivity.

In the analysis of the data, it was considered useful to employ a mathematical model to describe the distribution of the number of bites per host. The simplest model which might conceivably fit the observations is the Poisson, given by:

(1) $$f(x) = e^{-m} m^x / x! \qquad x = 0, 1, \ldots, n,$$



— Local battery field telephone line; □ telephone

○ On-site meteorological sensing station

□ Boxed recording instruments

× Test fixture

• Hosts

Figure 1

where x is the number of bites received by an individual host, $f(x)$ is the probability (or relative frequency) of x bites, and m is an unknown parameter equal to the "long-run" average number of bites per host. If the individuals of a strain are randomly distributed throughout a given area, and if hosts are equally attractive, then the Poisson model should be appropriate.

Previous studies, however, have indicated that the spatial distribution of insects released in this manner is not random (perhaps being influenced by the wind direction, for example), nor are all hosts equally attractive. As a result of these tendencies, the distribution of bites will be "over-dispersed" relative to the Poisson distribution, i.e., the number of hosts receiving a very large number of bites and the number of hosts receiving a very small number of bites will both be larger than the number predicted by the Poisson model, while the number receiving near-average numbers of bites will be smaller.

One of the simplest and most frequently used "over-dispersed" statistical models is the negative binomial, which has the general term:

$$(2) \qquad f(x) = \left(q^{-k}\right)\frac{(k+x-1)!}{x!\,(k-1)!}\,\frac{p^x}{q}, \qquad x = 0, 1, 2, \ldots, n,$$

where q equals $1 + p$, and p and k are unknown parameters. Various rationales may be given for the negative binomial.[1] One of the simplest is that the negative binomial is produced by a mixture of Poisson distributions in which the parameter, m, varies according to a "gamma" distribution. While no rationale appears to be particularly compelling in the present problem, the relative simplicity of the negative binomial model and the success with which other investigators have applied it to biological data are taken to justify its use, at least as a working hypothesis.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for over-dispersion with respect to a Poisson distribution by the χ^2 statistic:

[1]Bliss, C. F. Fitting the Negative Binomial Distribution to Biological Data. <u>Biometrics</u> Vol. 9, (2) pp. 176-196.

$$\chi^2_{n-1} = (n-1)s^2 / \bar{x}, \tag{3}$$

where n is the sample size (number of hosts, s^2 is the sample variance of the number of bites, and $\bar{x}$ is the sample mean number of bites.[2] The calculated statistic was tested for significance by comparison with the 20 per cent upper tail value of the χ^2 distribution.[3] If the test did not indicate over-dispersion, the data were subsequently fitted to a Poisson distribution and subjected to a χ^2 "goodness-of-fit" test. If the test did indicate over-dispersion, the data were fitted by the method of maximum likelihood[4] to a negative binomial distribution, and then subjected to a χ^2 goodness-of-fit test. All of the above calculations were performed on the IBM 1620 Computer, using a specially prepared FORTRAN program. Fifty-five of 96 sets of 5-minute data showed close agreement with the Poisson distribution. For each of these sets of data, however, the variance was usually larger than the mean, and, consequently, a further comparison with the negative binomial distribution would generally have shown even closer agreement.[5] Therefore, it was decided that, for the purposes of the analysis of variance, an appropriate transformation to stabilize variance for these data would be that derived for the negative binomial:[6,7]

---

[2]Ibid.

[3]For confeni ence of internal calculation on a digital computer, the 20 per cent upper tail χ^2 value was obtained from the approximation:

$$\log_e \frac{\chi^2 n-1}{n-1} -1 = -0.038 - 0.452 \log_e(n-1)$$

[4]Fisher, R. A. Notes on the Efficient Fitting of the Negative Binomial. <u>Biometrics</u>, Vol. 9(2), pp. 196-200, 1953.

[5]The Poisson is, in fact, a limiting case of the negative binomial, from which it follows that a negative binomial must fit data at least as well as the Poisson.

[6]Bartlett, M. S. The Use of Transformations. <u>Biometrics</u>, March 1947, Vol. 3(1) pp. 39-52.

[7]Kempthorne, O., <u>Design and Analysis of Experiments</u>. Chapter 8. John Wiley and Sons, Inc., 1952.

(4) $$y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2}).$$

Figure 2, showing a plot of mean versus variance on log-log paper, illustrates the closer agreement with the negative binomial distribution. The diagonal line represents the square root transformation, appropriate for variance stabilization of Poisson distributed data, and the curved line represents the transformation $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ has the value 1.0. (After several guesses of λ, the value of 1.0 was selected since, by eye-fitting, it appeared to reasonably minimize the deviations from the curve. Using the value $\lambda = 1.0$, 47 of the data points lie above the curve, and 49 below.) Subsequent analysis of the data of Experiment 1 using a method of Bliss and Owen[8] for the estimation of a common k, resulted in the estimate

$$k_c = 0.51.$$

Since $\lambda^2 \doteq (1/k)$, the value of λ appropriate for this estimate of k is

$$\lambda = k_c^{-0.5} = 1.4.$$

As shown in Figure 2, the value of $\lambda = 1.0$ obtained graphically agrees reasonably well with that estimated by the method of Bliss and Owen. It can easily be seen in Figure 2 that the data follow more closely to the curved line. However, the dashed line indicates that the logarithmic transformation may be as suitable as the inverse hyperbolic sine. Furthermore, analysis of logarithmically transformed data permits interpretations of results, in terms of ratios of treatment effects, while no such interpretation arises directly from the negative binomial transformation. Therefore, separate analyses of variance were performed, using the two transformations.

An analysis of variance, based on the three-way cross classification of trial, strain, and time period, was performed on each of the following four sets of data:

1. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ equals 1.0 and x is the total number of bites received by a host during a given time period,

---

[8]Bliss, C. I. and A. R. G. Owen, "Negative Binomial Distributions With a Common K", Biometrika 45, pp. 37-58, 1958.

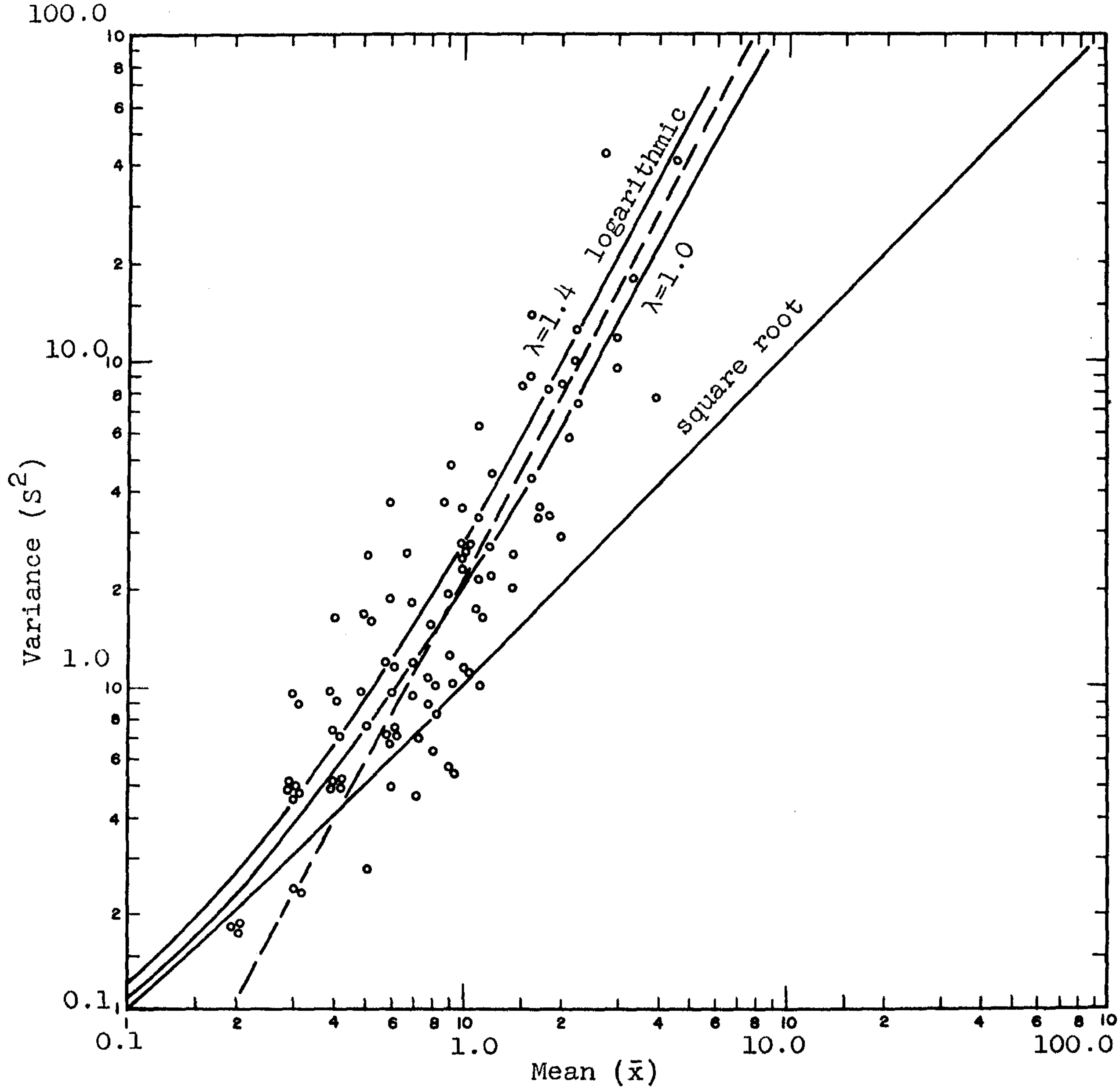

100.0
10.0
1.0
0.1
Variance (S^2)
λ=1.4
logarithmic
λ=1.0
square root
0.1
1.0
10.0
100.0
Mean ($\bar{x}$)

Figure 2

2. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ equals 1.0 and x is the total number of bites received at a circle during a given time period,

3. The values of $y = \log (x + 1)$, where x is the total number of bites received by a host during a given time period, and

4. The values of $y = \log (x + 1)$, where x is the total number of bites received at a circle during a given time period.

The results of each of these analyses are presented in Table 1.

As shown in Table 1, the results obtained in the four analyses of variance were essentially the same. Each analysis indicated that the total numbers of bites obtained during the six times periods were significantly different, and that no significant difference could be detected between strains. In every analysis, however, Error (a) was relatively large, so that the F test, comparing strain effects, was undoubtedly insensitive. As mentioned earlier, the insensitivity of the analyses for strain differences follows unavoidably from the design of these trials, in which strain comparisons could only be made between circles (rather than within circles), and, hence, are subject to the greater variability found from circle to circle as measured by Error (a).

The purpose of the second experiment was to compare the dispersal of two strains of a species of insect as measured by their biting activity.

For this experiment, it was greatly desired that the ambient air temperature and windspeed range of an A-B strain pair of trials be as similar as possible. However, because of the small number of men available concurrent testing of the two strains could not be accomplished. Therefore, whenever possible, two trials were conducted each day--one trial using the A strain and the other, following as soon after as practicable, employing the B strain.

In each trial, four concentric circles were used, designated Circles A, B, C, and D with radii equal to 100, 200, 300 and 400 feet. Eight men were positioned equidistantly around each circumference of Circles A, B, and D, and 16 men were positioned equidistantly around the circumference of Circle C. (See Figure 3.) At function time, 1000 individuals of the appropriate strain were released at the center of the concentric configuration, and the men, seated and facing the release point, recorded

This page intentionally left blank.

Table 1

SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMATION*							
		1		2		3		4	
		Mean Square	F Value	Mean Square	F Value	Mean Square	F Value	Mean Square	F Value
Trial, T	5	0.076396		0.054999		0.279571		0.216414	
Strain, S	1	0.240379	3.27	0.240754	2.39	0.894409	3.30	0.944946	2.48
T x S	5	0.017313	0.236	0.25430	0.253	0.065268	0.241	0.095473	0.251
Error (a)	12	0.073442		0.100546		0.270806		0.380320	
Time period, P	5	0.571952	22.8**	0.716051	51.8**	2.073743	60.4**	2.743388	54.1**
P x T	25	0.028863	1.15	0.026287	1.90	0.105622	3.08**	0.101577	2.00*[3]
P x S	5	0.011180	0.445	0.007883	0.570	0.041707	1.21	0.027349	0.540
P x T x S	25	0.016094	0.641	0.019936	1.44	0.058558	1.71	0.075415	1.49
Error (b)	60	0.025108		0.013830		0.034306		0.050667	
Sub-total	143								
Hosts/Circles, C	216					0.147148			
Hosts/P x C	1080					0.029697			
Total	1439								

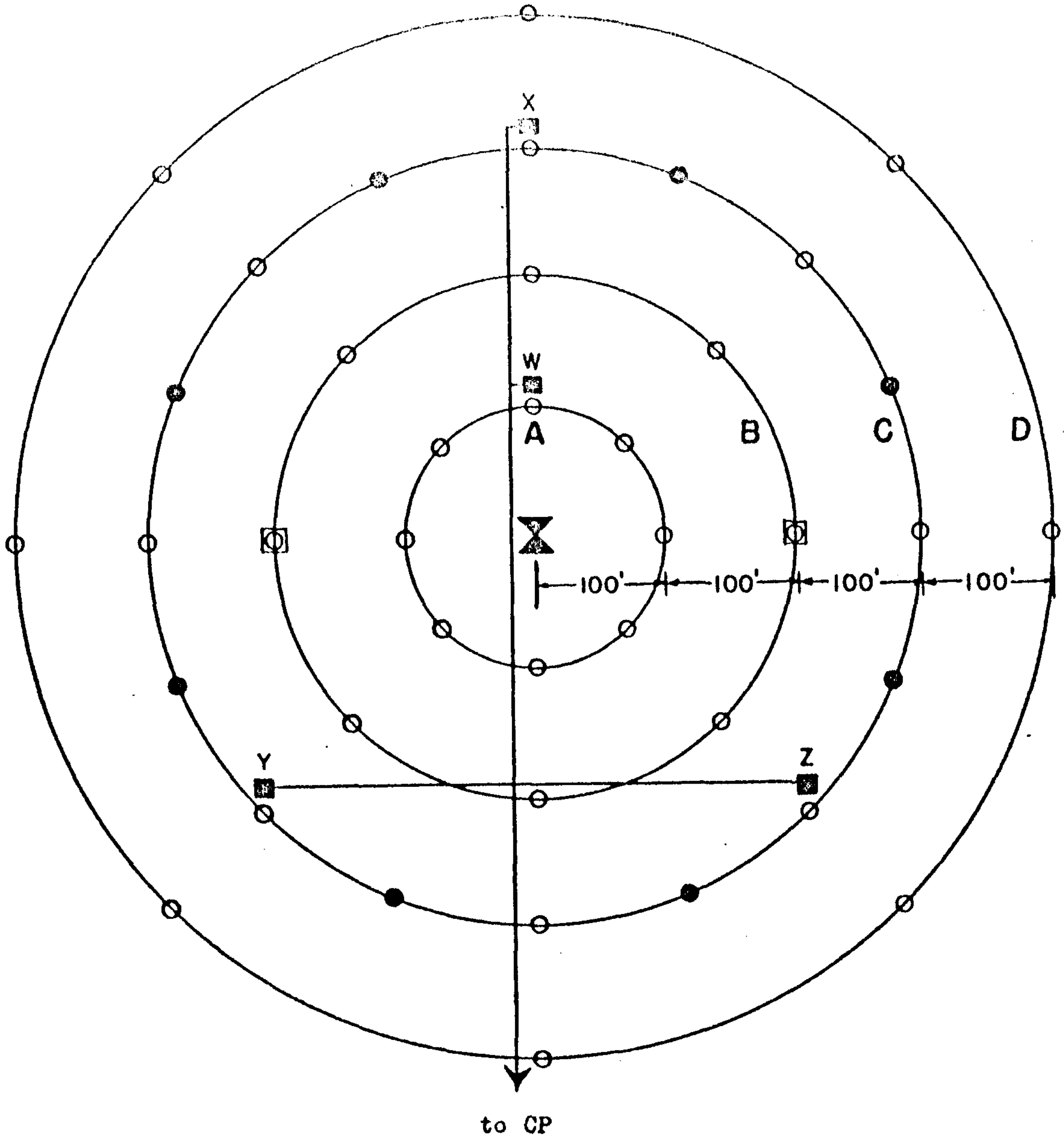
Trials A-3, A-8, and A-9 were omitted from this analysis because of insufficient data.

*1 denotes $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$ transformation of 5-minute total bites received by each host; 2 denotes $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$ transformation of 5-minute total bites received at each circle; 3 denotes log (x + 1) transformation of 5-minute total bites received by each host; and 4 denotes log (x + 1) transformation of 5-minute total bites received at each circle.

**Significant at the 1.0 per cent level.

*[3]Significant at the 5.0 per cent level.

This page intentionally left blank.



■ Field telephone

— Field telephone line

○ Sampling station

● Augmenting sampler position

□ 2-meter wind speed and direction station

Release point

Figure 3

This page intentionally left blank.

biting activity for at least 30 minutes. The sampling proceedures were the same as those used for the first experiment.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for "over-dispersion" with respect to a Poisson distribution in the same manner as in the first experiment.

The results indicated that there was nearly always a departure from the Poisson distribution, in the direction of higher variance and "over-dispersion." In addition, 45 of the 70 sets of 5-minute data showed agreement with the negative binomial distribution at a nomial 95 per cent confidence level. Further, from an examination of the plot of the mean versus the variance (see Figure 4), it did not appear that the data would fit any other distribution more consistently. Therefore, it was decided that, for the purposes of the analysis of variance, a suitable transformation to normalize these data would be:

$$y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2}).$$

After several guesses of λ, and, subsequently, fitting the data by eye to $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, it appeared that a reasonable estimate that would minimize the deviations from the curve was $\lambda = 1.0$. Using this value, 42 of the data points lie above the curve, and 43 below.

An analysis of variance, based on the four-way cross classification of day, strain, circle, and time period, was performed on each of the following sets of data:

1. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ where $\lambda = 1.0$, and x is the total number of bites received during a given time period at Circles A, B, and D, and one-half the total received at Circle C.

The results of these analyses are presented in Table 2.

This page intentionally left blank.

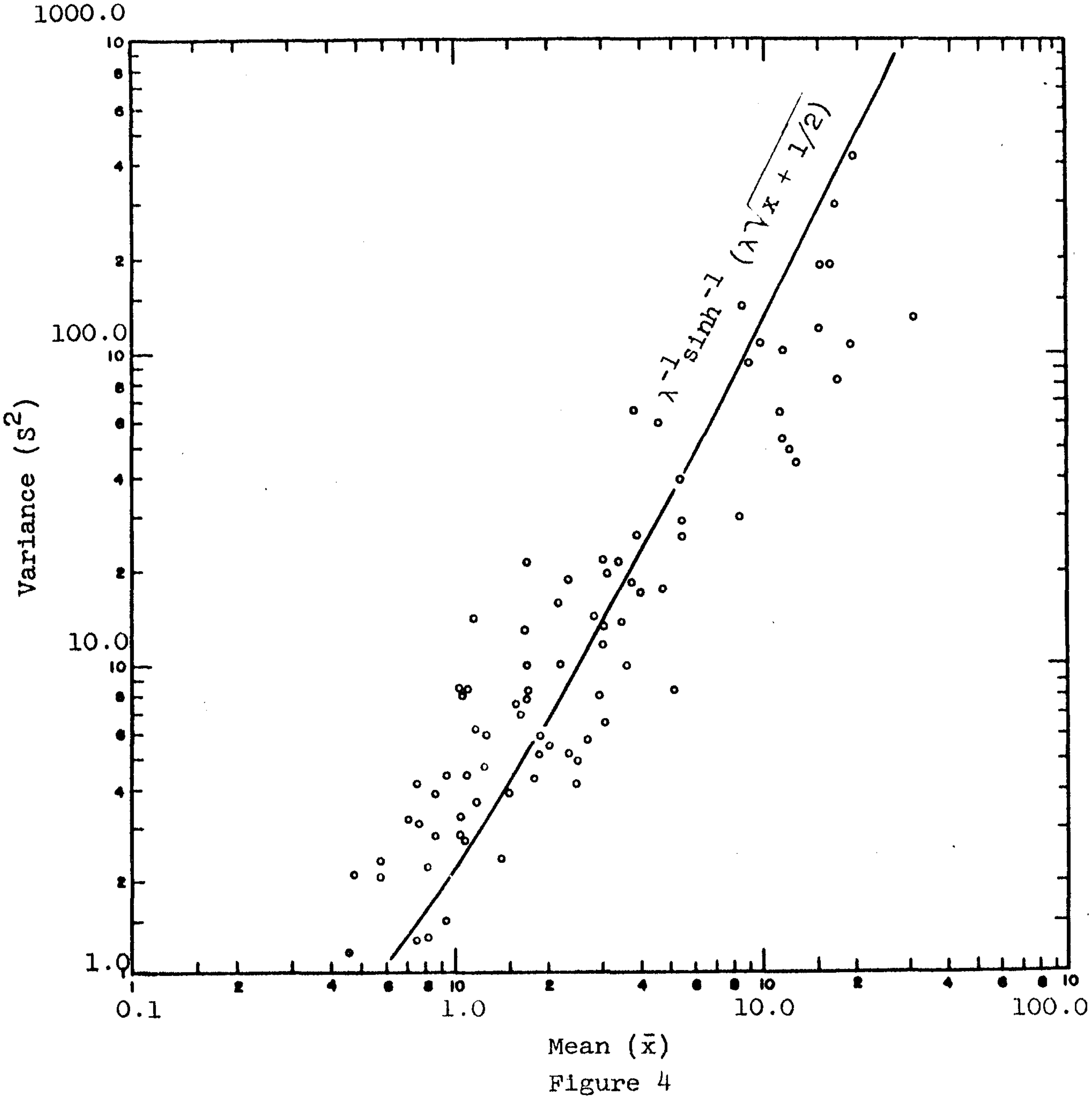

1000.0
100.0
10.0
1.0
Variance (s^2)
0.1
1.0
10.0
100.0
Mean ($\bar{x}$)
$\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$

Figure 4

Table 2: Results of the Analyses of Variance of Transformed Bite Data

SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMED DATA* 1 Mean Square	1 F Value	2 Mean Square	2 F Value
Day, D	1	0.342065		0.331695	
Strain, S	1	0.265038		0.241103	
Error (a)	1	0.271097		0.225376	
Circle, C	3	1.020387	50.5 **	1.199593	71.9 **
C x S	3	0.044502	2.20	0.044751	2.68
Time Period, T	9	0.670518	33.2 **	0.623410	37.4 **
T x S	9	0.115344	5.70**	0.108279	6.49**
T x C	27	0.044332	2.19**	0.043864	2.63**
T x C x S	27	0.011870	0.587	0.014230	0.853
Error (b)	78	0.020225		0.016677	
Total	159				

*1 denotes the data resulting from the $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ transformation of total bites received at a circle during a given time period; and 2 denotes the data resulting from the $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ transformation of total bites received during a given time period for Circles A, B, and D, and one-half the total bites received at Circle C.

** Significant at the 1.0 per cent level.

As shown by the F values in Table 2, the second analysis, adjusting for the augmented sampling on Circle C, was the more sensitive. Both analyses, however, showed circle, time period, T x S, and T x C to be highly significant.

Using the second set of transformed data, a further investigation of circle, T x S, and T x C was made in the following way. From the analysis of the transformed values:

(5) $$y_{ijk} = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x_{ijk} + 1/2}) , \quad \lambda = 1.0$$

$$= \sinh^{-1}\sqrt{x_{ijk} + 1/2},$$

where x_{ijk} equals the total number of bites received by the i-th host at the j-th circle during the k-th time period. The mean values, $\bar{y}_{jk}$, of the transformed variables were obtained. These mean values are related to the estimated true average number of bites received at the j-th circle during the k-th time period, $\hat{m}_{jk}$, by

(6) $$y_{jk} = \sinh^{-1}\sqrt{\hat{m}_{jk} + 1/2}, \text{ hence}$$

$$\hat{m}_{jk} = (\sinh \bar{y}_{jk})^2 - 1/2.$$

Relationships between $\bar{y}_{jk}$ and circle radius R_j, were sought. The best simple relationship found was:

(7) $$\bar{y}_{jk} = a - b \log_e R_j,$$

where a and b are regression constants determined by the method of least squares.

(8) Then,
$$\hat{m}_{jk} = \left[\sinh (a - b \log_e R_j)\right]^2 - 1/2,$$

(9) $$= \left[1/2(e^{a-b \log_e R_j} - e^{-a+b \log_e R_j})\right]^2 - 1/2,$$

(10) $$= \left[1/2(e^a e^{-b \log_e R_j} - e^{-a} e^{b \log_e R_j})\right]^2 - 1/2$$

$$(11) \qquad = \left[1/2\left(e^{a} R_j^{-b} - e^{-a} R_j^{b}\right)\right]^2 - 1/2,$$

which can be approximated by:

$$(12) \qquad \hat{m}_{jk} = \frac{e^{2a}}{4R_j^{2b}} - 1/2,$$

since $e^{-a}\left(R_j^{b}\right)$ is small relative to $e^{a}\left(R_j^{-b}\right)$.

For each 5-minute time period, the transformed data were summed with respect to circle and strain. These values were then fitted to the above regression model, and the average values of the various a's and b's determined. Subsequently, for each strain, the true average number of bites was estimated for each circle during the various time periods. These latter values are presented in Table 3.

Table 3: Estimated True Average Number of Bites of A and B Strain at the Various Circles During Given Time Periods.

STRAIN	TIME PERIOD (Minutes)	ESTIMATED TRUE AVERAGE NUMBER OF BITES AT INDICATED CIRCLE DURING GIVEN TIME PERIOD			
		Circle A (100 feet)	Circle B (200 feet)	Circle C (300 feet)	Circle D (400 feet)
A	0 - 5	123	15	4	1
	5 - 10	145	26	10	5
	10 - 15	119	28	12	7
	15 - 20	120	29	12	7
	20 - 25	82	20	8	4
	25 - 30	39	15	8	5
	30 - 35	22	10	6	4
	35 - 40	16	8	5	4
	40 - 45	8	6	5	4
	45 - 50	6	4	4	3
B	0 - 5	99	25	11	6
	5 - 10	132	35	16	9
	10 - 15	99	32	16	10
	15 - 20	59	21	12	8
	20 - 25	39	19	13	10
	25 - 30	19	9	5	4
	30 - 35	9	6	5	4
	35 - 40	2	2	2	2
	40 - 45	4	2	1	1
	45 - 50	0	0	0	0

As shown in Table 3, the expected number of A strain bites at Circle A during each of the various time periods is greater than that for B strain; however, the difference, in general, is not appreciable. At Circles B, C, and D, there appears to be no important difference between the number of bites. It was, therefore, concluded that the spatial dispersion of the two strains was comparable, as indicated by the nonsignificance of the C x S interaction.

The significance of the $T \times S$ interaction indicates a difference between strains with respect to the temporal dispersion characteristics. Generally speaking, the biting activity of strain B appeared to exhibit a more pronounced peak in time and a slightly earlier decline.

Unfortunately, no satisfactory model for the characterization of the biting activity as a function of time has been found by the authors. It is hoped that such a model may yet be developed.

COMPARISON OF APPROACHES TO OBTAINING A TRANSFORMATION MATRIX EFFECTING A FIT TO A FACTOR SOLUTION OBTAINED IN A DIFFERENT SAMPLE

Cecil D. Johnson
U. S. Army Personnel Research Office

BACKGROUND. The importance of physical proficiency measures to the selection and evaluation of Army personnel can scarcely be questioned. Examination of the duty assignments prevalent in various Army jobs indicates clearly that physical strength, endurance and coordination are often highly important factors in job success. At the United States Military Academy in particular, considerable attention has been devoted to physical training and to the measurement of various aspects of physical abilities or physical proficiency among cadets at West Point. Various tests of physical proficiency were introduced in the physical aptitude entrance examination procedure or studied for possible use. They have been examined both as individual measures and as component parts of various batteries. Several factor analyses of large batteries of physical proficiency measures, physical education, grades, and other variables were accomplished in previous studies. These studies had as their objective the identification of basic underlying physical ability variables that possess the simplifying statistical characteristics frequently referred to as simple structure. These basic variables, or factors, aid in understanding the nature of the scores, in eliminating duplicating measures, and in suggesting new tests.

The several factor solutions available for comparison contain numerous variables in common, other similar variables (as when a 25-yard dash is substituted for a 30-yard dash), and still other variables which are unique for a particular solution. This paper considers several methods for comparing solutions obtained in these separate studies involving physical proficiency and related measures.

The problem of approximating in a second sample, a rotated factor solution originally obtained in a previously analyzed sample is also present in another Army Personnel Research Office research study currently in the final computing phase. This study involved thirty-one psychological tests. Some of these tests are measures of intellectual ability, others are measures of cognitive information, and others are non-cognitive measures in the "personality" domain. As is typical with Personnel Research Office factor analysis studies, the objective was the identification of constructs which would predict the performance of soldiers on the job. In this case the job was that of an enlisted Infantryman and the measure of performance was obtained from ratings by superiors and peers at the close of maneuvers in

Germany. The tests had been administered to the enlisted men on their entry into the Army.

An initial principal component factor solution was transformed by an orthogonal matrix so as to provide simple structure. The initial factor solution can be described as a matrix whose elements are the correlations between the tests and standard length orthogonal reference vectors. This solution usually provides parsimony in that a relatively small number of reference vectors is needed to closely approximate the test correlation matrix when the factor matrix is post-multiplied by its transpose. However, the psychologist wants the reference vectors, or factors, to have additional properties implied by the concept of a simple structure. If simple structure is present among the reference vectors, each reference vector has high correlations with a few tests and approximately zero correlations with the remainder. Furthermore, the tests with which a particular reference factor has a high relationship will be relatively independent of the other reference vectors. It is apparent that the presence of simple structure permits the psychologist to interpret the reference vectors in terms of his test, and if the orthogonality of the reference vectors is retained, as when the transformation matrix is orthogonal, all the original parsimony of the initial principal component factor solution is retained. Psychologists usually refer to the process of transforming a solution to simple structure as rotation, and call the transformed solution a "rotated" solution.

In the factor analysis of psychological tests described above, the rotated solution, when extended to the rating variables, displayed a very interesting relationship between the rotated factors and the performance measures. One cognitive factor and one non-cognitive factor predicted performance while all other factors had a zero relationship with performance. It became a matter of considerable interest to determine whether these relationships could be verified in an independent sample where properties of the sample had not been used to determine the particular transformation used to obtain the rotated factor solution. Both factors retained their validity in the cross (independent) sample, but an additional factor (previously non-valid) also displayed a smaller amount of validity. In this study the factor validities in the first sample were fairly well replicated in the second sample.

Thus, both studies, the one involving physical proficiency variables and the one involving psychological tests, require an initial factor solution in a second independent sample, the transformation of this solution to one approximating the rotated solution in the first sample, and finally, the extension* of the transformed solution to non-overlapping predictor variables and/or criterion variables. The criterion variables may well overlap across the two studies but should be withheld from the initial factor analysis for two reasons:

(1) It is desirable that the factors be entirely defined by predictor variables.

(2) The validity of the transformed factors are being determined in the independent, or cross sample. Thus, the definition of the factors in the cross sample must be independent of the criterion variables.

B. F. Green has reported a method for computing an orthogonal transformation matrix which will minimize the sum of squares of the differences between the transformed matrix and the matrix to be fitted. However, his derivation does not generalize so as to provide an orthonormal transformation that can utilize more reference vectors in the cross sample than are in the matrix to be fitted.

If the investigator is confident that the initial cross sample factor solution does not have a rank which exceeds the rank of the solution to be fitted, this orthogonal transformation is clearly suitable. On the other hand, if in the cross sample there is likely to be considerable variance common to two or three variables that is not explained by the more general common factors utilized in the initial sample, the advantages of an orthonormal solution become apparent.

APPROACH AND RESULTS. Thus, in obtaining the transformation matrix necessary for fitting K factors, the investigator has a choice of using a method which obtains the best orthogonal transformation matrix applicable to the first K factors, or he can choose to use a non-square orthonormal transformation matrix which can be applied to a full factorization, i.e., to as many factors as there are variables. The first method, using an orthogonal transformation, requires the fitted solution to reproduce

---

*This factor extension is accomplished by post multiplying the m x n matrix of correlation coefficients (between the overlapping and non-overlapping) by $A_y D_y^{-1/2}$, where R_y is the matrix of correlation coefficients among the n overlapping variables and $A'_y R_y A_y = D_y$; $A'_y = A_y^{-1}$; D_y = eigen values.

the cross sample correlation matrix to the full extent possible with a principal component solution. The second method, using an orthonormal transformation, provides a less exact reproduction of the correlation matrix, but permits a better fit to the reference factors -- if the dimensionality of the experimental variable space exceeds the number of factors.

Since in the physical proficiency study the number of common variables, for some of the comparisons, was large as compared to the number of factors being fitted, and the initial solutions had been obtained using a correlation matrix involving even more variables, the non-square, orthonormal transformation was utilized. A description of this technique is provided in the hand-out.

The three rotated factor solutions to which the initial solutions in the cross-samples were fitted, tended to have smaller communalities than the four cross-sample fitted solutions. This is possibly explained by two things. The methods of obtaining the initial solutions (that were subsequently rotated in the reference samples) were less efficient that the principle component method used for the initial cross sample solutions. Also, the initial factor solutions in the first sample were obtained to span the non-common variables as well, whereas the cross sample solutions were obtained on the common variables only.

In comparing the use of the orthonormal as compared to an orthogonal transformation, it becomes a trade off between the better fit to the inter-correlation matrix obtained by using an orthogonal transformation and the better fit to the rotated factor solution possible under certain circumstances with the orthonormal transformation. The differences between the two methods in regard to fitting the rotated factor tend to diminish as the number of factors involved increases. On the other hand, the advantage possessed by the orthogonally transformed solution in reproducing the inter-correlation matrix increases. Thus it is clear that the value of the orthonormal transformation as compared to the orthogonal transformation is least likely when the number of factors in the initial rotated solution is large. However, the number and nature of the non-overlapping variables in the two studies is also important.

Since the extension of the transformed solution to the non-common variables is an important aspect of these studies, the reproduction of the intercorrelations between the extended factors and these additional variables is an important consideration.

It is interesting that the advantage of the orthogonally transformed solution for reproducing the intercorrelation matrix did not always hold among the non-overlapping variables. This underlines the fact that the advantage of the orthogonal transformation matrix in reproducing the correlation matrix in the cross sample is partly due to its more efficient capitalization on sampling error. This sampling error effect is further underlined by the fact that for the orthonormal transformed solution, for one sample, the elements of the residual matrix involving the non-common variables were smaller than the elements of the corresponding matrix involving the common variables. This was, of course, just the opposite of the results obtained from the orthogonally transformed factor solution. However, while the initial unrotated solution extended to the non-common variables, necessarily possesses the maximization properties of the initial principal component solution for only the common variables, the advantage, while reduced, was still present for non-common variables in the larger samples.

The two following questions were raised at the conclusion of the two USAPRO presentations:

(1) Have factors (i.e., factor pure tests) proved to be good predictors of Army performance criteria?

(2) What is the advantage, for prediction, of using orthogonal predictors over the original correlated predictors if optimal weights are applied?

The two questions are closely related in that they are both concerned with the immediate application of factor analysis results to the practical problem of predicting personnel performance. USAPRO has considerable research evidence indicating that tests developed to measure factors do not predict performance as well as factorially complex tests developed to predict a specific Army performance measure. We have very little evidence bearing on our own factor measures, since, on theoretical grounds, we have not expected factor measures to have immediate use as predictors. Factors are useful constructs because of their simplified (i.e., more easily understood) relationships with psychological or physiological measures and the other factors. Thus, factor scores are useful in experimentally testing hypotheses relating carefully defined psychological content of a measure to human performance, and the factor concept has general usefulness for the better understanding of the psychological content of a battery of tests. It is not expected that factors will have immediate usefulness as operational predictors.

APPENDIX I

Formulae and Notation

I. Certain letters will be used consistently to denote specific kinds of matrices. Different matrices of the same type will be discriminated by their subscripts.

R	a gramian matrix whose elements are product moment correlation coefficients.
P	a principal component factor solution.
A	an orthogonal eigen vector matrix derived from a gramian matrix.
D	eigen value matrix.
F	a factor solution other than a principal component factor solution.
T	a transformation matrix whose elements are cosines of the angles between reference vectors (factors).

II. The following formulae relate several of the above matrices:

$$A'RA = D, \quad AD^{1/2} = P, \quad PP' = R, \quad P'P = D$$

$$FF' = R, \quad F'F \neq D$$

$$F_Y T_1 = F_X \,; \quad T_1 = (F'_y F_y)^{-1} F'_y F_x$$

III.

Table 1

A Sectioned Matrix Whose Elements are Projections* Involving the Row and Column Variables

(As computed in the second sample)

		Rotated Factors from 1st Sample	PC Factors in Space Defined by Rotated Factors from 1st Sample	Experimental Variables
		$F_{x1} \ldots\ldots F_{xk}$	$F_{o1} \ldots\ldots F_{ok}$	$Y_1 \ldots\ldots Y_n$
Rotated Factors From First Sample ($k \leq n$)	F_{x1} ⋮ F_{xk}	R_L $R_L = T_2' T_2$ $P_y T_2 = F_x$ $T_2 = D_y^{-\frac{1}{2}} A'$ **	P_{XL} $P_{xL} = A_L D_L^{\frac{1}{2}}$ $P_{xL} = R_L A_L D_L^{-\frac{1}{2}}$ $A_L' R_L A_L = D_L$	F_X'
Principal Component Factors (PC Factorization of R_y)	P_{y1} ⋮ P_{yn}	T_2 $T_2 = D_y^{-\frac{1}{2}} A_y' F_x$ $T_2 = P_y^{-1} F_x$	T_3 $T_3 = T_2 A_L D_L^{-\frac{1}{2}}$	P_Y $P_y' = D_y^{\frac{1}{2}} A_y'$ $A_y' R_y A_y = D_y$
Experimental Variables	Y_1 ⋮ Y_n	F_X $F_x = F_y T_2$ $F_x = P_x A_x'$ $A_x (F_x' F_x) A_x = D_x$	F_{YL} $F_{yL} = F_y T_3$ $F_{yL} = F_x A_L D_L^{-\frac{1}{2}}$	R_Y

* These projections are cosines when both row and column vectors are of unit length.

** If $k = n$, $R_L = F_x' R_y^{-1} F_x$. However, in almost all practical situations, k, the number of rotated factors, will be considerably smaller than n, the number of variables.

IV. Green's Procedure

Problem: In order to fit a factor solution of k factors (i. e., P_{yk}) in the second sample to a rotated factor solution, F_x, in the first sample (i. e., $P_{yk}T_o = F_x$), compute T_o such that tr $(P_{yk}T_o - F_x)'(P_{yk}T_o - F_x)$ is minimized, while meeting the side constraint that $T_o' = T_o^{-1}$.

Solution:

(a)* $T_o = (P_{yk}' F_x F_x' P_{yk})^{-1/2} P_{yk}' F_x$, P_{yk} contains the k columns of P_y, the complete principal component factorization of R_y, corresponding to the k larger roots of D_y. $P_y = A_y D_y^{1/2}$ where $A_y' R_y A_y = D_y$ and R_y is the product moment correlation matrix for the experimental variables in the second sample.

(b) T_o can also be computed from the least square transformation,

$$T_1 = D_{yk}^{-1} P_{yk}' F_x$$

$$T_o = (D_y T_1 T_1' D_y)^{-1/2} D_y T_1$$

(c) A slightly different orthogonal transformation matrix, T_p, can be derived by directly orthogonalizing the T_1 as follows:

$$T_1'X = T_p' \quad , \quad T_p'X = T_1^{-1}$$

$$T_1'X = T_1^{-1} X^{-1} \quad , \quad X = (T_1 T_1')^{-1/2}$$

---

*Green, B. F. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, 1952, 17, 429-440.

$$T_p = (T_1 T_1')^{-1/2} T_1$$

Properties of T_o:

The orthogonally transformed factor solution, $P_{yk} T_o$, retains the maximum reproduction of R_y (i.e., $P_{yk} T_o T_o' P_{yk}' = P_{yk} P_{yk}'$). When communalities are substituted for ones in the diagonals of R_y and if the rank of R_y becomes K, this is undoubtedly the best procedure for fitting F_x.

V. Alternative Procedure

A non-square (orthonormal) transformation matrix permitting the full utilization of P_y can be developed as follows:

a. Whereas T_1 in the previous model provided a least square fit, T_2 in the model below provides an exact fit (since P_y has an inverse while P_{yk} does not when $k < n$.)

$$P_y T_2 = F_x ; \quad T_2 = P_y^{-1} \cdot F_x = D_y^{-1/2} \; A_y' F_x$$

b. The factor matrix F_x, computed and rotated in Sample 1, cannot usually be obtained by an orthogonal or orthonormal transformation of P_y, computed in Sample 2. The transformation of $P_y T_2$ into a solution within an orthogonal frame F_{yL}, can be accomplished as follows:

$$F_{yL} = P_y T_2 \; (A_L D_L^{-1/2})$$

c. While the matrix F_{yL} contains K column vectors (factors) spanning the same space as the oblique factors in F_x, the orthogonalization was not accomplished in such a way as to maximize the fit of F_{yL} to F_x. The additional transformation required to effect this fit can be accomplished by using Green's procedure to obtain orthogonal matrix T_{o2} as follows:

$$F_{yL} T_{o2} = F_x$$

$$T_{o2} = (F'_{yL} F_x F'_x F_{yL})^{-1/2} F'_{yL} F_x$$

$$T_{o2} = \left[D_L^{-1/2} A'_L (F'_x F_x)^2 A_L D_L^{-1/2} \right]^{-1/2} D_L^{-1/2} A'_L (F'_x F_x)$$

d. Thus the ortnonormal transformation matrix T_m which minimizes the trace $(P_y T_m - F_x)'(P_y T_m - F_x)$, where T_m is an $n \times k$ orthonormal matrix is,

$$T_m = T_2 A_L D_L^{-1/2} \left[D_L^{-1/2} A'_L (F'_x F_x)^2 A_L D_L^{-1/2} \right]^{-1/2} D_L^{-1/2} A'_L (F'_x F_x)$$

VI. An orthonormal transformation matrix T_n providing a least square fit of T_n to T_2, as compared to the fit of F_{yL} to F_x in Part V, can be provided as follows:

a. The conversion of T_2 to an orthonormal matrix, T_3, spanning exactly the same space can be accomplished as follows:

$$T'_2 T_2 = R_L$$

$$T_3 = T_2 A_L D_L^{-1/2} \quad ; \quad T'_3 T_3 = I$$

b. While T_3 is an orthonormal matrix, it is not the orthonormal matrix with the best least square fit to T_2. There is still the need to minimize the trace of $(T_3T_{o2} - T_2)'(T_3T_{o2} - T_2)$, where T_{o2} is an orthogonal matrix. This can be accomplished by making use of Green's procedure described under IV(a).

$$T_{o2} = (T_3'T_2T_2'T_3)^{-1/2}T_3'T_2$$

$$T'_3T_2 = P_{xL}^{-1}T_2'T_2 = P_{xL}^{-1}R_L = P'_{xL} \; ; \quad \text{since } P_{xL}^{-1} = D_L^{-1/2}A_L'$$

$$T_{o_2} = (P'_{xL}\,P_{xL})^{-1/2}P'_{xL} = D_L^{-1/2}\;P'_{xL}$$

$$T_{o2} = A_L'$$

c. Thus the $n \times k$ orthonormal transformation matrix T_n which minimizes (in the least square sense) trace $(T_n - T_2)'$ $(T_n - T_2)$ is equal to,

$$T_3T_{o2} = T_2A_LD_L^{-1/2}A_L' = T_2\,(T'_2T_2)^{-1/2}$$

$$T_n = T_2\,(T'_2T_2)^{-1/2}$$

Note the similarity in the form of the computing formulae used to describe T_n and T_p. $T_p = (T_1T_1')^{-1/2}\,T_1$.

APPENDIX II

The Comparison of Transformation Matrices

I. Method

The variables included in the rotated factor solution in the first sample are designated by x and the rotated factor solution by F_x. The same (i.e., overlapping) variables in the second sample are designated by y and the non-overlapping variables by z. The factorization of R_y and R_z are accomplished as $P_y = R_y A_y D_y^{-1/2}$ and $F_z = R_{yz} A_y D_y^{-1/2}$. The transformed factor matrices in Sample 2 are $F_{yr} = P_y T$ and $F_{zr} = F_z T$. Each transformation matrix, T, computed by the methods described in Appendix I, is evaluated by determining the fit of F_{yr} to F_x and the reproduction of R_y by $F_{yr}F'_{yr}$, R_{yz} by $F_{zr}F'_{yr}$ and R_z by $F_{zr}F'_{zr}$. This is determined by comparing (for the different T matrices) the traces of the following product matrices:

$$(F_{yr} - F_x)'(F_{yr} - F_x)\ ,\ (F_{zr}F'_{yr} - R_{yz})'(F_{zr}F'_{yr} - R_{yz})$$

$(F_{zr}F'_{zr} - R_z)'(F_{zr}F'_{zr} - R_z)$, and, after setting diagonal elements of $F_{yr}F'_{yr}$ and R_y equal to zero, $(F_{yr}F'_{yr} - R_y)'(F_{yr}F'_{yr} - R_y)$.

II. Results

The sums of squares of the residual matrices, computed as the traces of the matrices indicated in Part I above, are provided in Table 1 for a study involving physical proficiency measures. The x sample consisted of 254 West Point Cadets of the class of 1949. The y sample contained 294 West Point Cadets of the class of 1964. Table 2 relates to a study involving the following x and y variables (in samples 1 and 2 respectively): 15 "Personality" tests, 9 information tests, and 8 mental aptitude tests. Five rating variables based on performance as Infantryman, make up the z variables. Sample one (x variables) had 550 examinees and sample two(y variables) had 375 examinees.

The rank ordering of the magnitudes for the various entries in Tables 1 and 2 can be readily predicted from the algebraic formulations of the T's. The relatively efficiency for fitting F_x, going from high to low, is T_m, T_n, T_o, T_p. The relative efficiency for reproducing the R matrices is the same for all T's which are either orthogonal or capable of being linked by an orthogonal transformation. Thus all the orthogonal T matrices have more efficiency for reproducing R_y, when applied to PC solutions of R_y, than do the orthonormal transformations.

Table 1

Comparison of Transformation Matrices Computed on a Sample of 294 Examinees

(Physical Proficiency Variables)

Residual Matrices	Number of Elements Contributing to the Sums of Squares Reported as Entries in this Table	Total Sums of Squares of Elements in Residual Matrices						
		ORTHOGONAL Transformations				ORTHONORMAL Transformations		
		Communalities in Diagonals of R_y		Ones in Diagonals of R_y		Ones in Diagonals of R_y		Communalities in Diagonals of R_y
		T_o	T_p	T_o	T_p	T_m	T_n	T_m
$(F_{yr} - F_x)$	19 x 9 = 171	2.1381	3.1934	5.2665	5.7615	1.0943	1.5233	1.6758
$(F_{yr}F'_{yr} - R_y)$, other than diagonal elements	19 x 18 = 342	.2268	.2268	.8496	.8496	5.1851	5.1851	1.2382
$(F_{zr}F'_{yr} - R_{zy})$	19 x 11 = 209	.2953	.2953	.4129	.4129	3.0342	3.0342	.8050
$(F_{zr}F'_{zr} - R_z)$	$(11)^2$ = 121	2.1925	2.1925	3.6127	3.6127	6.2471	6.2471	2.3205

This page intentionally left blank.

Table 2

Comparison of Transformation Matrices Computed on a Sample of 375 Examinees

(Personality, Mental Aptitude, Information, and Rating Variables)

Residual Matrices	Number of Elements Contributing to the Sums of Squares Reported as Entries in this Table	Total Sums of Squares of Elements in Residual Matrices						
		ORTHOGONAL Transformations				ORTHONORMAL Transformations		
		Communalities in Diagonals of R_y		Ones in Diagonals of R_y		Ones in Diagonals of R_y		Communalities in Diagonals of R_y
		T_o	T_p	T_o	T_p	T_n	T_m	T_m
$(F_{yr} - F_x)$	32 x 8 = 256	1.3258	1.3667	1.8747	1.9054	.2133	.1790	1.4270
$(F_{yr}F'_{yr} - R_y)$, other than diagonal elements	32 x 31 = 992	.6612	.6612	2.0005	2.0005	3.5448	3.5448	4.4173
$(F_{zr}F'_{yr} - R_{zy})$	32 x 5 = 160	.2162	.2162	.1965	.1965	.3405	.3405	.4163
$(F_{zr}F'_{zr} - R_z)$	$(5)^2$ = 25	8.898	8.898	9.1654	9.1654	9.8459	9.8459	8.0030

This page intentionally left blank.

SOME LEAST-SQUARES TRANSFORMATIONS OF REGRESSION ESTIMATORS OF ORTHOGONAL FACTORS

Emil F. Heermann
U. S. Army Personnel Research Office

This paper provides a brief introduction to the factor analysis model used by psychologists and presents some methods for transforming regression estimators of orthogonal factors to a more useful form.

I. <u>THE FACTOR ANALYSIS MODEL.</u> Factor analysis is a mathematical method for representing n correlated measures obtained on N individuals by means of a linear combination of hypothetical measures, called factors. Psychologists draw a distinction between common factors, which are related to two or more measures, and specific factors which are related to only one of the original measures. In general, then, the linear model postulates n specific factors and r common factors where r is less than the number of measures n. The sampling theory for such a model is not yet thoroughly understood and this paper will confine itself to algebraic rather than statistical considerations.

The linear model of factor analysis can be most conveniently presented in the form of matrix equations:

(1) $$Z = Z_C F' + Z_S S' \text{ , where}$$

$Z = N \times n$ matrix of observed measures (zero mean, unit variances)

$Z_C = N \times r$ matrix of common factor scores

$F' = r \times N$ matrix of common factor weights

$Z_s = N \times n$ matrix of specific factor scores

$S' = n \times n$ diagonal matrix of specific factor weights

Terms on the right side of the equation are all unknown. In (1) it is assumed that:

(2) $$\frac{1}{N} Z'_c Z_c = I_r$$

(3) $$\frac{1}{N} Z'_s Z_s = I_n$$

(4) $$\frac{1}{N} Z'_c Z_s = 0$$

These assumptions allow us to find F' and S' from the observed measures, but before we examine this procedure it is important to indicate a rather serious difficulty with our linear model. When we postulate $n + r$ factors to explain n measures this invariably means that there exist multiple solutions for an individual's factor scores. Solutions exist because the linear equations are consistent but the rank of the coefficient matrix does not allow unique solutions. To get around this difficulty psychologists have utilized least-squares procedures for estimating the factor scores from the observed measures, and we shall examine the defects of these estimators at a later point.

We now return to the problem of finding F' and S' from the observed data. Because of (2), (3) and (4) we may write:

$$R = \frac{1}{N} Z'Z = FF' + SS', \text{ or}$$

$$R - SS' = FF'$$

Since SS' is a diagonal matrix, we begin by estimating values for SS' and then obtain the eigenvectors and eigenvalues of

$R - SS' = A K A'$, where

A = n x r matrix of eigenvectors

K = r x r matrix of eigenvalues

Now, $F = AK^{1/2}$, and the sums of squares of the rows of F are used to obtain new values of SS', and the process is continued until $R - SS'$ is fit with the minimum rank F. This whole process may be thought of as finding a set of values for SS', such that $R - SS'$ is of minimum rank.

It will be noted that F is arbitrary up to an orthogonal transformation and it is necessary to postulate some method for finding a psychologically meaningful F. Psychologists follow L. L. Thurstone here and attempt to transform the arbitrary F, so that it approximates "simple structure", i. e., has a maximal number of zero or near-zero entries. To accomplish this we find some orthogonal transformation, λ , such that

$$F\lambda = F_R$$

where $\lambda'\lambda = I$, and

F_r approximates "simple structure". A considerable number of computer programs exist which determine by analytical means the best transformation λ.

II. <u>ESTIMATION OF FACTOR SCORES.</u> Assuming now that F_r is adequately fixed, our problem is then to find the values of Z_c and Z_s that will satisfy the linear model specified in (1). Since an infinite set of such factor scores exist, psychologists commonly turn to a regression method for estimating Z_c for a fixed Z_s. We are primarily interested here in finding values for Z_c, and the least-squares solution for Z_c is easily found to be

$$ZR^{-1}F_R = Z_c \; .$$

But unlike the factor scores Z_c, these estimators are intercorrelated, because

(5) $$\frac{1}{N}\hat{Z}'_c\hat{Z}_c = F'_R R^{-1} F_R,$$ the

covariance matrix of the least-squares estimator is not diagonal. Adjusting the covariance matrix so that we have an intercorrelation matrix gives:

$$R_L = D_e^{-1} F'_R R^{-1} F_R D_e^{-1} ,$$

where

D_e = r x r diagonal matrix, formed from the square roots of diagonal entries of (5), the covariance matrix.

A further practical difficulty is that regression estimators are not univocal, i. e., each estimator is correlated with more than one common factor. This can be seen by examining the matrix of correlations between the factor scores and the least-squares estimators, which is

$$R_{fL} = D_e R_L.$$

It is not too difficult to derive transformations of the beta weights used in the regression estimators that remove these defects, but it is not possible by means of a single transformation to simultaneously remove both defects.

To adjust for non-orthogonality, we first find an arbitrary set of orthogonal vectors that serves as a vector basis for the regression estimators. To do this we find an $r \times r$ T, such that

$$R_L = TT'$$

The elements of T represent the correlations of the least-squares estimators with the r orthogonal axes. The correlations of the factors with the orthogonal axes are given by

$$R_{fT} = D_e T,$$

and to find the best set of orthogonal axes we must transform T so that $D_e T$ most closely approximates the diagonal matrix D_e. In matrix algebra, we wish to find a λ, such that if

$$R_{fT}\lambda - D_e = E, \text{ then}$$

$$\text{trace } E'E = \text{minimum, and}$$

$$\lambda' \lambda = I.$$

From a theorem proved by B. F. Green (1) we find

$$\Lambda = (T'D_e^4T)^{-1/2}\, T'\, D_e^2 \quad .$$

Now if the requirements of the psychologist are such that it is more important to have univocal rather than orthogonal estimators than a somewhat different procedure is required. In this case we determine so that if

$$R_{fT}\Lambda - D_e = E \, ,$$

$$\text{trace } E'E = \text{minimum, and}$$

$$\Lambda'_{.i}\Lambda_{.i} = 1, \; (i = 1, 2, \ldots r), \text{ where}$$

$$\Lambda_{.i} = i^{th} \text{ column of } \Lambda .$$

As can be seen the orthogonality restraint on Λ has been relaxed. The normal equations for these conditions are

(8) $$(T'D_e^2T - \Upsilon)\Lambda = T'D_e^2 \quad ,$$

where Υ = r x r matrix of Lagrangian Multipliers. The above equation does not allow us to find a matrix expression for Λ which does not also involve Υ . Hence we pursue a simpler approach which involves a least-squares solution for Λ followed by imposing the restraints that each column of Λ have sums of squares equal to unity. With this approach we find that

$$\Lambda_n = T^{-1}D, \text{ where}$$

D is the diagonal matrix of constants needed to adjust T^{-1} so that the column sums of squares equal unity. When the least-squares estimators are transformed by Λ_N we find that the matrix of correlations between Z_c and our transformed estimators is $R_{fN} = D_eD$, which is quite obviously a diagonal matrix.

III. AN APPLICATION OF THE TRANSFORMATIONS. In 1954 APRO carried out a factor analysis of visual acuity tests administered during dark adaption. The examinees in this experiment were 100 soldiers from Fort Myer, Virginia. The design of the experiment called for pre-adapting subjects to a high brightness level and then during the period of dark adaption the examinees were tested on visual acuity targets presented at scotopic, mesopic, and low photopic brightness. The various acuity tests (Modified Landolt Ring and Chevron Contrast Adoption tests) were presented in a modified Armed Forces Vision Test.

The 35 variable intercorrelation matrix obtained in this experiment yielded 8 orthogonal factors which were rotated to orthogonal simple structure. To illustrate the derivations presented in this paper, a submatrix was selected from the 35 x 8 complete factor matrix. Table 1 gives this submatrix.

Table 1

Illustrative Factor Matrix

Variable Name		I	II	III	IV
Landolt Scotopic	1	.84	.07	.13	.34
Landolt Low Photopic	2	.16	.85	.03	.13
Chevron Scotopic	3	.28	.32	.12	.75
Chevron Mesopic	4	.01	.03	.79	.29

Factor I was interpreted to represent "Rod-Adapted Resolution"; Factor II was interpreted to represent "Cone Adapted Resolution"; Factor III was interpreted to represent "Cone Adapted Brightness Discrimination", and Factor IV was interpreted as "Rod-Adapted Brightness Discrimination".

Table 2 compares the classic regression equations approach to estimating these four factors with the methods derived in this paper. As noted previously the least-squares estimators are intercorrelated and fail to be univocal. The univocal estimators achieve an ideal pattern of correlations with the factors but they are intercorrelated somewhat more than the least-squares estimators. The orthogonal estimators, while uncorrelated, are not univocal although they are closer to being univocal than the least-squares estimators. Inevitably then, choice of any one solution means that certain defects must be tolerated. From the results presented in Table 2 it would appear that the orthogonal estimators represent something of a compromise between the maximum validity of the least-squares estimators and purity of the univocal estimators.

(1) Green, B. F. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, 1952, 17 429-440.

This page intentionally left blank.

A RELIABILITY TEST METHOD FOR "ONE-SHOT" ITEMS

H. J. Langlie
Aeronutronic Division, Ford Motor Company

1. INTRODUCTION. As a result of many reliability problems which have plagued procuring agencies in the missile and space programs, increased emphasis has been placed on the development of improved reliability demonstration test methods. In this connection, the Army has advocated increased use of the "test-to-failure" concept to establish the existence of satisfactory margins of reliability with respect to critical factors. This philosophy has the advantage in that statistical statements can be made regarding reliability on the basis of relatively small samples. This paper will discuss the application of this general concept to a particular class of hardware.

In essence, the "test-to-failure" concept involves submitting a test specimen to an increasing environment or load stress until failure is detected. By observing the statistical behavior of the stresses at which failure occurs, the lower limit of stress below which the probability of failure is very small can be selected by using the mean and standard deviation of these data. Robert Lusser (reference 1) advocated the safety margin concept in interpreting these data. As is shown in Figure 1, the larger the value of k (which is the distance from the mean strength to the upper limit of the operating applied stress divided by the standard deviation of the data), the greater the reliability of the specimen with respect to the stress involved.

In establishing the reliability objectives for the Shillelagh Program, the Army Missile Command required that safety margins be demonstrated in the test laboratory for Shillelagh components with respect to critical environmental stress factors. Many of these components are of a "go-no-go" type such as thermal batteries, electrical relays, and other short-lived equipment items. In most cases, little information was available prior to test regarding the nature of the standard deviation of the distribution of strenghts for these parts. Furthermore, it was desired to perform a laboratory test involving a minimum number of samples. A review of attribute sensitivity testing techniques such as the Up and Down method (reference 2) and the Probit method (reference 3) indicated that these methods cannot be applied satisfactorily under the sample size and technical limitations imposed. As a result, a study was made to develop a method for selecting stress levels for testing which required no a'priori assumption regarding the standard deviation of the unknown strength distribution and could be performed satisfactorily with sample sizes of the order of fifteen or twenty.

This page intentionally left blank.

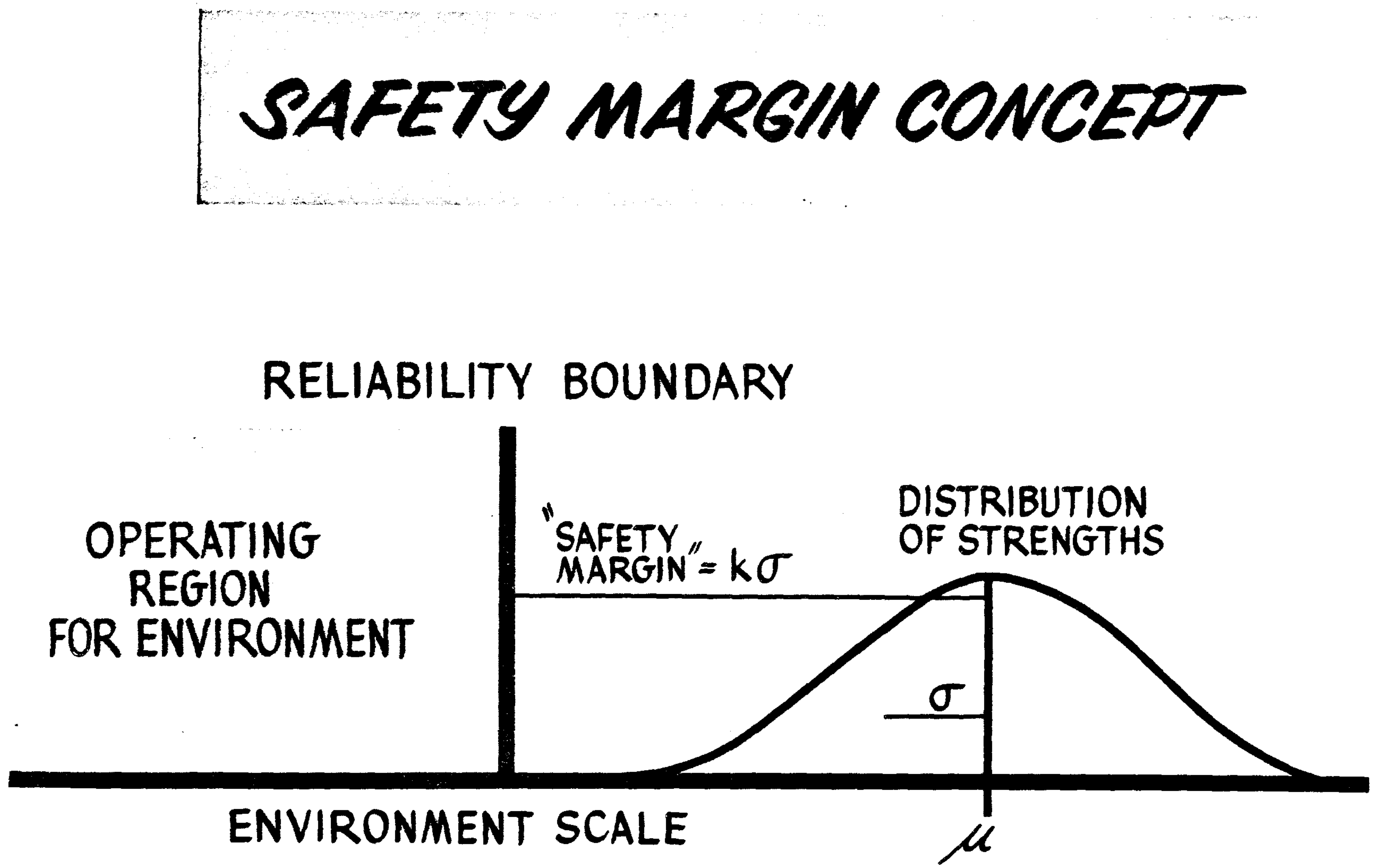

SAFETY MARGIN CONCEPT
RELIABILITY BOUNDARY
OPERATING REGION FOR ENVIRONMENT
"SAFETY MARGIN" = kσ
DISTRIBUTION OF STRENGTHS
σ
ENVIRONMENT SCALE
μ

FIGURE 1.

This page intentionally left blank.

Detailed empirical investigations were performed, using Monte Carlo methods with a high speed digital computer, to develop a satisfactory algorithm for determining successive stress test levels as the experiment proceeds. Exact maximum likelihood equations were used to calculate the statistical estimates, μ_e, and σ_e, of the population parameters, μ and σ, of the distribution of part strengths. By repeated simulations of experiments for sample sizes ranging from ten to 900, empirical curves were obtained showing the variance in the calculated estimates vs sample size. After this study was complete, a technical report was prepared (reference 4). The paper today will discuss some significant results contained in this report.

Since the completion of the study, several applications have been made in reliability testing electrical and mechanical components. One such application is presented in this paper by way of illustration.

2. <u>DISCUSSION</u>. In order to proceed with the discussion of the test method, a specific definition is given regarding the terms "stress" and "strength" as follows:

> <u>Stress</u> is a test factor, such as environmental level or force level which is applied to the test specimen. Operational stresses represent the mix of environmental or load conditions that can be expected to be imposed on a typical specimen during its life. During the conduct of a "one-shot" test, the stress represents the applied test factor which is varied in magnitude from specimen to specimen in a systematic manner.
>
> <u>Strength</u> is a property ascribed to a specimen such that if the stress imposed on the part is greater than the strength of the part, the part will fail. Conversely, if the stress is less than the strength of the part, the part will not fail.

Failure in the sense used above is a general term referring to unsatisfactory completion of function, out of tolerance performance, breakage, or other evidence of malfunction. For each test attempt wherein a stress is applied to a specimen, there is associated an <u>outcome</u> which is a binary variable: success or failure.

It is assumed that, given a homogenous sample of replicate specimens, the part strengths are distributed normally with an unknown mean and standard deviation. The purpose of the test method is to select stress levels in such a way as to generate outcomes which can then be used to calculate statistical estimates of the parameters of the distribution of part strengths.

3. PERFORMING THE TEST. In undertaking to perform the "one-shot" test, there are three steps to be taken which should be followed regardless of the nature of the application. They are:

1. Establish the criteria of failure, or acceptance.
2. Determine the test interval.
3. Select the stress levels.

The last of these three steps, selecting the stress levels, proceeds concurrent with the actual performance of the test. These three steps are discussed below.

3.1 Establishing Failure Criteria.

It is very important that painstaking care be given to the set of ground rules which will be followed for differentiating between a success or a failure. For purposes of reliability testing, this means that a careful enumeration should be made of all undesirable responses of a test specimen, such as particular modes of failure, out of tolerance performance, and any other mode of unacceptable product performance. These criteria are the very basis for establishing product assurance in the laboratory and should be reviewed and approved by all qualified parties having a technical interest in the product.

3.2 Determining the Test Interval.

In order to proceed with a generation of stress levels for testing purposes, it is necessary to choose a test interval which is used as a basis for the stress sequencing method. This interval should be selected large enough to include all possible ranges of strengths of the parts to be tested. This interval can be made conservatively large, since the "one-shot" method has been designed to cause the stress levels to be generated in the vicinity of interest (i.e., in the vicinity of the distribution of strengths) as the test proceeds. As a sample illustration, the range for a drop height test for glass containers designed to withstand say, a six inch drop, could be chosen to have a lower limit of zero and an upper limit of three feet. The method of analysis of the data is such that the particular choice of the endpoints of the test interval do not have an appreciable effect on the results for sample sizes of fifteen or more. In the event that the test interval turns out, as the test proceeds, to be inappropriately chosen, then the stress levels will tend to converge towards one limit or the other. In such an event, particularly in the case of reliability testing, convergence towards the lower level is usually indicative of a totally unsatisfactory product, whereas convergence towards the upper limit can be shown to be statistically acceptable by use of the likelihood ratio test.

In Figure 2 there is represented the results of an actual "one-shot" test on thermal batteries to determine the reliability with regard to high temperature. In this instance, the batteries were designed to perform reliably at 145°F. On the basis of conservative engineering judgement and some limited development test data, the lower limit was selected to be 100°F (the level at which all thermal batteries would be expected to perform satisfactorily) and the higher limit was selected to be 350°F (the level at which all thermal batteries would be expected to fail).

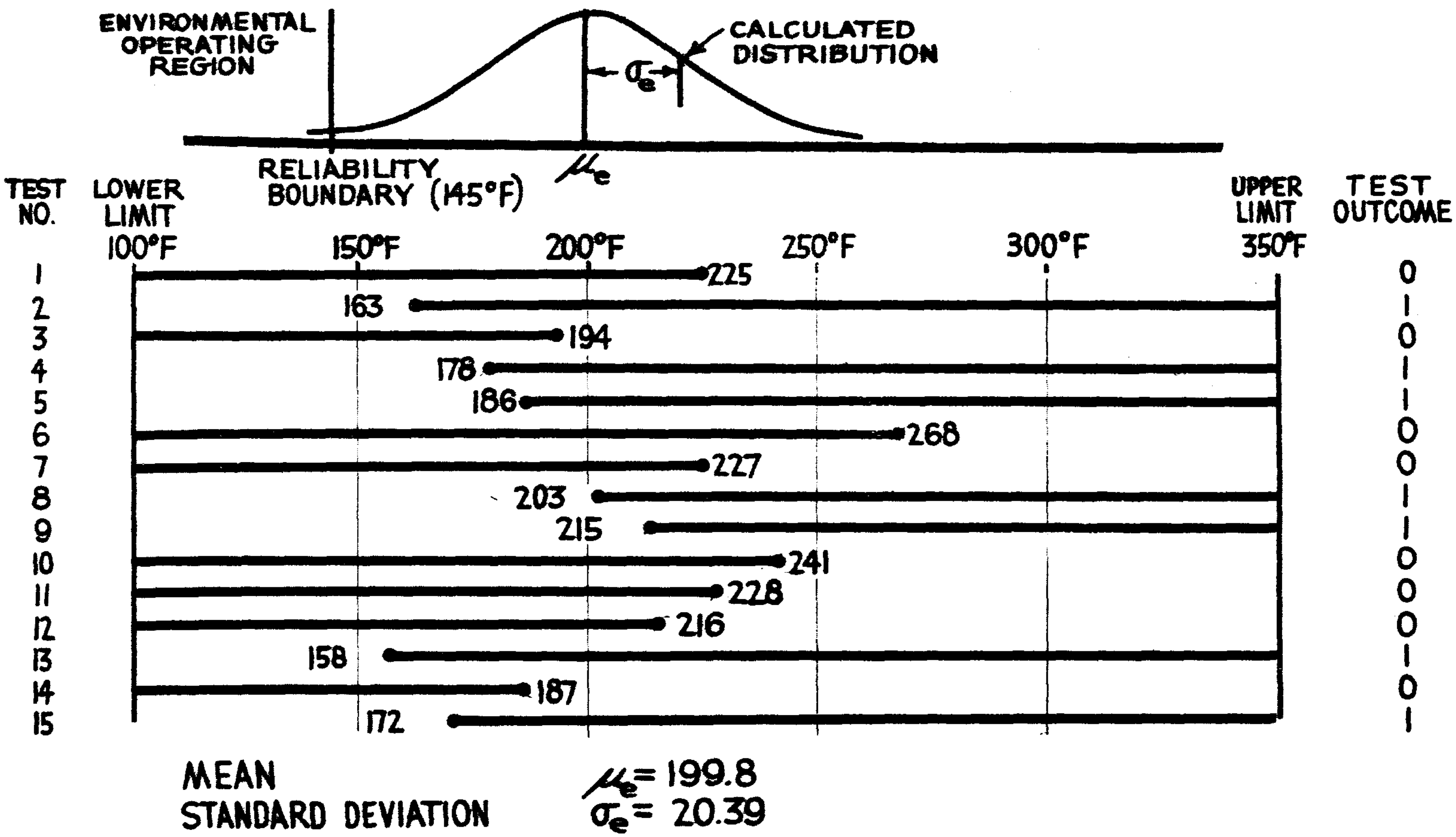
3.3 Selecting the Stress Levels.

Once the test interval and failure criteria have been established, the test commences by selecting the first stress level at the midpoint of the interval. After exposing the first specimen to this environmental level and activating it, a one or zero is recorded to indicate the outcome as a success or failure respectively (see Figure 2).

The general rule for obtaining the $(n+1)^{st}$ stress level, having completed n trials, is to work backward in the test sequence, starting at the n^{th} trial, until a previous trial (call is the p^{th} trial) is found such that there are as many successes as failures in the p^{th} through the n^{th} trials. The $(n+1)^{st}$ stress level is then obtained by averaging the n^{th} stress level with the p^{th} stress level. If there exists no previous stress level satisfying the requirement stated above, then the $(n+1)^{st}$ stress level is obtained by averaging the n^{th} stress level with the lower or upper stress limits of the test interval according to whether the n^{th} result was a failure or a success.

To illustrate, suppose it is desired to find the second stress level in Figure 2. Since there was only one previous observation (i.e., first unit failed) it is not possible to find a stress level where all intervening results even out. That is, the second stress level is obtained by averaging the first with the lower limit. To find the eighth stress level, it is observed that results from test 4 through 7 (i.e., the last four results) cancel each other out. Thus, the eighth stress level is obtained by averaging the fourth.

SAMPLE "ONE-SHOT" TEST



TEST-TO-FAILURE OF THERMAL BATTERIES IN TEMPERATURE

FIGURE 2.

stress level with the seventh.

As a final example, it is observed that after the twelfth test has been completed, there again exists no previous stress level for which the number of failures equals the number of successes. Since the twelfth test was a failure, the thirteenth stress level is obtained by averaging the twelfth stress level with the lower limit.

As an aid in identifying the important parameters of the test, the stress level is designated by the letter s and the outcome is designated by the letter u. The lower limit of the test interval is designated A and the upper limit is designated B. Upon the conclusion of the test, the stress values, $(s_1, s_2, \ldots, s_N)$, and the corresponding outcomes, $(u_1, u_2, \ldots, u_N)$, where N equals the test sample size, are used to perform a complete analysis for hardware reliability.

4. DERIVATION OF MAXIMUM LIKELIHOOD EQUATIONS. Consider the random sample $X = (x_1, x_2, \ldots, x_N)$ of N observations where the x_i are independent random variable from a Gaussian distribution, $g(x;\mu, \sigma)$, with mean μ and standard deviation σ. Consider also an N-dimensional vector $S = (s_1, s_2, \ldots s_N)$ where $A \leq s_i \leq B$. From this, construct a third vector $U = (u_1, u_2, \ldots, u_N)$ where

$$u_i = 1 \text{ if } s_i < x_i$$
$$u_i = O \text{ if } s_i \geq x_i$$

The variable x_i is called the strength of the i^{th} part; s_i is called the applied stress level for the i^{th} part, and u_i is called the outcome of the "test" on the i^{th} part. The outcome $u = 1$ is called a success (i.e., the applied stress was less than the part strength) and, conversely, $u = O$ is called a failure.

The object of this development is to obtain formulas for calculating the estimates μ_e and σ_e of μ and σ, given only S and U. This will be accomplished by obtaining values of μ_e and σ_e which maximize the likelihood (i.e., probability) of obtaining the outcome U given S.

The probability of outcome u_i given s_i can be written

$$(1) \qquad p_i = \text{Prob}\left[u_i \mid s_i\right] = u_i \int_{s_i}^{\infty} g(v;\mu,\sigma)dv + (1 - u_i)\int_{-\infty}^{s_i} g(v;\mu,\sigma)dv$$

The probability of outcome U can be written as the product of the probabilities of the individual u_i since the x_i are independent.

$$(2) \qquad P\left[U\right] = \prod_{i=1}^{N} p_i = L(\mu, \sigma)$$

The expression $P\left[U\right]$, when regarded as a function of the population parameters μ and σ, becomes the likelihood function for outcome U (ref. 5).

To find the values of μ and σ (now regarded as variables) which maximize L, we differentiate (2) with respect to μ and σ and solve the system

$$(3) \qquad \begin{aligned} \partial \ln L/\partial \mu &= 0 \\ \partial \ln L/\partial \mu &= 0 \end{aligned}$$

where the logarithm of L is used to simplify the algebra.

Letting $t_i = (s_i - \mu)/\sigma$, $g_o(v) = (2\pi)^{-1/2} \exp(-v^2/2)$, the normalized Gaussian, and remembering that u_i can take on only values of 0 or 1, equation (1) can be re-written

$$(4) \qquad \ln p_i = u_i \ln\left[\int_{t_i}^{\infty} g_o(v)dv\right] + (1 - u_i)\ln\left[\int_{-\infty}^{t_i} g_o(v)dv\right]$$

The following definitions ($\stackrel{d}{=}$) and derivations will be helpful:

$$(5) \qquad G(t) \stackrel{d}{=} \int_{-\infty}^{t} g_o(v)\, dv$$

$$(6) \qquad dg_o(t)/dt = -tg_o(t)$$

$$(7) \qquad \partial t/\partial \mu = -1/\sigma \qquad (\sigma > 0)$$

$$(8) \qquad \partial t/\partial \sigma = -t/\sigma \qquad (\sigma > 0)$$

Since $\ln L = \Sigma \ln p_i$, equations (3) become

$$(9) \qquad \begin{aligned} \partial \ln L/\partial \mu &= \Sigma\, \partial \ln p_i/\partial \mu = 0 \\ \partial \ln L/\partial \sigma &= \Sigma\, \partial \ln p_i/\partial \sigma = 0 \end{aligned}$$

From (4)

$$(10) \qquad \partial \ln p_i/\partial\mu = \frac{-u_i g_o(t_i)}{1 - G(t_i)} \frac{\partial t_i}{\partial\mu} + \frac{(1-u_i)g_o(t_i)}{G(t_i)} \frac{\partial t_i}{\partial\mu}$$

$$= \frac{g_i}{\sigma}\left[\frac{u_i}{1-G_i} - \frac{1-u_i}{G_i}\right]$$

where the arguments (and subscript "o") have been omitted to simplify writing. Similarly,

$$(11) \qquad \partial \ln p_i/\partial\sigma = \frac{t_i g_i}{\sigma}\left[\frac{u_i}{1-G_i} - \frac{1-u_i}{G_i}\right]$$

Denoting by h_i the expression in brackets and eliminating the constant σ, equations (9) become

$$\left.\begin{aligned} \partial \ln L/\partial\mu &\overset{d}{=} p(\mu,\sigma) = \sum_{i=1}^{N} g_i h_i = 0 \\ \partial \ln L/\partial\sigma &\overset{d}{=} q(\mu,\sigma) = \sum_{i=1}^{N} t_i g_i h_i = 0 \end{aligned}\right\}$$

Equations (12) are valid only if the value of σ which satisfies the maximum likelihood equations is non-zero. A quick examination of the data can be made to determine if a non-zero σ is a maximum likelihood solution. If the maximum stress level at which a success occurred is greater than the minimum level at which a failure occurred, than a non-zero σ satisfies equations (12). If this statement is not true, then $\sigma = 0$ represents a maximum likelihood estimate for the standard deviation and the maximum likelihood estimate of the mean is a connected interval contained between the maximum failure stress level and the minimum success stress level which represent the lower and upper bounds of the interval respectively. The latter situation illustrates an outcome which in fact must be achieved if all of the part strengths were concentrated at a mass point within the above mentioned interval. The maximum likelihood corresponding to this outcome is one. Although unique estimates cannot be obtained for μ_e and σ_e in such an instance, it is possible to provide a basis for decision-making, using the likelihood ratio statistic along with a suitably constructed hypothesis (reference 4). A result of this type is referred to as degenerate. It should be mentioned that, for a fixed population standard deviation, $\sigma \neq 0$, the probability of obtaining a degenerate result approaches 0 as the sample size becomes large. The proof of these statements is contained in reference 4 and is beyond the scope of this paper.

5. STATISTICAL PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATES. The preceding sections describe the method for selecting stress levels and the likelihood equations for calculating μ_e and σ_e. In order to make practical applications of the method, however, information is required regarding the statistical properties of the maximum likelihood estimates corresponding to various sample sizes. To facilitate obtaining this information, an extensive computer simulation was carried out using Monte Carlo techniques. In this way, hundreds of values of μ_e and σ_e could be obtained corresponding to hundreds of simulated experiments using random numbers for part strengths. Statistical summaries were then obtained and the variance of the estimates of the parameters empirically derived.

To perform the simulation, a standard interval, A = 1, B = 1, was chosen. The sampling of strengths was simulated by converting the sum of twelve two-digit numbers, constructed from a file of one millions random digits, to a random deviate with population mean μ and standard deviation, σ. Two populations were employed: $\mu = 0$ and $\sigma = 0.25$, and $\mu = 0.2$ and $\sigma = 0.1$. For each population, one hundred runs, each consisting of N samples, were made for N = 4 through 15, 20, 25, 30 and 35, with an additional four hundred runs for N = 15 and N = 30 for additional information on the distributions of the estimates. Finally, four runs of N = 900 were made for each population to empirically investigate the asymptotic convergence of (μ_e, σ_e) to (μ, σ).

For each set of 100 runs at a fixed sample size, the mean and variance were averaged separately and plotted as shown in Figures 3 and 4. Straight lines were then fitted to the data, recognizing that some spurious effects are introduced for small sample sizes (i. e., 15 or less) due to the discreteness of the admissible outcomes which are possible. (For sample size of N, only 2^N outcomes are possible corresponding to the 2^N possible configurations of 0 and 1.)

Based on the results of Figures 3 and 4, the variance of the estimates are given by:

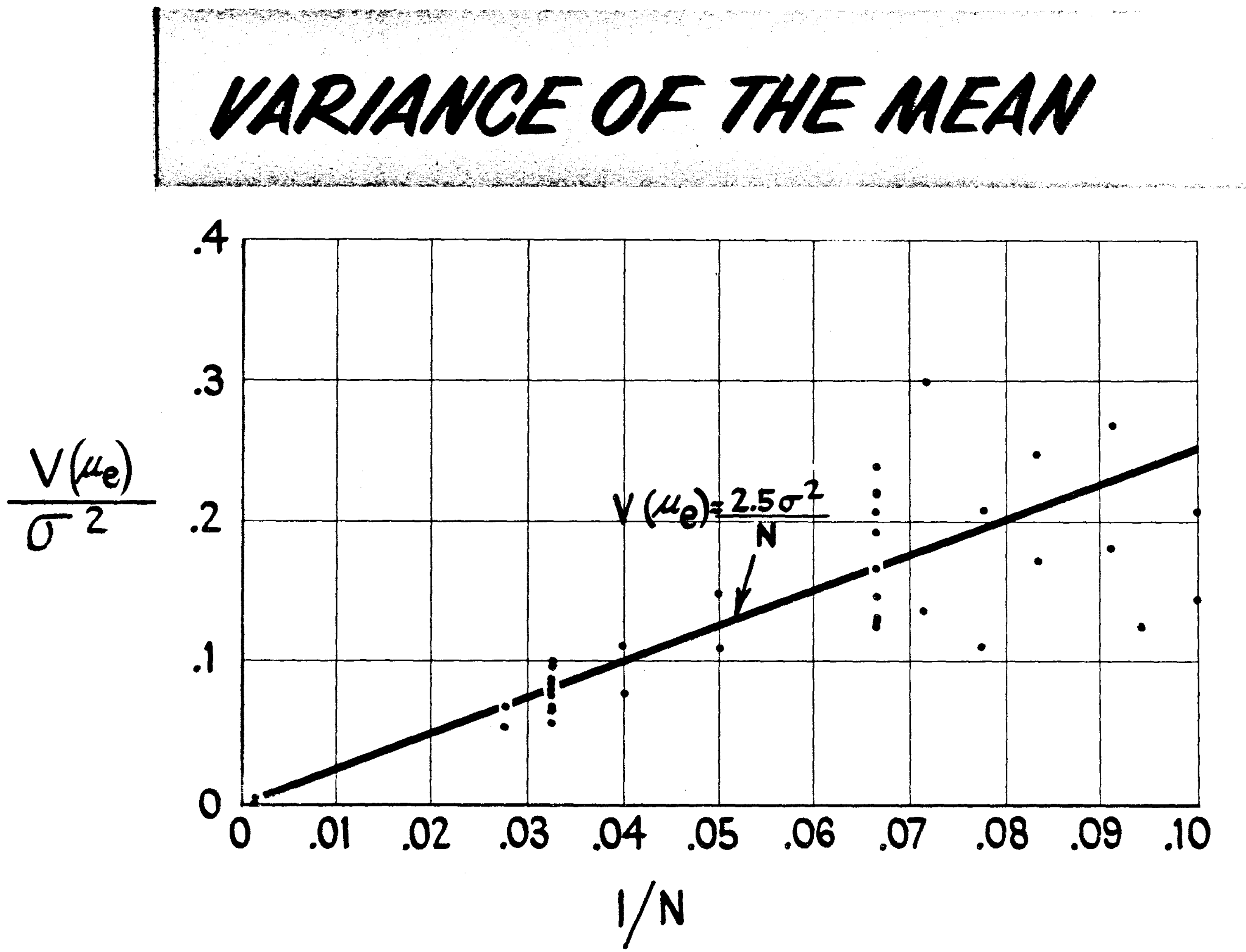

VARIANCE OF THE MEAN
$\frac{V(\mu_e)}{\sigma^2}$
.4
.3
.2
.1
0
$V(\mu_e) = \frac{2.5\sigma^2}{N}$
0
.01
.02
.03
.04
.05
.06
.07
.08
.09
.10
1/N

FIGURE 3.

This page intentionally left blank.

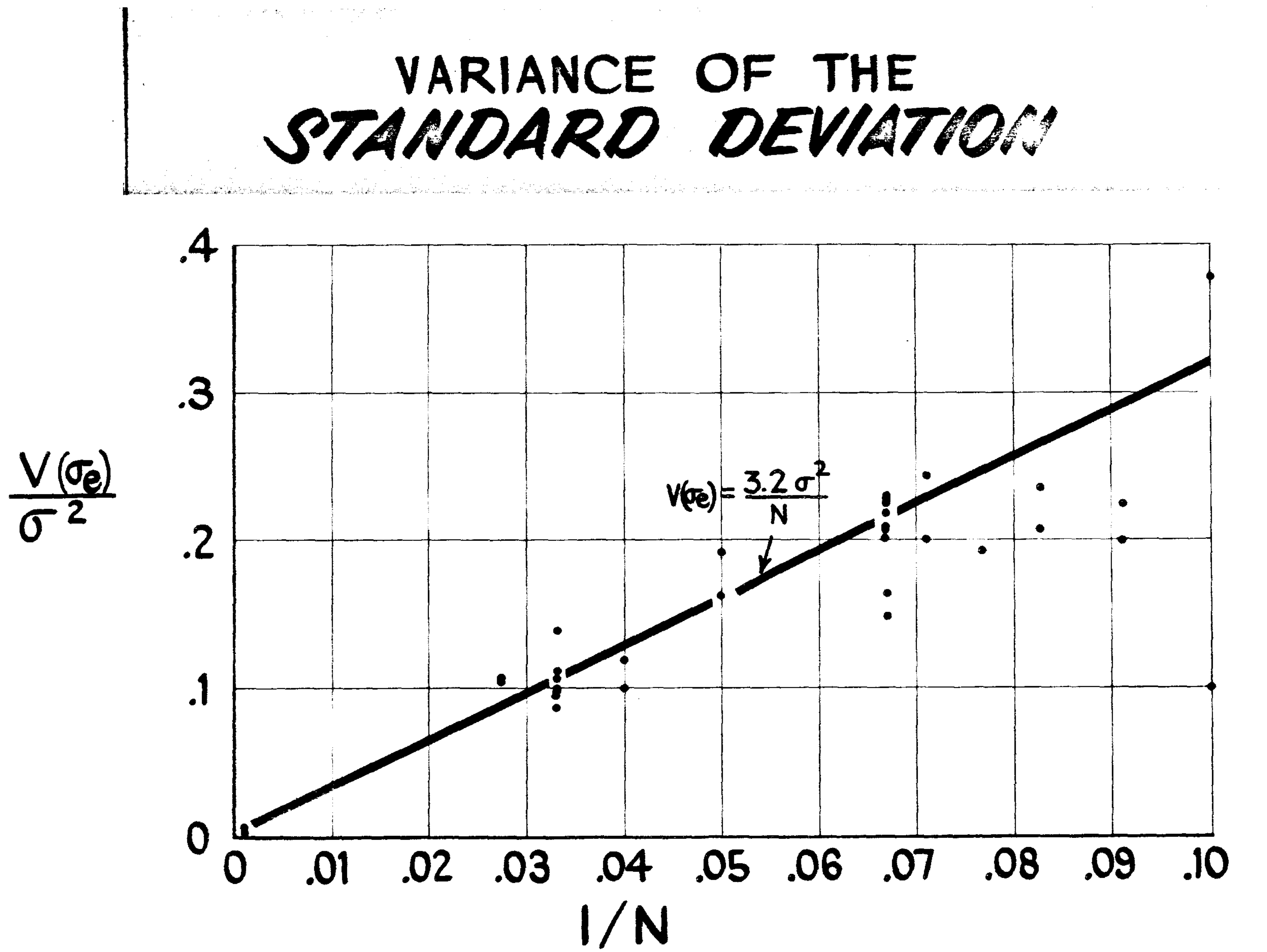

VARIANCE OF THE
STANDARD DEVIATION
.4
.3
.2
.1
0
V(σe)/σ²
V(σe) = 3.2 σ²/N
0 .01 .02 .03 .04 .05 .06 .07 .08 .09 .10
1/N

FIGURE 4.

This page intentionally left blank.

$$V(\mu_e) = 2.5\sigma^2/N$$

$$V(\sigma_e) = 3.2\sigma^2/N$$

In order to use these formulas in practical applications, the unbiased estimate of σ, denoted $\hat{\sigma}$, is substituted for σ in the above formulas. The relationship between the unbiased estimate and the maximum likelihood estimate is given empirically in Figure 5. The variance of the unbiased estimate can be calculated using the relation

$$V(\hat{\sigma}) = \frac{V(\sigma_e)}{\beta^2} = \frac{3.2\hat{\sigma}^2}{N\beta^2} = \frac{3.2\sigma_e^2}{N\beta^4}$$

The above formula is sufficiently accurate for sample sizes on the order of 50 or greater, wherein the distribution of the estimate of the standard deviation approaches the normal distribution. For smaller size samples it was observed from the empirical study that $n\hat{\sigma}/\sigma$ approximately follows the chi-square distribution with n degress of freedom where n is given by

$$n = 0.625\beta^2 N \text{ (reference 4)}$$

where N is the sample size and β is given by Figure 5.

This page intentionally left blank.

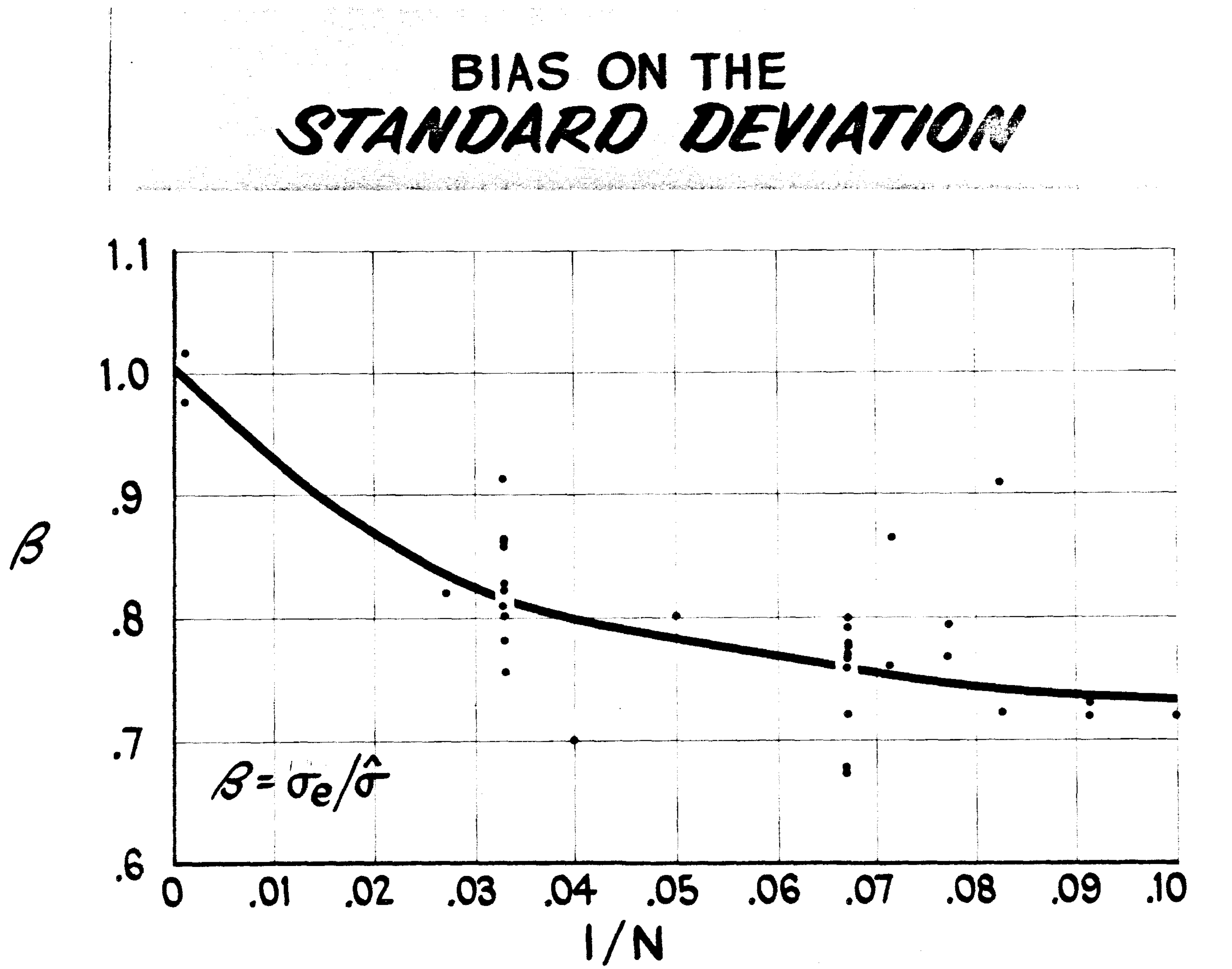

BIAS ON THE
STANDARD DEVIATION
1.1
1.0
.9
.8
.7
.6
β
β = σe/σ̂
0
.01
.02
.03
.04
.05
.06
.07
.08
.09
.10
1/N

FIGURE 5.

This page intentionally left blank.

REFERENCES

1. Lusser, R., Reliability Through Safety Margin, Research and Development Division, Army Rocket and Guided Missile Agency, Redstone Arsenal, Alabama, October 1958.

2. Dixon and Massey, Introduction to Statistical Analysis, McGraw-Hill 2nd Edition, 1957, Chapter 6.

3. Finney, D. J., Probit Analysis, Cambridge University Press

4. Langlie, H. J., A Reliability Test Method for "One-Shot" Items, Aeronutronic Division of Ford Motor Company Publication No. U-1792, August 10, 1962.

5. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, 6th printing, 1954, p. 478.

This page intentionally left blank.

INVESTIGATION IN TEMPERATURE CONTROL OF HYDRAULIC SYSTEMS IN ROUGH TERRAIN FORK TRUCKS

Irving Tarlow
Quartermaster Research and Engineering Command
Natick, Massachusetts

The problem of overheating in hydraulic systems of mobile equipment has not been given full consideration nor have the ramifications been completely understood by designers of such equipment. The effect of excessive heat can be injurious to the operator, the equipment and to persons in the proximity of the equipment.

In industrial applications of hydraulic equipment, space has not been one of the problems. Hydraulic systems designed on this premise provided for a reservoir tank of sufficient fluid capacity to dissipate the heat or a system in which a heat exchanger has been incorporated.

This approach cannot be applied to mobile hydraulic systems. Space is not available for large fluid reservoirs or heat exchangers. The situation is usually more aggravated by the radiation of heat from the engine as well as crowding components of the system into a small enclosed area in which the heat from the system itself has to be considered.

The designer, evaluating the heat effect, takes into consideration the type of operation for which the equipment was intended. For commercial materials handling equipment this approach has been for the most part valid. The hydraulic system is not taxed to the degree that heat has been a factor. Idle time between cycles has been sufficient to allow cooling of the fluid to a safe temperature level. However, in Military applications, where unloading operations are on a 24-hour per day basis and long road marches are a required operational feature, temperature control must be built in.

The condition of high fluid temperature was detected in an All Purpose Fork Lift Truck having a 6,000-lb capacity at 24-inch load center. This vehicle is a Military item presently in the supply system in large quantities. The purpose of this vehicle is to handle supplies, whether palletized, unitized or containerized, over rough terrain, in open storage areas and in all types of climatic environments as well as all types of terrain. This vehicle must be capable of operating under conditions of limited visibility, inclement weather and blackout and be capable of negotiating open rolling and hilly terrain consisting of mud, snow and sand. In addition, this vehicle must also be capable of highway operations.

Prior to actual investigation and subsequent remedial action, the possible effects of the problem were analyzed. The analysis included but was not limited to:

(1) The effect on humans.

(2) The effect on the system.

The location of the operator's compartment is directly above the reservoir tank. Therefore, operator discomfort is a factor. Since the tank is mounted outboard of the chassis, personnel coming in contact with the tank surface could suffer burns. While either and both of the above are serious, far more important were the injuries and possible fatalities which could result from hydraulic line ruptures and component failures. The unleashing of hydraulic fluid which was under pressure of up to 1400 psi and whose temperature was in excess of 250 F is considered to be the most serious of those previously mentioned.

The second area to be considered in evaluating the effect of high temperature rise is the effect upon the system itself. Excessive temperatures of petroleum fluids will cause lubricating qualities to be reduced. For gear pumps whose shafts run in journal bearings, the design is such that the shaft "floats" on a film of oil so that there is no contact between the shaft and the bearings. Extreme heat will cause the film to break down. When this occurs, the shaft comes in contact with the bearing and causes galling and seizing. Galling and seizing will also occur when unequal expansion occurs between two parts of dissimilar material. Directional control valves are particularly susceptible to the problem of unequal expansion which causes spools to stick resulting in an inoperative system.

Petroleum fluid deterioration can have additional harmful effects upon the system. When a petroleum fluid deteriorates sludge forms, this sludge will be deposited on surfaces which are subjected to localized heating. Filters and lines become clogged and valves become stuck. If a sump filter is used in the reservoir tank and it should become clogged due to the formation of sludge, cavitation of the pump will in all likelihood result.

The particular hydraulic system on which the investigation was conducted is of the open center type and consists of a fixed displacement pump coupled to the engine input shaft. Fluid passes from the pump to a directional control valve and is then directed either to reservoir tank when no work is required or to a working cylinder.

It was necessary to first isolate the cause of excessive heating. A work cycle was devised based on a simulated field operation. This consisted of lifting a 6,000-lb rated load, transporting it to a new site, stacking the load and then returning to the starting position. Temperature was recorded in the reservoir tank approximately 2 inches from the bottom and 18 inches away from suction outlet.

Through observation, it was detected that the rise in temperature occurred during the transporting phase of the cycle. Temperature would drop during the work cycle and then rise again during transport. By analyzing the circuitry, it was determined that the heat rise was in all probability due to a restriction and further that this restriction was located in the directional control valve.

In order to substantiate the analysis, the directional control valve was by-passed and the fluid path was controlled so that it passed from the pump directly to the reservoir. This resulted in a substantial reduction in temperature and stabilization point in the system thereby validating the analysis.

The problem of controlling the fluid temperature became twofold:

(1) A method had to be devised by which vehicles already issued to the field could be modified quickly and economically and still permit satisfactory operation during the service life of the vehicle. (Service life is considered to last from 8 to 11 years for this type of equipment.)

(2) Redesign of the hydraulic system in order to produce specification changes which would permit future procurement of a vehicle capable of satisfying all of the operational requirements.

Prior to development of a solution, it was first necessary to establish acceptable operating parameters. To do this, the vehicle was instrumented to record reservoir tank temperature, pressure between pump and directional control valve and pressure on the return line between the control valve and reservoir tank.

Since the pump was of a fixed-displacement type directly coupled to the engine, the controlled variable became engine speed. The pump output in gallons per minute was determined at various engine speeds. After this, the back pressure in pounds per square inch at various flow rates (gpm) was determined for the directional control valve. The next step was to determine the acceptable back pressure and relative temperature rise limits. This was necessary since it already had been determined that 100 deg. temperature rise above ambient could be tolerated for the particular oil being used in the system and for areas of the world where the vehicle would be used. In order to ascertain, therefore, what back-pressure could be tolerated which would result in no more than 100 deg. temperature rise, the following tests were conducted:

The vehicle was operated at maximum engine speed until the temperature in the tank stabilized for a minimum period of 20 minutes. Engine speed was varied and successive runs at different speeds were made. Through observation and analysis of the stabilized temperature and back pressure, it was determined that the naximum open center pressure drops across the directional control valve should not exceed 100 psig.

In explanation of the above let me, at this time, cite a few of the specifics:

a. The majority of test runs were made in low gear and at full engine rpm. This will reproduce all conditions anticipated during convoy with the exception of the cooling effect from a high velocity wind passing by the hydraulic tank. When tested in high gear at 25 mph, a 20 degree drop in maximum temperature below that of low gear was observed.

b. Pump specifications required a minimum oil viscosity of 45 SSU. For the oil used in the system 45 SSU occurred at 200 F.

c. All convoy runs were made with the engine side panels off. When the engine side panels were in place, the result was additional drop in stabilization temperature of 24 F.

d. An additional factor which, although not measurable, and which has a direct bearing on the temperature is the nature of a convoy. Convoys have what is known as an accordion effect. At high speeds on long runs vehicles have a tendency to crowd together. When this happens, the line has to slow down and vehicles separate in order to maintain proper spacing between them. When the fork lift truck slows down, engine speed decreases. The pump will circulate fluid at a lower rate and the oil will cool.

These factors just listed will permit safe operation in ambient temperatures up to 125 F when the temperature rise measured in the hydraulic tank with the transmission in lowest gear and engine at maximum governed speed is limited to 100 F.

While temperature could be used as the criterion to control the heat rise in the system, it is still necessary to provide the manufacturer with some guide for the selection of the components. Since it had been determined that temperature rise could be controlled by controlling the pressure drop across the main control valve, it became necessary to provide a tolerance for this factor. Analysis of the data indicated that the maximum allowable pressure which would maintain the 100 degree rise in hydraulic tank was 100 psig. This additional information would control the maximum amount of restriction which could be tolerated in the main control valve.

The limitations placed on temperature and pressure would suffice for specification requirements. However, it was still necessary to modify the equipment already issued to the field.

While there are many possible solutions to correcting the deficiency in the equipment, only four approaches were tested. These were:

a. Redesign of the hydraulic tank. As previously stated, space was at a premium, therefore little could be done in this area of increasing the reservoir volume. However, it was possible to redesign the tank so that the flow of fluid from discharge to suction would be along the tank walls thereby taking maximum advantage of surface radiation. While there was a temperature reduction, it was not of significant magnitude by itself to control temperature rise.

b. Control the flow into the directional control valve. By dividing the fluid as it was delivered from the pump it was possible to control the amount of fluid delivered to the valve. Excess fluid could be shunted directly to the reservoir tank. By controlling the amount of fluid which passed through the control valve it was possible to control back pressure.

c. Replace the hydraulic pump. By varying engine speed, it was possible to control the amount of fluid into the directional control valve. However, at a point where back pressure was reduced sufficiently and temperature stabilized within acceptable limits, engine speed was reduced to the point where other characteristics were affected materially enough to render the vehicle not operationally suitable. The effect, hydraulically, could be achieved by replacing the pump with one having smaller delivery rates.

d. Replace the directional control valve. This is the most obvious of those listed. Since the valve was the cause of heat by having internal porting such that flow was being restricted, it could be replaced with another model which had larger porting with decreased restriction.

It should be mentioned that there are various combinations of the four listed which could also offer satisfactory solutions. Each of the four solutions has disadvantages and by themselves do not offer the ideal solution. When certain combinations are made the disadvantages of each by itself are often nullified.

Since a number of vehicles had already been issued, other factors had to be taken into consideration and trade-offs had to be made before the final solution was agreed upon. Some of the factors were:

a. Economy: The pumps and directional control valves are costly units in the price range upwards of $300 each. Installation costs for directional controls would be high as it required extensive repiping.

b. Parts availability: The parts required for the item would have to be available within a reasonable time frame since operation of these vehicles affected our combat readiness.

c. Labor availability: The skills required to apply the remedy would have to be available at the level at which the actual work would be done.

d. Simplicity of installation: Consideration had to be given to the desired echelon which would apply remedy. The world wide dispersion of vehicles dictated that the remedy should be applied at the lowest possible echelon in order to avoid massive transportation costs to any single location. Therefore, if the remedy could be readily applied without the need of skilled labor, the echelon would necessarily be a low one.

e. Ease of maintenance: This item was a basic consideration. The parts required to correct the deficiency could not create an added burden on those responsible for the maintenance of the item.

Applying the factors just listed to the possible solution, it was decided that the application of a by-pass and flow divider type valve which would govern the flow from the pump to directional control valve on a priority basis would satisfy these requirements.

a. Economy: The complete installation would approximate 20% of the cost of installation of either a new pump or new directional control valve.

b. Parts availability: The item was an off-the-shelf item and readily available.

c. Labor availability: The item could be applied by an ordinary mechanic.

d. Ease of installation: The unit could be sent to the field in kit form and would require only three connections.

e. Ease of maintenance: The operation of this valve is relatively simple -- fluid from the pump goes to the inlet of the divider and through a fixed orifice in a sliding valve. The oil passing over the orifice creates a pressure differential that pushes the valve against a spring. Since the spring exerts essentially a constant force, the pressure differential is constant and the flow to create the differential is constant. If the flow through the orifice tends to increase, the pressure differential is increased and the valve moves to uncover a by-pass port. The size of the orifice may be selected for any desired amount and can be controlled within one gallon per minute. Due to the simplicity of this type valve, no additional maintenance burden would result from its installation.

Testing of this valve against the operational requirements indicated that concessions in this area had to be made as follows:

a. Engine speed had to be governed down slightly.

b. Back pressure up to 140 psig would have to be tolerated resulting in

c. A temperature rise up to approximately 143 F.

As for future procurement, the design of the hydraulic system has been changed so that a new pump and directional control valve will be used which during testing met the 100 psi back pressure and 100 **F** temperature rise limitations.

PRECISION OF SIMULTANEOUS MEASUREMENT PROCEDURES

W. A. Thompson, Jr.
National Bureau of Standards
Washington, D. C.

1. ABSTRACT. We consider the problem of measurement under the following conditions: The process of gathering the data is such that on any given item only one opportunity for measurement occurs, but it can be observed simultaneously by several instruments. The items to be measured are variable so that one cannot obtain replicate observations with the same instrument which would show directly the variance of the instrument readings. Procedures are discussed for estimating the precisions of the instruments and the variability of the items being measured.

An example due to Simon and Grubbs is helpful in fixing ideas. The burning times of thirty similar fuzes are determined by several different observers. We limit our discussion to the data taken by observers A and C; hence there are two determinations of the burning times of thirty different fuzes or sixty observations in all. If each of the fuzes had the same running time (which is the manufacturer's goal) and if both of the observers were absolutely accurate, then all sixty observations would be equal. However, considerable inequality in such data always occurs due to variation in the manufacturing process and inaccuracy of the observations. It then becomes desirable to use the sixty observations to answer as many questions as possible about measurement bias and precision, mean fuze running time, and variability of burning times about their mean.

2. THE MODEL. With the verbal description of the previous section in mind, consider the following mathematical formulation. Let $x_1, \ldots, x_N$ denote the true values of the items to be measured. Assume that $x_1, \ldots, x_N$ constitute a random sample of size N selected from a population having mean μ and variance σ^2. Each of the items in this sample is then measured by p instruments. y_{ij} is the measurement of the i^{th} item $(i=1, \ldots, N)$ according to the j^{th} instrument $(j = 1, \ldots, p)$. The consequence of this measurement is that an instrumentation error e_{ij}, chosen at random from the j^{th} instrument's population of errors, is added to the true value of the i^{th} item:

$$y_{ij} = x_i + e_{ij}. \tag{2.1}$$

The instrumentation errors are taken to be uncorrelated among themselves and also uncorrelated with the items selected for measurement. The mean and variance of e_{ij} are β_j and σ_j^2, respectively; β_j may be called the bias of the j^{th} instrument.

Denoting the vector $(y_{i1}, \ldots, y_{ip})$ by Y_i, we may think of $Y_1, \ldots, Y_N$ as constituting a sample of size $N = n+1$ from a p-variate distribution with mean vector $(\mu + \beta_1, \ldots, \mu + \beta_p)$ and dispersion matrix

$$\Sigma = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma^2 & \ldots & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_2^2 & \ldots & \sigma^2 \\ \vdots & \vdots & & \vdots \\ \sigma^2 & \sigma^2 & & \sigma^2 + \sigma_p^2 \end{pmatrix}. \tag{2.2}$$

Notice, in passing, that if all instrument variances are equal, then the model becomes a completely random one-way layout and may be analysed by the methods which appear, for example, in Scheffé [5].

A paragraph on notation will perhaps be helpful. ξ_j will be used as a more succinct notation for $\mu + \beta_j$, $j = 1, \ldots, p$. We will frequently write $\Sigma = (\sigma_{jj'})$ when we mean that $\sigma_{jj'}$ is the element in the j^{th} row and j'^{th} column of Σ. In the same spirit $\Sigma^{-1} = (\sigma^{jj'})$, $A = (a_{jj'})$ and $S = (s_{jj'})$ will be common notations. Here

$$a_{jj'} = \sum_{i=1}^{N} (y_{ij} - \bar{y}_{\cdot j})(y_{ij'} - \bar{y}_{\cdot j'}) \quad , \tag{2.3}$$

and $s_{jj'}$ is the usual unbiased estimate of $\sigma_{jj'}$, i.e. $s_{jj'} = a_{jj'} / n$.

3. <u>POINT ESTIMATES.</u> In [2], Grubbs recommends certain estimates of item and instrument variance. For p = 2 instruments, these estimates are

(3.1) $\sigma^2 \sim s_{12}$, $\sigma_1^2 \sim s_{11} - s_{12}$ and $\sigma_2^2 \sim s_{22} - s_{12}$,

where $\sim$ is to be read "is estimated by". For $p \geq 3$, Grubbs recommends

$$\sigma^2 \sim \frac{2}{p(p-1)} \sum_{j<j'} s_{jj'},$$

$$(3.2) \qquad \sigma_1^2 \sim s_{11} - \frac{2}{p-1} \sum_{j=2}^{p} s_{1j} + \frac{2}{(p-1)(p-2)} \sum_{2 \leq j<j'} s_{jj'},$$

with an analogeous estimate of the other instrument variances. Gaylor [1] shows that Grubbs' estimate of σ^2 is equivalent to a familiar variance component estimate. These estimates are reasonable in that they have the correct dimensionality, are unbiased, and have appropriate symmetry properties when the instrument labels are interchanged. Further, if the underlying distribution is normal, then Grubbs' estimates are simple functions of the sufficient statistics and in the case p=2 they have a maximum likelihood property. However, as Grubbs has verbally pointed out his estimates are unreasonable in that they frequently assume negative values even though the parameters themselves must be non-negative by their very definition.

For p=2 this objectional characteristic has been eliminated in [8]; here the altered estimates of table 1 have been proposed. The top line of this table yields the estimates (3.1) under conditions where they are non-negative. The remaining entries show how Grubbs' estimates can be modified when negativity would result from using (3.1). These modified estimates have been derived, under normality assumptions, from a principle of restricted maximum likelihood which is fairly well accepted in other brances of statistical practice. A tilda placed over a parameter indicates its restricted maximum likelihood estimate.

Table 1. Non-negative estimates in the two instrument-case.

Conditions	$\tilde{\sigma}^2$	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$
$s_{11} \geq s_{12}$ $s_{22} \geq s_{12} \geq 0$	s_{12}	$s_{11} - s_{12}$	$s_{22} - s_{12}$
$s_{22} \geq s_{12} > s_{11}$	s_{11}	0	$s_{11} + s_{22} - 2s_{12}$
$s_{11} \geq s_{12} > s_{22}$	s_{22}	$s_{11} + s_{22} - 2s_{12}$	0
$s_{12} < 0$	0	s_{11}	s_{22}

4. RELATIVE PRECISION. It is clear that if the instrumentation of an experiment is to be effective then the instruments must be precise relative to the variability of the quantity being measured. A frequently quoted rule of thumb is that the instrument precision should be an order of magnitude greater than that of the item being measured. Such a statement has no firm meaning unless a measure of instrument precision and a measure of item variability have been agreed upon. Here we adopt $\Delta_1 = \sigma/\sigma_1$ as a measure of the relative precision of the first instrument. Then, for example, the above mentioned rule of thumb would become $\Delta_1 \geq 10$.

In the two-instrument case, assuming normality, we may use a result of Roy and Bose [3] to make inferential statements of a statistical nature concerning the parameter Δ_1. In our terminology their result states that

$$(4.1) \qquad a_{11}\left[\frac{n-1}{|A|}\right]^{\frac{1}{2}}\left(\frac{a_{12}}{a_{11}} - \frac{\sigma_{12}}{\sigma_{11}}\right)$$

has the t-distribution with n - 1 d.f. where $|A| = a_{11}a_{22} - a_{12}^2$. Noting that $\sigma_{12}/\sigma_{11} = (1 + \Delta_1^2)^{-1}$, we may verify that the quantity (4.1) is less than t_α, if and only if

$$(4.2) \qquad \Delta_1^2 > \frac{a_{12} - t_\alpha\left(\frac{|A|}{n-1}\right)^{\frac{1}{2}}}{a_{11} - a_{12} + t_\alpha\left(\frac{|A|}{n-1}\right)^{1/2}} .$$

Hence if t_α is the upper α percentage point of the t- distribution with n-1 d.f. then the square root of the right hand side of (4.2) provides a lower confidence bound for Δ_1, the confidence coefficient being $1-\alpha$. The inequality (4.2) can also be used for the purpose of hypothesis testing For example, we may reject $\Delta_1 \geq 10$ at the significance level α if (4.2) is violated with $\Delta_1^2 = 100$.

5. A SIMULTANEOUS CONFIDENCE REGION. For some purposes it may not be enough to consider relative precision; we may be interested in the actual non-relative precisions and the item variability. Estimation of

the parameters $\sigma^2, \sigma_1^2, \ldots, \sigma_p^2$ was discussed in section 3; but how reliable are estimates? This question is dealth with in [9], again under assumptions of normality.

In the two-instrument case, the probability is at least $1 - 2\alpha$ that the following three relations hold simultaneously

$$\left| \sigma^2 - a_{12} K \right| \leq M(a_{11} a_{22})^{\frac{1}{2}} ,$$

(5.1)

$$\left| \sigma_1^2 - (a_{ii} - a_{12}) K \right| \leq M \left[a_{ii} (a_{11} + a_{22} - 2a_{12}) \right]^{\frac{1}{2}} ; \quad i = 1, 2.$$

Here K and M are to be found in Table 2 under the desired value of 2α.

Table 2. The table gives values of K and M which yield 1 - 2α confidence regions when used in conjunction with the relations (5.1).

2α	.01		.05	
n	K	M	K	M
3	99.78	99.72	19.79	19.71
4	12.38	12.33	4.146	4.077
5	3.980	3.931	1.726	1.665
6	1.903	1.353	.9636	.9088
7	1.120	1.078	.6290	.5786
8	0.7459	.7076	.4516	.4032
9	0.5389	.5031	.3453	.3022
10	0.4120	.3782	.2761	.2357
11	0.3282	.2963	.2230	.1901
12	0.2698	.2395	.1932	.1573
13	0.2272	.1983	.1663	.1328
14	0.1951	.1675	.1464	.1140
15	0.1702	.1438	.1301	.09925
16	0.1505	.1251	.1169	.08738
17	0.1344	.1100	.1060	.07767
18	0.1213	.09772	.09632	.06962
19	0.1103	.08752	.08904	.06287
20	0.1009	.07896	.08237	.05713
22	.08610	.06546	.07152	.04795
24	.07484	.05538	.06311	.04098
26	.06605	.04763	.05641	.03554
28	.05901	.04152	.05096	.03121
30	.05328	.03660	.04644	.02768
35	.04272	.02778	.03796	.02127
40	.03556	.02200	.03205	.01700
45	.03040	.01797	.02771	.01398
50	.02652	.01503	.02440	.01176
60	.02109	.01110	.01967	.00875
70	.01748	.00862	.01646	.00684
80	.01492	.00694	.01415	.00553
90	.01300	.00575	.01241	.00460
100	.01152	.00486	.01104	.00390

This page intentionally left blank.

For more than two instruments a similar result is valid. We have the following relations with probability at least $1 - 2\alpha$

$$\max_{j \neq j'} \left[a_{jj'} K - M (a_{jj} a_{j'j'})^{\frac{1}{2}} \right] \tag{5.2}$$

$$\leq \sigma^2 \leq \min_{j \neq j'} \left[a_{jj'} K + M(a_{jj} a_{j'j'})^{\frac{1}{2}} \right] ,$$

$$\max_{j \neq 1} \left\{ (a_{11} - a_{1j}) K - M \left[a_{11} + a_{jj} - 2a_{1j} \right]^{\frac{1}{2}} \right\}$$

$$\leq \sigma_1^2 \leq \min_{j \neq 1} \left\{ (a_{11} - a_{1j}) K + M \left[a_{11} + a_{jj} - 2a_{1j} \right]^{\frac{1}{2}} \right\},$$

plus $p - 1$ similar inequalities involving $\sigma_2^2, \ldots, \sigma_p^2$. Unfortunately, for p in excess of two, tables of K and M are unavailable. The only result which is currently ready for use is an approximation valid for large n: Choose ℓ to satisfy $P(\ell \leq \chi^2_{n-p+2}) = 1 - 2\alpha$, write $K = M = 1/2\ell$ and substitute this common value in (5.2). I feel obliged to point out that for $p = 2$, the only case where exact values are available, this approximation is rather poor.

6. <u>NUMERICAL EXAMPLE</u>. Returning to the fuze burning time data, we may identify observer A as the first instrument and observer C as the second. From Table I of Grubbs' paper [2] we obtain $a_{11} = 1.3671$, $a_{22} = 1.3227$, $a_{12} = 1.3320$ and $n = 29$. From the third row entry of our table 1, we estimate $\tilde{\sigma} = .21$, $\tilde{\sigma}_1 = .03$ and $\tilde{\sigma}_2 = 0$. By the method of section 4 we obtain, for example, that the relative precision of observer C exceeds 5.1 with a confidence of 95%. Alternatively, from a hypothesis testing point of view we would reject the rule of thumb requirement, $\Delta_2 \geq 10$, at the 5% level. The relations (5.1) and table 2 yield the following 95% simultaneous confidence region: $.16 \leq \sigma \leq .32$, $0 \leq \sigma_1 \leq .09$ and $0 \leq \sigma_2 \leq .07$. In calculating these simultaneous confidence intervals we have replaced all negative lower bounds by zero. Notice that the confidence intervals bracket their respective estimates and hence, in the confidence region sense, indicate the uncertainty of these estimates.

I would like to express my appreciation to the National Academy of Sciences-National Research Council for granting me the opportunity to participate in their postdoctoral research program, and in particular to the Natïonl Bureau of Standards for its support of my research under this program. Also I wish to thank Mrs. Marion Carson for her help in carrying out the computations for table 2.

REFERENCES

[1] D. W. Gaylor, "Equivalence of two estimates of product varinace", J. Amer, Stat. Assoc., vol. 51 (1956), pp. 451-453

[2] Frank E. Grubbs, "On estimating precision of measuring instruments and product variability", J. Amer. Stat. Assoc., vol 43 (1948), pp. 243-264.

[3] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," Ann. Math. Stat., vol. 24 (1953), pp. 513-536.

[4] Thomas S. Russel and Ralph Allan Bradley, "One-way variances in a two-way classification", Biometrika, vol. 45 (1958), pp. 111-129

[5] Henry Scheffe, The Analysis of Variance, John Wiley and Sons, Inc., N. Y., 1959.

[6] L. E. Simon, "On the relation of instrumentation to quality control", Instruments, vol. 19 (1946), pp. 654-656.

[7] H. Fairfield Smith, "Estimating precision of measuring instruments", J. Amer. Stat. Assoc., vol. 45 (1950), pp. 447-451.

[8] W. A. Thompson, Jr., "The problem of negative estimates of variance components", Ann. Math. Stat., vol 33 (1962), pp. 273-289.

[9] W. A. Thompson, Jr., "Estimation of dispersion parameters", in preparation.

This page intentionally left blank.

APPLIED MICROSCOPY

George I. Lavin
Terminal Ballistics Laboratory
USA Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

The Aberdeen Proving Ground in general and the Ballistic Research Laboratories in particular are interested in the manner by which absorbed energy is utilized by various systems. The technique describes the possibility of an evaluation of the traumatic effect brought about by the absorption of energy at the cellular level.

This paper (presented, in brief, to the Eighth Conference on the Design of Experiments in Army Research, Development and Testing, The Walter Reed Army Institute of Research, Washington, D. C., October 1962) has to do with the ultraviolet microscopy of biological tissues.

There is much controversy having to do with the structure and function of biological cellular configurations. This controversy will continue to exist until methods are employed for the analysis and evaluation of experimental results which do not (by preparation and/or examination) change the materials during the evaluation procedure.

A paper describing a photomicrographic technique was published by Köhler [(1)] in 1904. The optical system was quartz instead of the usual glass. The purpose of this was to take advantage of the increase in resolution which is realized when ultraviolet light is used as the source of illumination instead of the visible region of the spectrum.

The light source employed by Köhler was a condensed electrical spark between metallic electrodes. This is unsatisfactory for the illumination of a microscope. The electrodes burn away non-uniformally which means that it is difficult to keep an image of the source centered on the microscope condenser. Also the noise of the spark is objectionable. In spite of these and other difficulties, some work has been published [(2)].

The early workers were attracted to this field because the technique offered the possibility of considerable ultimate magnification (since the resolution was about twice that obtained with visible light) although the initial magnification might be low.

In other words, the increase in resolution made possible the visualization of microscopic structure which was not evident until it was revealed by the enlarging camera.

It is to be pointed out that the obvious lack of light intensity was due to the inefficiency of the monochromator system used to isolate the wave length of light (2750A) for which the objectives were corrected.

It was also realized that tissue components such as proteins, nucleic acids and nucleoproteins have specific absorption bands in this region of the spectrum so that it should be possible to obtain micrographs of fresh and unstained material.

Thus, in a sense of the word, it can be said that the chemical constituents of the tissue can act as their own specific light absorbing medium -- the stain.

It is well understood that the fixation process (necessary for the manipulation of tissue cutting) may contribute to the micrographical picture. We do not always know what contribution this represents to our over-all picture. By means of frozen sections we can minimize this uncertainty. We will also need the frozen section-type of section for our analytical procedures.

The inherent objection to previous procedures was that it was not possible to obtain sufficient light intensity of the proper wave length (in the plane of the photographic plate) so that a field and focus could be localized and imaged with the wave length of light which is specifically absorbed by the material under examination.

In some cases, a fluorescent plate was placed directly over the microscope eyepiece. Thus another adjustment had to be made in order to bring the image in focus in the same plane of the photographic plate. In other words, the focus must be in the plane of the photographic plate.

In 1943 a short paper on the subject of ultraviolet microscopy was published by Lavin[3]. A procedure was described whereby an image of the material on the stage of a microscope could be visualized by a fluorescent screen which was temporarily placed in the position usually occupied by a photographic plate. That is, in the plate holder.

The 2537 Angstrom mercury line was the light source (loc. cit.). It has been previously pointed out that this wave length of light is in the region of absorption of those chemical compounds which are to be associated with tissue structure.

The photographs which are now shown will serve to illustrate this technique and some of the results obtained by application of this procedure. The original plates were taken at a magnification of about 200 diameters.

DESCRIPTION OF PLATES

1. Ultraviolet photomicrographic apparatus described by Köhler. Note the spark source, the monochromator and the eyepiece focusing attachment.

2. Apparatus used in the present work. Light source, liquid filter, willemite focusing screen.

3. Light source -- quartz resonance radiation lamp.

4. Spectrum of light source, with and without filter.

5. Fresh Hamster muscle -- teased out (not cut), fresh, unstained.

6. Muscular dystrophy -- cross section.

7. Muscular dystrophy -- longitudinal.

8. Fresh smear of cells from a chicken egg.

9. Liver, normal -- fixed, unstained.

10. Liver, infectious hepatitis, fixed, unstained.

11. Cross section of a plant root (Sorrell), unstained.

12. Cross section, skin (mal del pinta), unstained.

13. Salivary gland of a mosquito, fresh -- unfixed, unstained.

14. Enlargement of a portion of the salivary gland shown above.

15. Kidney, cross section -- unstained.

16. Red cells, smear, monkey, dried.

17. Red cells, smear, chicken, dried, showing nuclei.

18. Arbacia eggs, showing the "relayering" on rupture.

19. Arbacia eggs, showing the results of photography in the visible, ultraviolet and infrared regions of the spectrum.

20. Muscle photographed using the 2537 A^o mercury line, unstained, visible -- stained, desicated.

21. Absorption spectrum curve of 1, 2 -- Benzanthracene.

22. Absorption spectrum of the same substance using the continuous light from a hydrogen discharge tube as the light source.

23. Apparatus for photomicroscopy -- three light sources, infrared, visible, ultraviolet.

THE USE OF CONTINUUM AS A LIGHT SOURCE FOR ABSORPTION SPECTRA. As has been pointed out those compounds such as proteins nucleic acids, nucleoproteins etc. have broad absorption bands when measured by spectrophotometers.

It has also been shown by Lavin and Northrop[(4)] and others that a considerable amount of band structure can be demonstrated (at room temperature) by the use of a spectrograph of comparative low dispersion. It was also indicated that these bands can be interpreted in terms of the component parts of the molecule by Lavin, Loring and Stanley[(5)]. This technique has also been applied to complicated mixtures (body fluids) by Dobriner, Lavin, and Rhoads[(6 & 7)].

Photographs illustrating the above are shown in Plate 22. It is thought that efforts to apply this technique in obtaining the absorption spectra of materials on the stage of a microscope might be worthwhile. This could be a clue to the chemical composition of the various sections of material under examination.

SUMMARY. Photomicrographs of fresh and of unstained tissues obtained with the 2537 A° mercury line as the light source are shown and some of the implications of this technique are discussed.

The possibility of a microspectrographic application to the problem is considered.

ACKNOWLEDGEMENT. The photomicrographs of tissue which were made while the author was at the Rockefeller Institute for Medical Research are used with the kind permission of Dr. Detlev W. Bronk.

REFERENCES TO PREVIOUS WORK

(1) A. Kohler, Zeits. wiss. Mik. 21, 273 (1904).

(2) F. F. Lucas, J. Franklin Inst. 217, 661 (1934).
R. W. G. Wyckoff, Cold Spring Harbor Symposia 2, 39 (1936).
J. E. Barnard, J. Roy. Microscop. Soc. 56, 365 (1936).
T. Caspersson, Skand. Arch. Physiol. 73, Suppl. 8 (1936).
J. S. Mitchell, Brit. J. Exptl. Path. 23, 285 (1942).

(3) G. I. Lavin, Rev. Sci. Instr. 14, 375 (1943).

(4) G. I. Lavin and J. H. Northrop, J. Am. Chem. Soc. 57, 874 (1935).

(5) G. I. Lavin, H. S. Loring, and W. M. Stanley, J. Biol. Chem. 130, 259 (1940).

(6) K. Dobriner, G. I. Lavin, and C. P. Rhoads, Cancer Research 2, 79 (1942).

(7) K. Dobriner, C. P. Rhoads, and G. I. Lavin, Cancer Research 2, 95 (1942).

This page intentionally left blank.

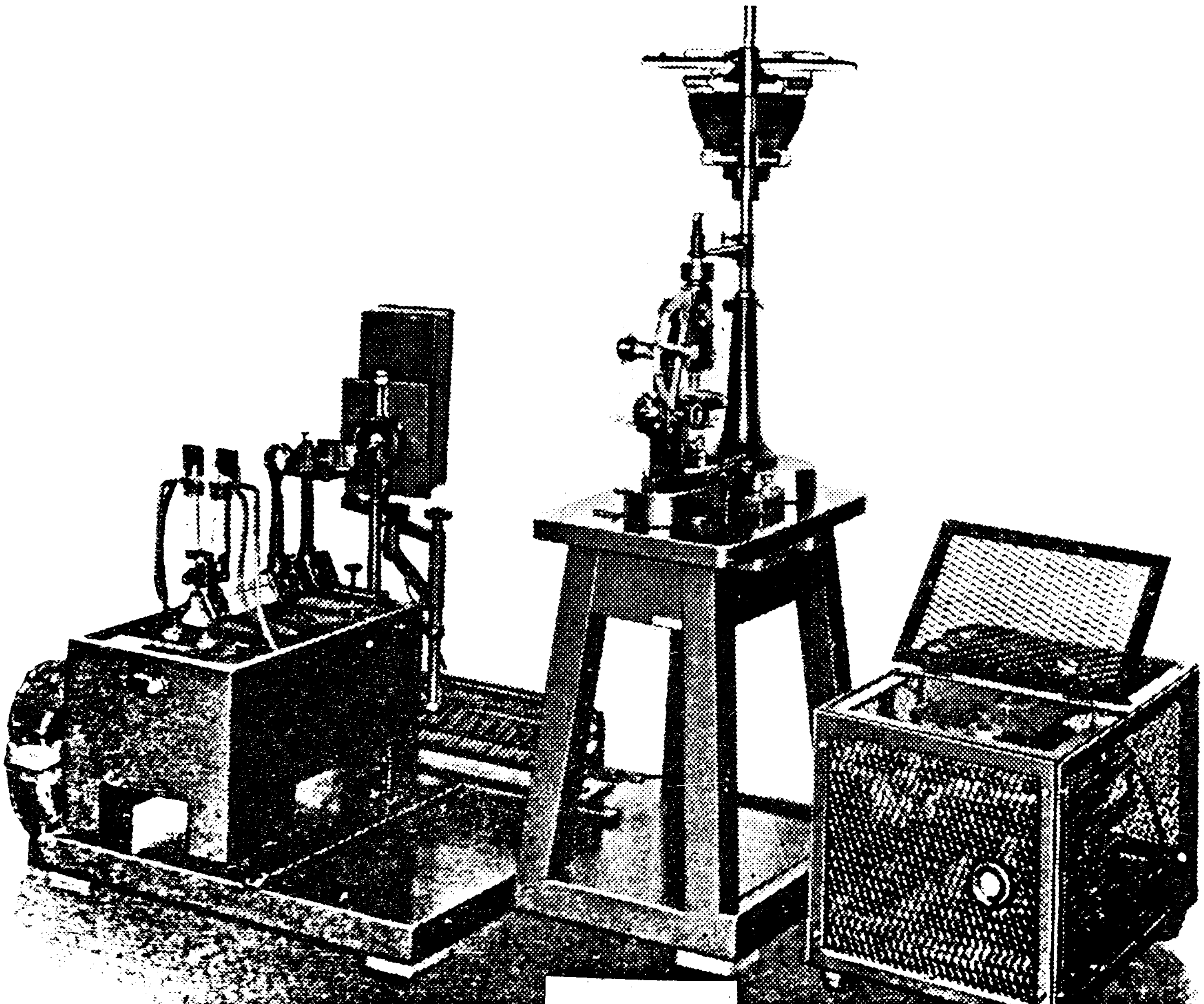


Plate No. 1

This page intentionally left blank.

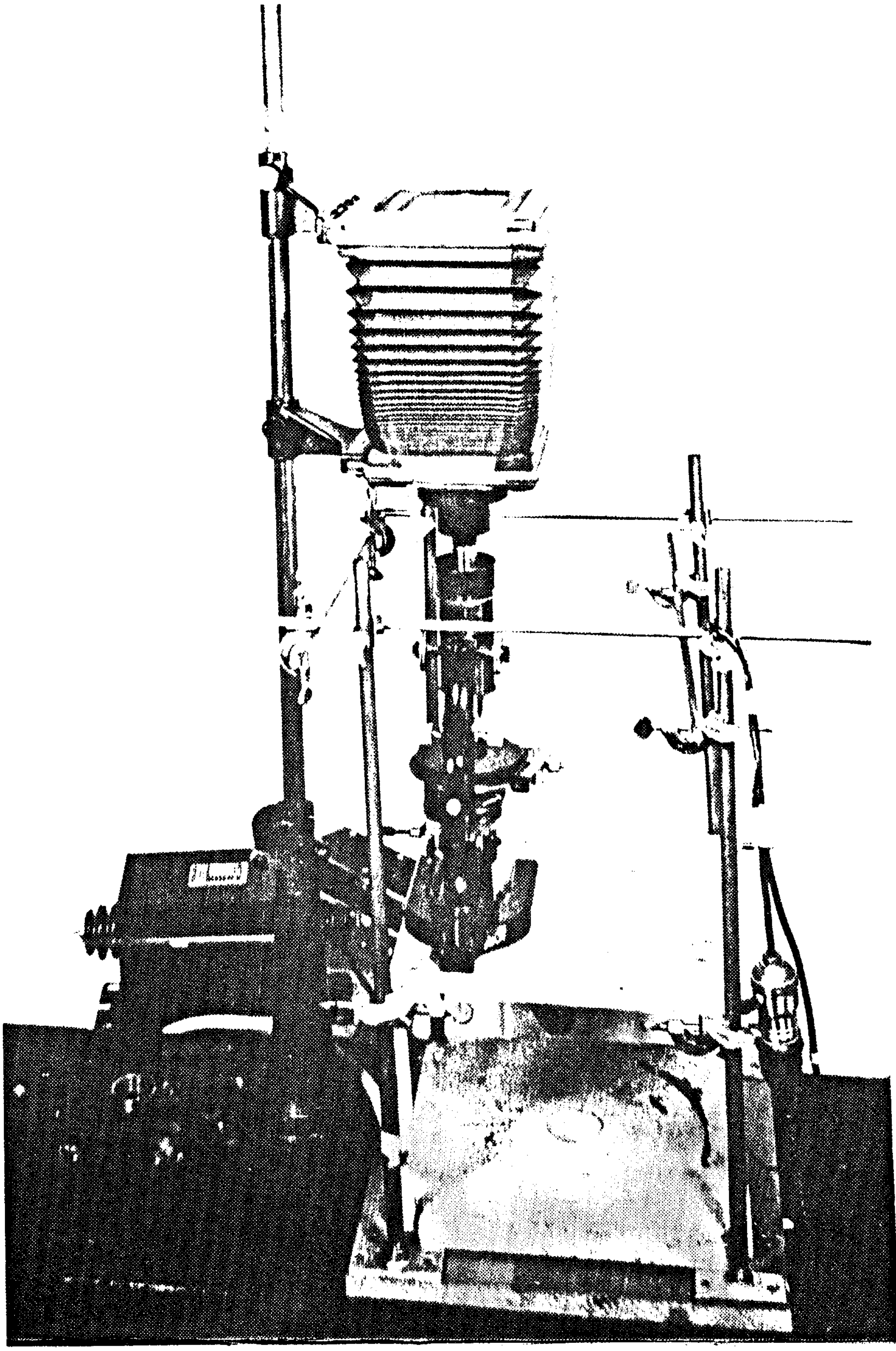


Plate No. 2

This page intentionally left blank.

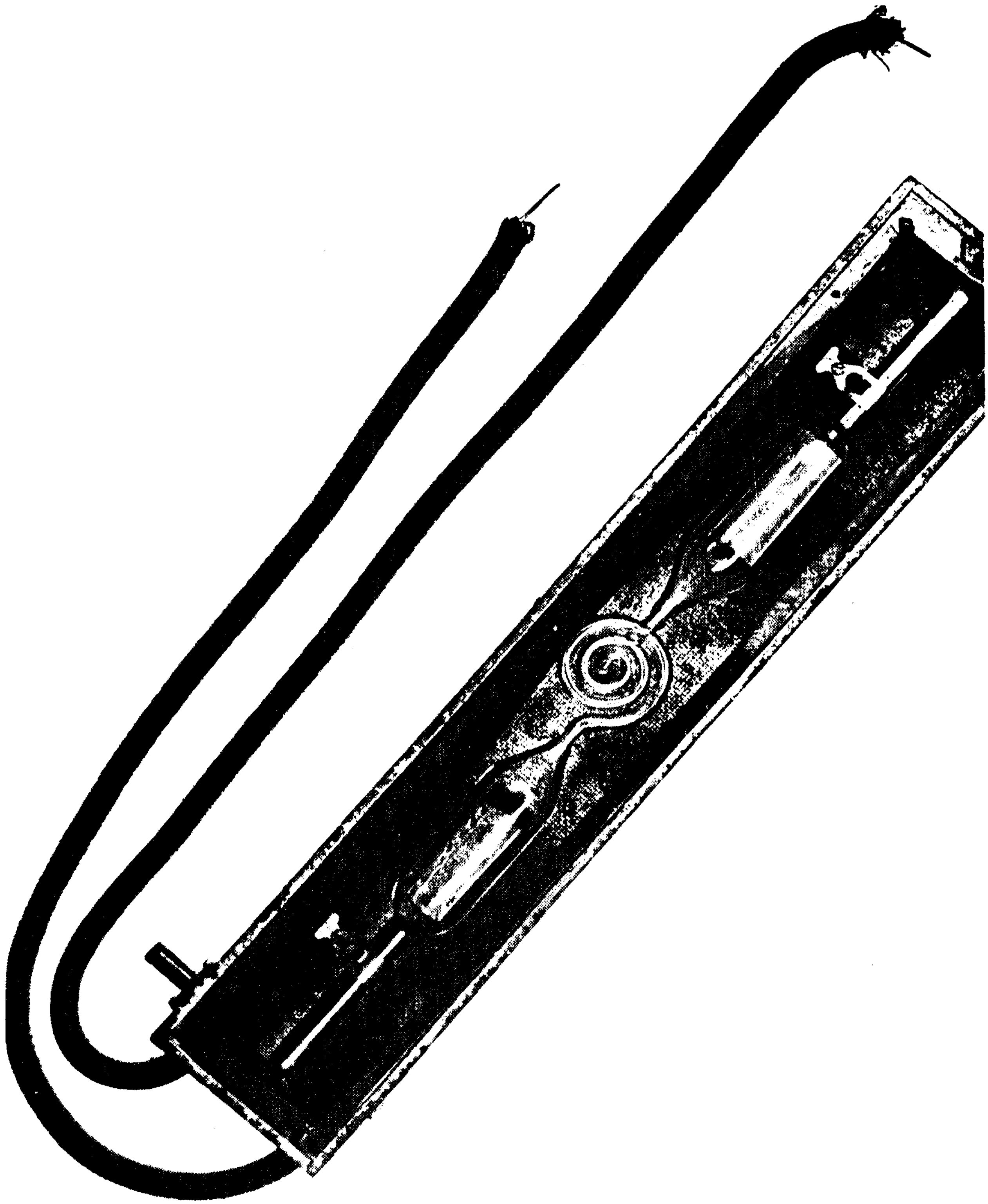


Plate No. 3

This page intentionally left blank.