[redacted]

OORR 59-2

c. 1

Office of Ordnance Research

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

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

OFFICE OF ORDNANCE RESEARCH
Report No. 59-2
August 1959

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

Sponsored by the Army Mathematics Steering Committee
conducted at
The Quartermaster Research and Engineering Center
Natick, Massachusetts
22-24 October 1958

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

TABLE OF CONTENTS

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

TABLE OF CONTENTS (Cont'd) Page

FOREWORD

The Army Mathematics Steering Committee (AMSC) at its 1958 April meeting accepted the invitation, issued by Dr. John K. Sterrett on behalf of the Quartermaster Corps, to hold the Fourth Conference on the Design of Experiments in Army Research, Development and Testing at the Quartermaster Research and Engineering Center at Natick, Massachusetts. This meeting, held 22-24 October 1958, was the first in this series of Army-wide conferences to be conducted outside the Washington, D. C. area. Through these symposia the AMSC hopes to introduce and encourage the use of the latest statistical and design techniques into the research, development, and testing conducted by Army scientific and engineering personnel. It is believed that this purpose can be pursued best by holding these meetings at various government installations through the country.

The five invited speakers at the Fourth Design Conference were C. I. Bliss, A. C. Cohen, A. W. Kimball, C. F. Kossack, and L. H. C. Tippett. Various aspects of preference studies, information on restricted samples, errors of the third kinds, and the American Association of State Highway Officials road test were the topics discussed by the first four of these men. The fifth speaker, L. H. C. Tippett, of the Shirley Institute, Manchester, England, talked to the group on some of the statistical methods now being applied in the textile industry. In addition to these addresses there were nine papers presented in the Clinical Sessions and eight in the Technical Sessions. Characteristics of sensitivity data, trajectory smoothing, performance criteria, and advanced scheduling were a few of the topics that came up for discussion in the Clinical Sessions. The papers presented in the Technical Sessions covered a wide range of topics; examples of problems dealt with included interface resistance in cathode tubes, properties of armor plate, bio-assay with pathogens, field tests, radar systems, and complex weapon systems.

The Fourth Conference was attended by 94 registrants and participants from 45 organizations. Speakers and panelists came from Boston University, Connecticut Agricultural Experiment Station, Harvard University, Oak Ridge National Laboratory, Princeton University, Purdue University, RCA Service Company, Shirley Institute, University of Georgia, University of Michigan, Virginia Polytechnic Institute, and 11 Army facilities. The present volume is the Proceedings of this conference, and it contains 19 of the 21 presented papers. The papers are being made available in the present form in order to encourage wider use of modern statistical principles of the design of experiments in research, development, and testing work of concern to the Army.

The members of the Army Mathematics Steering Committee take this opportunigy to express their thanks to the many speakers and other research workers who participated in the meeting; to Major General C. G. Calloway, Commanding General of the Quartermaster Research and Engineering Center at Natick, for making available the excellent facilities of his organization for the Conference; and to Mr. J. Schaller who handled the details of the local arrangements for the Conference.

Finally, the Chairman wishes to express his appreciation to his Advisory Committee, W. G. Cochran, F. G. Dressel (Secretary), Churchill Eisenhart, Landis Gephart, Frank Grubbs, Clifford Maloney, and J. K. Sterrett for their help in organizing the conference.

S. S. Wilks
Professor of Mathematics
Princeton University

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

22 - 24 October 1958

Quartermaster Research and Engineering Center

22 October 1958

REGISTRATION: 0930 - 1000 (Eastern Daylight Saving Time)
Lobby of the Administration Building

MORNING SESSION: 1000 - 1215 - Auditorium in the Administration Building

Chairman: Colonel George F. Leist, Ordnance Corps
Commanding Officer of the Office of
Ordnance Research

Introductory Remarks: Dr. J. Fred Oesterling, Acting
Scientific Director, Quartermaster Research
and Engineering Center, Quartermaster Research
and Engineering Command

Errors of the Third Kind in Statistical Consulting
Dr. A. W. Kimball, Oak Ridge National Laboratory

The AASHO Road Test as an Example of Large Scale Tests
Professor Carl F. Kossack, Purdue University

LUNCH: 1215 - 1345 - Cafeteria in the Administration Building

There will be two Technical Sessions conducted Wednesday afternoon. The security classification of Session II is CONFIDENTIAL. No clearances will be required for Session I.

TECHNICAL SESSION I: 1345 - 1600 - Conference Room in the Field House

Chairman: Walter Pressman, U. S. Army Signal Research
and Development Laboratory

Multiple Correlation of Mechanical with Ballistic
Properties of Armor Plate
Olga Sipes, Frankford Arsenal

Analysis of Cathode Interface Resistance Equipment
M. H. Zinn, U. S. Army Signal Research and
Development Laboratory

Experimental Designs for Bio-Assay with Pathogens
Ira A. DeArmon, Jr., Biological Warfare Laboratories,
U. S. Army Chemical Corps

TECHNICAL SESSION II: 1345 - 1600 - Room 200 in the Development Building

Security Classification - CONFIDENTIAL

Chairman: L. F. Nichols, Picatinny Arsenal

TECHNICAL SESSION II: (Cont'd)

The Application of Experimental Designs to Radar Systems Data
E. Biser, Harvey Eisenberg, and George Millman, Systems Division, Surveillance Department, U. S. Army Signal R & D Laboratory

Effects of Ballistic and Meteorological Variations on the Accuracy of Artillery Fire
O. P. Bruno, Weapon Systems Laboratory

SOCIAL HOUR: 1630 - 1730 - Conference Room in the Field House

23 October 1958

Clinical Sessions A and B will run concurrently on Thursday morning. The General Session Thursday afternoon will be followed by Technical Sessions III and IV. No clearances are required for any of the papers in the Thursday sessions.

CLINICAL SESSION A: 0930 - 1215 - Room 200 in the Development Building

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

Panel Members: A. C. Cohen, Jr., University of Georgia
A. Golub, Weapon Systems Laboratory
F. E. Grubbs, Weapon Systems Laboratory
Boyd Harshbarger, Virginia Polytechnic Institute

Characteristics of Various Methods for Collecting Sensitivity Data
A. Bulfinch, Picatinny Arsenal

Causes of Excess Dispersion in, and Optimum Components for, 20 mm HEI Accuracy Firing
Benjamin Shratter, Lake City Arsenal

Establishing and Testing Criteria for Trajectory Smoothing
Paul C. Cox, Reliability and Statistics Office, Ordnance Mission, White Sands Missile Range

Problems of Analysis. (1) Individual Variability (2) Interaction Effects
A. M. Galligan, Quartermaster Research and Engineering Center

CLINICAL SESSION B: 0930 - 1215 - Conference Room in the Field House

Chairman: D. H. K. Lee, Quartermaster Research and Engineering Center

CLINICAL SESSION B (Cont'd):

Panel Members: G. E. P. Box, Princeton University
W. G. Cochran, Harvard University
G. E. Noether, Boston University
L. H. C. Tippett, Shirley Institute

Evaluation by Indirect Means of Effects of Bacteria on an Unchallenged Host
Morris A. Rhian, Biological Warfare Laboratories, U. S. Army Chemical Corps

Determination of Performance Criteria for Quartermaster Corps Functions
John K. Sterrett, Research and Engineering Division, Office of the Quartermaster General

Program for the Interlaboratory Determination of Compression Set of Elastomers at Low Temperatures
S. L. Eisler, Rock Island Arsenal

An Appraisal of Sequential Analysis Under Conditions Restricted by the Requirements for Advanced Scheduling and Programming
E. W. Larson and W. D. Foster, Biological Warfare Laboratories, U. S. Army Chemical Corps

LUNCH: 1215 - 1315 - Cafeteria in the Administration Building

GENERAL SESSION: 1315 - 1500 - Auditorium in the Administration Building

Chairman: Dr. John K. Sterrett
Office of the Quartermaster General

Simplified Computational Procedures for Estimating Parameters of a Normal Distribution from Restricted Samples
Professor A. C. Cohen, University of Georgia

Statistical Problems Associated with Missile Testing
Dr. Charles L. Carroll, Jr., RCA Service Company

TECHNICAL SESSION III: 1515 - 1600 - Room 200 in the Development Building

Chairman: Ernest M. Kenyon
Quartermaster Research and Engineering Center

Application of Sequential Type Design and Analysis to Field Tests
Harold R. Rush, Quartermaster Field Evaluation Agency, Quartermaster R & D Command

TECHNICAL SESSION IV: 1515 - 1600 - Conference Room in the Field House

TECHNICAL SESSION IV (Cont'd):

Chairman: P. J. Loatman, Watervliet Arsenal

A Discourse on a Sequential Observational Program Use(in a Study of a Response Surface for a Complex Weapon: System
William J. Wrobleski, The University of Michigan

24 October 1958

Two invited speakers are scheduled to address the group on Friday morning. Right after the noon meal your host for this conference will conduct a tour of their installation.

GENERAL SESSION: 0930 - 1200 - Auditorium in the Administration Buildi

Chairman: Dr. Clifford J. Maloney, Chemical Corps Research and Development Command

Some Statistical Aspects of Preference Studies
C. I. Bliss, Connecticut Agricultural Experiment Station

Statistical Methods Applied to the Textile Industry
L. H. C. Tippett, Shirley Institute, Manchester, England

LUNCH: 1200 - 1330 - Cafeteria in the Administration Buildin

TOURS: 1330 - 1500 - Lobby of the Administration Building

Tours of the recently dedicated Solar Furnace, the Climatic Chambers, and special interest areas of the Chemicals and Plastics, Environmental Protection Research, Mechanical Engineering, Pioneering Research Textile, Clothing and Footwear Divisions will be arranged for Friday afternoon. The number and extent of the tours will be dependent on the time available.

WELCOME TO FOURTH CONFERENCE ON DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT, AND TESTING

J. Fred Oesterling
Headquarters Quartermaster Research and Development Command

Evolution in technical capability resembles biological evolution. In the last analysis, significant advances on the broad scale are dependent upon numerous small advances often occurring independently and sometimes almost in random fashion. But progress historically appears to proceed by saltation, rather than in the steady fashion that multiple independent events might be expected to produce. It seems that there is a certain dependence of discoveries on each other, or an interaction between discoveries independently made, which leads at times to vigorous upsurge in total effect, and occasionally to spectacular results.

It is very difficult to determine the historical significance of events while they are actually taking place, since one's field of vision scarcely exceeds the probable error of the events; but one might be pardoned, I think, for feeling that we now stand in the presence of one of these upsurges. The classical scientific procedure has been, and probably must remain, predominantly analytical. In the past, analytical procedures have largely had a deterministic base. One has tried to arrange events so that only one variable is at work, and this by controlled and precisely determined intervals. The limitations of this procedure were readily apparent, but methods for circumventing its limitations have been slow in coming. The turn of the century saw a rapid development in the basic probabilistic handling of data, and permitted the use of probabilistic experimental design. Today we see an expanding application of the probabilistic outlook to a wide variety of problems, including those of "naturally " occurring and uncontrolled events. No method can abstract from the data information which they do not contain; but there has been vast improvement in wringing from given data the maximum of information that they do contain. Where it is not possible to control variables to the extent that might be desired, or to make a fresh collection of better data, the information contained in the available observations can be largely abstracted and used as guides until better data are available. The degree of reliability to be placed upon the emergent information is probably as important a contribution as the information itself in furnishing guidance.

We are well aware that the subject you are meeting to discuss over the next three days, in spite of the esoteric titles of some papers, is methodologically of very great importance to the Quartermaster R&E Command. Insofar as we can, we do design and conduct highly controlled experiments; but so much of the QM operations is not susceptible to this type of examination, and the application of experimental results to actual operation is often far from a straightforward matter. Dr. Sterrett represents the spearhead of a modern mathematical attack upon our problems within the QM R&E structure. This will be extended, and we look forward to considerably better solutions to our numerous problems whether research, developmental, or operational. To this objective your deliberations cannot fail to make a material contribution.

ERRORS OF THE THIRD KIND IN STATISTICAL CONSULTING*

A. W. Kimball
Oak Ridge National Laboratory

Because graduate students in statistics are given little, if any, preparation for actual consulting, they are prone, particularly in their early years, to commit errors of the third kind, many of which could be avoided if the students were properly trained. Errors of the third kind are defined and are illustrated with actual examples from consulting experience. The cases used represent types of error which result from different situations that arise frequently in practice. Some discussion is included of possible remedies for this problem that are suggested by the experience of educators in other fields.

INTRODUCTION

At a relatively early age in their graduate academic life, students of statistics become familiar with certain risks associated with what they come to know as the first and second kinds of error in the theory of testing hypotheses. They soon learn that in many widely used statistical tests the first kind of error is easy to control but that often the risk of the second kind of error is difficult to compute and more often neglected entirely in practice. The importance of these errors is constantly brought to their attention through emphasis in their course work on such things as uniformly most powerful tests and sequential procedures which control the risks of both kinds of error. More recently the theory of decision making, the natural sequel to hypothesis testing, has elevated the notion of risk to an even higher place in the hierarchy of ideas passed on from professor to student.

As a result of these teachings many of today's statistics graduates come away from the warm comfort of university complacence into the coldly realistic outside world imbued with the idea (and probably rightly so) that the statistician's only real function in this world is to compute risks of error for other people who have to make decisions. To be sure, there is a vast amount of planning (design of experiments, model building) and intermediate adjustment (missing data, extreme observations) necessary before the statistician can estimate these risks, but essentially this is his main task, and the student finds it out usually before the end of the first semester.

Consider then the embryo statistician who has been released from the university's uterus with a shiny new degree and who proceeds on his mission as a risk computer fully equipped with the tools of his trade and the mental wherewithal to apply them. Let us assume that during the first few years of his initiation as a consulting statistician he is lucky, from a mathematical statistics point of view, and computes correctly the risks of error for all problems he tackles. The chances are, speaking nonmathematically, that during this time he will commit the third kind of error more often than he or anyone else realizes. What is even more tragic is that, although as a student

*This paper was originally published in the Journal of the American Statistical Association, vol. 52, no. 278 (June, 1957). Permission to reproduce it here is greatly appreciated by the editors.

he was constantly reminded of the importance of the first two kinds of error ar duly vowed always to keep sight of them, he was probably never made aware of th existence of a third kind of error, let alone told what to do about it.

The purpose of this paper is to draw attention to the third kind of error by quoting actual examples in which the error was made and later rectified. The hope is that the paper will serve simultaneously as a warning and as a moderator for newly trained consultants who tend to descend on research workers with the sometimes frightening enthusiasm and confidence of a freshman at his first football practice, and that perhaps it will help stimulate responsible educators to move more rapidly in filling this wide gap in graduate statistics training. Most conscientious teachers of statistics recognize this need and are searching for effective methods of correcting the situation, but very little real progress has been made.

In this connection there is an interesting analogy between graduate statistical training and medical training. The physician of today, after he completes internship and residency, is well trained to practice medicine but not so well trained to do research. This fact is recognized by many schools in which the M. D. who wants to do research in physiology is advised to get a Ph.D. in this field after he completes medical school. The emphasis in medical school is on practice since most medical graduates never see the inside of a research laboratory. The graduate statistician, on the other hand, is for the most part well trained to "go into practice," that is, to do statistical consulting. A safe guess is that over half of the graduates in statistics each year are lured into industry or government where their principal work is consulting, and those who do go to universities frequently find their nonteaching time fully occupied wit consulting both on and off campus. It is of utmost importance, therefore, that the third kind of error in statistical consulting be emphasized and brought out into the open. Otherwise nothing may ever be done about it.

THE ERROR OF THE THIRD KIND

A simple and almost ludicrous definition of the error of the third kind is the error committed by giving the right answer to the wrong problem. In definin it this way we are allowing the statistician the benefit of the doubt by rejecting the possibility that he would give the wrong answer to the wrong question. We are also protecting ourselves against the occurrence of a false positive, th is, the situation in which the wrong answer to the wrong problem turns out to b the right answer to the right problem. At this point the reader who finished the introduction without succumbing to the temptation to look ahead for a defin tion may well feel like the reader of a murder mystery who on the last page discovers that the victim committed suicide. Why, he may ask, should we concern ourselves with any consulting statistician who could be stupid enough to commit such an error? Admittedly, there may be many mature statisticians who prefer t take this attitude rather than face the consequences of accepting its alternati If this is so, the situation is indeed a grave one.

There is no way of knowing how many of us, particularly in our early years as consultants, were guilty of errors of the third kind, but it is almost certain that few have escaped an occasional mistake of this nature. The reason is simple enough. Many of us, in good faith, have helped research workers make t-tests, or compute analyses of variance, or design experiments thinking we were giving the right answer to the right problem; and usually we do give the right answer to the question that is asked. Unfortunately it often happens that the question asked has little bearing on the real problem, and we are led into committing the third kind of error.

A stranger to the intimacies of statistical consulting might well doubt that such ridiculous events could ever occur, but the experienced statistician knows that they do occur and will probably never be completely eliminated. Basically, errors of the third kind are caused by inadequate communication between the consultant and the research worker. In some instances, the research worker is at fault for failing to discuss his problem in complete perspective. He may feel that the statistician is weak in the subject matter field and that any attempt at a complete explanation would be a waste of time; or he may not have his ideas completely crystallized and may not want to be "confused" by a mathematician; or he may know a little statistics and feel that he can state the question adequately himself; or he may simply not want to take up too much of the consultant's time. At the same time the statistician is at fault for not becoming sufficiently familiar with the problem to enable him to advise intelligently. With proper preparation, sufficient patience, and persistent questioning of the experimenter, the consultant should be able to avoid most errors of the third kind, but not until he recognizes that they exist. In the next section an attempt is made to show that such errors can happen and under circumstances that ordinarily would not be regarded as unusual or bizarre.

EXAMPLES OF ERRORS OF THE THIRD KIND

The material for these examples is drawn for the most part from the author's own experience, with the natural result that most of the problems come from the field of biology. The main theme of the paper, however, is not biological and except for weakness in the subject matter field, either on the part of the author or the reader, the message should be clear. It should not be inferred that the errors illustrated are necessarily those of the author, although he would not deny this possibility.

Example I. An engineer was engaged in particle size determinations in connection with corrosion studies. He wanted to estimate the particle size distribution, which he was willing to assume normal, but his method prevented him from observing particle sizes below a certain diameter. He knew very little about statistics but he had heard that there were ways of estimating distributions when samples are restricted. There was no statistician in his own group to whom he could turn for help, but there was one nearby who, although very busy, might give him a reference.

So he visited the statistician and presented him with the following sample of particle sizes: 25.6, 7.1, 5.1, 4.2, 3.7, 3.0, 2.6, 2.0, 1.8, 1.6, 1.5, 1.4, 1.3, 1.2, 1.1, 1.0, 0.9, 0.8, 0.7 - and pointed out that his method would not allow him to determine particle sizes less than 0.7. Assuming the distribution normal, he wanted to know how he could estimate its mean and variance. The statistician was indeed quite busy and not inclined to spend much time on a problem he knew very little about and which did not originate in his group. On the other hand he did not want to cause any ill feelings by refusing to give any help at all. An easy way out was simply to hand the engineer one of his many reprints on truncated normal distributions (after all the engineer had asked for a reference), and this he did. Both participants in this short conference went away happy, the engineer because he thought he had an answer to his problem and the statistician because he disposed of an uninteresting proble in short order. But, as any reader who carefully inspected the "sample" of particle sizes already knows, an error of the third kind was committed. It might easily have gone unnoticed indefinitely, as do many others, but fortunately thi error was caught.

The engineer returned to his desk armed confidently with the newly acquire reprint and began to apply the method with the help of his 1935 model calculato He had not gotten very far along before he found that one of the statistics he computed was far outside the range of a key table given in the reprint to facilitate solution of the equations. After checking for and finding no arithmetical inaccuracies, he reluctantly returned to the statistician who inwardly was not too happy to see the engineer back. This conference lasted longer than the first, and with great chagrin the statistician finally realized what a stupid blunder had been made.

Among the methods used in particle size determination is one known as the sedimentation method. Briefly, it consists of the preparation of a liquid suspension of the material to be analyzed and the measurement of the decrease in concentration of particles at or above a particular level in the suspension as sedimentation proceeds. Under suitable conditions, Stoke's law can be used to compute the percentage of particles in the suspension having diameters greater than $\underline{d}$, say, where the value of $\underline{d}$ is determined by the time elapsed after sedimentation starts. Thus the random variables are the percentages, and $\underline{d}$ is a fixed or independent variate. It was this technique that the engineer had used. The appropriate method of estimation is, of course, probit analysis or one of i counterparts, and the "truncation" is not a problem except insofar as it increa the errors of estimate.

If the statistician had been familiar with particle size methods, or even i he had carefully scrutinized the "sample" that was presented to him, the error could never have occurred. It might be argued that both parties to this near-fiasco were the victims of circumstance and not really responsible, but if we are honest we must admit that the statistician has a duty to be more careful in avoiding this kind of error than perhaps any other. If he commits an error of the third kind, he is no less at fault than the physician who inadvertently administers arsenic instead of asperin.

Example II. A geneticist working in the field of radiation biology became interested in the relative biological effects of different kinds of radiation. In one experiment he hoped to compare the effects of gamma radiation and neutror radiation by exposing two groups of organisms separately to graded doses of eack kind of radiation and then determining the frequency of mutations at each dose. In previous experiments it had been found that mutation frequencies increase linearly with dose, so he planned to evaluate the relative biological effect by a comparison of the two slopes for the two kinds of radiation.

After the experiment was completed, he visited a newly hooded statistician and asked him to estimate the two slopes and make a statistical test of the difference between them. He explained that the gamma source used in the experiment was radioactive cobalt which provided an essentially pure source of gamma rays, but that the neutron experiment was carried out in a cyclotron and he had "corrected" the neutron doses for a known gamma ray contamination of about 7 per cent. The young statistician, who had little or no experience with radiation experiments and who at the moment was not particularly interested in learning about radiation, proceeded promptly and, as it turned out, rashly with his analysis. From the biologist he had obtained the following data:

Gamma experiment $(i = 1, \ldots, n)$

y_i = proportion of mutations
x_i = dose of gamma radiation

Neutron experiment $(j = 1, \ldots, m)$
u_j = proportion of mutations
v_j = "corrected" dose of neutron radiation.

Originally there were several replications at each dose point and the statistician had carefully tested for homogeneity. Finding no significant departure from binomiality, he pooled the replications and proceeded with a weighted linear regression for each experiment. He ended up with the two equations

$$\hat{y} = a + b_\gamma x$$
$$\hat{u} = a + b_n v,$$

for the gamma and neutron experiments, respectively. Finally he made the requested test of significance and chalked up (he thought) another successfully completed problem.

The third kind of error made by this statistician was most certainly avoidable. He had only to question the geneticist about the nature of the "correction" of the neutron dose, and without having to learn much at all about radiation dosimetry, he would have discovered his error. The consulting statistician, particularly in the physical science and engineering fields, soon learns to question any "corrections" applied by the experimenter before the data are presented for analysis. In the problem at hand it turned out that the geneticist had simply reduced the original neutron dose by 7 per cent intending thereby to evaluate the effect of neutrons uncontaminated by gamma rays. Overlooked was the fact that the corresponding biological effect still included the gamma component. When the error was uncovered, a somewhat different approach was taken. The two experiments were analyzed simultaneously by minimizing

$$\sum_{i=1}^{n} \lambda_i (y_j - \hat{y}_j)^2 + \sum_{j=1}^{n} v_j (u_j - \hat{u}_j)^2,$$

where

$$\hat{y}_i = a' + b_\gamma' x_i$$

$$\hat{u}_i = a' + b_\gamma' \, (0.07 w_i) + b_n' \, (0.93 w_i),$$

where the uncorrected neutron doses (w_i) were determined from the relation, $v_i = 0.93\, w_i$, and where λ_i and v_i are the appropriate weights. Needless to say, the second approach yielded estimates and standard errors somewhat different from those of the first approach, and the new significance test had to allow for the covariance between b_γ' and b_n'.

Once again in this example the blame must rest primarily with the statistician. Perhaps in his eagerness to apply his newly acquired skills to a problem which he thought fell into a pattern he had seen in graduate school, he temporarily lost his common sense. Whatever the explanation it is hard to draw any conclusion other than one which reflects the fact that he was just not ready to do statistical consulting on his own.

Example III. This example illustrates in a sort of general way a situation which must occur many times in the life of every consulting statistician. It might be called "Consulting by remote control," or "Communication without representation." Frequently the situation arises in a manner similar to the one in this example.

A research worker who, mostly through experience, had become fairly adept with many text-book statistical methods, encountered a problem which was new to him and which he could not find in his elementary text-book. He had computed two product-moment correlation coefficients and wanted to test the hypothesis that the population correlations were equal. He was reasonably sure that the t-test would not be appropriate, but he was also sure that some method must exist. The research organization to which he belonged did not employ a statistician, but he had a statistician friend in the same city who he felt would certainly have the answer. For such a minor problem the trip across town was hardly worthwhile, but thanks to Alexander Graham Bell, he knew he could solve his problem without leaving his desk. The phone call was made and the statistician, not wanting to be impolite or difficult by suggesting a meeting in person, and being allergic to long telephone conversations, quickly told his friend about the z-transformation and where to find an example of its use.

Sometime later both men happened to attend the same local seminar, and upon seeing his friend, the research worker rushed over to thank him for the useful advice about the z-transformation. During the course of the conversation, the statistician discovered to his horror that the experimenter had taken N simultaneous observations on three mutually correlated variables, x, y and z, and the two correlation coefficients which had been the subject of the aforementioned telephone conversation turned out to be the correlations between x and z and between y and z. With much embarrassment he realized that he had recommended a t-test between two z-transformed correlation coefficients which were not independent. Summing up all his courage he confessed his mistake and referred the experimenter to the paper by Hotelling [1] in which it is shown that under the null hypothesis, $\rho_{xz} = \rho_{yz}$,

$$t = \frac{\sqrt{N-3}(r_{xz} - r_{yz})\sqrt{1 + r_{xy}}}{\sqrt{2D}}$$

is distributed approximately as "Student's" t with N - 3 degrees of freedom, where

$$D = \begin{vmatrix} 1 & r_{xz} & r_{xy} \\ r_{xz} & 1 & r_{yz} \\ r_{xy} & r_{yz} & 1 \end{vmatrix}$$

The experimenter tried to accept the blame for this mistake contending that he should have taken the time to explain the actual problem more completely. Actually in this error of the third kind it would appear that both parties were at fault and for essentially the same reason - neither wanted to take the time to find out what the other was really doing.

Example IV. It seems desirable to include, as one of the examples of errors of the third kind, an error of omission. Essentially these errors occur when the statistician fails to do the best job possible simply because he has not taken enough time to question the research worker thoroughly about his experiment. In these cases, the answer given is often the right answer to the right problem but not always the best right answer. The following example illustrates an error of this kind.

A geneticist was engaged in a series of recombination experiments with bacteriophage T4. He was interested in testing for independence of the occurrence of two markers, r and tu. Under the hypothesis of independence, in an experiment in which plaques are counted for all four types of progeny, the observed and expected plaque counts can be represented as shown in the following table:

PLAQUE COUNT FREQUENCIES

Frequency	Type of Progeny				Total
	Parental	r^+	tu^+	r^+tu^+	
Observed	a_1	a_2	a_3	a_4	M
Expected	$M\ q_1q_2$	$M\ p_1q_2$	$M\ q_1p_2$	$M\ p_1p_2$	M

where p_1 and p_2 are the probabilities of events leading to recombinants r^+ and tu^+, respectively, and $q_1 = 1 - p_1$, $q_2 = 1 - p_2$. Typical experiments of this type yield about 90 per cent of parental type progeny and 10 per cent recombinants.

The geneticist who was doing these experiments had had some experience using chi-square in testing for independence with genetic frequently data, but since there were two parameters to be estimated in this case, he was not quite sure how to proceed. So he visited a young biometrician and presented him with data of the type shown in the above table. After explaining the experiment, he mentioned casually that he had much more data from another replication of this experiment but that it would probably be of little use since not all of the four classes of progeny were counted.

Perhaps it was too early in the morning, or perhaps the biometrician had his mind on something else. In any event he ignored the experimenter's casual remark about the other replication, proceeded to obtain maximum likelihood estimates of the parameters p_1 and p_2 from the complete experiment and correctly computed a chi-square with one degree of freedom which provided the required test for independence.

The results of the test were somewhat inconclusive, at least in the mind of the experimenter, and he began to reflect on why he had done the second replication in the first place. The greatest labor in experiments of this type is the counting of plaques, and since about 90 per cent of them represent parental type progeny, most of the work is done in counting plaques which provide little information about independence. It seemed reasonable to him, therefore, to do an experiment in which only the recombinants were counted. This was the second replication which he had mentioned to the statistician and it was about twice the size of the first.

With these points in mind he returned to the statistician and asked specifically if there wasn't some way in which the information from the second replication could be combined with the first so as to provide a more sensitive test for independence. As a result of this gentle prodding by the experimenter, who was obviously thinking more clearly than our young biometrician, an approach was found which would make use of all the data. The result of the second experiment was representable as:

PLAQUE COUNT FREQUENCIES

Frequency	Type of Progeny				Total
	Parental	r+	tu+	r+tu+	
Observed	-	a_5	a_6	a_7	N
Expected	-	$\frac{N\,p_1q_2}{(1-q_1q_2)}$	$\frac{N\,q_1p_2}{(1-q_1q_2)}$	$\frac{N\,p_1p_2}{(1-q_1q_2)}$	N

Under the hypothesis of independence the joint probability of both samples is

$$\frac{M!}{a_1!a_2!a_3!a_4!}\,(q_1q_2)^{a_1}(p_1q_2)^{a_2}(q_1p_2)^{a_3}(p_1p_2)^{a_4}$$

$$X\ \frac{N!}{a_5!a_6!a_7!}\,(1-q_1q_2)^{-N}(p_1q_2)^{a_5}(q_1p_2)^{a_6}(p_1p_2)^{a_7}.$$

The maximum likelihood equations for p_1 and p_2 can be reduced to a quadratic equation in p_2 with only one admissible root, and an equation in p_1 which is linear in p_2. A chi-square with three degrees of freedom is then easily computed. In this particular experiment the added strength of the second replication was sufficient to convince the geneticist that he had no reason to suspect lack of independence, whereas the significance level of chi-square based on the first replication alone had left him in doubt.

Perhaps there are only a few young statisticians who would commit an error of this kind, but the temptation must be great in many practical situations for the new consultant to discard extra observations which make the pattern of an experiment look different from what he has been accustomed to seeing in class examples. We so often hear it said that many research workers never come to the statistician until after the experiment is completed, and that frequently much of the data is worthless for statistical analysis. Certainly this does happen more often than it should, but in many apparently hopeless cases it also happens, as in the foregoing example, that a little extra effort on the part of the consultant will yield a workable, relatively simple method of analysis. A feel for these situations comes only with experience, but the graduate student should be given a chance to get some of this experience before he starts out completely on his own.

A POSSIBLE SOLUTION TO THE PROBLEM

Many readers may object to the examples which were chosen to illustrate errors of the third kind as being unrealistic and unlikely to happen in actual practice. To a large extent they are right because all of the errors discussed were eventually corrected and hence no longer qualify as errors. But it should be obvious that the only errors of the third kind which become known are those which are corrected, and for every one which is corrected there must be many which we will never know about. If we are ready to admit that these errors are committed and perhaps in large numbers, then we should also be ready to do something about it.

The obvious place to start is in graduate schools where degrees in statistics are awarded to students who expect to do statistical consulting. For some time to come these institutions will provide the largest part of the supply of consulting statisticians. If the consulting statistician were required by law to obtain a license before he could go into practice, we could take our cue from the medical profession. Every statistics graduate who expects to consult would be required to intern for, say, one year, and at the end of this time would be required to take an examination to obtain his license. This arrangement might or might not prove satisfactory but most people would admit that it is not practicable, at least not in the foreseeable future.

Let us turn then to the teaching profession. In many states licenses to teach are either not required or can be obtained merely by payment of a fee, and the teachers colleges, in addition to providing a comprehensive curriculum of course work must somehow prepare students for actual teaching. They accomplish this by the long established requirement of practice teaching. Every conscientious teachers college includes as part of its curriculum a period in which the student leaves the campus and under the direction of an experienced teacher learns to teach by teaching. In some schools practice teaching begins at the junior level, and college administrators have found that there is absolutely no substitute for it. Why then should not the statistics student be required to learn to consult by consulting?

Some statistics departments have attempted to achieve this goal by having the student "sit in" on consultations held by members of the staff. This undoubtedly helps to some extent, but frequently the student participates very little in the discussion and some staff members complain that their clients are reluctant to talk in the presence of graduate students. Whereas attendance at staff consultations may serve to introduce the student to the complexities of consulting, he can never learn to cope with them until he tries it on his own. To achieve this opportunity it is imperative that he leave the campus and "intern" in the field.

Exactly how this can best be accomplished is anybody's guess. As a start it would seem that graduate schools should attempt to obtain affiliations with consulting groups in government and industry, much as medical schools are affiliated with hospitals, or teachers colleges with practice schools. Universities contribute heavily to government and industry through the medium of the research contract. Both parties benefit, of course, even under the present system, but certainly both would benefit more in the long run if programs of student participation could be arranged. There must be many instances in which essentially this sort of arrangement has been made and proved successful, but only for an isolated student here and there. To be really effective such a program would have to be made an integral part of the graduate curriculum and listed in the catalog as one of the requirements for a degree.

Those of us in the profession of statistical consulting who take honest pride in our work face a real challenge. Two avenues are open to us. One is to ignore the presence of this situation and to continue along our narrow paths of individual self-satisfaction, oblivious of the effect it might have on the future of our profession. If this course is followed, when the production rate of new statisticians begins to catch up with the demand, we will face loss of prestige and public confidence, and possibly even virtual extinction. The other avenue is to recognize the problem, to appreciate that it is constantly increasing in intensity and to push hard for positive action as soon as practicable. We should have begun yesterday; today we are only thinking about it; tomorrow we must act.

REFERENCE

[1] Hotelling, Harold, "The selection of variates for use in prediction with some comments on the general problem of nuisance parameters," Annals of Mathematical Statistics, 11 (1940), 271-83.

THE AASHO ROAD TEST AS AN EXAMPLE OF LARGE SCALE TESTS

Carl F. Kossack
Purdue University

The AASHO Road Test is an extensive study of phenomena which arise when highway pavements of varying structural designs are subject to specified traffic. The experiment is sponsored by the American Association of State Highway Officials, the AASHO, and is administered by the National Academy of Science through its Highway Research Board.

The Road Test may be the largest experiment in history in which statistical designs have been attempted. Geographically, the Road Test is contained in an area about 300 feet wide and eight miles long near Ottawa, Illinois. Treatments and observations average to cost about fifty thousand dollars per experimental unit since there is a budget of over twenty million dollars for more than four hundred pavement units. Controlled treatments were begun in 1956 and will be continued into 1960. Many groups serve to advise the administrators and staff of the Road Test. One such group is the Statistical Advisory Panel, of which I am chairman, with W. J. Youden, National Bureau of Standards, and K. A. Brownlee, University of Chicago, the other two members. Dr. Paul Irick is the full time senior statistician associated with the Field Office of the Road Test.

In considering the experimental design for such an extensive experiment as the Road Test, it seems to be quite necessary for one to first consider the formal structure of an experiment and then to attempt to describe the Road Test with respect to these more general views.

Following the approach used by Dr. Irick, one can consider the structure of an experiment from five different but interrelated aspects:

(1) Objectives

(2) Designs for Data Acquisition

(3) Experimental Data

(4) Models for Association among experimental variables

(5) Analyses of the data

Let us consider first the problem of Objectives. Although experimental objectives generally call for the discovery or demonstration of associations among observable phenomena, explicit objective must often be inferred from general statements of purpose. This inference often makes any consideration of the consistency of objectives with the remaining aspects of the experiment a matter of interpretation. This point can bear careful consideration in most experiments since all too often the general purposes are vague and ambiguous in their expression if in fact they are even stated. One trouble with a large scale experiment in this connection is that the investment of so much time and energy in a large-scale experiment makes the interpretation of these objectives most critical since one usually does not have the option of simply modifying the experimental set up on the next time around if it is discovered that one has misinterpreted the objectives. We can thus note the first characteristic of a large-scale experiment that distinguishes it from other experiments. That is, the interpretation of the general purpose into specific objectives can rarely be evolved sequentially as th experiment progresses but must be clearly determined in advance of the actual acquisition of the experimental data.

In the AASHO Test the following general purposes were involved by a national advisory committee.

Purposes: The AASHO Road Test is intended to develop engineering facts and criteria which can be used

(1) In the design and construction of new pavements

(2) In the preservation or betterment of existing pavements and to evaluate the load carrying capabilities of existing highways

(3) As an engineering basis for the enactment of adequate and equitable legislation covering allowable loadings and highway taxation structure

(4) To provide information to assist vehicle manufacturers as to the types and capacities of highway vehicles which they design, construct, and offer as equipment to obtain overall economy of highway transportation

(5) To provide basic information as to engineering problems and the correlated costs of highways of different load carrying capabilities, and the proper taxation to cover cost of structural standards for highways which may be related to the cost of vehicle operation.

If one reflects for a minute over these general purposes I am sure that he will be impressed with the fact that they represent a coverage of problems in the highway transportation field that is, to say the least, breath-taking. The problem in the first stage of the experimental program then is to take such high sounding and general purposes and to translate them into more meaningful and concrete objectives. Such a task for large experiments of this type is a formidable one and requires a deep appreciation of the "state of the art" of the area involved as well as an appreciation of the research capabilities of the program being developed. In the case of the AASHO Road Test, this interpretation of purposes into objectives took more than three years and in fact the following objectives were not completely formulated until the experimental design itself was completed.

The official objectives are as follows:

Objective 1: "To determine the significant relationships between the number of repetitions of specified axle loads of different magnitude and arrangement and the performance of different thicknesses of uniformly designed and constructed asphaltic concrete, plain portland cement concrete, and reinforced portland cement concrete surfaces on different thicknesses of bases and subbases when on a basement soil of known characteristics."

Objective 2: "To make special studies dealing with such subjects as paved shoulders, base types, pavement fatigue, tire size and pressure, and heavy military vehicles, and to correlate the findings of these special studies with the results of the basic research."

Objective 3: "To provide a record of the type and extent of effort and materials required to keep each of the test sections or portions thereof in a satisfactory condition until discontinued for test purposes."

Objective 4: "To develop instrumentation, test procedures, data, charts, graphs, and formulas which will reflect the capabilities of the various test sections, and which will be helpful in future highway design in the evaluation of the load carrying capabilities of existing highways and in determining the most promising areas for further highway research."

Because of the time restriction, let us leave the objective phase of the experiment and consider the Designs for Data Acquisition. Let us look at the general layout of the Road Test Experiment in order to facilitate our consideration of this design phase of the experiment. Figures 1, 2, and 3 show the general layout of the test. The fact that each loop must be separated from the other loops created some design complications, but essentially the experimental unit involved are sections of pavements varying from 120 to 240 feet in length within each loop. As mentioned earlier there were available some 400 such experimental units to use in the design.

FIGURE 1: SCHEMATIC LOOP-TANGENT LAYOUT - AASHO TEST ROAD

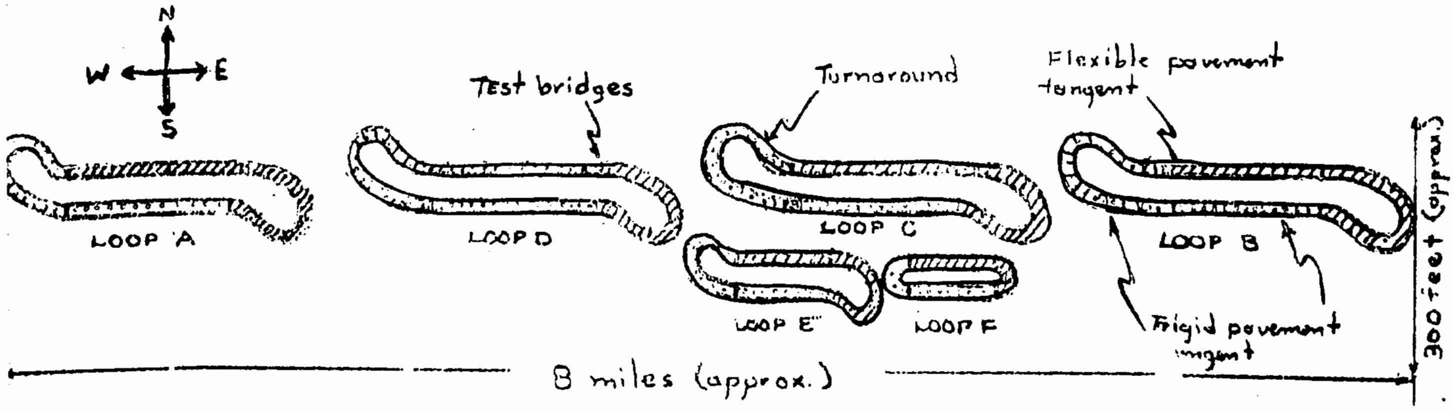
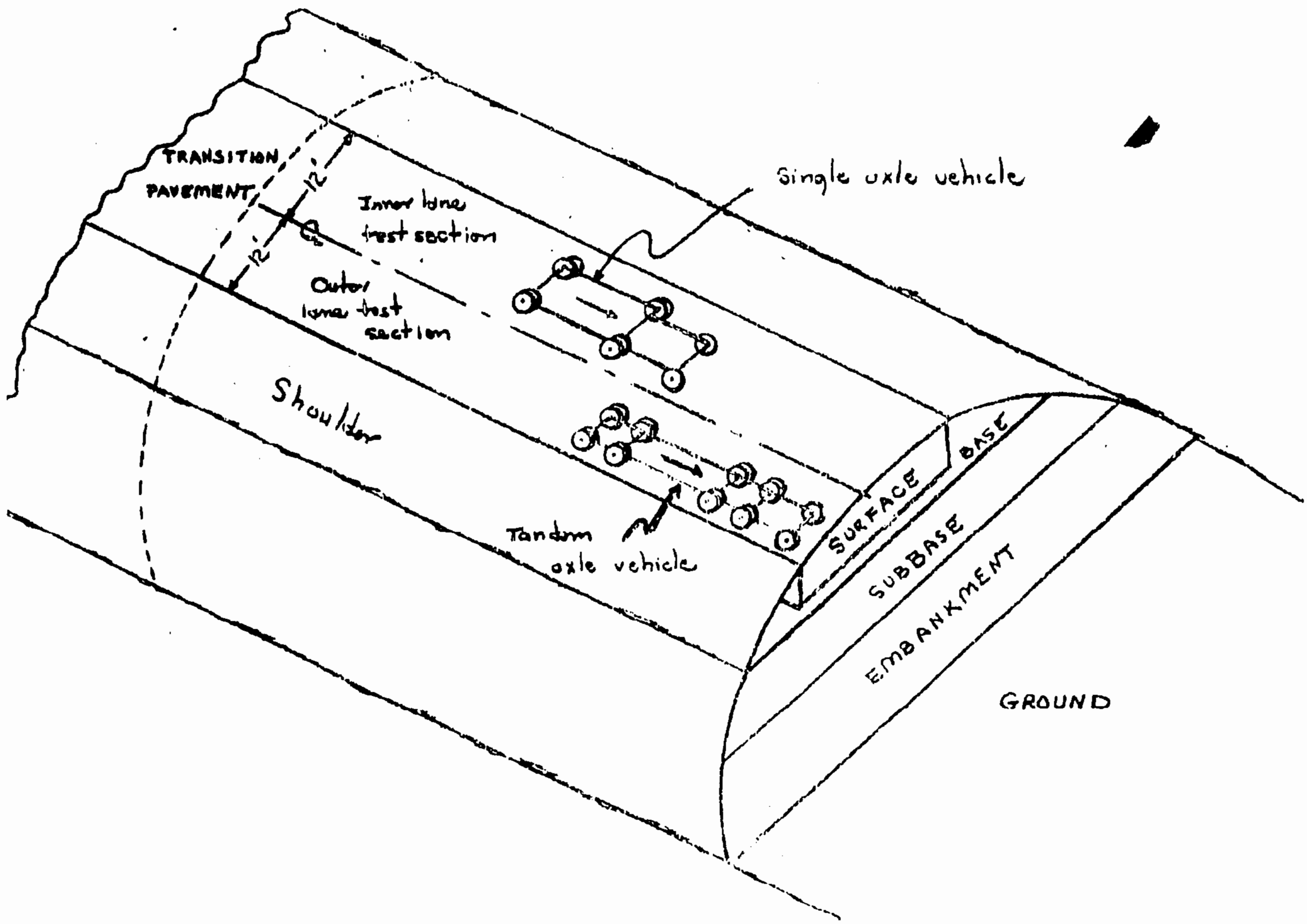


FIGURE 2: SCHEMATIC LOADED STRUCTURAL SECTION



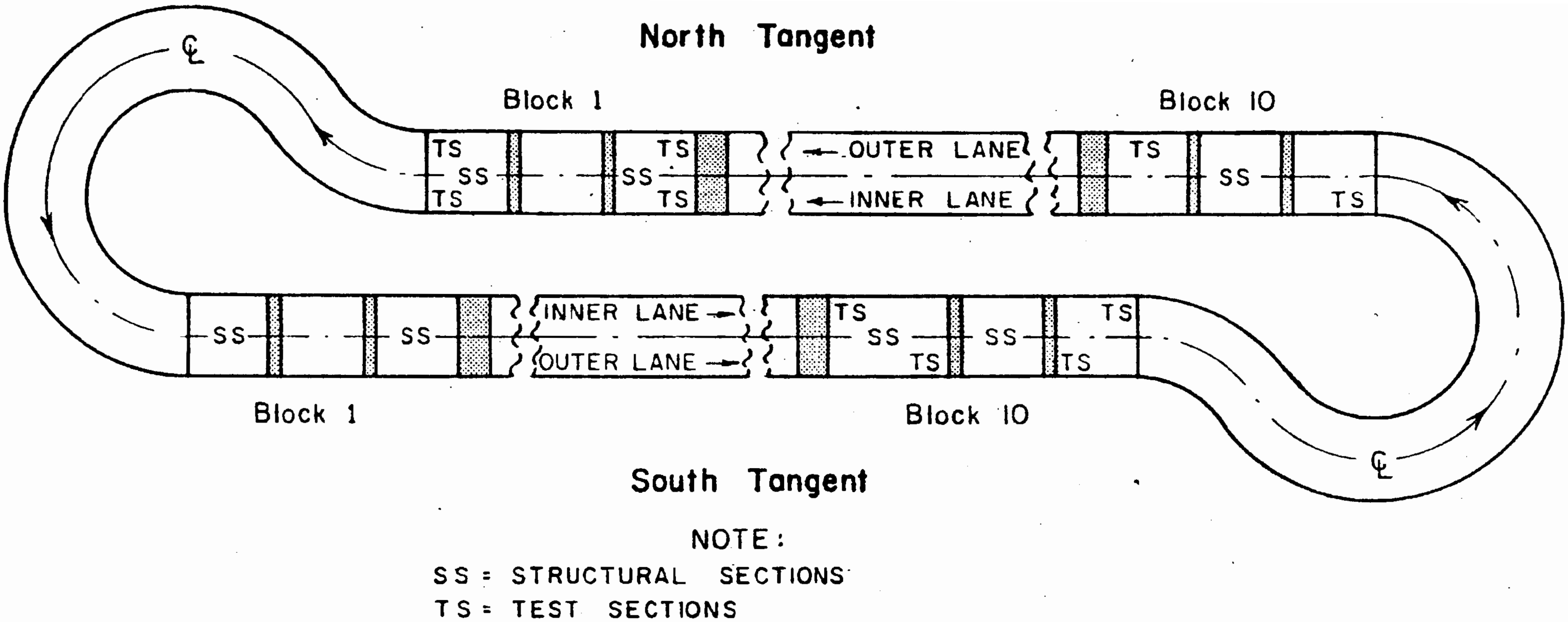


Fig. 3. TYPICAL LAYOUT OF TRAFFIC LANES, CONSTRUCTION BLOCKS, STRUCTURAL SECTIONS AND TEST SECTIONS FOR ONE LOOP OF THE AASHO ROAD TEST.

To evolve the design, it was decided that the main experiment was to center around a study of the interrelationship that involved the following factors:

Type of pavement

Surface thickness

Base thickness

Subbase thickness

Axle type

Axle load

The design used for these variables was to first divide the two principal pavement types, flexible (asphalt) and rigid (concrete) into separate experiments and within each of these experiments to use essentially a factorial design. That is in the flexible case the factorial was taken as a 3x3x3, three surface thickness, three base thickness and three subbase thicknesses. Figure 4 shows in a schematic diagram the type of factorial design used and how the various levels of each factor were assigned to the various loops. The surface types were divided between the two tangents of the loops, the two axle types were divided between the two lanes making up the loops and the varying loads were divided over the loops.

Test traffic for each main loop will consist of twelve vehicles, six in outer lanes and six in inner lanes. Vehicles will proceed counter clockwise around each loop at 30 m.p.h. and in a prescribed distribution of lateral placement. Load applications are scheduled to occur simultaneously in all traffic loops so that each structural section receives approximately 800 vehicle application per eighteen hour day, six days per week. The test traffic will continue for about two years and will involve considerably more than ten million miles of traffic. Figure 5 shows how these types of vehicles and loads were assigned to the various loops.

RIGID PAVEMENT

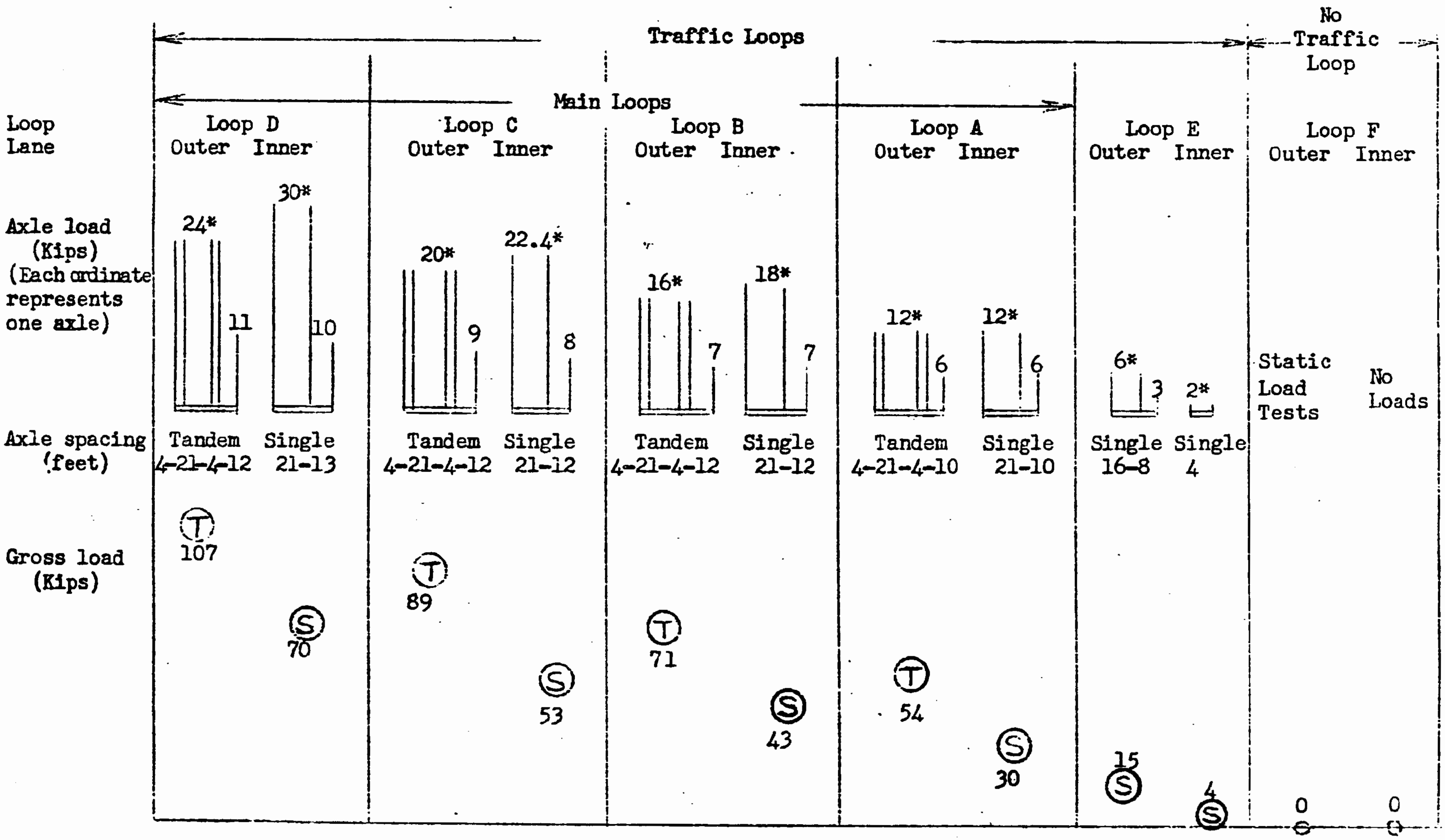
FACTORIAL EXPERIMENT DESIGN

Loop	Subbase Thick.	3.5 R	3.5 N	5.0 R	5.0 N	6.5 R	6.5 N	8.0 R	8.0 N	9.5 R	9.5 N	11.0 R	11.0 N	12.5 R	12.5 N
LOOP A 12 Kips Single	3	x	x	⊗	x	x	⊗	x	x						
	6	x	x	x	⊗	⊗	x	x	x						
	9	x	x	x	x	x	x	x	x						
LOOP A 24 Kips Tandem	3	x	x	⊗	x	x	⊗	x	x						
	6	x	x	x	⊗	⊗	x	x	x						
	9	x	x	x	x	x	x	x	x						
LOOP B 18 Kips Single	3			x	x	⊗	x	x	⊗	x	x				
	6			x	x	x	⊗	⊗	x	x	x				
	9			x	x	x	x	x	x	x	x				
LOOP B 32 Kips Tandem	3			x	x	⊗	x	x	⊗	x	x				
	6			x	x	x	⊗	⊗	x	x	x				
	9			x	x	x	x	x	x	x	x				
LOOP C 24 Kips Single	3					x	x	⊗	x	x	⊗	x	x		
	6					x	x	x	⊗	⊗	x	x	x		
	9					x	x	x	x	x	x	x	x		
LOOP C 40 Kips Tandem	3					x	x	⊗	x	x	⊗	x	x		
	6					x	x	x	⊗	⊗	x	x	x		
	9					x	x	x	x	x	x	x	x		
LOOP D 30 Kips Single	3							x	x	⊗	x	x	⊗	x	x
	6							x	x	x	⊗	⊗	x	x	x
	9							x	x	x	x	x	x	x	x
LOOP D 48 Kips Tandem	3							x	x	⊗	x	x	⊗	x	x
	6							x	x	x	⊗	⊗	x	x	x
	9							x	x	x	x	x	x	x	x

Notes: x Represents one test section.
⊗ Replicate section.
R Reinforced section, 6 panels at 40 ft. equals 240 ft.
N Non-reinforced section, 8 panels at 15 ft. equals 120 ft.

Figure 4: Factorial Design - Rigid Pavement

Figure 5: Axle load and Axle spacing levels for the AASHO Road Test
(All given values are approximate unless marked by *)



With this description of the Road Test let us give some attention to the more detailed description of the four additional aspects of an experimental investigation over and beyond the setting up of objectives. .

Under the Designs for Data Acquisition we can mention the following sub areas:

1) Selection of environmental and experimental units

In the Road Test we had such problems as to why locate at Ottawa, when to begin traffic, how long to make the sections, what shape should they have, how much spacing to have between sections, etc.

2) Selections for one level factors, design factors, co variables

In this area, for example, the use of a single aggregate in the construction was a major consideration. But the problems involved are numerous since it is at this time that one starts to consider the characteristic of the model to be evolved.

3) Selections for dependent variables

One could spend some time on this problem. Just to recall the range of interest expressed in the general purposes leads one to recognize that perhaps no single, simple dependent variable would suffice. At present a condition index is being evolved using several variables in the hope that through such an index one can measure the overall performance of a section under repeated application of loads. How to develop such an index is a major problem in itself.

4) Selection of transducer system for all measured variables

In this area one encounters the problems of actual making the physical measurements of the variables involved in an experiment. In the case of large scale experiments the transducer systems often need to be automatic which introduce problems of both validity and reliability. The danger is that one will become so wrapped up in developing the transducer systems that he will almost forget the main purpose of the experiment.

5) Selections for replication factors

What is needed is a decision as to the extent of replication that should be made associated with each of the several design variables. In the case of a large scale experiment in which the cost of each individual observation is considerable, a complete replication of the experiment is often uneconomical as well as politically not feasible. However, the fundamentals of scientific experimental design require both replication and randomization. In the Road Test Experiment a partial replication was evolved, see Figure 4, so that within each loop there appeared some replicated sections. One should remember that without replication no true error variance can be obtained and thus any relationship that is evolved from non-replicated experimental data must be taken at face value since confidence limits can not be determined. The other requirement, that of randomization, frequently meets resistance from the experimental worker or engineer on the grounds that it is simply busy work and only tends to complicate the operation of the experiment. I feel it significant that in the Road Test the Statistical Panel stood firm on the principal of complete randomization and had this principal adopted by the National Advisory Committee consisting of the foremost highway engineers of the country. Thus the Road Test design is a randomized design. The acceptance of this principal in a case such as this should put the lie to anyone who complains that to randomize a design is not possible. In building these highway sections the randomization often made it necessary to build the thin sections adjacent to thick sections according to the way the randomization came out. This randomization feature is one of the main requirements coming under the final area of Design and can be considered as item six:

6) <u>Space-time layouts for units, factors and observations.</u>

To turn our attention now to the third aspect, that of <u>Experimental Data</u> we have:

(1) <u>Values for dependent variables</u>

Here again, I could dwell at some length on the problems associated with this aspect of an experiment. In my experience, this country is filled with persons busily engaged in making observations on the wrong dependent variable. We seem to have in operation an unwritten axiom which says that as long as a dependent variable has been defined and is available that such a variable will meet the requirements of the experiment. In the case of the Road Test, concrete was actually being poured before an adequate dependent variable was finally evolved. Even at this time with the trucks actually beginning to run one it is not certain that satisfactory values of the dependent variables have been evolved.

(2) <u>Levels for design factors</u>

When a balanced design such as the complete factorial is used, the limitation encountered as to the number of levels available for the design factors is truly trying. One cannot simply throw a couple of extra levels into the experimental design to increase the assurance that the interesting range for the variable is covered for each factor. The expense is overpowering due to the multiplicative nature of most designs. In the Road Test we wanted to create a design so that the probability of a section failing sometime during the test would be about 2/3. This required that the section design straddle the point of adequate design for the given type of traffic. It should be noted the always occurring dilemma that if one knew how to design a highway we could thus design the experiment to find out how to design a highway. To meet this situation we called upon the best experience in the country on highway design to aid in the determination of the levels to be used in the design.

Dr. Box and his associates have recently considered this problem and have evolved some fairly significant results, but most of these require some sequential programming. In a large scale experiment especially those covering a long time period the inability to run preliminary experimental trials makes the problem more critical.

(3) <u>Levels for replication factors</u>

You can imagine the type of problem involved in this area when you realize that the same difficulties are present here as in the design levels. However, there usually is not as much freedom of action in selection of replication levels as in design levels. One often used the general idea in such situations that the replication should be spread over the entire sample space so as to yield a good estimate of the error variance. As one can note from Figure 4, the Road Test design followed this general pattern.

(4) Values for covariables

Often in the case of a large scale experiment there are many variables that may be measured which have some relationship with the dependent variable. The danger here is that one may lose sight of the main goal of the program in his zeal to measure all the variables that are available. In fact here again one encounters an axiomatic concept in existence. Namely, if one simply faithfully measures and records everything that happens in an experiment the analyses of the results are bound to be fruitful. The Road Test may be characterized in some respect as an outstanding opportunity for engineers to attempt to measure variables with increasing precision and automation. I have really lost tab of the extent of the data acquisition involved but the daily rate is in the millions of digits, all of which are needing storage and perhaps eventual analyses.

Let us proceed to the next aspect, that of Models for Associations where there are three areas delineated:

(1) Definition of experimental universe

In this case one must carefully consider what types of generalizations are to be made. On the one extreme the results obtained in the experiment can be simply stated to represent the particular and peculiar situation present at the time and place and conditions of the experiment. While at the other extreme, the experimental worker may attempt to conclude that his findings are applicable over all time and conditions. Still another problem is the determination of the sets of variables that will be used in attempting to explain a given phenomonon. I can only mention these problems in passing since their careful consideration would require more time than is available.

(2) Forms to represent associations

After giving a considerable amount of time to this problem, I find that it is in this area in which one frequently has difficulties. All too often tests are made of the data in which there is implicity involved some given form of association which is not applicable, but the routine of the test is carried out with little concern for these restrictions.

Even when one directly attacks the form of association problems he finds difficulties that are deeply rooted. One needs to consider many questions such as: Can the "existing state of the art" provide the necessary form? How should boundary value conditions be introduced into the form? Is one interested in an interpolative form or an extrapolation form? Can one use a routine polynomial model for the association and obtain satisfactory results? It should be mentioned that when the experiment must not only provide information which will yield the form of the association, but must also yield estimates of the constants appearing in the determined form, that such a dual requirement is most exacting.

(3) Allocation of assumptions and hypotheses

It is apparent that care must be given as to which relationships will be tested for their validity and which will be considered as assumptions and not amenable to testing. Much attention has been given in recent years to the difficulties encountered in sequential testing of hypotheses. One knows that the usually assumed procedures that are valid for single tests fail when applied to a sequential situation. Thus the failure to give proper allocation to assumptions and hypotheses will often lead into the pit-fall of sequential hypothesis testing.

The final aspect is that of Analysis. Here it may be noted that analyses especially those associated with large scale experiments are as often non-mathematical as they are mathematical in their nature.

Frequently an analysis will simply consist of a free hand sketch of a curve through some plotted points, or simply a visual comparison of the distribution of different sets of data. The more mathematical tests or analyses are reserved for those items deserving or requiring more careful methods.

Two areas can be noted under this aspect:

(1) Transformation of data into specific associates - the estimation problem

(2) Inferences with respect to the objectives

Since I believe these two areas are fairly well appreciated I would like to summarize my paper by giving some impressions I have received from serving as chairman of the Statistical Panel of the AASHO Road Test especially as they are related to the general problem of designing large scale experiments.

I will simply itemize these impressions without comment.

(1) There is a distinct problem of going from the general purpose of such experiments to the objectives and finally to the design and analyses. This becomes especially critical when one considers the more economic aspects of the problem.

(2) The emphasis upon instrumentation development work as the end in itself in such experiments needs to be modified.

(3) The dependent variable need exists in most problems.

(4) The danger of obtaining too much data is a real one.

(5) The education of the large number of individuals involved in the test in at least the rudiments of scientific method is essential.

(6) The interrelationship between the main problem and related problems need to be carefully studied.

(7) Committees can only do certain things in experimental work; the main decision making must be left to one or two individuals.

(8) Large scale experiments often introduce the need for Robust designs since one cannot risk the validity of the results to some "high-power" assumption.

MULTIPLE CORRELATION OF MECHANICAL WITH BALLISTIC PROPERTIES OF ARMOR PLATE

Olga Sipes
Research and Development Group, Metallurgy Research Laboratory
Frankford Arsenal

List of Symbols

N	Number of samples
r	Simple correlation
$r_{1.25}$, $r_{1.26}$, $r_{1.56}$ $r_{12.56}$, $r_{15.26}$, $r_{16.25}$	Partial coefficient correlation for variables designated by subscripts
$\bar{X}_1$, $\bar{X}_2$, $\bar{X}_5$, $\bar{X}_6$	Arithmetic Means
Σx^2	Total Variation
Σx_c^2	Explained variation
Σx_s^2	Unexplained variation
σ	Standard deviation
σ_s	Standard error of estimate
R	Multiple correlation coefficient

INTRODUCTION. A great deal of effort is currently being expended by the personnel of Metallurgy Research Laboratory at Frankford Arsenal in an attempt to find an adequate specification for aluminum armor plate. To do this by means of ballistic testing is costly and time consuming; therefore, the desirability of finding one or two simple mechanical or metallurgical tests whose results would correlate closely with the ballistic test is obvious. In any case it was suggested that a statistical analysis, using the techniques of multiple correlation be investigated in order to provide a quantitative index of the relative importance of the mechanical properties, singly or in combination, as they would pertain to the ballistic limits of an alloy.

In this particular study four aluminum alloys, from the Al-Mg family, made to the same specification but supplied by different manufactures, were considered; and six mechanical properties were correlated with the ballistic limit for specimens within this alloy family.

To determine whether or not the alloys should be studied individually or as a group, assuming the importance of variation in alloy chemical composition, the "t-test" was used. This analysis indicated that the difference observed in the means of the ballistic limits, were non-significant:

$$t_{X_1} = 0.5226 < t_{.05.21} = 2.080.$$

Hence, the data for all the aluminum alloys were treated as a group independent of composition.

The mechanical properties to be correlated with the ballistic limit (X_1) were as follows:

Yield Strength .... X_2

Ultimate Strength .... X_3

Modulus of Resilience .... X_4

Area Under Stress-Strain Curve ..... X_5

% Elongation .... X_6

% Contraction in Area .... X_7

To work with these six independent variables would make the task extremely complex and time consuming. Two methods were employed to reduce the number of variables. First: those properties which were related to each other or which measured similar characteristics were considered in order to eliminate one of them. Thus, modulus of resilience which is related to yield strength was eliminated. In turn, of the two common measures of ductility, percent contraction in area was eliminated in favor of percent elongation. The second method was to compute the simple correlation of each mechanical property with the ballistic limit and to confirm the reliability of these correlations by use of the "t-test" for "r" and the table of 1% and 5% points for "r". If the correlation was found to be non-significant, the mechanical property under consideration was dropped. When the simple correlation was found to be of borderline significance, as in the case of the ultimate strength, the decision to drop or retain this variable was based upon additional considerations. Thus the final decision to eliminate the ultimate tensile strength was based upon physical reasoning as supplied by the metallurgist, as well as statistics.

By this process of elimination the problem was reduced to the manageable task of considering three mechanical properties: yield strength, area under stress-strain curve, and percent elongation as the independent variables to be correlated with the ballistic limit as the dependent variable.

STATISTICAL PROCEDURE AND DISCUSSION OF RESULTS. It was convenient to compute at one time all the values of the sums and product sums that were needed in different formulae throughout the work and arranged them in tabular form, for ease of manipulation (Table II). Proceeding from this point the relationship or simple correlation between the dependent and each independent variable was determined. One variable was chosen to begin the evaluation and the remaining variables were introduced one at a time to note their effect individually and totally on the correlation. It was possible to study these effects through the changes of the explained and unexplained variations, the changes in correlations and in the standard error of estimates. Partial correlation was used to a great degree to note these changes and to determine the weights of each independent variable. Finally the estimating equation was derived including the four variables, also the standard error of estimate for these variables and the multiple correlation were obtained. Table I shows the data for the 23 specimens of the alloy under consideration. To determine the individual effects of the three factors X_2, X_5 , and X_6 on the ballistic limit, X_1, refer to Table III.* From this table it appears that elongations is the most important of the three independent variables. It has the biggest explained squares, explained variations and coefficient of correlation and it has the smallest unexplained variations and standard error of estimate. Figure I shows the scatter diagram of the simple relationship between the ballistic limit and each of the independent variables, the lines of best fit with a band of plus or minus one standard error of estimate, and the coefficient of simple correlations are also shown at the bottom. From this figure it appears that per cent elongations is the most important factor: the slope of the line approaches minus one, it has the largest simple correlation and the values are more concentrated around the line of best fit indicated by the narrower band or smaller standard error of estimate. Area under the stress-strain curve is about the same or of slightly lesser importance than % elongation and yield strength the least important. At this point it is tempting to say "Why go any further, we have established the important variables, what more do we want?" But this ranking does not necessarily hold true when other variables are introduced. The problem is to determine that it does.

The calculation used yield strength as the first factor, then % elongation and area under the stress-strain curve were considered in the order named. More information is obtained by a careful study of Figure 2. Section A indicates deviations of X_1 from their mean, that is, the total deviations; while section B shows the deviations in the estimates of the ballistic limit from their mean, that is the individual explained variations. Roughly, a small number of the bars in section B appear about the same as those in section A. Section C indicates the individual variations that have not yet been accounted for, that is, the deviations of the actual ballistic limit values from the estimated values. Inspection of the bars in section C will roughly verify the magnitude of unexplained variations or "residuals" (since they are obtained for each sample by algebraically subtracting the value of estimate from the actual).

In general, the bars in section C are smaller than those in section A, but there are exceptions. There is yet to explain why the ballistic limit is so low in sample 12, 13, moderately low in 19, 21 and 32, and so high in 11 and moderately high in 2, 15, and 30. Some clue to this difficulty is given by reference to Figure 3. In each section of this figure the dependent variable is the individual unexplained variations in the ballistic limit which are taken from section C of Figure 2. Section A (Figure 3) shows samples 12 and 32 with large and moderately large negative residual in the ballistic limit but a high reading in area under the stress-strain curve. Sample 11 is very high with respect to positive residual and moderately low in its reading of area under the stress-strain curve and the latter for sample 19 appears to be moderately above average but low in residual. Section B shows that samples

---

* Tables and Figures are placed at the end of this article.

12 and 19 are low in residual for % elongation also, and sample 11 shows a very high residual but is closer to the average in % elongation.

From an examination of the two sections of Figure 3 it appears that area under stress-strain curve and % elongation have approximately the same effect on the ballistic limit. It may be possible to reduce the errors in the estimate and improve the correlation, by including one or the other as a second factor. Percent elongation was taken as the second factor since its values are a little more concentrated around the mean, (r_{16} is slightly larger than r_{15}).

The results of some of the calculations using two independent variables are shown in Figure 4. Section A is the same as Figure 2 but in section B the line lengths have increased. Statistically, the explained variation has increased from $\Sigma x^2_{c1.6} = 71,031$ to $\Sigma x^2_{c1.26} = 100,450$. The increase in the explained variation is approximately 29,500. Since this increase is larger the correlation increased considerably from $r_{12} = 0.637$ to $r_{1.26} = 0.757$ but the increase from $\Sigma x^2_{c1.6} = 97,623$ to $\Sigma x^2_{c1.26}$ is not as large (2830), hence the increases from $r_{16} = -0.744$ to $R_{1.26} = 0.757$ is small. Likewise the unexplained variation in section C has been reduced. Correspondingly, the standard error of estimate $\sigma s_{1.26} = 57.01$ ft/sec. has declined from $\sigma s_{12} = 67.3$ ft/sec. and slightly declined from $\sigma s_{16} = 58.01$ ft/sec. Figure 5 shows the scatter diagram of values for area under stress-strain curve adjusted for yield strength and % elongation. In this figure the ballistic limit is considerably below the estimate for sample 13 and far above for sample 11 and moderately above for 30; however, these samples are close to the average value of area under the stress-strain curve for the group. It remains to be seen whether area under stress-strain curve, as such, is an important explanation of the ballistic limit. In reference to Figure 7, note that as the coefficient of multiple correlation becomes larger the standard error of estimate becomes smaller. The R's which include % elongation as a factor are the larges and the σ's which include this factor are the smallest. From this calculation it appears that the variation in the ballistic limit is influenced more by percent elongation than by the area under the stress-strain curve and by yield strength last.

Before introducing the last independent variable, (area under stress-strain curve), the partial correlation was considered to learn whether the relative importance of the different independent variables in explaining variations in the dependent variable was the same as already found by multiple correlation. This was done by finding the extent to which correlation was increased by addition of another factor. By definition, partial correlation is a measure of the relationship between the dependent variable and one independent variable, when the influence of the other independent variables have theoretically been removed from both. More precisely it is the square root of the ratio between the increase in the variation of the computed values of the dependent variable resulting from introducing another variable, and the variation that has not been explained before the introduction of the new factor.

One of the formulas for partial correlation is:

$$r_{12.6} = \sqrt{\frac{\Sigma x^2_{c_{1.26}} - \Sigma x^2_{c_{1.26}}}{\Sigma x_1^2 - \Sigma x^2_{c_{1.6}}}}$$

The values of the partial correlations found are as follows:

$$r_{12.6} = -0.194 \qquad r_{16.5} = -0.337$$

$$r_{12.5} = 0.069 \qquad r_{15.6} = 0.125$$

$$r_{15.2} = -0.463 \qquad r_{16.2} = -0.531$$

It might be thought that as additional factors were held constant the dependent variable would become progressively less closely associated with a given independent variable. For instance, the simple correlation between the ballistic limit and yield strength was found to be $r_{12} = 0.637$, but when the area under stress-strain curve and % elongation were each brought into the picture (technically, when the ballistic limit and the yield strength were adjusted for variation in area under stress-strain curve and % elongation) $r_{12.5}$ was 0.069 and $r_{12.6}$ was -0.194. What appeared to be a relationship between yield strength and ballistic limit was in fact largely a relationship between area under stress-strain curve and ballistic limit, and between ballistic limit and % elongation. The other partial coefficients were interpreted analogously. The relationship between ballistic limit and area under stress-strain curve, and between ballistic limit and % elongation exists to a much larger degree than that between ballistic limit and yield strength. Thus it appeared that the independent variables, taken together, produced a partial correlation that was fairly constant from sample to sample. There were some exceptions to this statement notably sample 11.

The results of the computation of partial correlation led to the same conclusions as did multiple coefficients: that % elongation was more closely related to the ballistic limit than either area under the stress-strain curve or yield strength, and that of the latter two, area under stress-strain curve was more influential than yield strength. The results of the computation using two independent variables are shown in Table IV.

It remained to be seen whether the conclusions concerning the relative importance of the three independent variables remained the same when all four were considered simultaneously, rather than as different combinations of three variables.

Having added the area under stress-strain curve into the calculation it was noted that the explained and unexplained variations did not change very much from the value obtained with yield strength and % elongation. Mathematically $\Sigma x_{c_{1.256}} = 102{,}584$ and $\Sigma x_{c_{1.26}} = 100{,}450$; their difference is 2160. The values for the partial correlation were:

$r_{12.56}$ = -0.221, $r_{16.25}$ = -0.410, and $r_{15.26}$ = 0.170.

Figure 7 is illustrative of progress made thus far. It may be noted that the coefficient of multiple correlation, $R_{1.256}$ = 0.764, steadily became larger and the standard error of estimate, $\sigma_{s1.256}$ = 56.33 ft/sec. steadily became smaller. By substituting the numerical values obtained for the constants in the normal equations computed by the method of least squares and solved by Doolittle's technique, the estimating equation became:

$$X_{c_{1.256}} = a_{1.256} + b_{12.56}X_2 + b_{15.26}X_5 + b_{16.25}X_6$$

$$X_1 = 2859 - .00567X_2 + .02737X_5 - 57.460X_6$$

To test the significance of the multiple correlation the analysis of variance was used. The F table indicated that for the .001 level of significance and with 3 and 19 degrees of freedom (n_1 and n_2), F should equal 8.28. Since the computed value for F (8.95) was larger than the tabular value, it can be said that $R_{1.256}$ is significant.

Finally, Figure 6 shows at this point that the addition of the area factor, has not, on the whole, improved the estimate very much as compared with the first two independent variables used.

SUMMARY. From the foregoing discussion it may be concluded that it would have been sufficient to work with two independent variables only: yield strength and % elongation or yield strength and area under the stress-strain curve, since the latter and per cent elongation show practically the same effect on the ballistic limit.

For this particular set of experimental data and statistical treatment the multiple correlation of 0.764 indicates that, there is relationship among the mechanical properties considered and the ballistic limit, and that about 60% of the variance in the ballistic limit is accounted for by the three factors considered. Moreover, if the readings of the mechanical properties of each sample are multiplied by the optimum weight factor indicated by the partial regression coefficients in the estimating equation, these readings would predict a ballistic limit within plus or minus 56.3 ft/sec. about 68% of the time, and within plus or minus 112.7 ft/sec. about 95.3% of the time.

Very much of the remaining unexplained variation could be due to the variability of the ballistic test. The standard error of 56.3 ft/sec. obtained above is not far from the variation normally observed in ballistic testing.

TABLE I

Item	Ballistic limit (X_1) (ft/sec.) *	Yield Strength (X_2) (lbs/sq.in.)	Area (X_5) Under Stress-Strain Curve	% Elongation (X_6)
2	2320	38200	11492	12.0
7	2240	38500	11218	12.2
10	2140	24600	15273	18.0
11	2400	35200	11053	12.5
12	2080	33100	14634	16.7
13	2120	39600	11882	12.8
15	2280	35400	12152	12.8
16	2235	33600	11402	12.6
17	2240	34200	13214	15.2
19	2140	35300	12334	14.6
20	2345	46800	7633	7.9
21	2180	37400	11710	13.4
22	2215	39600	11026	12.5
30	2345	42200	11090	11.6
31	2090	24600	14527	17.5
32	2125	32100	13871	15.4
33	2220	37200	12659	13.9
34	2150	29600	13046	14.4
35	2220	38600	10630	11.6
36	2275	37000	11146	11.8
37	2310	47100	9132	9.0
45	2330	45000	10414	10.0
46	2230	41700	9437	10.1

* Each value in column one is obtained by taking the average of three highest partial and three lowest complete penetrations with a grouping of six falling within 125 Ft/sec.

TABLE II. Computation of Deviation Product Sums Required for Measures of Relationship between Ballistic Limit and Yield Strength, Area under Stress-Strain Curve and Per Cent Elongation of 23 Samples of Aluminum Armor Plate

Sums and Means

$\sum X_1 = 51230$ $\bar{X}_1 = 2227.3913043$	$\sum X_2 = 846600$ $\bar{X}_2 = 36808.6956522$	$\sum X_5 = 270975$ $\bar{X}_5 = 11781.5217391$	$\sum X_6 = 298.5$ $\bar{X}_6 = 12.9782609$	$\sum X_Z = 1169103.5$ $\bar{X}_Z = 50830.5869565$

Product Sums, Correction Factors, and Deviation Product Sums

$\sum X_1^2 = 114,284,450$ $\bar{X}_1 \sum X_1 = 114,109,256.5193$	$\sum X_1 X_2 = 1,893,056,500$ $\bar{X}_2 \sum X_1 = 1,885,709,478.2622$	$\sum X_1 X_5 = 601,018,700.$ $\bar{X}_5 \sum X_1 = 603,567,358.6941$	$\sum X_1 X_6 = 661,152.$ $\bar{X}_6 \sum X_1 = 664,876.3059$	$\sum X_1 X_Z = 2,609,020,802$ $\bar{X}_1 \sum X_Z = 2,604,050,969.7815$
$\sum x_1^2 = 175,193.4807$	$\sum x_1 x_2 = 7,347,021.7378$	$\sum x_1 x_5 = -2,548,658.6941$	$\sum x_1 x_6 = -3724.3059$	$\sum x_1 x_Z = 4,969,832.2185$
	$\sum X_2^2 = 31,922,140,000$ $\bar{X}_2 \sum X_2 = 31,162,241,739.1525$	$\sum X_2 X_5 = 9,765,674,200$ $\bar{X}_2 \sum X_5 = 9,974,236,304.3545$	$\sum X_2 X_6 = 10,685,240.$ $\bar{X}_2 \sum X_6 = 10,987,395.6522$	$\sum X_2 X_Z = 43,591,555,940$ $\bar{X}_2 \sum X_Z = 43,033,174,917.4218$
	$\sum x_2^2 = 759,898,260.8475$	$\sum x_2 x_5 = -208,562,104.3545$	$\sum x_2 x_6 = -302,155.6522$	$\sum x_2 x_Z = 558,381,022.5782$
		$\sum X_5^2 = 3,265,059,203$ $\bar{X}_5 \sum X_5 = 3,192,497,853.252$	$\sum X_5 X_6 = 3,616,272.1000$ $\bar{X}_5 \sum X_6 = 3,516,784.2391$	$\sum X_5 X_Z = 1,363,536,837.51$ $\bar{X}_5 \sum X_Z = 13,773,818,300.5407$
		$\sum x_5^2 = 72,561,349.7474$	$\sum x_5 x_6 = 99,487.8609$	$\sum x_5 x_Z = -138,449,925.4407$
			$\sum X_6^2 = 4016.1900$ $\bar{X}_6 \sum X_6 = 3874.1088$	$\sum X_6 X_Z = 14,966,680.2900$ $\bar{X}_6 \sum X_Z = 15,172,930.3060$
			$\sum x_6^2 = 142.0812$	$\sum x_6 x_Z = -206,250.0160$

x_1	x_2	x_5	x_6	x_Z	
$\sum x_1^2 = 175,193.4807$	$\sum x_1 x_2 = 7,347,021.7378$	$\sum x_1 x_5 = 2,548,658.6941$	$\sum x_1 x_6 = -3724.3059$	$\sum x_1 x_Z = 4,969,832.2185$	x_1
	$\sum x_2^2 = 759,898,260.8475$	$\sum x_2 x_5 = -208,562,104.3545$	$\sum x_2 x_6 = -302,155.6522$	$\sum x_2 x_Z = 558,381,022.5782$	x_2
		$\sum x_5^2 = 72,561,349.7474$	$\sum x_5 x_6 = 99,487.8609$	$\sum x_5 x_Z = -138,449,925.4407$	x_5
			$\sum x_6^2 = 142.0812$	$\sum x_6 x_Z = -206,250.0160$	x_6

TABLE III

X_1 vs X_2 (Yield Strength)	X_1 vs X_5 (Area)	X_1 vs X_6 (% elongation)
Estimating Equations:		
$X_{c1.2} = 1871.525 + .00967X_2$	$X_{c1.5} = 2641.208 - .03512\ X_5$	$X_{c1.6} = 2567.584 - 26.212\ X_6$
Sum of explained squares (explained by gross estimating equations):		
$\Sigma X^2_{c1.2} = 114,180,286$	$\Sigma X^2_{c1.5} = 114,198,381$	$\Sigma X^2_{c1.6} = 114,206,908$
Explained variations (explained by gross estimating equations):		
$\Sigma x^2_{c1.2} = 71,031$	$\Sigma x^2_{c1.5} = 89124$	$\Sigma x^2_{c1.6} = 97,623$
Unexplained variations (sum of squares deviation from estimates):		
$\Sigma x^2_{s1.2} = 104,163$	$\Sigma x^2_{s1.5} = 86,069$	$\Sigma x^2_{s1.6} = 77,570$
Standard errors of estimates:		
$\sigma_{s1.2} = 66.58$	$\sigma_{s1.5} = 61.13$	$\sigma_{s1.6} = 58.01$
Coefficients of simple correlation:		
$r_{12} = 0.637$	$r_{15} = -0.715$	$r_{16} = -0.744$

TABLE IV

X_1 with X_2 (Y.S.) & X_5 (Area)	X_1 with X_2 (Y.S.) & X_6 (% elong.)	X_1 with X_5 (Area) & X_6 (% elong.)
	Estimating Equations:	
$X_{c1.25} = 2631.28 + .000134\ X_2 - .0347\ X_5$	$X_{c1.26} = 2882.87 - .00488\ X_2 - 36.60\ X_6$	$X_{c1.56} = 2512.57 + .0204\ X_5 - 40.51\ X_6$
	Sum of explained squares:	
$\Sigma X^2_{c1.25} = 114{,}198{,}795$	$\Sigma X^2_{c1.26} = 114{,}209{,}744$	$\Sigma X^2_{c1.56} = 114{,}207{,}558$
	Explained variations:	
$\Sigma x^2_{c1.25} = 89{,}423$	$\Sigma x^2_{c1.26} = 100{,}450$	$\Sigma x^2_{c1.56} = 98{,}302$
	Unexplained variations:	
$\Sigma x^2_{s1.25} = 85{,}770$	$\Sigma x^2_{s1.26} = 74{,}760$	$\Sigma x^2_{s1.56} = 76{,}360$
	Standard errors of estimates:	
$\sigma_{s1.25} = 61.02$	$\sigma_{s1.26} = 57.01$	$\sigma_{s1.56} = 57.72$
	Partial coefficient of correlation:	
$r_{1.25} = 0.059$	$r_{1.26} = -0.193$	$r_{1.56} = 0.125$
	Multiple coefficient of correlation:	
$R_{1.25} = 0.715$	$R_{1.26} = 0.757$	$R_{1.56} = 0.749$

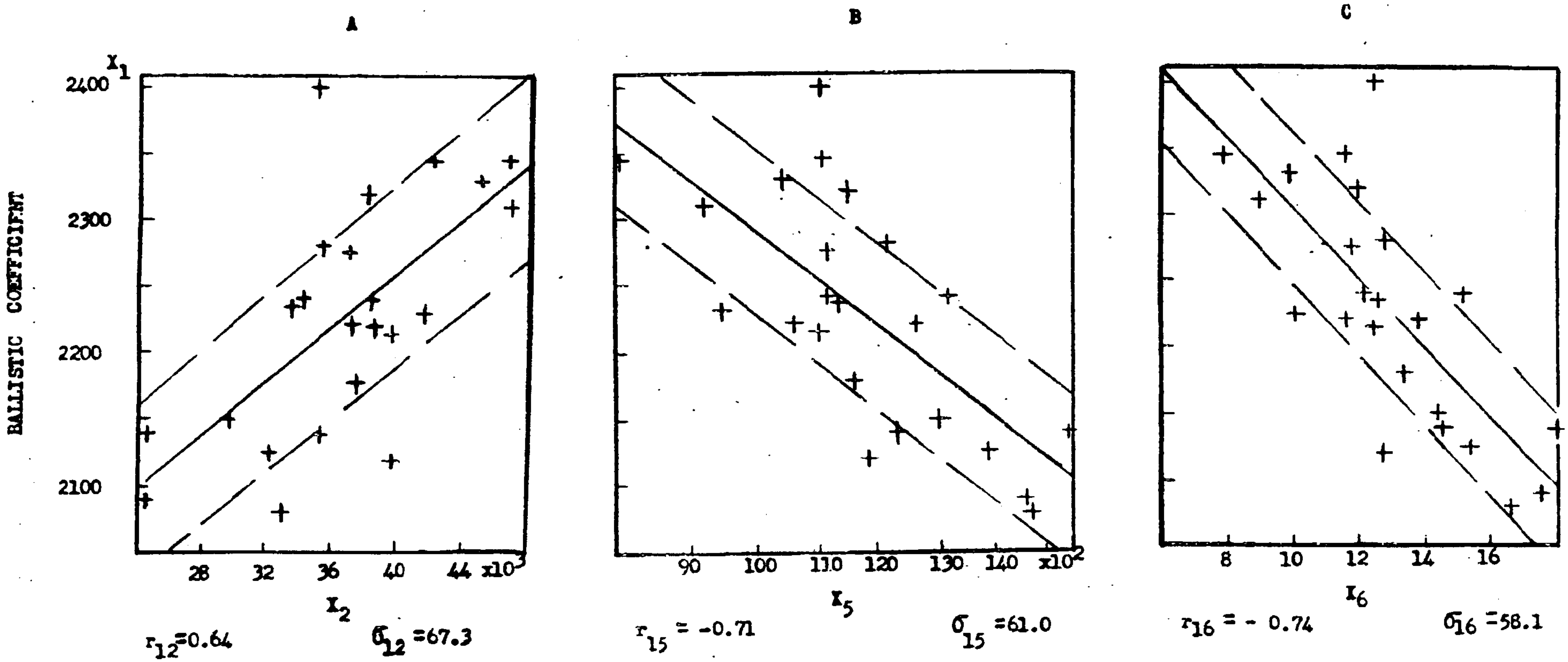


Figure 1. Scatter diagrams of gross relationship of yield strength (X_2), area (X_5), under Stress-Strain Curve and per cent elongation (X_6) for 23 specimens

E

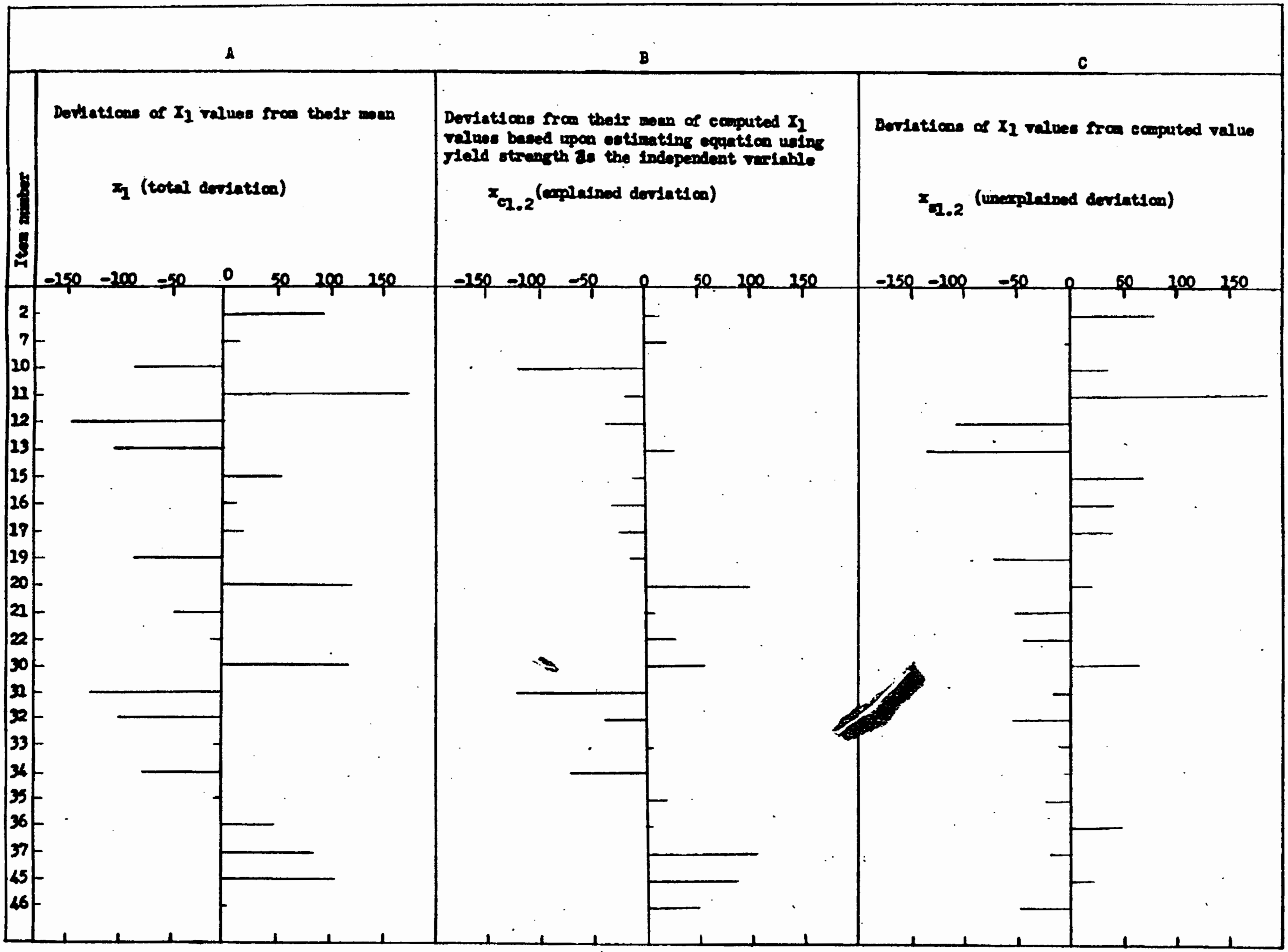

A
B
C
Item number
Deviations of X1 values from their mean
x1 (total deviation)
Deviations from their mean of computed X1 values based upon estimating equation using yield strength as the independent variable
xc1.2 (explained deviation)
Deviations of X1 values from computed value
xs1.2 (unexplained deviation)
-150 -100 -50 0 50 100 150
-150 -100 -50 0 50 100 150
-150 -100 -50 0 50 100 150
2
7
10
11
12
13
15
16
17
19
20
21
22
30
31
32
33
34
35
36
37
45
46

Figure 2.

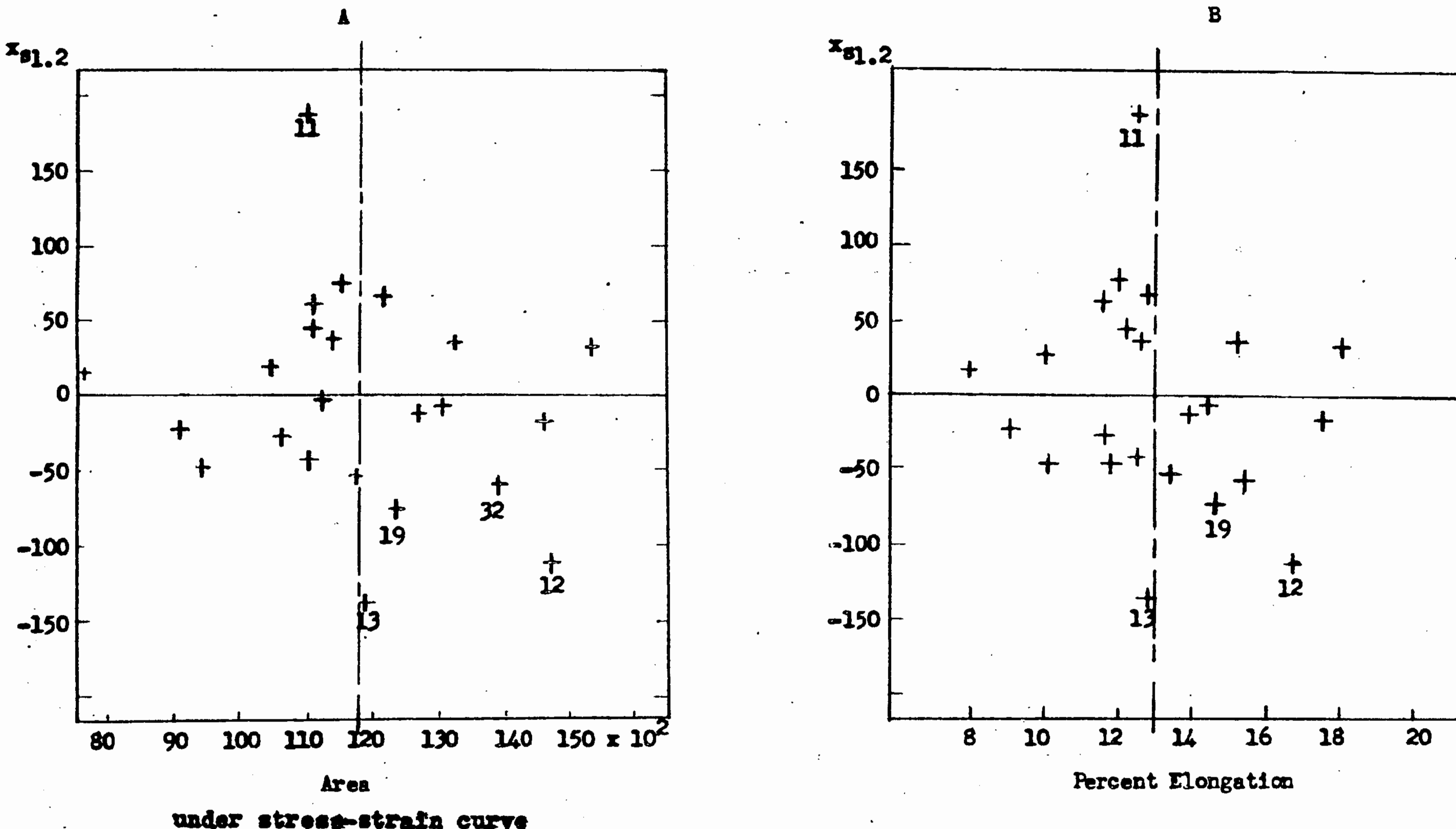


Figure 3. Scatter diagrams of area under stress-strain curve (X_5) and percent elongation (X_6), compared with the ballistic limit (X_1) and adjusted for yield strength (X_2)

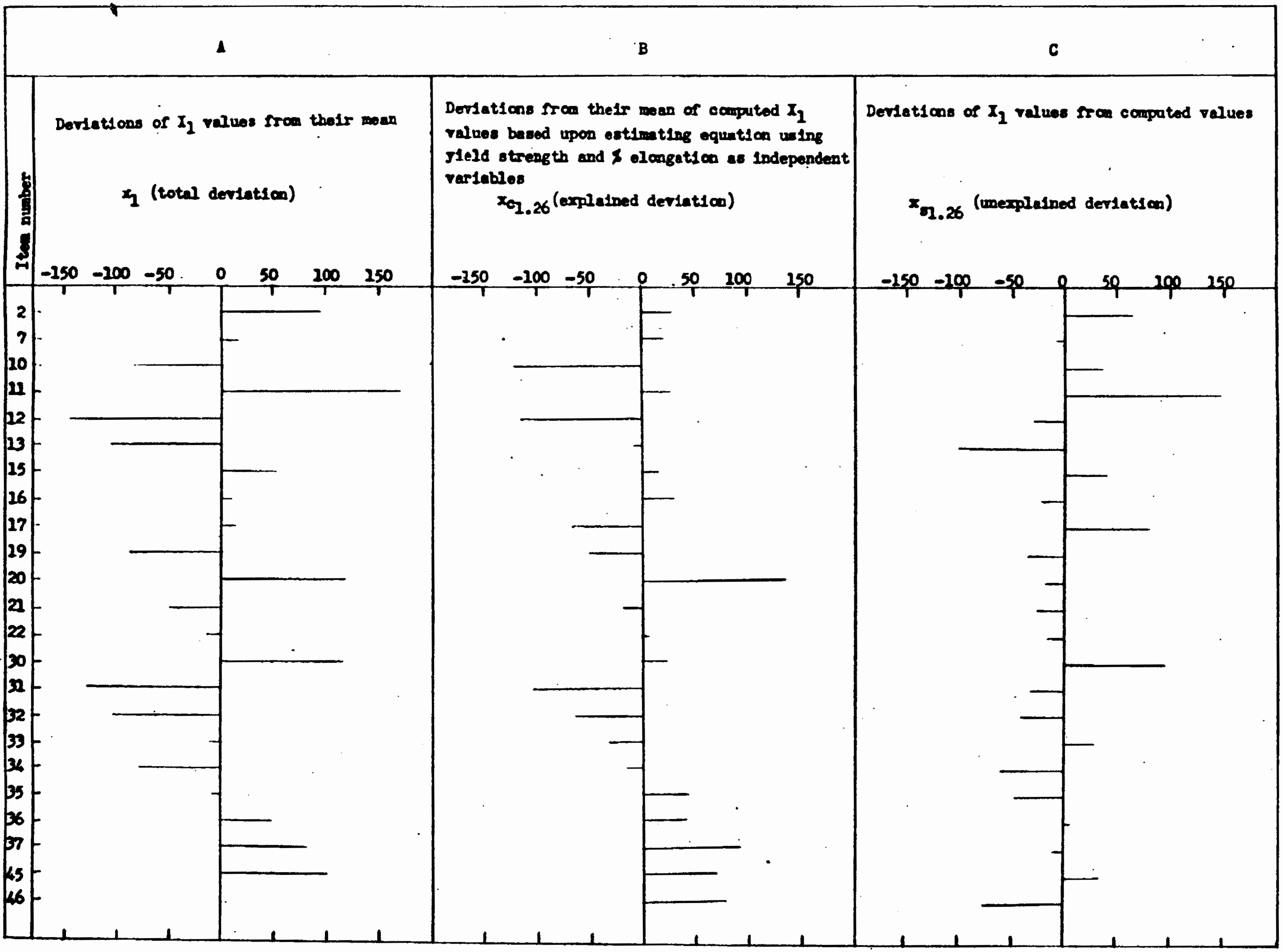

A
B
C
Item number
Deviations of X_1 values from their mean
x_1 (total deviation)
Deviations from their mean of computed X_1 values based upon estimating equation using yield strength and % elongation as independent variables
$x_{c1.26}$ (explained deviation)
Deviations of X_1 values from computed values
$x_{s1.26}$ (unexplained deviation)
-150 -100 -50 0 50 100 150
2
7
10
11
12
13
15
16
17
19
20
21
22
30
31
32
33
34
35
36
37
45
46

Figure 4.

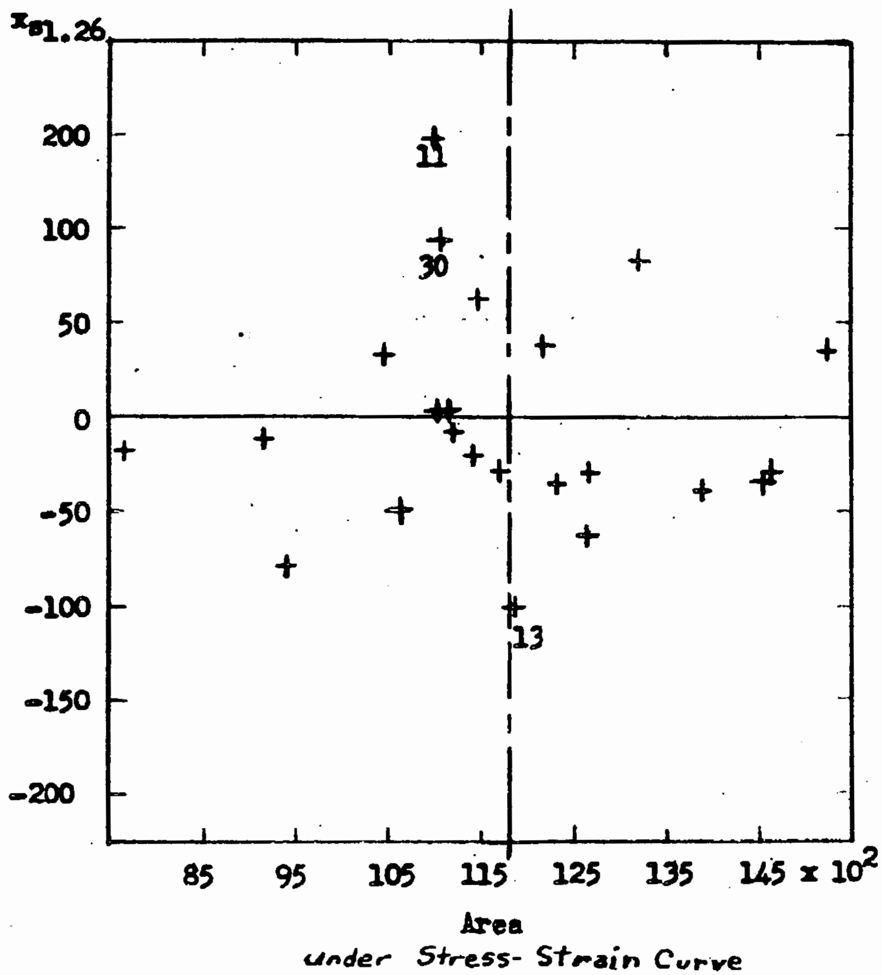


Figure 5. Scatter diagram of area (X$_5$) and X$_1$ values adjusted for yield strength and percent elongation

36-221-54711-ORD-58

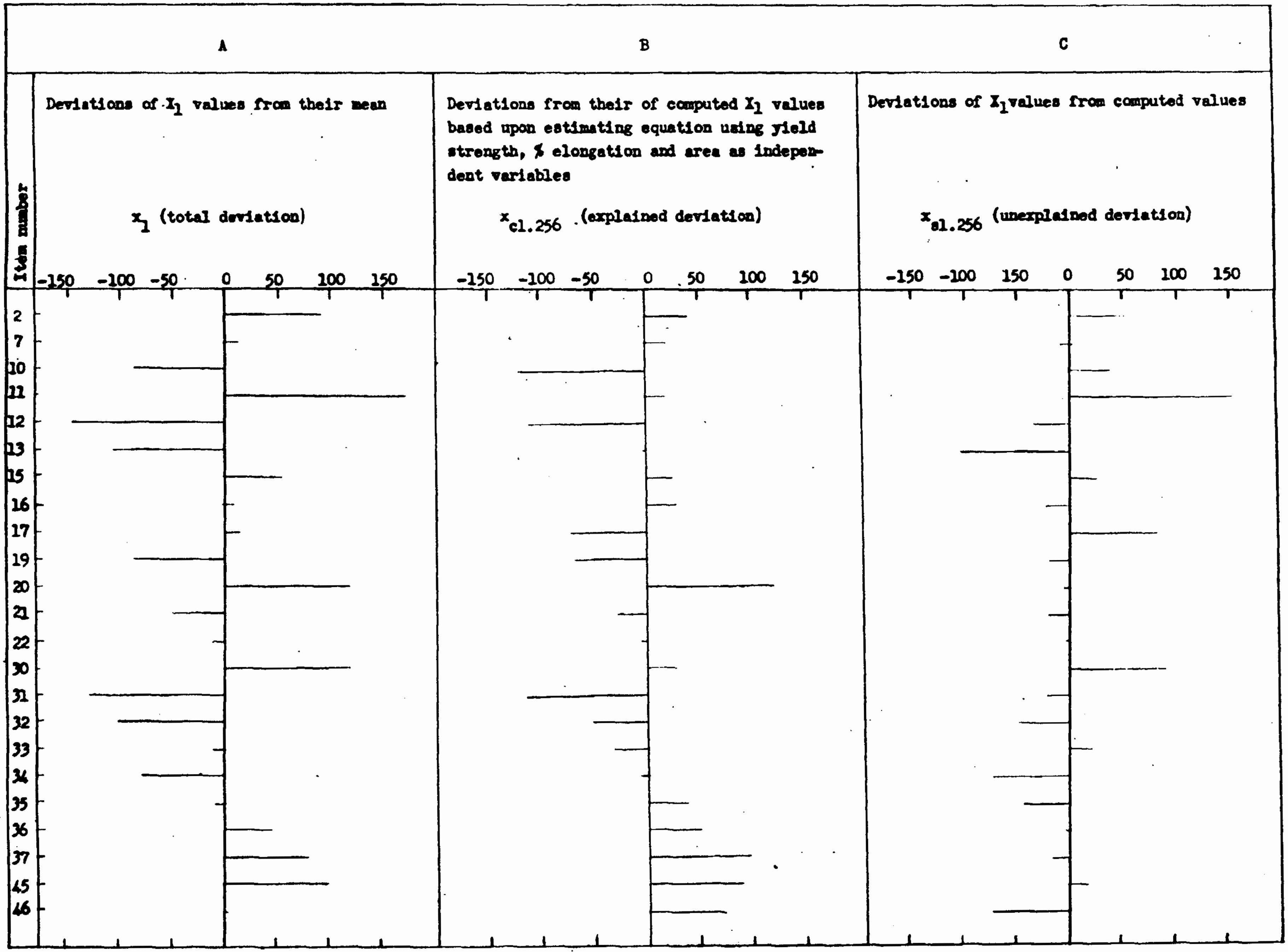

A
B
C
Item number
Deviations of X_1 values from their mean
x_1 (total deviation)
Deviations from their of computed X_1 values based upon estimating equation using yield strength, % elongation and area as independent variables
$x_{c1.256}$ (explained deviation)
Deviations of X_1 values from computed values
$x_{s1.256}$ (unexplained deviation)
-150 -100 -50 0 50 100 150
-150 -100 -50 0 50 100 150
-150 -100 150 0 50 100 150
2
7
10
11
12
13
15
16
17
19
20
21
22
30
31
32
33
34
35
36
37
45
46

Figure 6.

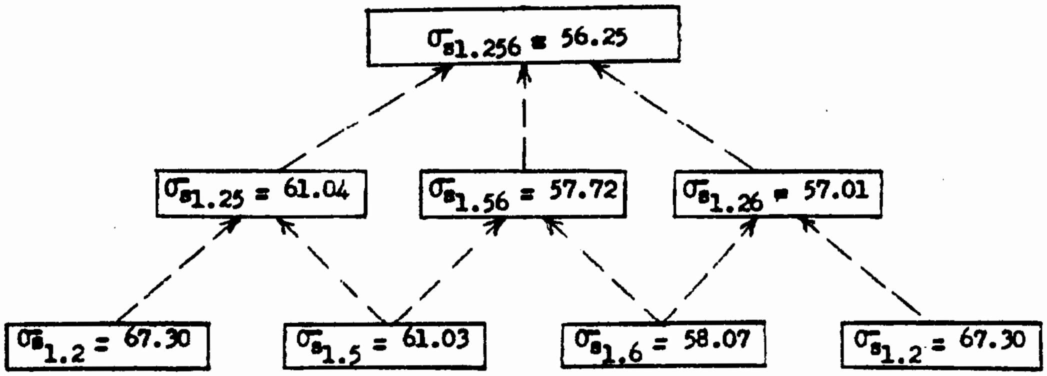


Figure 7a. As the number of variables increase the standard error of estimate becomes smaller

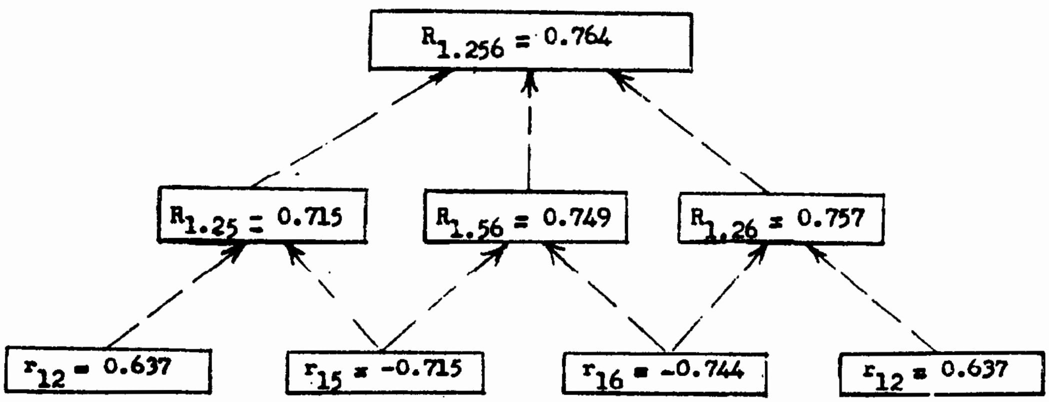


Figure 7b. As the number of variables increase the coefficient of correlation increases

Figure 7

ANALYSIS OF CATHODE INTERFACE RESISTANCE EXPERIMENT[1]

M. H. Zinn
U. S. Army Signal Research and Development Laboratory

At the Third Conference on Design of Experiments the general problem of electron tube experiments was discussed in a clinical paper.[2] The discussion was illustrated with a particular experiment concerning the study of cathode interface resistance growth during life of receiving-type electron tubes. It is the purpose of this paper to review the experimental design and discuss the statistical techniques utilized in the analysis of the data.

Cathode interface resistance is caused by the formation of a layer at the interface between the barium-oxide coating and the nickel base of an oxide-coated cathode. The layer is formed by a chemical reaction between the barium oxide and impurities in the nickel, such as silicon, magnesium, manganese, aluminum, tungsten, etc. Silicon impurities react to form barium-orthosilicate, which is considered by many workers in the cathode field to be responsible for the high resistance type of layer. The growth of the layer is influenced by the temperature of operation of the cathode and the conditions of operation of the tube. The experimental design set up to test the effects of these various factors and the influence of different manufacturing processes are shown in Figure 1. As can be seen, a complete factorial design was used. Four types of nickel alloy were selected for the test, and a quantity of each alloy selected from a particular melt was sent to a single cathode manufacturer to be formed into cathode sleeves. The finished cathode sleeves were then divided among three tube manufacturers who used them in the construction of a common tube type selected for the test. Each alloy-manufacturer lot was tested at three levels of filament voltage corresponding approximately to three levels of cathode temperature and three levels of plate current. The sample sizes or number of replications of the experiment was chosen based on the equation

$$\sqrt{n} = (\mu_{1-\beta} + \mu_{1-\frac{\alpha}{2}}) \frac{\sigma}{\delta}$$

$$n = (1.645 + 1.960)\frac{\sigma}{\delta}$$

for $\beta = .05$

and $\alpha = .05$

where $\sigma = \sqrt{2}\,\sigma_e$

and σ_e^2 = residual variance of a homogeneous group

δ = desired minimum resolution between two groups

---

1. Work carried out for Signal Corps by Briggs Associates, Inc. under contract DA36-039 sc-72336.

2. Problems in Analysis of Electron Tube Experiments - M. H. Zinn.

A ratio of 1.5 between means on an arithmetic base was arbitrarily selected as the minimum resolution desired yielding a value of 0.176 for δ on a logarithmic base. An estimated value of 0.358, based on a limited amount of data, was used for the standard deviation of log interface of a homogeneous group. This result in a minimum sample size of 107 tubes required to detect a significant differenc between two homogeneous groups. Since this sample size, if it were assigned to each individual cell in the factorial experiment, would lead to a huge number of tubes to be tested over a period of 5000 hours each, a compromise was reached by using this number as the approximate size of an alloy-manufacturer group. Actually 13 tubes were tested under the 9 conditions of operation for a total of 117 tubes for each alloy-manufacturer group. This resulted in a total of 156 tubes under each of the 9 life test conditions.

Life tests were initiated in accordance with this experimental design with the modification that two burning runs were made. The first burning run was made with 7 tubes placed on life under each cell condition for 5000 hours. At the end of this period 6 tubes were placed on life for another 5000-hour period to constitute the second run.

While the life hours were being accumulated, the opportunity was provided to study the problems involved in the analysis of the data. Three problems were considered to be of great importance; one problem is peculiar to this experiment, while the other two problems are common to most electron tube experiments involving life tests.

The first problem resulted from the choice of a twin triode as the test vehicle. (A twin triode consists of two triode sections each with individual cathodes in the same envelope.) If the two sections of each tube behaved as independent samples from the same tube population, the effective size of the sample would be doubled. A doubling of the sample size would tend to minimize the effects of higher order interactions than the two-way interactions for which we had made provisions to detect, if present, in the experimental design. It was possible, however, that the close environmental conditions of two triodes in the same envelope would cause a mirroring effect, which would result in such small differences between samples that one would not be justified in using the sections as individual replications. In addition, a third possibility existed that there would be some bias existing between the sections because of methods of processing or due to fixed filament voltage differences that would require the sections to be treated as two separate groups. This would add a new factor to the experiment, which could result in additional interactions to weaken the power of the analysis. The solution to this problem is discussed below.

A second problem considered prior to the collection of the complete data was methods of overcoming variance and drift in the true levels of real factors that could not be adequately controlled. In this experiment this problem was due to the inability to directly control the cathode temperature, which represents the real variable affecting the growth of interface rather than the filament voltage, which could be controlled. The effects of this lack of control would mean a wider spread in residual variance than would otherwise be present. No solution was found for this problem during the course of the experiment, and this contribution to the residual error had to be accepted.

The third problem was the method of treatment of readings at various periods during life. Should these be treated as another factor in the analysis of variance or should other methods of treatment be utilized? This point will also be touched on later in the discussion.

In performing the actual analysis of the data, a standard analysis of variance was performed. The problem of the treatment of the sections was resolved by considering them as another factor in the analysis, as advised by Professor Hartley at the Third Conference on Design of Experiments, with the added feature that a two-sided test of the F ratio was to be made for the sections rather than the usual one-sided test for significantly large differences in variance. The inclusion of the section factor in the analysis of variance and a run factor due to the fact the cell lots were divided approximately in two resulted in the following overall analysis requirements:

(R S A M I E) Replications

where R = Runs

S = Sections

A = Alloys

M = Manufacturers

I = Plate Current Operating Conditions

E = Filament Voltage Operating Conditions

or $(2 \times 2 \times 4 \times 3 \times 3 \times 3)7 = 3024$

readings to be analyzed for each of 10 reading periods taken during the 5000-hour test.

A search was instituted to find a machine program that could handle this number of factors. It was determined that, even though each alloy group was analyzed separately, which appeared to be desirable based on initial analyses showing large differences between alloys, and each run was handled separately with the run analysis done by manual methods at a later time, the cost of programming the remaining SMIE four-factor analysis was higher than our budget could handle. Further study of the problem by Mr. R. Dickson, the statistician on the program, indicated that the analysis could be carried out completely on a manual basis using statistical clerks to perform the calculations, provided that the manual program was organized properly. If the program was handled on this basis, it would be possible to remain within the costs budgeted for the analysis and obtain the desired results. The manual analysis would also make it possible to perform additional graphical treatment of the data, since all of the subtotals would be automatically available, compared to having to pay for these subtotals in programming and machine time, if automatic machine calculations were used. The lower cost of the manual program is made possible only by being able to stop the calculation process at appropriate points and make a decision based on preliminary plotting of average values that there is nothing to be gained by continuing the calculation. Thus, all of the early-life readings, where a simple plot of the data could show that there were no significant differences, did not need to be put through a complete analysis of variance.

The organization of the manual program was based on the use of a three-way table such as is illustrated in the Appendix. A step-by-step procedure for using this table is also included. As can be seen, the table covers a three-factor analysis of Section, Plate Current, and Filament Voltage (SIE) effects for a single manufacturer and a single alloy and covers the first run of seven tubes out of the 13-tube sample per cell. Similar tables were used to enter the combined calculations for the two runs and then to combine the results of tests for the three manufacturers. No attempt was made to combine the results of different alloys since, as previously mentioned, the differences between alloys were large.

It should be noted that the data shown in the table included in the Appendix represent early calculations that were performed when the organization of the data was being worked out. At that time the data were carried in the averages to only three decimal places, which resulted in negative values being calculated for some of the variance estimates. This was corrected in the later analyses where the sums and averages were carried out to five decimal places, thus eliminating the calculation error resulting in fictitious negative values of variance. The method is open to criticism in that use is made of averages at early stages of the calculation rather than in carrying partial sums. The use of the procedure can be justified on the basis that the numbers carried through the procedure are relatively simple, with an ordered level of magnitude. This permits a relatively untrained calculator to spot his own calculation errors or transpositions of entries as he goes along, and it simplifies the checking procedures. It also simplifies the treatment of missing entries caused by failure of a tube for reasons other than interface resistance prior to a reading period. These failures were few enough in number to permit use of the section-burning-condition average value for the missing tube, thus allowing a constant sample size to be used for variance estimates. The advantages more than offset the error introduced by taking premature averages. It is possible, however, to use the same basic organization and carry total sums, if one so desires.

The data resulting from the analysis of variance will be presented in a final report on the program. It will be in tabular and graphical form. The methods used for presentation of these results, rather than the results themselves, should be of interest to this audience. The tabular summary will contain the calculated variances for each of the main effects and interactions and the residual variance due to error or uncontrolled effects. The average residual variance for the overall experiment was determined to be 0.085 on a logarithmic base. The measured value of σ_e was, therefore, 0.292 compared to the estimated value of 0.358 used to calculate the sample sizes. The results indicate that the planned statistical power of the experiment to detect differences between alloy-manufacturer samples equivalent to an arithmetic ratio of 1.5 was obtained. Significant differences between smaller samples representing the individual cell groups were detected because the differences that arose due to the test conditions were in many cases greater than the minimum detectable limit selected.

In addition to the tabular summary, the data have been presented in various graphical forms. Figure II is illustrative of one of the methods of presenting the life data. It is a plot of the results of the four alloys for one manufacturer. The final presentation will include an individual plot for each manufacturer and alloy of the average curve similar to each of these curves with one sigma limits for the grand average of all burning conditions for the sample size used and one sigma limits for the average of burning conditions. Note the rather large differences in alloys that were experienced; these resulted in the alloy analyses being handled separately.

Figure III also shows life data. In this figure curves are shown for three of the four alloys for all three manufacturers. The curves for alloy 220 have been omitted to eliminate confusion since they would fall close to the 330 alloy curves. Note the large alloy-manufacturer interaction that is apparent even without resorting to analysis of variance. The P-50 alloy shows no significant differences between manufacturers while the A-32 shows a pronounced difference all through life. The 330 alloy, however, only shows a significant difference at the later stages of the 5000-hour tests.

Figure IV illustrates the method used to graphically show the effects of the burning conditions. These plots have been made for individual alloys and for all manufacturers, when no significant manufacturer effects were present, or for each manufacturer when required. The crosses represent the burning conditions of filament voltage and plate current. The figure in parentheses represents the average value expressed in arithmetic values of interface resistance. Contour lines have been drawn onto this matrix showing the placement of a contour corresponding to the average of all of the burning conditions and contours corresponding to one sigma limits due to the residual error of the test. A contour plot, such as represented by this figure, for the P-50 alloy and all manufacturers, showing a very-flat topographical structure, is indicative of no significant effects due to the various levels of burning conditions.

Figure V, covering alloy 220 and manufacturer 1, shows a little more topographical structure. The presence of more contour lines between burning points indicates that some significant effects are present. If all of the contour lines were straight lines essentially parallel to each other, this would mean that a main effect was present. With the curvature as shown, it is indicative that a voltage-current interaction is present, raising a slight but significant hill at the 5.7-volt, 0.9-milliampere condition.

Figure VI illustrates the interaction effect even more graphically, showing data for alloy A-32, manufacturer 1. The number of contour lines has increased considerably, showing highly significant differences existing, with a severe interaction effect as shown by the large curvature, plus a significant plate current effect. The presentation of contour data for different periods of life for the same alloy would show a general lifting of the elevation levels at all points with a hill beginning to appear near middle life and a shift in the apex of the hill as life progresses. The contour maps thus represent a rather graphic moving-picture of the life history of the effect of burning conditions on the growth of interface resistance.

The experiment revealed many effects that had not previously been suspected, one of which might be of interest to those of you involved in the operation of electron tube computers. This effect showed that the cut-off condition, zero plate current, is not necessarily the worst condition of operation of electron tubes as far as interface resistance growth is concerned. The cure for the so-called "sleeping sickness" of tubes in computers operating for long periods of time at cut-off, by maintaining a low current drain, is not necessarily the best action to take since it has been demonstrated that extremely high values of interface can be formed at the low current drain compared to high current or zero current for some alloys. A surer cure of the problem is to use tubes with passive alloys equivalent to the P-50 alloy tested in this experiment and obtain relatively low values of interface resistance over the life of the tube and freedom from the effects of operating conditions.

The successful conclusion of this experiment is due in large measure to the use of the statistical approach to the Design of Experiments. The results obtained are conclusive in the areas covered by the experimental design. Those points that are missing, such as the effects of sampling within a given manufacturer's production over a period of time, can now be obtained with a fairly simple experimental design. The contour data obtained for the burning effects need checking both as to the reproducibility of the data for the operating conditions used for these tests and the interpolation of the data between points. A partial factorial design covering an experiment to obtain this additional data is presently being examined. While the debt owed to the field of statistics is great, it is hoped that at least a partial repayment has been made to this field through the techniques of analysis evolved during the program.

ACKNOWLEDGMENTS

The author wishes to thank the following personnel of Briggs Associates, Inc.: R. Dickson, who developed the statistical techniques discussed in this paper, and T. A. Briggs for making available the figures used.

Fig I DESIGN OF LIFE OPERATING CONDITIONS

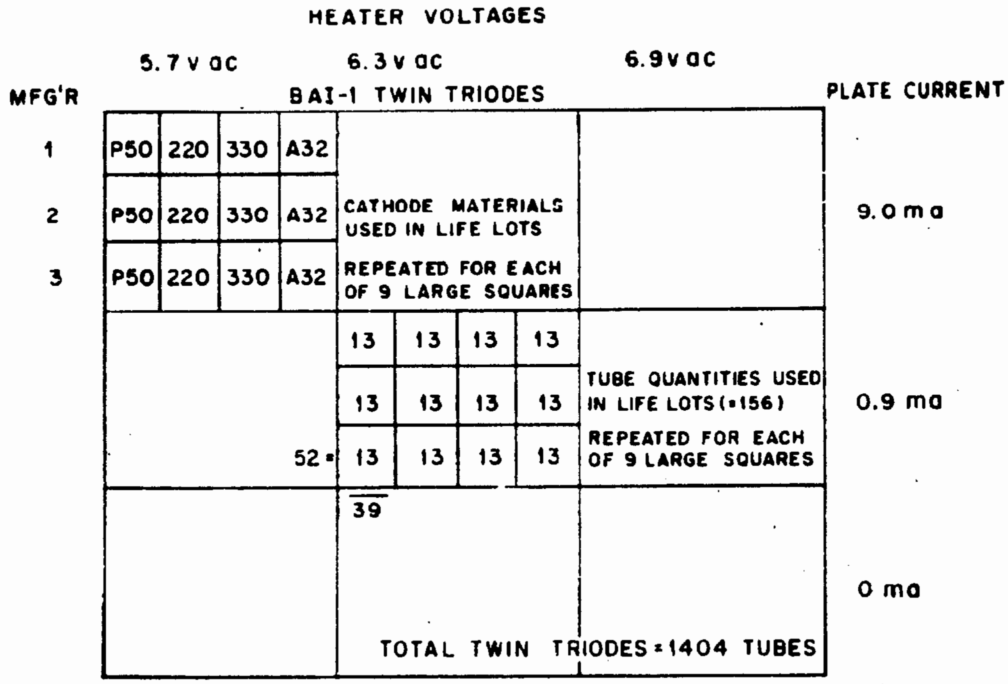


Fig. II - INTERFACE RESISTANCE AS A FUNCTION OF CATHODE ALLOY

8AI-1 TRIODES MFGR 1 ALL LIFE CONDITIONS

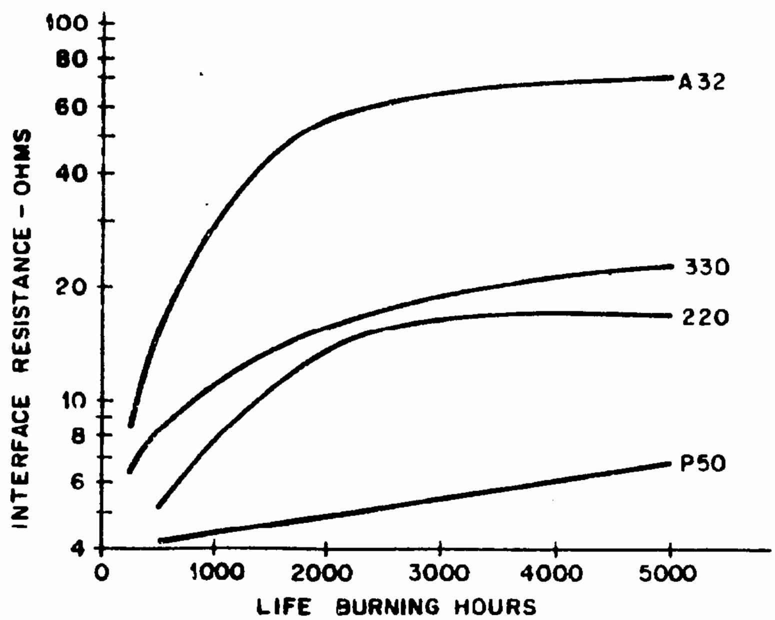


Fig. III - INTERFACE RESISTANCE AS A FUNCTION OF ALLOY AND MANUFACTURER

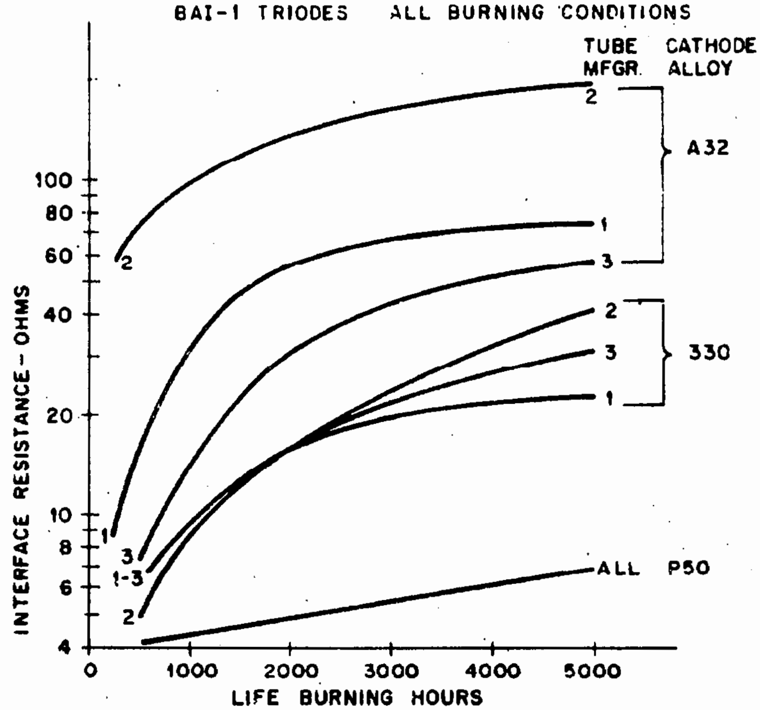
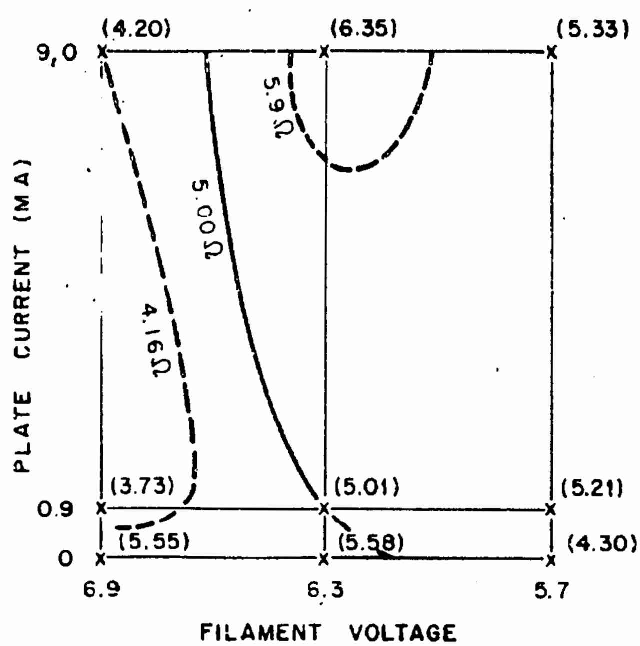


Fig. IV - INTERFACE READING CONTOURS - TRIODES
ALLOY P-50 ALL MANUFACTURERS
2000 LIFE HOURS

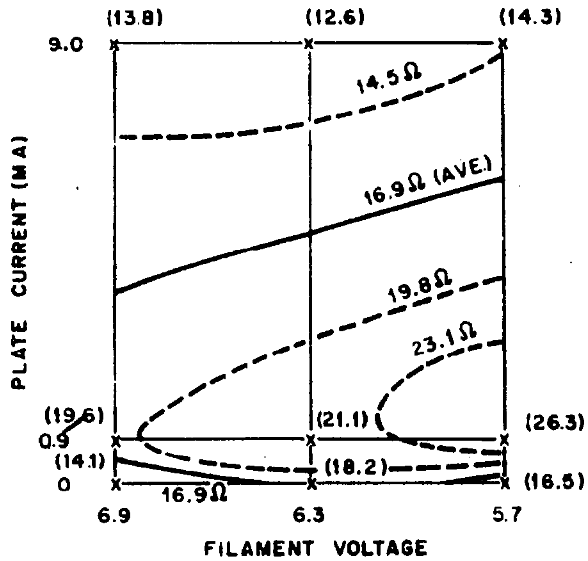


$$\sigma \text{ LOG} = \frac{0.616}{\sqrt{72}} = \frac{0.616}{8.49} = 0.0726 \text{ (18\% READING SHIFT)}$$

Fig. V - INTERFACE READING CONTOURS-TRIODES

ALLOY 220 MANUFACTURER 1

5000 LIFE HOURS

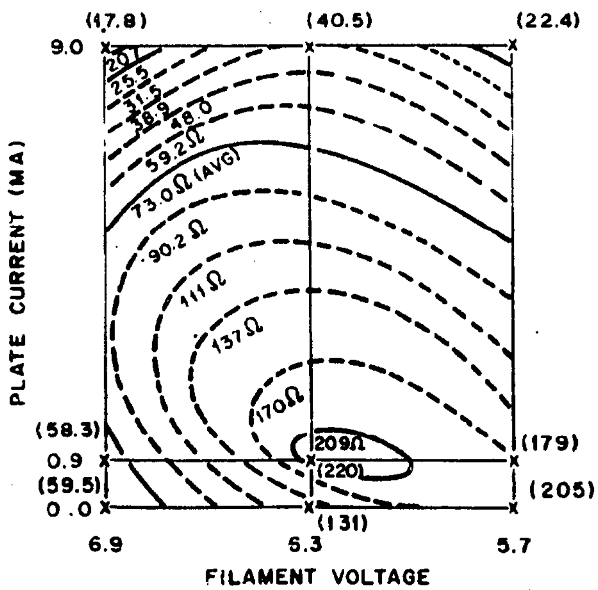


$$\sigma \text{ LOG} = \frac{.331}{\sqrt{24}} = \frac{.331}{4.9} = 0.0678 \text{ (17\% READING SHIFT)}$$

Fig. VI - INTERFACE READING CONTOURS-TRIODES

ALLOY A-32 MANUFACTURER 1

5000 LIFE HOURS



$$\sigma \text{LOG} = \frac{0.447}{\sqrt{24}} = \frac{0.447}{4.9} = 0.0913 \text{ (23\% READING SHIFT)}$$

APPENDIX

USE OF DICKSON SINGLE-TABLE METHOD FOR THREE-FACTOR ANALYSIS

Steps	Illustration from Sample Chart
1. Enter data for individual readings (1)	Enter 126 individual readings from 7 tubes X 2 Sections X 3 voltage levels X 3 current levels.
2. Calculate sum and sum of squares for individual readings in first box of chart covering replications of experiment at single level of each of three factors. Enter sum of squares in row labeled Σ^2 (2a) and average of sum in row labeled Ave (2b). Repeat for each 3-way level.	Sum of squares obtained for individual readings for Section (1) at $E_f = 6.9$ volts and $I_b = 9.0$ mA. The sum of these same readings is averaged Σ^2 -- $(1.176)^2+(1.230)^2+(1.146)^2+(1.672)^2+(0.954)^2+(1.079)^2+(1.204)^2 = 10.529$ Ave -- $\frac{1.176+1.230+1.146+1.672+0.954+1.079+1.204}{7} = 1.209$ The value 10.529 represents a partial sum of squares of individual readings, ΣX^2. The value 1.209 is $\bar{X}_{S_1 I_1 E_1} = \frac{\Sigma X_{S_1 I_1 E_1}}{7}$ Process is repeated 17 times to obtain total of 18 values of partial sum of squares and 18 values of $\bar{X}_{S_j I_k E_L}$
3. Calculate sum and sum of squares for average values obtained in row labeled Σ^2 under column labeled Ave (3a). Enter average of sum in row labeled Ave (3b). Repeat for each level of two remaining factors.	Sum of squares of the two averages over Sections (1) and (2) for $E_f = 6.9$ volts and $I_b = 9.0$ mA is obtained. The average of these two averages is also calculated. Σ^2 -- $(1.209)^2+(1.075)^2 = 2.617$ Ave -- $\frac{1.209+1.075}{2} = 1.142$ The value 2.617 represents a partial sum of squares of $\bar{X}_{SIE}$ values or a partial sum of (X_{SIE}) The value 1.142 is $\bar{X}_{I_1 E_1} = \frac{\Sigma X_{I_1 E_1}}{7x2}$ Process is repeated 8 times to obtain total of 9 values of partial sums of square and 9 values of $X_{I_k E_L}$

(All entries in the body of table have been made. The next step represents the first of the peripheral calculations.)

4. Calculate sum and sum of squares for average values single level of first factor and second factor across levels of third factor. Enter sum of squares in row labeled Σ^2 and column (1) under Σ^2 on righthand periphery of chart (4a). Enter average of sum of averages in row labeled Ave and column (1) under Σ^2 on righthand periphery of chart (4b). Repeat for each level of second factor.	Sum of squares of the averages for Section (1) at I_b = 9.0 mA for three levels of E_f is obtained. The average of these three averages is also calculated Σ^2 -- $(1.209)^2 + (1.206)^2 + (1.020) = 3.957$ Ave -- $\frac{1.209+1.206+1.020}{3} = 1.145$ The value 3.957 represents a partial sum of squares of $\bar{X}_{SIE}$ values or a partial sum of $\frac{(X_{SIE})^2}{7}$ These data are a duplication of the partial sums of squares calculated in Step 2 and can be used as a computation check of the total sum of $(X_{SIE})^2$. The value 1.145 is the average of a partial sum of X_{SI}. Process is repeated 2 times to obtain a total of 3 values of partial sum of squares and 3 averages of partial sums of $X_{S_j I_k}$
5. Repeat Step 4 for all levels of first factor entering data in appropriate column of first factor in righthand peripheral area. (5a and 5b)	This step obtains additional partial sums of squares of X_{SIE} values and the remaining average of the partial sum of $X_{S_j I_k}$ values

6. Repeat Step 4 for single level of first factor and third factor across levels of second factor. Enter sum of squares in row in bottom peripheral area labeled Σ^2(6a). Enter average of sum of averages in row labeled Ave in bottom peripheral area (6b).	Sum of squares of the averages for Section (1) at E_f = 6.9 volts for three levels of I_b is obtained. The average of these three averages is also calculated. Σ^2 -- $(1.209)^2+(1.291)^2+(1.184)^2$ = 4.530 Ave -- $\frac{1.209+1.291+1.184}{3}$ = 1.228 The value 4.530 represents a partial sum of squares of $\bar{X}_{SE}$ values or a partial sum of $\frac{(X_{SIE})^2}{(7)^2}$ The value of 1.228 is the average of a partial sum of X_{E_1} Process is repeated 2 times to obtain a total of 3 values of partial sums of squares and 3 averages of partial sums of $X_{S_jE_L}$
7. Repeat Step 6 for all levels of first factor, entering data in appropriate column of first factor in bottom peripheral area. (7a and 7b)	This step obtains additional values of partial sums of squares $\bar{X}_{SIE}$ values and the remaining average of the partial sum of $X_{S_jE_L}$ values
8. Calculate sum and sum of squares of averages found in Step 3 across first level of the second factor. Enter sum of squares in row labeled Σ^2 under column labeled Ave in right-hand peripheral area (8a).	Sum of squares for the $\bar{X}_{IE}$ values across I_b = 9.0 mA is obtained. The average of these same values is also calculated. Σ^2 -- $(1.142)^2+(1.192)^2+(1.190)^2$ = 4.141 Ave -- $\frac{1.142+1.192+1.190}{3}$ = 1.175 Note that (4b) and (5b) values can be used to obtain the same average $\frac{1.145+1.204}{2}$ = 1.175 The value 4.141 represents a partial sum of squares of $\bar{X}_{IE}$ values or a partial sum of $\frac{(X_{IE})^2}{(7X2)^2}$

	Enter average of the sum of averages in the row labeled Ave under column labeled Ave in the righthand peripheral area (8b). Repeat across each level of the second factor.	The value of 1.175 is the average of the total sum of X_I Process is repeated 2 times to obtain a total of 3 values of partial sums of squares and 3 averages of the total sum of X_{I_k}
9.	Calculate sum and sum of squares of averages found in Step 3 across first level of third factor. Enter sum of squares in row labeled Σ^2 in bottom peripheral area (9a). Enter average of the sum of averages in the row labeled Ave in bottom peripheral area (9b). Repeat across each level of the third factor.	Sum of squares for the $\bar{X}_{IE}$ values across E_f = 6.9 volts is obtained. The average of these same values is also calculated Σ^2 -- $(1.142)^2+(1.302)^2+(1.204)^2 = 4.366$ Ave -- $\frac{1.142+1.302+1.204}{3} = 1.204$ Note that (6b) and (7b) values can be used to obtain the same average $\frac{1.228+1.180}{2} = 1.204$ The value 4.366 represents a partial sum of squares of $\bar{X}_{IE}$ values or a partial sum of $\frac{(X_{IE})^2}{(7X2)^2}$ The value of 1.204 is the average of the total sum of X_{E_1} The process is repeated 2 times to obtain a total of 3 partial sums of squares and 3 averages of the total sum of X_{E_L} The partial sums of squares are essentially a duplication of data obtained in Step 8 and are used as a computation check of the total sum of $(\bar{X}_{IE})^2$

10.	Calculate the sum of squares of the averages found in Steps 4 and 5 for the first level of the second factor. Enter in box labeled (10) on Step Procedure chart. Repeat for each level of the second factor.	Sum of squares for the $\bar{X}_{SI}$ values for I_b = 9.0 mA is obtained $(1.145)^2+(1.204)^2 = 2.761$ The value of 2.761 represents a partial sum of squares of $\bar{X}_{SI}$ values or a partial sum of $\frac{(X_{SI})^2}{(7X3)^2}$ The process is repeated 2 times to obtain a total of 3 partial sums of squares
11.	Calculate the sum of squares of the averages found in Steps 6 and 7 for the first level of the third factor. Enter in box labeled (11) on Step Procedure chart. Repeat for each level of the third factor.	Sum of squares for the $\bar{X}_{SE}$ values for E_f = 6.9 volts is obtained $(1.228)^2+(1.180)^2 = 2.900$ The value 2.900 represents a partial sum of squares of $\bar{X}_{SE}$ or a partial sum of $\frac{(X_{SE})^2}{(7X3)^2}$ The process is repeated 2 times to obtain a total of 3 partial sums of squares
12.	Sum the values of partial sums of squares found in Step 2 (6 values) across the first level of the second factor. Enter in box labeled (12) on Step Procedure chart. Repeat for each level of the second factor.	Sum of partial sum of X^2 values is obtained for I_b = 9.0 mA 10.529+8.303+10.252+9.812+7.067+13.087 = 59.590 The value of 59.590 is a further summing of the partial sum of X^2 values. The process is repeated 2 times to obtain a total of 3 partial sums of squares.

13. Sum the values of partial sums of squares found in Step 2 (6 values) across the first level of the third factor. Enter in box labeled (13) on Step Procedure chart. Repeat process for each level of the third factor.	Sum of the partial sum of X^2 values is obtained for E_f = 6.9 volts 10.529+8.303+11.689+12.115+9.837+9.466 = 61.939 The value 61.939 is a further summing of the partial sum of X^2 values. The process is repeated for 2 times to obtain a total of 3 partial sums of squares. The partial sums of squares are a duplication of data obtained in Step 12 and are used as a computation check of the sum of X^2
14. Sum the values obtained in Step 12 or Step 13. Enter in box labeled (14) on Step Procedure chart.	The total sum of X^2 values is obtained 59.590+77.091+71.655 = 208.336 or 61.939+69.966+76.431 = 208.336 The value 208.336 is the total sum of X^2
15. Calculate the sum and sum of squares of the averages found in Step 4. Enter the sum of squares in the box labeled (15a). Enter the average of the sum of averages in the box labeled (15b). Repeat the process for the averages found in Step 5.	The sum of squares of the $\bar{X}_{SI}$ values is obtained. The average of these same values is also calculated. $(1.145)^2+(1.312)^2+(1.219)^2 = 4.518$ $\frac{1.145+1.312+1.219}{3} = 1.225$ The value 4.518 represents a partial sum of squares of $\bar{X}_{SI}$ values or a partial sum of $\frac{(X_{SI})^2}{(7X3)^2}$ The value 1.225 represents the average of the total sum of X_{SI} The process is repeated to find a total of 2 values of partial sum of squares and 2 averages of the total sum of X_{Sj}

16.	Repeat Step 15 for the averages found in Step 6. Enter the sum of squares in the box labeled (16a). Enter the average of the sum of averages in the box labeled (16b). Repeat the process for the averages found in Step 7.	The sum of squares of the $\overline{X}_{SE}$ values is obtained. The average of these same values is also calculated $(1.228)^2+(1.235)^2+(1.212)^2 = 4.502$ $\frac{1.228+1.235+1.212}{3} = 1.225$ The value 4.502 represents a partial sum of $(\overline{X}_{SE})^2$ values or a partial sum of $\frac{(X_{SE})^2}{(7X3)^2}$ The value 1.225 represents the average of the total sum of X_{SI}. As a check on the computation process, it should be equal to the average for the sum of X_{SI} found in Step 15. The process is repeated to find a total of 2 values of the partial sum of squares and 2 averages of the total sum of X_{S_j}.
17.	Calculate the sum and sum of squares of the average values found in Step 8. Enter the sum of squares in the box labeled (17a). Enter the average of the sum of averages in the box labeled (17b).	The sum of squares of the $\overline{X}_I$ values is obtained. The average of these same values is also calculated $(1.175)^2+(1.336)^2+(1.272)^2 = 4.784$ $\frac{1.175+1.336+1.272}{3} = 1.261$ The value 4.784 represents the total sum of squares of the $\overline{X}_I$ values or the total sum of $\frac{(X_I)^2}{(7X3X2)^2}$ The value 1.261 represents the grand average.
18.	Repeat Step 17 for the average values found in Step 9. Enter the sum of squares in the box labeled (18a). Enter the average in the box labeled (18b).	The sum of squares of the $\overline{X}_E$ values is obtained. The average of these same values is also calculated $(1.204)^2+ (1.270)^2+(1.308)^2 = 4.773$ $\frac{1.204+1.270+1.308}{3} = 1.261$ The value 4.773 represents the total sum of squares of the $\overline{X}_E$ values or the total sum of $\frac{(X_E)^2}{(7X3X2)^2}$

	The value 1.261 represents the grand average and should check the value found in Step 17.
19. Calculate the sum and sum of squares of the average values found in Steps 15 or 16. Enter the sum of squares in the boxes labeled 19. The average of the sum of averages is not entered but should check the value entered in boxes 17b and 18b.	The sum of squares of the $\bar{X}_S$ values is obtained. The average of these same values is also calculated $(1.225)^2+(1.296)^2 = 3.180$ $\frac{1.225+1.296}{2} = 1.261$ The value 3.180 represents the total sum of squares of the $\bar{X}_S$ or the total sum of $\frac{(X_S)^2}{(7X3X3)^2}$ The value 1.261 represents the grand average and should check the value found in Steps 17 and 18.
20. Sum the values of the sum of squares found in Steps 4 and 5. Enter in the box labeled (20) on the Step Procedure chart.	The sum of squares of the $\bar{X}_{SIE}$ values is obtained. $3.957+4.393+5.162+5.550+4.461+5.305 = 28.829$ The value 28.829 represents the total sum of squares of $\bar{X}_{SIE}$ values of the sum of $\frac{(X_{SIE})}{(7)^2}$
21. Sum the values of the sum of squares found in Steps 6 and 7. Enter in the box labeled (21) on the Step Procedure chart.	The sum of squares of the $\bar{X}_{SIE}$ values is obtained as a computation check $4.530+4.209+4.586+5.131+4.464+5.909 = 28.829$ This value should check the value found in Step 20. The sum of the 9 values of sum of squares found in Step 3 should also check this value.

22. Sum the values of the sum of squares found in Step 8. Enter in the box labeled (22) on the Step Procedure Chart.	The sum of squares of the $\bar{X}_{IE}$ values is obtained. 4.141+5.357+4.868 = 14.366 The value 14.366 represents the total sum of squares of the X_{IE} values or the sum of $\frac{(X_{IE})^2}{(7X2)^2}$
23. Sum the values of the sum of squares found in Step 9. Enter in the box labeled (23) on the Step Procedure chart.	The sum of squares of the $\bar{X}_{IE}$ values is obtained. 4.366+4.849+5.151 = 14.366 This value should check the value found in Step 22
24. Sum the values of the sum of squares found in Step 10. Enter in the box labeled (24) on the Step Procedure chart.	The sum of squares of the $\bar{X}_{SI}$ values is obtained 2.761+3.571+3.239 = 9.570 The value 9.570 represents the total sum of squares of the $\bar{X}_{SI}$ values or the sum of $\frac{(X_{SI})^2}{(7X3X3)^2}$ This value should check the sum of the 2 values obtained for sum of squares in Step 15.
25. Sum the values of the sum of squares found in Step 11. Enter in the box labeled (25) on the Step Procedure chart.	The sum of squares of the $\bar{X}_{SE}$ values is obtained 2.900+3.228+3.437 = 9.566 The value 9.566 represents the total sum of squares of the $\bar{X}_{SE}$ values or $\frac{(X_{SE})^2}{(7X3X3)^2}$ This value should check the sum of the 2 values obtained for sum of squares in Step 16.

SUMMATION OF DATA

The calculated values are now entered in a Summary Table which is normally present at the bottom of the Three-Factor Analysis Table. Only one operation on the calculated values is required in the transfer of data to the Summary Table, i.e., find the square of the grand average in box 17b or 18b for entry in the row labeled Correction Factor in the Summary Table. The steps involved in the completion of the Summary Table are enumerated below:

Step 1. Enter the appropriate values of sums of squares from the Three-Factor Analysis Table in the first column.

Step 2. Enter the number of sections (or number of readings) involved in the individual terms of the summation of sums of squares. Thus, the residual sum of squares consists of the sum of squares of individual readings and, therefore, the number of sections involved is one. For the three-factor interaction term, the square of the sum over 7 sections is involved, so this number is entered in the second row. The two-factor interaction terms involve the square of the sum over the number of replications times the number of levels of the third level, i.e., for SI interaction terms 7 X 3 (7 replications X 3 levels of E). Likewise, the main effect terms calculated from the number of replications and the number of levels of the two remaining factors, i.e., for the S effect 7 X 3 X 3 represents the number of sections involved (7 replications X 3 levels of E X 3 levels of I). Finally, the correction factor involves the square of the sum over the total number of readings or 7 X 3 X 3 X 2.

Step 3. Multiply the value of the sum of squares listed in Column 1 by the number of sections listed in Column 2. This step is required to adjust for the method of calculation in terms of averages rather than the conventional method of direct summation. Thus, the sum of squares for the three-factor interaction term was found from

$$(\bar{X}_{S_1I_1E_1})^2+(\bar{X}_{S_2I_1E_1})^2+ \ldots\ldots (\bar{X}_{S_jI_kE_L})^2 = \Sigma(\bar{X}_{SIE})^2$$

The term actually desired for the analysis of variance is

$$(\Sigma X_{S_1I_1E_1})^2+(\Sigma X_{S_2I_1E_1})^2+ \ldots\ldots (\Sigma X_{S_jI_kE_L})^2 = \frac{\Sigma(\Sigma X_{SIE})^2}{n}$$

where n = the number of readings involved in each summation

$$\text{Since } \bar{X}_{S_1I_1E_1} = \frac{\Sigma X_{S_1I_1E_1}}{n}$$

$$\Sigma(\bar{X}_{SIE})^2 = \frac{\Sigma(\Sigma X_{SIE})^2}{n} = \frac{\Sigma(\Sigma X_{SIE})^2}{n^2}$$

or

$$\frac{\Sigma(\Sigma X_{SIE})^2}{n} = n\Sigma(\bar{X}_{SIE})^2$$

Step 4. The final values of the sum of squares are found from the following equation

Final Sum of Squares for S = Adjusted Sum of Squares for S - Correction Factor or

$$\Sigma^2 S_F = \Sigma^2 S_A - C.F.$$

$$\Sigma^2 I_F = \Sigma^2 I_A - C.F.$$

$$\Sigma^2 E_F = \Sigma^2 E_A - C.F.$$

$$\Sigma^2 IE_F = \Sigma^2 IE_A - C.F. - \Sigma^2 E_F - \Sigma^2 I_F$$

$$\Sigma^2 SE_F = \Sigma^2 SE_A - C.F. - \Sigma^2 S_F - \Sigma^2 E_F$$

$$\Sigma^2 SI_F = \Sigma^2 SI_A - C.F. - \Sigma^2 S_F - \Sigma^2 I_F$$

$$\Sigma^2 SIE_F = \Sigma^2 SIE_A - C.F. - \Sigma^2 SI_F - \Sigma^2 SE_F - \Sigma^2 IE_F - \Sigma^2 E_F - \Sigma^2 S_F$$

$$Res_F = Res_A - \Sigma^2 SIE_F$$

The final values of sum of squares are equivalent to the values normally tabulated in an Analysis of Variance Table and the remainder of the table is conventional.

	Sum of Squares*	Number of Sections	Adjusted Sum of Squares	Final Sum of Squares	of Freedom	Variance	F	Critical F
Residual	208.336 (14)	1	208.336	6.533	108	0.065	-	-
SIE	28.829 (20 or (21)	7	201.803	0.343	4	0.086	1.42	3.50
SI	9.570 (24)	21	200.970	***	2	0.0	0.0	4.81
SE	9.566 (25)	21	200.886	0.294	2	0.147	2.43	4.81
IE	14.366 (22) or (23)	14	201.124	***	4	0.0	0.0	3.50
E	4.773 (18a)	42	200.466	0.252	2	0.126	2.08	4.81
I	4.784 (17a)	42	200.928	0.714	2	0.357	5.90†	4.81
S	3.180 (19)	63	200.340	0.126	1	0.126	2.08	6.88
Correction Factor	1.589 **	126	200.214					

* Number in parentheses indicates box location on the Step Procedure Chart.

** Square of 1.261 in boxes (17b) or (18b) of Step Procedure Chart.

*** Negative values of variance obtained due to premature rounding off of decimal places. Variance assumed to be negligibly different from zero.

† Significant at the 1.0% level of confidence.

SIGM/EL-ADP 629

SUMMARY TABLE

	Tube No.	Ef = 6.9V (1)	Ef = 6.9V (2)	Ef = 6.9V Ave	Ef = 6.3V (1)	Ef = 6.3V (2)	Ef = 6.3V Ave	Ef = 5.7V (1)	Ef = 5.7V (2)	Ef = 5.7V Ave	Σ^2 (1)	Σ^2 (2)	Σ^2 Ave	
I_b = 9.0	1	1.176	1.279											
	2	1.230	1.000											
	3	1.146	1.146											
	4	1.672	1.176											
	5	0.954	0.699											
	6	1.079	1.114											
	7	1.204	1.114											
	Σ^2	10.529	8.303	2.617	10.252	9.812	2.842	7.607	13.087	2.890	3.957	4.393	4.141	59.590 6(Σ^2)
	Ave	1.209	1.075	1.142	1.206	1.178	1.192	1.020	1.360	1.190	1.145	1.204	1.175	
													2.761	
I_b = 0.9	1													
	2													
	3													
	4													
	5													
	6													
	7													
	Σ^2	11.689	12.115	3.388	12.549	13.594	3.577	12.892	14.252	3.748	5.162	5.550	5.357	77.091 6(Σ^2)
	Ave	1.291	1.312	1.302	1.308	1.366	1.337	1.336	1.401	1.369	1.312	1.360	1.336	
													3.571	
I_b = 0	1													
	2													
	3													
	4													
	5													
	6													
	7													
	Σ^2	9.837	9.466	2.734	10.135	13.624	3.298	12.800	15.793	3.735	4.461	5.305	4.868	71.655 6(Σ^2)
	Ave	1.184	1.154	1.169	1.192	1.370	1.281	1.280	1.448	1.364	1.219	1.324	1.272	
													3.239	
Σ^2		4.530	4.209	4.366	4.586	5.131	4.849	4.464	5.909	5.151				
Ave		1.228	1.180	1.204	1.235	1.305	1.270	1.212	1.403	1.308				
				2.900			3.228			3.437				
			61.939 6(Σ^2)			69.966 6(Σ^2)			76.431 6(Σ^2)					

4.502	5.064	4.773	28.829	14.366
1.225	1.296	1.261		
		3.180		9.566

4.518	5.052	4.784	208.336
1.225	1.296	1.261	
		3.180	
	28.829	14.366	
		9.570	

SAMPLE DICKSON THREE-FACTOR ANALYSIS TABLE

NOTE: INDIVIDUAL SECTION READINGS HAVE BEEN OMITTED, EXCEPT FOR FIRST BOX, FOR THE PURPOSE OF THIS PRESENTATION.

	TUBE NO.	Ef=6.9V S(1)	Ef=6.9V S(2)	Ef=6.9V AVE	Ef=6.3V S(1)	Ef=6.3V S(2)	Ef=6.3V AVE	Ef=5.7V S(1)	Ef=5.7V S(2)	Ef=5.7V AVE	Σ² S(1)	Σ² S(2)	Σ² AVE	
I_b= 9.0	1	|	|											
	2	|	|											
	3	|	|											
	4	|	|											
	5	|	|											
	6	|	|											
	7	|	|											
	Σ^2	2a	2a	3a	2a	2a	3a	2a	2a	3a	4a	5a	8a	12
	AVE	2b	2b	3b	2b	2b	3b	2b	2b	3b	4b	5b	8b	
													10	
I_b= 0.9	1													
	2													
	3													
	4													
	5													
	6													
	7													
	Σ^2	2a	2a	3a	2a	2a	3a	2a	2a	3a	4a	5a	8a	12
	AVE	2b	2b	3b	2b	2b	3b	2b	2b	3b	4b	5b	8b	
													10	
I_b= 0	1													
	2													
	3													
	4													
	5													
	6													
	7													
	Σ^2	2a	2a	3a	2a	2a	3a	2a	2a	3a	4a	5a	8a	12
	AVE	2b	2b	3b	2b	2b	3b	2b	2b	3b	4b	5b	8b	
Σ^2		6a	7a	9a	6a	7a	9a	6a	7a	9a			10	
AVE		6b	7b	9b	6b	7b	9b	6b	7b	9b				
				11			11			11				
			13			13			13					

16a + 16a	18a	21	23
16b | 16b	18b		
	19		25

15a + 15a	17a	14
15b | 15b	17b	
	19	
20	22	
	24	

STEP PROCEDURE FOR FILLING IN
DICKSON THREE-FACTOR ANALYSIS TABLE

THE APPLICATION OF EXPERIMENTAL DESIGN TO A RADAR TARGET ACQUISITION SYSTEM

Dr. Erwin Biser, Harvey Eisenberg & George Millman
Systems Division
Surveillance Department, Evans Area
U. S. Army Signal R & D Laboratory
Fort Monmouth, N. J.

TABLE OF CONTENTS

ABSTRACT. 2^n factorial designs were applied to the problem of minimizing target acquisition time for standard and modified radar tracking systems in a series of tests at USASRDL.

With the aid of 2^4 and 2^3 factorial designs, the modification was shown to reduce target acquisition time and target transfer failure rates significantly. A determination was made of the dependence of acquisition time upon target velocity, course type, radar-crew combination, target range at destination, and time lapses.

INTRODUCTION.

a. The Problem. The radar tracking system under analysis required the transfer of target position information from an acquisition radar to a tracking radar. The latter radar was slewed from a random point.

(An acquisition radar is one that periodically scans a predetermined volume of space, searching for enemy targets. A track radar is one that closely follows a target, obtaining present position information and velocity for tracking and missile firing purposes.)

b. The Objective. The objective of the analysis is to determine the conditions under which target acquisition time is a minimum. This is the "yield of the process". The acquisition time function is presumed to be a function of many variables, such as:

1) The absence or presence of the modification, the Height Comparator, which presents the third target coordinate (height) to the target track radar operators during the acquisition process. (Note that this modification is not possible when the radar is operating by itself, and not as part of a defense system).

2) Proficiency of the four-man crew. (Two crews were used.)

3) Target range at designation, or, target slew-range.

4) Target velocity. (Slow, medium and fast aircraft were used.)

5) Altitude maneuver of the target.

6) Type of target course. (Radial or tangential courses.)

7) The effect of time lapses between sets of data.

8) Target transfer failure rate.

9) Operator overshoot.

10) Human Engineering aspects of target acquisition.

11) Initial designation failures ("warmup period").

TEST PLAN. Close control of the flight pattern was accomplished by means of the reference radar plotting board and UHF radio.

In order to avoid future pitfalls in future test planning, it is pertinent to add the following remarks. The final test plan and experimental design were quite different from the one originally conceived. Originally, it had been decided to acquire the target at definite points in space and in particular at:

Ranges: 8, 18, 28, 38 thousands of yards

Azimuths: 4800, 56,000, 6400 mils (W, NW, N)

Altitudes: Varying randomly among three altitudes, such as 6, 8, 10 thousands of feet.

The original concepts were revised when it became apparent that insufficient data would be produced. When the aircraft reached the point in space, the three radars were not always ready. When the radars were ready, the aircraft had often drifted off course. To increase efficiency it was decided to acquire targets in a random fashion, spotchecking afterwards to insure equal distribution in range and altitude.

VARIABLES USED IN TEST PLAN

1. MODIFICATION

   Level 1 : Modification in use during target acquisition.

   Level 2 : Modification not in use during target acquisition.

   (Standard mode of operation.)

2. RADAR - CREW COMBINATION

   Level 1: Radar #1 with its "permanent" crew.

   Level 2: Radar #2 with its "permanent" crew.

3. RANGE OF TARGET AT DESIGNATION

   Level 1 : Short range, e.g. , less than 20,000 yds.

   Level 2 : Long range, e.g. , greater than 20,000 yds.

Fig 1

VARIABLES USED IN TEST PLAN (CONT.)

4. TARGET COURSE

Level 1 : Radial course (e.g., azimuth angle constant).

Level 2 : Tangential course (e.g. , azimuth angle changing rapidly).

5. TIME LAPSE

Level 1 : October series of tests.

Level 2 : April series of tests.

6. AIRCRAFT (SIZE - VELOCITY - ALTITUDE COMBINATION)

Level 1 : Plane #1 , a slow propellor-driven plane used as a statistical control.

Level 2 : Plane #2.

Level 3 : Plane #3.

Fig 2

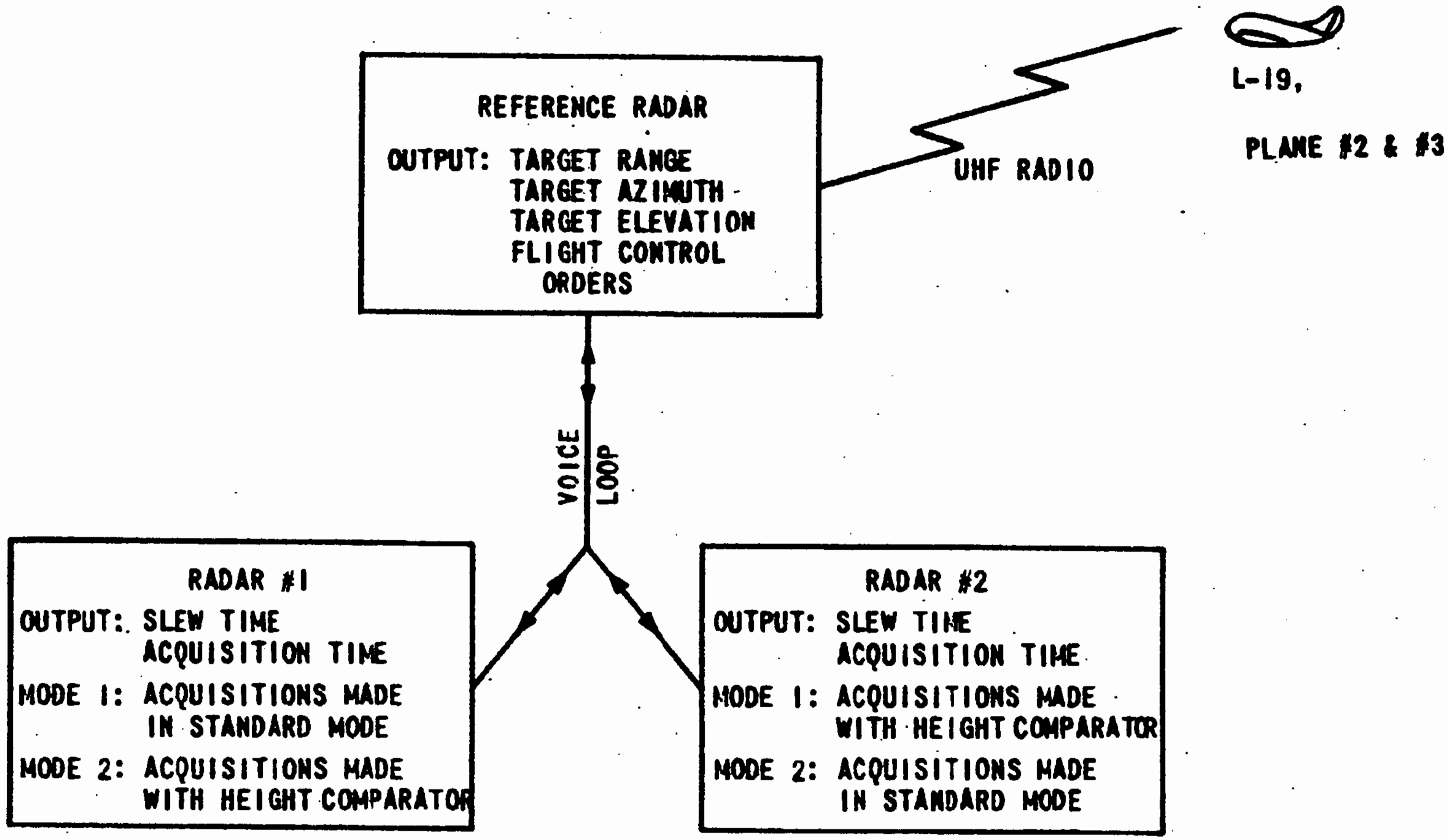


Fig 3 TEST CONFIGURATION

An additional revision was required when fast aircraft were flown. The fast aircraft consumed large amounts of fuel, reducing the data recording session, if flown at lower altitudes. In general, faster planes fly higher, and it would have been very expensive to separate the effects of velocity and altitude. This rapidly became evident after an initial try and the test plan was revised accordingly.

The final test plan considered the variables shown in Figs. 1 & 2.

TEST PROCEDURE. Three aircraft were flown against three tracking radars as shown in Fig. 3.

One radar was used as a target reference source and flight control center. The next radar, called radar #1 in the report, was used in the modified mode of acquisition when the third radar, called radar #2 in the report, was used in the standard mode. After approximately ten acquisitions, radars #1 and #2 changed their modes of acquisition.

The essential difference among the targets was that of velocity. Target #1 (plane #1) was the slowest; target #3 (plane #3) was the fastest. Average acquisition time and transfer failure rates were expected to increase with velocity, and this was verified by the analysis. The variable of target altitude was confounded with veloctiy, since faster planes tend to fly higher.

The two models of the Height Comparator (the modification) differed slightly. The model in radar #1 needed an operator to slew the antenna elevation, while the model in radar #2 was completely automatic. This effect was considered minor and is confounded with the radar-crew variable.

Plane #1, (L-19), the slowest plane, was varied continuously in altitude to prevent the elevation operators from anticipating the target elevation angle. It was found that this aspect of the test plan was rarely considered in the field. This can be easily accomplished with slow aircraft. Planes #2 and #3 were flown at constant altitude and were varied only slightly in altitude during the test since their speed and position change made each designation appear as a new target. In general, interdependence between any two successive acquisitions was reduced by varying the elevation angle randomly as much as $\pm$ 250 mils.

The "Count down to acquire" command was given only if both designation radars obtained good video on the PPI display. Prior to designation each track radar was off target for a minimum of one minute, standing by at a pre-determined range, zero azimuth, and zero elevation. Each designation was performed simultaneously by both radars. The designation time and target position were recorded. The time clock was activated when the designation operator pressed his designate button and was deactivated when the track operators threw the automatic track switch. Slew time was also recorded but was not used in the analysis except as a check. (Average slew time was 4 seconds.)

APPLICATION OF A 2^3 DESIGN. To elucidate the application of experimental designs, let us examine Data Set #6. Three factors, each with two levels, were studied. The factors, upper levels, and lower levels are defined in Fig. 4.

DATA SET 6 : A 2^3 EXPERIMENT

A : MODIFICATION FACTOR

a_1 = Modification in use during target acquisition.

a_2 = Modification not in use during target acquisition.

(Standard mode of operation.)

B : TARGET RANGE FACTOR

b_1 = Short range. (Range less than 20,000 yards.)

b_2 = Long range. (Range equal to or greater than 20,000 yards.)

C : AIRCRAFT - VELOCITY FACTOR

c_1 = Plane #1

c_2 = Plane #2

Fig 4

FACTORS AND TREATMENTS FOR A 2^3 EXPERIMENT

A : Modification Factor

B : Range Factor

C : Plane Factor

TABLE OF AVERAGE ACQUISITION TIMES (SEC.)

		c_1	c_2
a_1	b_1	7.76 ± 0.53	9.03 ± 1.12
	b_2	5.86 ± 0.47	9.18 ± 1.80
a_2	b_1	10.13 ± 0.74	14.59 ± 3.22
	b_2	8.76 ± 1.24	13.04 ± 2.58

Fig 5

AVERAGES FOR DATA SET #6 (20 REPLICATIONS)

Treatment Symbol	Average Acquisition Time (secs.)
(1) = $a_1b_1c_1$	7.76
a = $a_2b_1c_1$	10.13
b = $a_1b_2c_1$	5.86
c = $a_1b_1c_2$	9.03
ab = $a_2b_2c_1$	8.76
ac = $a_2b_1c_2$	14.59
bc = $a_1b_2c_2$	9.18
abc = $a_2b_2c_2$	13.04

	Average Acquisition Time (secs.)
$\bar{a}_1$	7.96
$\bar{a}_2$	11.63
$\bar{b}_1$	10.38
$\bar{b}_2$	9.21
$\bar{c}_1$	8.13
$\bar{c}_2$	11.46

Fig 6

DATA SET #6 : A 2^3 EXPERIMENT (CONT.)

$$\text{"A" Effect} = \frac{1}{4}\left[a - (1) + (ac - c) + (ab - b) + (abc - bc)\right]$$

At b_1c_1 (Sht Rgd,Rdr 1) :

a-(1) = 10.13 - 7.76 = 2.36 seconds improvement

At b_1c_2 (Sht Rgd,Rdr 2) :

ac - c = 14.59 - 9.03 = 5.56 seconds improvement

At b_2c_1 (Lg Rge, Rdr 1) :

ab - b = 8.76 - 5.86 = 2.90 seconds improvement

At b_2c_2 (Lg Rge, Rdr 2) :

abc - bc = 13.04 - 9.18 = 3.86 seconds improvement

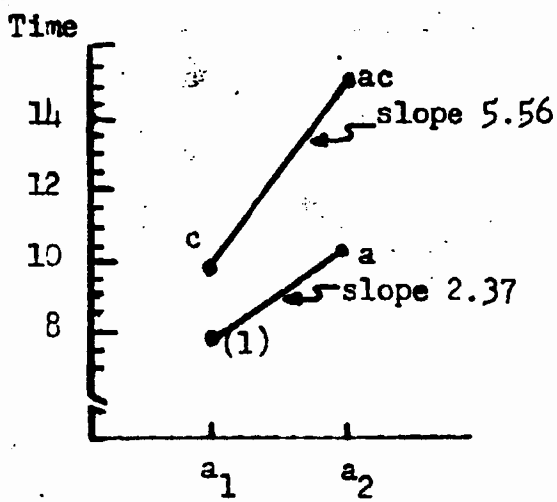
Average of four subeffects:

$\frac{1}{4}$ (2.37 + 5.56 + 2.90 + 3.86) = 3.67 seconds overall improvement attributable to the modification.

Fig 7

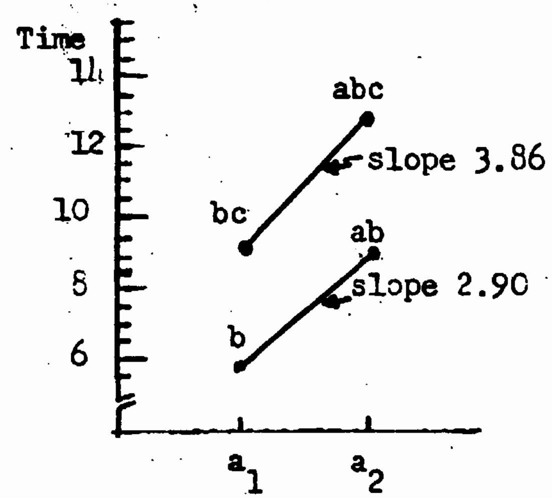
DATA SET # 6: A 2^3 EXPERIMENT (CONT.):

Illustration of a "A" Effect



"A" Effect Measured at

Lower Level of B (b_1)



"A" Effect Measured at

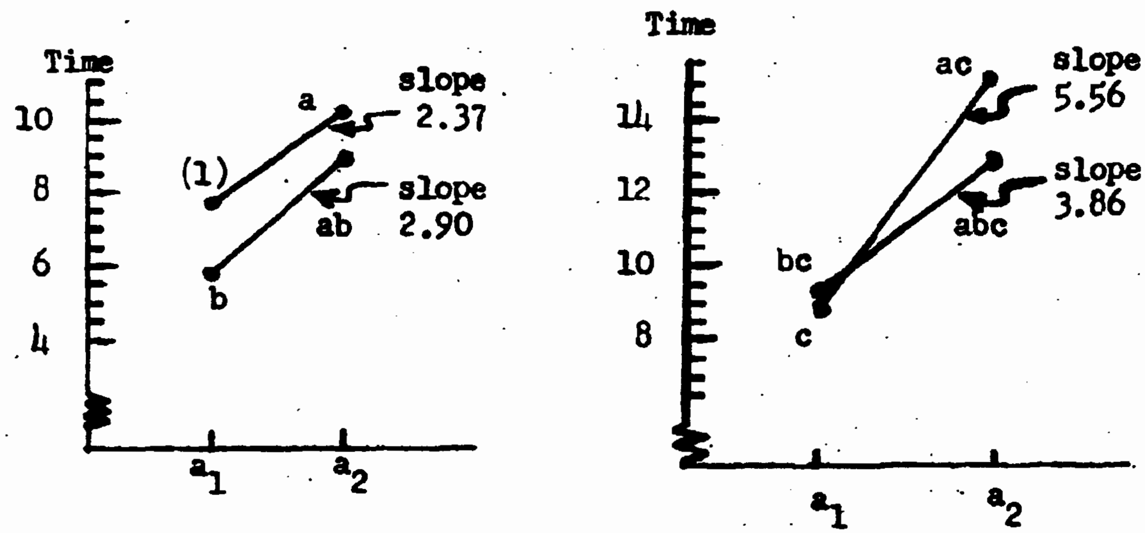
Upper Level of B (b_2)

Note the parallelism or independence between **a** and **c** .

Fig 8

DATA SET # 6 : A 2^3 EXPERIMENT (CONT.):

Illustration of "A" Effect



"A" Effect Measured at Lower Level of C (c_1)

"A" Effect Measured at Upper Level of C (c_2)

Fig 9

DATA SET #6 : A 2^3 EXPERIMENT (CONT.)

AB Interaction

$$AB = \frac{1}{4} \left[(ab-b) + (abc - bc) - (a - (1)) - (ac - c) \right]$$

If AB = 0, there is no interaction between A and B (Modification and Range.) Thus, the modification would have the same effect on acquisition time for any target range.

At b_2, the "A" Effect is:

$$(ab - b) + (abc - bc) = \frac{(2.90 + 3.86)}{2} = 3.38$$

At b_1, the "A" Effect is:

$$(a - (1)) + (ac - c) = \frac{(2.36 + 5.56)}{2} = 3.96$$

The AB Interaction is:

$$\frac{1}{2}(3.38 - 3.96) = -0.29$$

Fig 10

The experiment was replicated twenty times. The averages are illustrated in Figs. 5 & 6.

The effect of the modification, the "A" factor, can be determined by means of the formula shown in Fig 7.

The above slopes in Fig. 7 (2.36, 5.56, 2.90, 3.86) can be illustrated by considering either the pair of projections shown in Fig. 8 or the pair of projections shown in Fig. 9.

The interactions AB, AC, BC can be determined. AB (the effect on the modification factor of changing levels of the range factor) is given in Fig. 10.

For the remaining calculations, see Data Set #6.

NATURE OF THE ACQUISITION TIME FUNCTION: This series of tests sheds much light on the overall acquisition procedure. The acquisition time function can be discussed with respect to two points of view.

a. Mean Acquisition Time - The detailed discussions on mean acquisition time that follow can be generalized as follows:

1. When the slow radar target, plane #1, was flown as a control on a straight and level course, acquisition time averages were in the region of 8 to 10 seconds. The Height Comparator reduced acquisition time as much as 2 seconds depending upon the crew, training, length of test, and target course type. When the L-19 is not flown as a control, the average acquisition times can fall between 14 to 16 seconds.

2. Acquisition time averages on radar target #2, plane #2, were in the neighborhood of 14 to 16 seconds and were reduced significantly to approximately 9 seconds by the Height Comparator.

3. Acquisition time averages on radar target #3, plane #3, were in the vicinity of 16 to 19 seconds and were reduced significantly to about 14 or 15 seconds. The plane #3 data is biased in that a loss of skill had occurred in the five month interval between the plane #2 and plane #3 flights. Thus, the plane #3 averages were adjusted downwards when comparisons with the earlier data are made.

4. Target slew-range, time lapses, and the radar-crew combination had a considerable effect upon this time analysis.

b. Overshoot - The acquisition time function possesses a characteristic more commonly found in servo systems. If the radar return from the target is weak, or if the target is moving rapidly with respect to the slew time, the radar gates, or the crew's coordination capabilities, it is easy for the elevation operator to bypass, or overshoot, the target while slewing blindly in elevation. When this occurs, several seconds are needed to correct this error. From a mathematical point of view this produces a bimodal distribution, that is, the acquisitions are grouped about two means instead of one. If overshoot does not occur, the acquisition time averages fall within the interval of 8 to 12 seconds. If overshoot does occur, the acquisition averages fall within 14 to 18 seconds. The overall average must therefore be representative of both averages,

FREQUENCY DISTRIBUTION OF ACQUISITION
TIMES ON PLANE #3

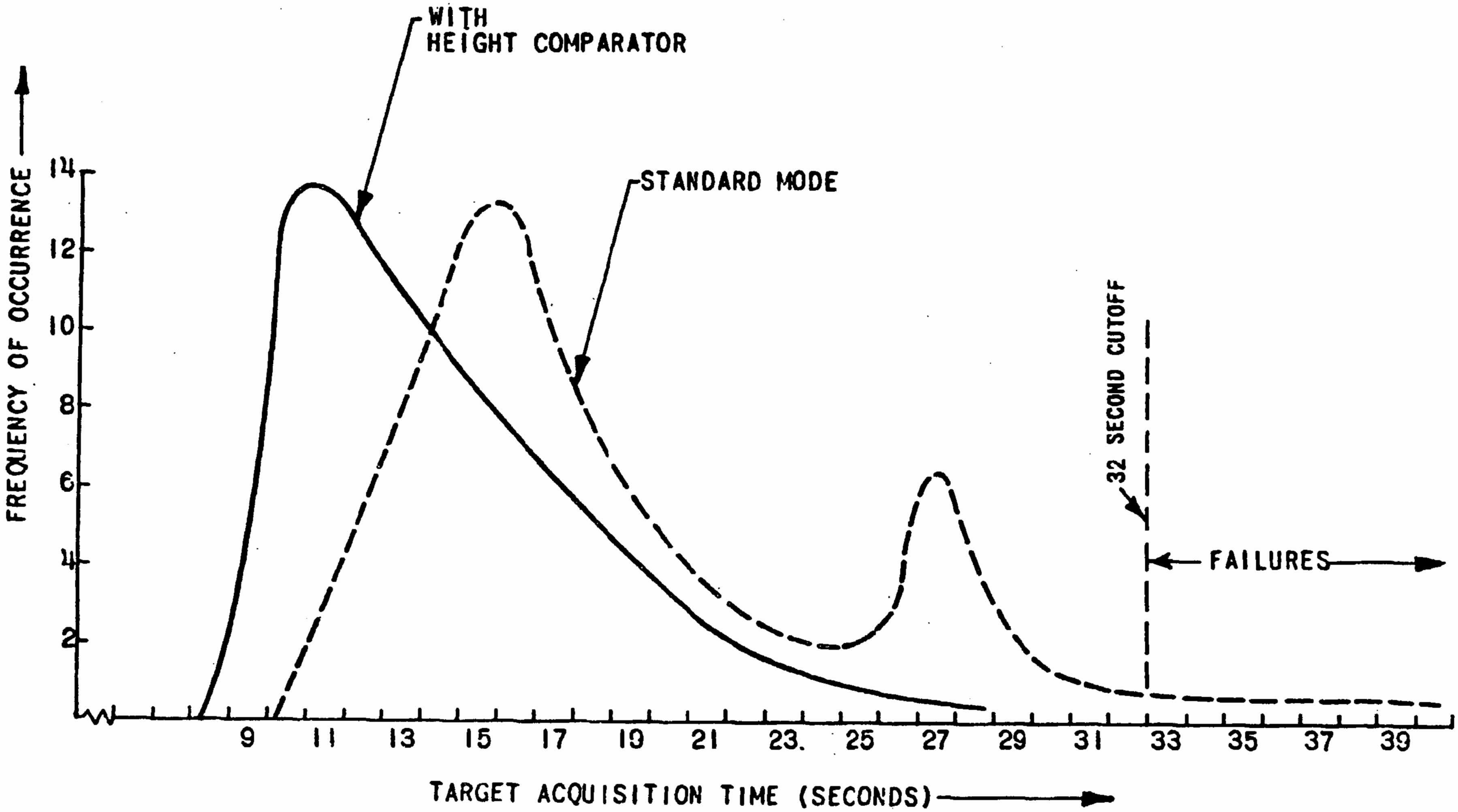


Fig 11

FREQUENCY DISTRIBUTION OF
ACQUISITION TIME ON PLANE #2

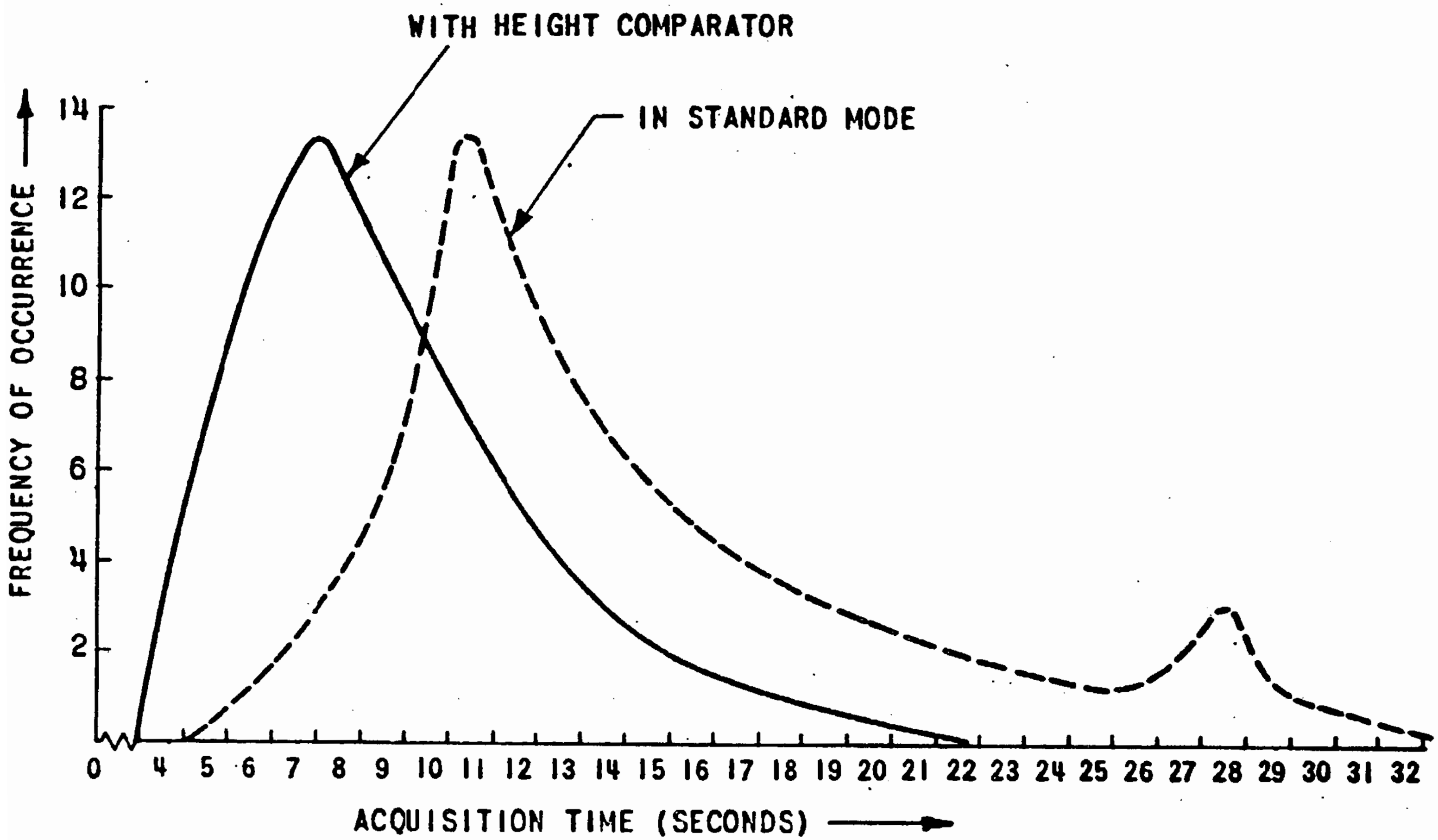


Fig 12

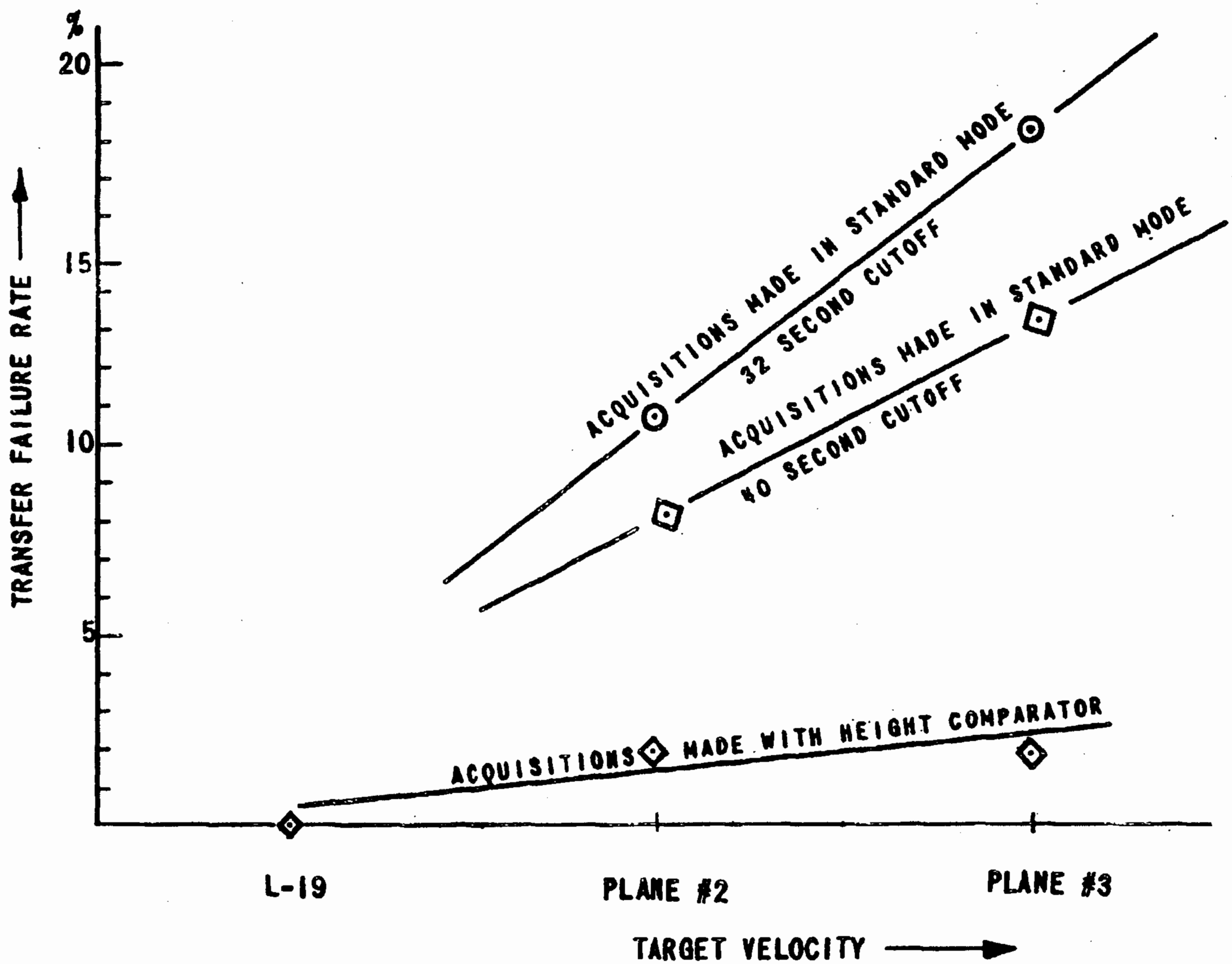

%
20
15
10
5
TRANSFER FAILURE RATE
ACQUISITIONS MADE IN STANDARD MODE
32 SECOND CUTOFF
ACQUISITIONS MADE IN STANDARD MODE
NO SECOND CUTOFF
ACQUISITIONS MADE WITH HEIGHT COMPARATOR
L-19
PLANE #2
PLANE #3
TARGET VELOCITY

Fig 13

93

indicating the percentage of the time that overshoot occurs. It is apparent from the data, and from the visual observations made during the test series, that the presence of the Height Comparator almost completely eliminates overshoot. The data on plane #3 indicates that the bimodal nature of the acquisition time function disappears at approximately 32 seconds, indicating that acquisitions made after 32 seconds are different in nature and suggesting that acquisitions after 32 seconds, not 40, should be called failures. Note that the overshoot nature is not as apparent for the slower aircraft, but the data indicates that, if present, it occurs in less than 20 or 25 seconds.

EFFECT OF VARIABLES ON ACQUISITION TIME: This series of tests was designed as a 2^n factorial experiment and the data was reduced accordingly. Although a large number of variables exists in a test of this type, the data reduction indicated definite trends and consistencies.

a. Statistical Control - In order to maintain consistent control of the data over an extended test period, the low velocity radar target (L-19 aircraft) was used to obtain statistical control data as well as data pertaining to low velocity aircraft. The initial tests were performed with the low velocity aircraft (L-19 propeller-driven aircraft), the second phase used a medium velocity, and the third phase used a high velocity target. In phases two and three, the low speed aircraft was utilized as a time check standard against men and equipment.

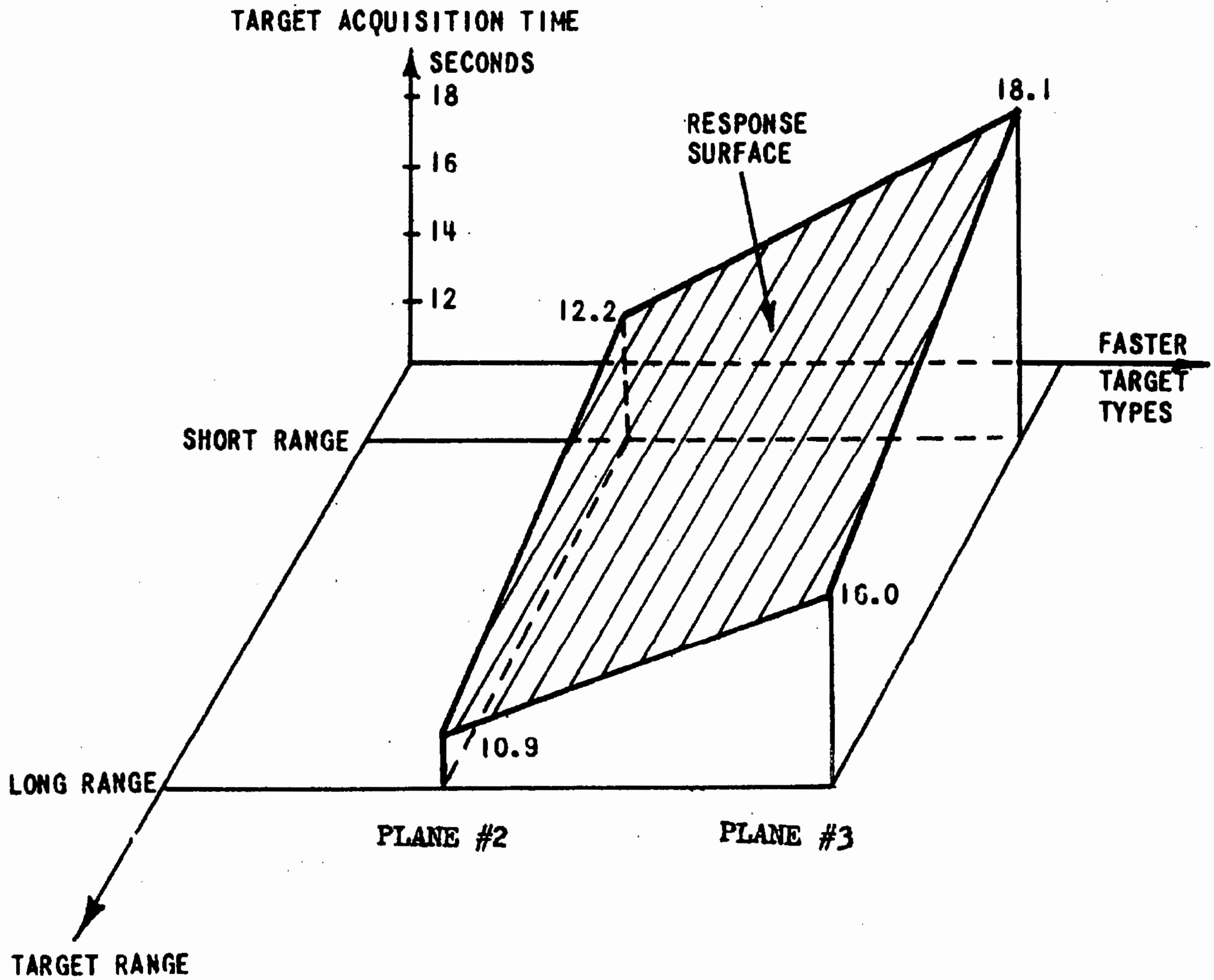
b. Fast Targets - Effects of Height Comparator, Target Course Type and Time Lapse on Acquisition Time - From data Sets 6 and 7 it is seen that the use of the Height Comparator resulted in reducing target acquisition time, for planes #2 and #3 , by 4.7 seconds (See Tables 2 and 3).

The data indicate (the small sample size resulted in large errors) that the course type, radial vs. tangential, may have some effect on acquisition time. But this effect, if present, is dependent upon the radar-crew combination. The effect of flying a fast plane radially and tangentially will be found in Data Set 10. The difference in average acquisition time is 1.5 seconds, but this figure cannot be considered statistically significant, since the interval of uncertainty is almost $\pm$ 2 seconds. Further details are given in table 4.

The effect of the radar-crew combination upon target acquisition time was not pronounced, and in fact was statistically insignificant.

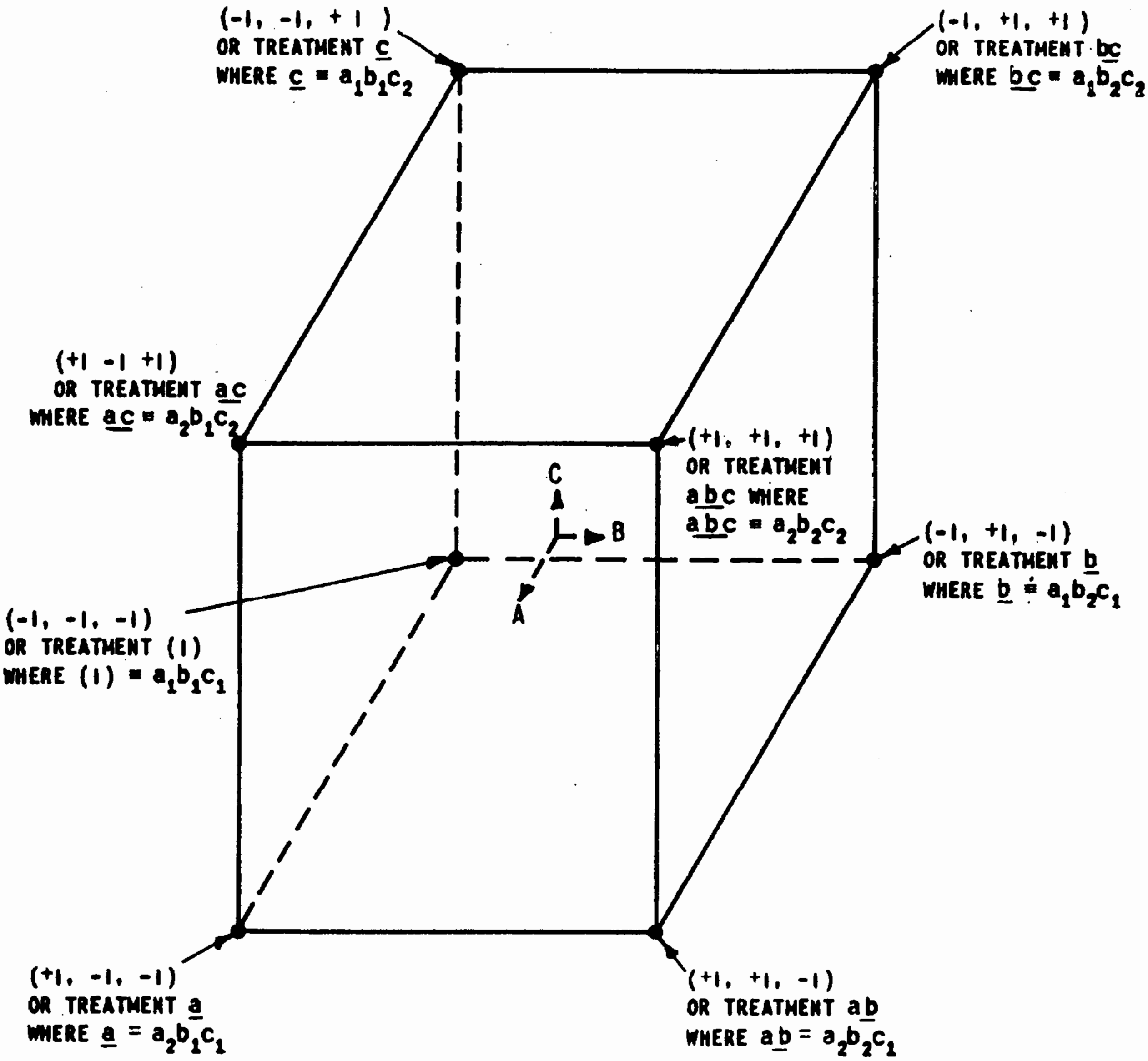
Due to the lack of data on radar-crew #2 for target or plane #2 acquisitions, only the data of radar-crew #1 could be used in table 5 comparing planes #2 and #3.

Referring to Table 5, the figures 3.26 and 2.35 seconds (in the last column) are confounded with the effect of a six months time lapse between the plane 2 and plane 3 flights. This effect has been shown to be statistically significant. From the data (discussed in Data Set 4) it is seen that the acquisition times on the L-19 increased by 1.4 seconds because of the six months time lapse when both crews were studied; 1.07 seconds of this increase was attributed to radar-crew #1. At this point, one of two assumptions can be made.



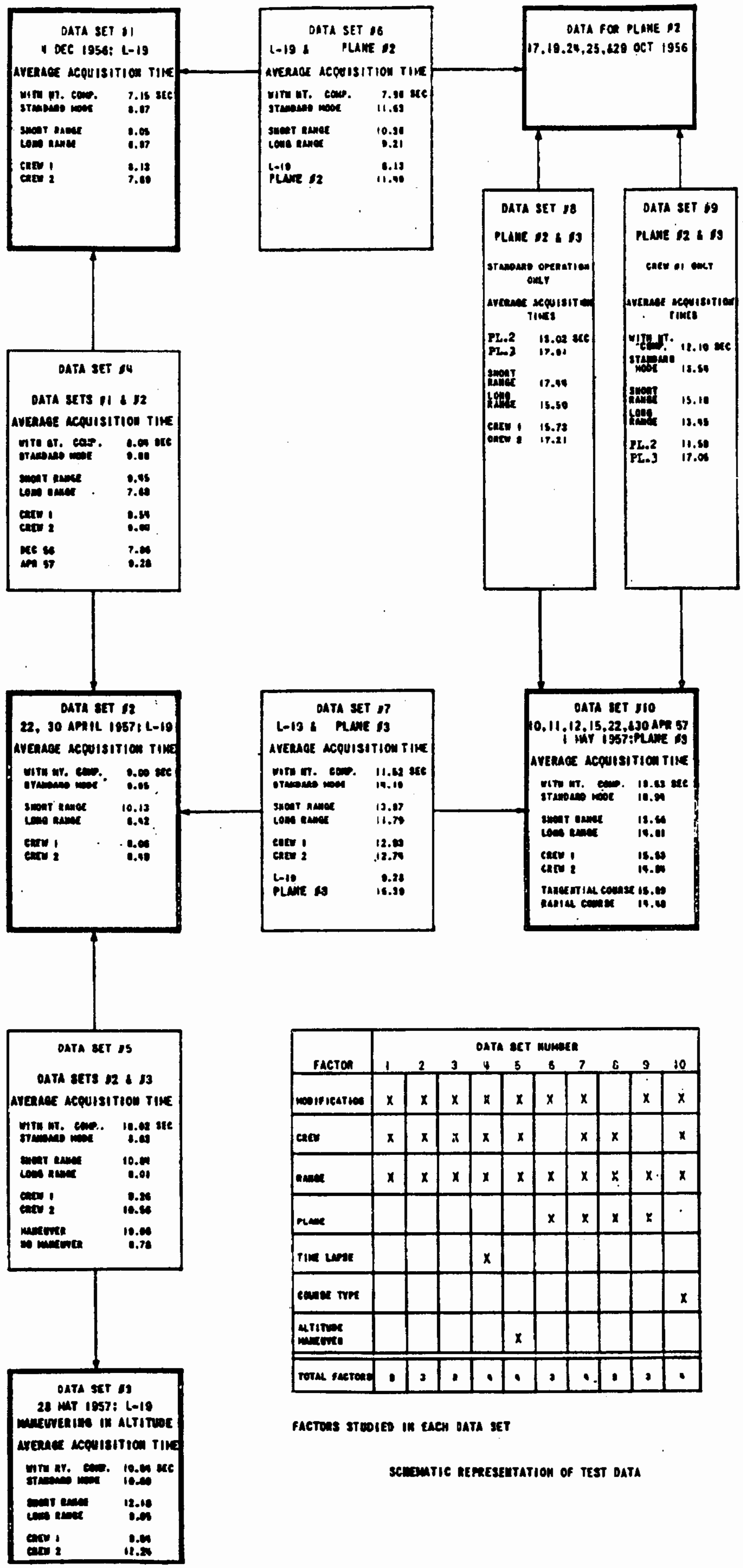
RESPONSE SURFACE OF ACQUISITION TIME
FROM DATA SET 9

Fig 14



LEVELS OF THE VARIABLES FOR A 2^3 FACTORIAL EXPERIMENT

Fig 15



FACTOR	DATA SET NUMBER 1	2	3	4	5	6	7	8	9	10
MODIFICATION	X	X	X	X	X	X	X		X	X
CREW	X	X	X	X	X		X	X		X
RANGE	X	X	X	X	X	X	X	X	X	X
PLANE						X	X	X	X	
TIME LAPSE				X						
COURSE TYPE										X
ALTITUDE MANEUVER					X					
TOTAL FACTORS	[illegible]	3	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	3	[illegible]

FACTORS STUDIED IN EACH DATA SET

SCHEMATIC REPRESENTATION OF TEST DATA

Fig 16

FORMAT FOR 2^n FACTORIAL DESIGN AND REDUCTION
TIME OF ACQUISITION
4 DEC 1956 VS 22 APR 1957

	a_1								a_2								ROW TOTALS			
	b_1				b_2				b_1				b_2							
	c_1		c_2		c_1		c_2		c_1		c_2		c_1		c_2					
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	Σt	Σt^2	$\bar{t}=\Sigma t/16$	$(\Sigma t)^2$
TREATMENT	(1)	d	c	cd	b	bd	bc	bcd	a	ad	ac	acd	ab	abd	abc	abcd				
REPLICATE 1																				
2																				
.																				
.																				
.																				
r																				
COLUMN TOTALS: $G=\Sigma t$																	G.T.=	X X X X X	X X X X X	$\Sigma((\Sigma t)^2)$=
$H=\Sigma t^2$																	X X X X X	CSS=	X X X X X	$\Sigma(\ )/2^n$=
$G/r=\bar{t}$																	X X X X X	X X X X X	G.A.	$\Sigma(\ )/2^n$-C.T.
$I=(\Sigma t)^2$																	→ Σ=	Σ/r=	(Σ/r)-C.T.=	C.T.=
$J=I/r$																				
$K=H-J$																				
$L=K/(r-1)$																				
$\sqrt{L}=\sigma_e$																				
$M.E.^*=t\,\sigma_e/\sqrt{r}$																				

NOTE:
G.A. = G.T./$2^n r$
C.T. = $(G.T.)^2/2^n r$ = (G.T.)x(G.A.)
TREATMENT EFFECT = Σ/r - C.T.
REPLICATE EFFECT = $\Sigma(\ )/2^n$ - C.T.

*NOTE: $t_{.05}$ is obtained from Student t-table, as column for (r-1) d.f.

LEVELS:
- a_1: ACQUISITIONS MADE WITH HEIGHT COMPARATOR
- a_2: ACQUISITIONS MADE WITHOUT HEIGHT COMPARATOR (STANDARD MODE)
- b_1: SHORT RANGE, e.g. <20,000 YARDS
- b_2: LONG RANGE, e.g. ≥20,000 YARDS
- c_1: RADAR-CREW 1 (RADAR #530)
- c_2: RADAR-CREW 2 (RADAR #360)
- d_1: TIME OF ACQUISITION AS OF 4 DEC 1956
- d_2: TIME OF ACQUISITION AS OF 22 APR 1957

CONTRAST AND TREATMENT VARIANCES

FACTORIAL EFFECT	EFFECT SUM	Q	MEAN DIFF. IN AVG. = EFF. SUM/Q	(EFFECT SUM)²	(EFF. SUM)²/$2^n r$
M	G.T.=	$2^n r$=	G.A.=		C.T.=
A		$2^{n-1} r$=	±L.S.D.*		
B		•	•		
C		•	•		
D		•	•		
AB		•	•		
AC		•	•		
AD		•	•		
BC		•	•		
BD		•	•		
CD		•	•		
ABC		•	•		
ABD		•	•		
ACD		•	•		
BCD		•	•		
ABCD		•	•		
	TOTAL TREATMENT EFFECT = Σ OF 2^n-1 EFFECTS				
	CSS = GSS - C.T.				

*NOTE: L.S.D. = $t_{.05}\sigma_e\sqrt{2/2^{n-1}r} = t_{.05}\sigma_e 2/\sqrt{r}$, where $t_{.05}$ is obtained from t-table for $(2^n r-1)$d.f.

TREATMENT ALGORITHM FOR 2^n FACTORIAL DESIGN

FACTORIAL EFFECT	TREATMENT COMBINATION APPLIED TO COLUMN TOTALS Σt															
	(1)	d	c	cd	b	bd	bc	bcd	a	ad	ac	acd	ab	abd	abc	abcd
M	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
B	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+
C	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
D	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
AB	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+
AC	+	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+
AD	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
BC	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+
BD	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+
CD	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+
ABC	-	-	+	+	+	+	-	-	+	+	-	-	-	-	+	+
ABD	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-	+
ACD	-	+	+	-	-	+	+	-	+	-	-	+	+	-	-	+
BCD	-	+	+	-	+	-	-	+	-	+	+	-	+	-	-	+
ABCD	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+

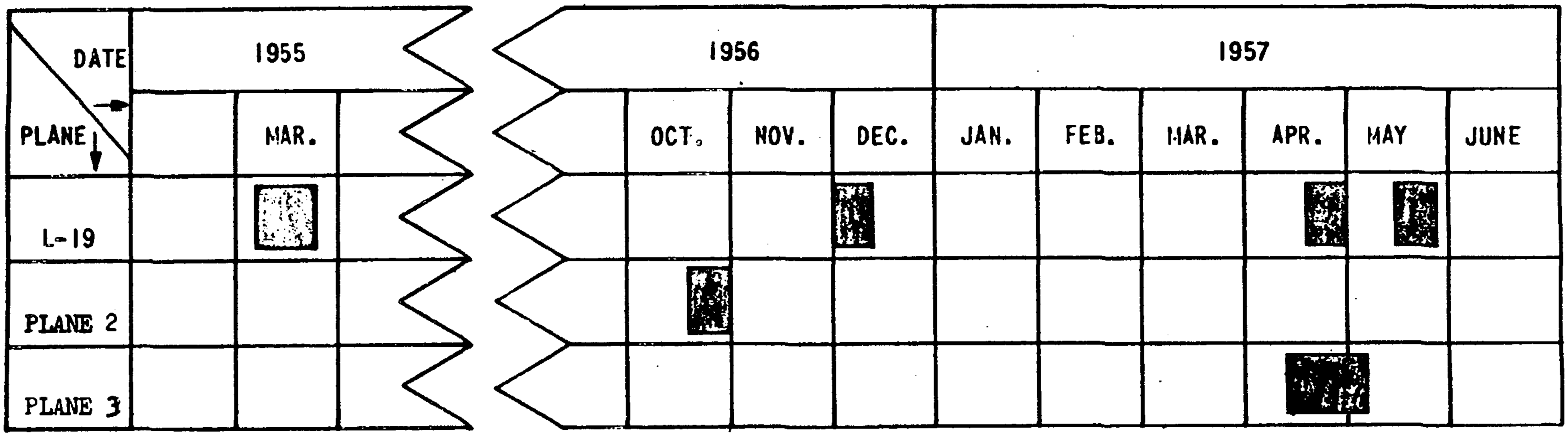
NOTE: The treatment algorithm for a 2^3 factorial design may be obtained from the above table by deleting all rows and columns involving the symbols D and d respectively.

ANALYSIS OF VARIANCE TABLE WITH F-TEST OF SIGNIFICANCE

EFFECT	SS COMPONENTS	TOTALS	D.F.	MSS =SS/D.F.	F_c =MSS÷R.E.MSS
TREATMENTS					
A			2-1=1		
B			1		
C			1		
D			1		
AB			1		
AC			1		
AD			1		
BC			1		
BD			1		
CD			1		
ABC			1		
ABD			1		
ACD			1		
BCD			1		
ABCD			1		
TOTAL TREATMENT			2^n-1=15		
REPLICATE EFFECT (R)			r-1		
RESIDUAL ERROR (R.E.)(BY SUBTRACTION)			15(r-1)	σ_e^2=	1.00
CSS = GSS - C.T.			$2^n r-1$	$\sigma_{G.T.}^2$=	X X X X X

OBTAIN:
- σ_e — $F_{.05}[1,15(r-1)]$ — $F_{.01}[1,15(r-1)]$
- $\sigma_{G.T.}$ — $F_{.05}[15,15(r-1)]$ — $F_{.01}[15,15(r-1)]$
- $F_{.05}[(r-1),15(r-1)]$ — $F_{.01}[(r-1),15(r-1)]$

Fig 17



EXACT DATES FOR START OF FLIGHTS

L-19: 4 DECEMBER 1956, 22 APRIL & 28 MAY 1957
PL.2: 17 OCTOBER, 1956
PL.3: 10 APRIL 1957

Fig 18 DATES OF AIRCRAFT FLIGHTS

If the degradation applies only to the L-19 acquisitions, and not to the fast plane, then the second figure of 2.35 seconds quoted above still holds. If the degradation does apply to the fast aircraft, then the difference in acquisition time improvement between the plane #2 and plane #3 is 1.28 seconds (2.35 minus 1.07). The improvement figure of 1.28 seconds is not statistically significant since the interval of uncertainty about it is almost ± 2 seconds. Thus, it cannot be determined from this data whether or not the improvement is greatest for the fastest aircraft. See Table 6 for further details.

The improvement in acquisition performance with respect to plane #3 is more outstanding in the reduction of the number of overshoots and the transfer failure rate.

c. Comparison of Slow and Fast Radar Targets for the Height Comparator: When the L-19 is compared to the fast aircraft, two important facts emerge:

1. Between the first and second groups of flights, a time lapse effect of 1.4 seconds was present.

2. There was a gap in data of plane #2. Acquisition times for radar-crew #2 with the Height Comparator was essentially missing due to radar malfunction.

Thus, only the performance of radar-crew #1 was considered. The tables on the following page from Data Sets 6 and 7 are relevant.

d. The Height Comparator and Slow Targets: The effect of the Height Comparator is closely related to the skill and proficiency of the radar-crew combination when consecutive acquisitions are made on a slow aircraft flying a constant altitude course. This difference (in acquisition time) could be attributed to the difference in radars, to the difference between the two models of the Height Comparator, or to the difference in operator skill of the two crews. One crew made better use of the Height Comparator while the other did not need it. As an illustration consider Table 7 taken from Data Set 1.

The average improvement in acquisition time averaged over both crews, is 1.72 seconds (one-half sum of 2.64 and 0.80).

From Data Set 2, it is seen that the crew-comparator interaction was also present in the April 1957 tests. The effect of the Height Comparator in the May 1957 tests was not statistically significant, but the average acquisition time of the crews varied by 1.35 seconds.

Though not statistically significant, there is an indication of interaction between target range and the Height Comparator. If this interaction is not a random fluctuation, then the results imply that the Height Comparator is more effective in reducing acquisition time for targets at short range. This interaction is to be expected, because acquisitions made at short range have greater range slew times.

e. Range and Slewing Effect: For all targets the designation range of the target consistently had a significant effect on acquisition time. This effect was expected because of the test procedure. The radar was always slewed inward from 40,000 yards range, the maximum computer range. However, this effect occurred with unexpected consistency. The difference in average acquisition time between short and long ranges varied from 1 to 3 seconds. For the L-19, the average short range was 12 to 15 thousand yards, while the average long range was 25 to 26 thousand yards. For the fast aircraft the range figures are somewhat larger. It is noteworthy to point out that slew time was also recorded. However, search time (acquisition time minus slew time) was not analyzed. It was felt that the extra effort was not warranted. A cursory examination showed that the average slew time appeared to run about 4 or 5 seconds. This was considered to be a reasonable amount of time--neither too great nor too small. The actual slew times ran greater than the theoretical slew times computed from the maximum slew rate when spot checks were performed. The slew time can be considered as range-slew time since azimuth slewing was relatively unimportant in this test series.

f. Effect of Altitude Maneuver: The Data Set 5 describes the effect of diving and climbing the L-19 during the April 1957 series. The results given therein appear to be at variance with those reported in prior tests. However, several factors must be considered. First, different crews were involved, and as shown previously, the effect of the Height Comparator is closely related to the radar-crew combination. Second, the operators were permanent ESL personnel with extensive radar experience. It was clear from the beginning of the test series that these radar crews possessed more skill and proficiency than the enlisted men used for the prior tests mentioned above. Third, the altitude maneuver in the April 1957 tests corroborates the conclusion that the radar crews were not making full use of the Height Comparator.

TABLE I: SHORT SUMMARY OF RESULTS (U)

FLIGHT NR.	TARGET VELOCITY	ACQUISITIONS MADE WITH HEIGHT COMPARATOR					ACQUISITIONS MADE IN STANDARD MODE					REMARKS
		SAMPLE SIZE	95th PERCENTILE* (sec)	FAILURE RATES**			SAMPLE SIZE	95th PERCENTILE*	FAILURE RATES**			
				40 sec. CUTOFF	32. sec. CUTOFF	16 sec. CUTOFF			40 sec. CUTOFF	32 sec. CUTOFF	16 sec. CUTOFF	
I		117	10.4	0%	NA	0%	120	17.5	0%	NA	7.5%	Flight I was flown as a statistical control, soon after flight IV.
II	SLOW	72	13.7	0	NA	1.4	72	15.8	0	NA	9.7	Flight II was flown as a statistical control, just before flight V.
III		31	13.4	0	NA	3.2	31	23.4	0	NA	3.2	Target was maneuvered in altitude for flight III.
IV	MEDIUM	96	16.3	2.1	2.1	NA	120	28.6	8.3	10.7%	NA	Flight IV was flown only on tangential courses.
V		110	22.9	1.8	1.8	NA	120	31.5	13.3	18.3	NA	Flight V was flown only on tangential courses.
VI	FAST	48	19.8	4.2	4.2	NA	48	34.9	8.5	18.9	NA	Flight VI was flown only on radial courses.

* 95% of the acquisition times fell in the interval between 0 seconds and the numbers indicated in this column. The numbers in the column were computed for a 40 seconds cutoff time.

** The failure rate is defined in terms of cutoff time. For example, if a 40 second cutoff time is used, all acquisitions consuming 40 seconds or more are called failures.

NA: Not applicable.

TABLE 2: PERFORMANCE ON PLANE #2
(RADAR CREW 1)

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)	
	WITH HT. COMP.	STANDARD MODE		
SHORT RANGE	9.03±1.12	14.59±3.22	5.56	4.71
LONG RANGE	9.18±1.80	13.04±2.58	3.86	
DIFFERENCE	0.15	1.55	X X	X X

*FOR FURTHER DETAILS SEE DATA SET 6

TABLE 3: PERFORMANCE ON PLANE #3

	AVERAGE ACQUISITION TIME (SEC)				AVERAGE IMPROVEMENT OF HT. COMP. OVER STANDARD MODE (SEC)	
	WITH HT. COMP.		STANDARD MODE			
	SHORT RANGE	LONG RANGE	SHORT RANGE	LONG RANGE		
RADAR CREW 1	16.55±2.63	13.04±1.44	19.21±2.80	18.37±2.99	4.00	4.71
RADAR CREW 2	14.56±1.60	11.99±1.70	20.11±3.65	17.29±2.60	5.42	
DIFFERENCE BETWEEN CREWS	1.49	1.05	-0.90	1.08	X X	X X

*FOR FURTHER DETAILS SEE DATA SET 7

TABLE 4: AVERAGE ACQUISITION TIME FOR DIFFERENT COURSE TYPES (SECONDS) PLANE #3

	COURSE TYPE		AVERAGE IMPROVEMENT
	RADIAL	TANGENTIAL	
RADAR CREW 1	16.87	14.20	2.67
RADAR CREW 2	15.12	14.76	0.36
DIFFERENCE	1.75	-0.56	

*HEIGHT COMPARATOR AND STANDARD MODE AVERAGED OUT

TABLE 5: ACQUISITION TIMES FOR FAST AIRCRAFT
(RADAR CREW COMBINATION #1)

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)
	WITH HT. COMP.	STANDARD MODE	
PLANE #2	8.78±1.27	14.39±2.13	5.61
PLANE #3	15.42±1.57	18.68±1.65	3.26(4.33*)
DIFFERENCE	-6.64	-4.29	2.35(1.28*)

*ESTIMATED IMPROVEMENT FIGURES IF A CORRECTION OF 1.07 SECONDS IS MADE FOR A TIME LAPSE FACTOR.

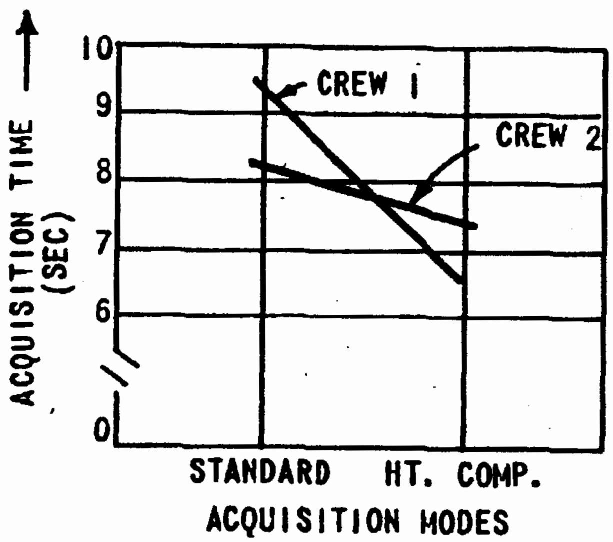
TABLE 6: THE EFFECT OF A SIX MONTH LAPSE (U)

	L-19 AVERAGE ACQUISITION TIME (SEC)				INCREASE OF BOTH MODES (SEC)
	WITH HT. COMP		STANDARD MODE		
	DEC 1956	APRIL 1957	DEC 1956	APRIL 1957	
RADAR CREW 1	6.88±0.55	8.15±0.92	9.14±0.83	9.99±0.68	1.06
RADAR CREW 2	7.31±0.46	9.85±0.94	8.12±1.08	9.13±0.56	1.77
INCREASE OF BOTH CREWS*	1.91		0.93		

*WITH TIME

TABLE 7 ACQUISITION TIMES FOR L-19 AIRCRAFT (U)

	AVERAGE ACQUISITION TIME (SEC)		IMPROVEMENT (SEC)
	WITH HT. COMP.	STANDARD MODE	
RADAR CREW 1	6.81±0.46	9.45±0.73	2.64
RADAR CREW 2	7.49±0.51	8.29±0.84	0.80
DIFFERENCE BETWEEN CREWS	-0.68	1.16	1.84



MODIFICATION-CREW INTERACTION (U)

Note that the first crew-radar was faster than the second when the Height Comparator was used. However, in the standard mode, the first was slower.

TREATMENT		ACQUISITION TIME (SEC) — HEIGHT COMPARATOR: a_1 — SHORT RANGE: b_1 — RADAR-CREW 1: c_1	HEIGHT COMPARATOR: a_1 — SHORT RANGE: b_1 — RADAR-CREW 2: c_2	HEIGHT COMPARATOR: a_1 — LONG RANGE: b_2 — RADAR-CREW 1: c_1	HEIGHT COMPARATOR: a_1 — LONG RANGE: b_2 — RADAR-CREW 2: c_2	STANDARD MODE: a_2 — SHORT RANGE: b_1 — RADAR-CREW 1: c_1	STANDARD MODE: a_2 — SHORT RANGE: b_1 — RADAR-CREW 2: c_2	STANDARD MODE: a_2 — LONG RANGE: b_2 — RADAR-CREW 1: c_1	STANDARD MODE: a_2 — LONG RANGE: b_2 — RADAR-CREW 2: c_2
		(1)	c	b	bc	a	ac	ab	abc
REPLICATE	1	8.5	10.8	6.8	5.7	14.0	14.7	7.8	7.4
	2	7.7	9.2	6.5	6.2	12.5	15.6	8.6	7.2
	3	8.4	8.7	6.1	5.7	11.8	13.3	7.5	6.9
	4	9.0	9.0	7.3	5.2	9.8	8.6	17.0	6.3
	5	6.5	8.9	6.2	6.7	10.8	9.9	10.0	6.9
	6	8.5	6.1	6.7	6.6	9.0	7.8	10.2	5.9
	7	7.1	7.2	5.8	6.8	7.5	11.0	7.3	5.5
	8	7.4	8.2	6.5	6.5	9.5	7.5	7.2	5.8
	9	5.8	7.4	5.2	7.2	9.0	7.6	7.0	10.6
	10	5.9	8.1	4.3	7.0	7.8	8.3	6.5	6.3
	11	6.2	7.0	6.3	6.9	10.2	7.6	7.2	5.6
	12	8.3	6.3	5.5	6.9	10.0	8.8	7.5	6.4
	13	7.7	8.5	6.8	7.3	10.2	10.1	8.9	7.0
	14	8.3	7.4	7.5	6.3	10.5	8.1	14.2	9.9
	15	7.8	6.3	5.8	9.3	9.2	14.6	8.2	5.6
	16	10.1	8.9	6.1	5.4	9.7	9.4	10.5	5.1
	17	6.8	10.9	4.4	6.2	9.5	9.2	6.5	6.6
	18	7.8	8.9	4.5	5.5	10.8	7.9	6.8	7.0
	19	8.2	12.0	4.4	5.7	12.2	9.7	8.7	7.0
	20	9.2	8.7	4.5	8.0	8.5	7.5	7.5	5.2

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
WITH HT. COMP. : $\bar{a}_1$	7.15±.34
STANDARD MODE : $\bar{a}_2$	8.87±.56
SHORT RANGE : $\bar{b}_1$	9.05±.45
LONG RANGE : $\bar{b}_2$	6.97±.44
CREW 1: $\bar{c}_1$	8.13±.51
CREW 2: $\bar{c}_2$	7.89±.49

DATA & RESULTS
OF
L-19 DEC. 56
FLIGHTS

ANALYSIS OF VARIANCE TABLE

FACTORS	S S COMPONENTS	TOTALS	DEGREES OF FREEDOM (D.F.)	MSS =SS/D.F.	F_c =MSS/R.E.MSS
MODIFICATION (A)	117.31		1	117.31	42.50 **
RANGE (B)	171.81		1	171.81	62.25 **
RADAR-CREW (C)	2.26		1	2.26	.82
MOD. X RANGE (AB)	1.41		1	1.41	.51
MOD. X RADAR-CREW (AC)	33.67		1	33.67	12.20 **
RANGE X RADAR-CREW (BC)	7.66		1	7.66	2.78
MOD X RANGE X RADAR-CREW (ABC)	8.19		1	8.19	2.97
TOTAL TREATMENT		342.31	7	48.90	17.72 **
REPLICATE EFFECT		91.47	19	4.81	1.74 *
RESIDUAL ERROR(R.E.), BY SUBTRACTION		366.63	133	2.76	1.00
CORRECTED SS = GSS - CT		800.41	159	5.03	

*SIGNIFICANT AT 5% LEVEL **SIGNIFICANT AT 1% LEVEL

AVERAGE ACQUISITION TIME (SEC)

	WITH HT COMP. (a_1) — SHORT RANGE (b_1)	WITH HT COMP. (a_1) — LONG RANGE (b_2)	STANDARD MODE (a_2) — SHORT RANGE (b_1)	STANDARD MODE (a_2) — LONG RANGE (b_2)
RADAR-CREW 1 (c_1)	7.76±.53	5.86±.47	10.13±.74	8.76±1.24
	S=1.13	S=1.01	S=1.59	S=2.66
RADAR-CREW 2 (c_2)	8.43±.73	6.56±.45	9.86±1.23	6.71±.64
	S=1.57	S=0.97	S=2.62	S=1.37

NOTE: SHORT RANGE < 20K YDS
LONG RANGE > 20K YDS

INTERACTION TABLE
MOD. X RADAR CREW

	AVERAGE ACQUISITION TIME (SEC) — WITH HT. COMP.	AVERAGE ACQUISITION TIME (SEC) — STANDARD MODE	AVERAGE IMPROVEMENT (SEC)
RADAR-CREW 1	6.81	9.45	2.64
RADAR-CREW 2	7.49	8.29	0.80
DIFFERENCE	-0.68	1.16	1.84

DATA SET I

		ACQUISITION TIME (SEC)							
		HEIGHT COMPARATOR: a_1				STANDARD MODE: a_2			
		SHORT RANGE: b_1		LONG RANGE: b_2		SHORT RANGE: b_1		LONG RANGE: b_2	
		RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2
TREATMENT		(1)	c	b	bc	a	ac	ab	abc
REPLICATE	1	7.5	16.0	10.8	9.9	10.8	8.2	10.8	8.5
	2	9.2	8.0	6.7	10.5	11.0	10.5	9.8	9.2
	3	7.8	12.0	8.2	9.5	11.0	9.4	7.8	9.7
	4	7.2	15.4	6.5	7.5	11.0	9.5	10.0	11.0
	5	7.9	9.7	8.5	8.0	9.5	9.2	10.0	8.2
	6	16.5	10.2	6.3	9.2	10.0	12.2	7.7	8.2
	7	7.8	11.9	7.8	7.4	11.0	11.0	12.8	7.2
	8	9.5	9.9	6.2	6.8	10.0	8.4	8.2	9.4
	9	7.5	9.2	7.4	7.0	12.0	11.4	9.1	7.5
	10	7.7	12.5	7.5	12.7	13.0	10.9	11.5	7.0
	11	10.0	12.0	5.7	8.0	9.8	8.6	13.0	7.9
	12	7.5	10.0	6.8	9.0	11.0	8.3	8.2	7.2
	13	8.0	8.9	5.2	6.5	9.1	8.3	6.8	7.4
	14	14.0	10.9	6.5	7.0	7.5	10.4	7.1	10.9

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC		
WITH HT COMP.	$\bar{a}_1$	9.00±.68
STANDARD: MODE	$\bar{a}_2$	9.55±.44
SHORT RANGE:	$\bar{b}_1$	10.13±.57
LONG RANGE:	$\bar{b}_2$	8.42±.48
CREW 1:	$\bar{c}_1$	9.06±.60
CREW 2:	$\bar{c}_2$	9.49±.54

DATA & RESULTS
OF
L-19 APR. 57
FLIGHTS

ANALYSIS OF VARIANCE TABLE

FACTORS		SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c= MSS/R.E.MSS
MODIFICATION	(A)	8.75		1	8.75	2.72
RANGE	(B)	83.15		1	83.15	25.82 **
RADAR-CREW	(C)	4.93		1	4.93	1.53
MOD X RANGE	(AB)	10.75		1	10.75	3.34
MOD X RADAR-CREW	(AC)	45.39		1	45.39	14.10 **
RANGE X RADAR-CREW	(BC)	1.44		1	1.44	.45
MOD X RANGE X RADAR-CREW	(ABC)	.38		1	.38	.12
TOTAL TREATMENT			154.76	7	22.11	6.87
REPLICATE EFFECT			61.63	13	4.74	1.47
RESIDUAL ERROR			293.39	91	3.22	1.00
CORRECTED SS=GSS-CT			509.78	111	4.59	

*SIGNIFICANT AT 5% LEVEL **SIGNIFICANT AT 1% LEVEL

TABLE OF AVERAGES (SEC)

		WITH HT. COMP. (a_1)		STANDARD MODE (a_2)	
		(b_1) SHORT RANGE	(b_2) LONG RANGE	(b_1) SHORT RANGE	(b_2) LONG RANGE
RADAR-CREW 1 (c_1)		9.15±1.59	7.15±.81	10.43±.76	9.49±1.15
		S=2.76	S=1.40	S=1.32	S=2.00
RADAR-CREW 2 (c_2)		11.19±1.34	8.50±1.00	9.74±.77	8.52±.77
		S=2.32	S=1.74	S=1.33	S=1.33

INTERACTION TABLE: MOD X RADAR CREW

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)
	WITH HT. COMP.	STANDARD MODE	
RADAR-CREW 1	8.15	9.96	1.81
RADAR-CREW 2	9.85	9.13	-.72
DIFFERENCE	-1.70	.83	2.53

DATA SET 2

	ACQUISITION TIME (SEC)							
	HEIGHT COMPARATOR: a_1				STANDARD MODE: a_2			
	SHORT RANGE: b_1		LONG RANGE: b_2		SHORT RANGE: b_1		LONG RANGE: b_2	
	RADAR CREW 1: C_1	RADAR CREW 2: C_2	RADAR CREW 1: C_1	RADAR CREW 2: C_2	RADAR CREW 1: C_1	RADAR CREW 2: C_2	RADAR CREW 1: C_1	RADAR CREW 2: C_2
TREATMENT	(1)	c	b	bc	a	ac	ab	abc
1	12.1	13.0	8.8	12.0	14.8	12.8	8.6	9.8
2	17.1	13.7	7.8	11.2	11.2	16.4	5.2	9.7
3	11.0	12.4	8.8	8.7	10.0	12.3	8.5	9.2
4	8.5	11.0	8.5	8.3	11.8	11.4	8.6	9.4
5	8.5	10.9	7.9	9.3	9.0	13.4	8.5	10.8
6	13.8	10.6	8.5	12.6	15.8	13.0	7.5	10.1
7	9.5	12.4	8.0	10.2	13.2	11.5	6.8	10.1

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
WITH HT. COMP.: $\bar{a}_1$	10.54±.62
STANDARD MODE: $\bar{a}_2$	10.69±.71
SHORT RANGE: $\bar{b}_1$	12.18±.59
LONG RANGE: $\bar{b}_2$	9.05±.41
CREW 1: $\bar{c}_1$	9.94±.76
CREW 2: $\bar{c}_2$	11.29±.49

DATA & RESULTS
OF
L-19 MAY 57
FLIGHTS

INTERACTION TIME: MOD X RANGE

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT
	WITH HT COMP. (a_1)	STANDARD MODE (a_2)	
SHORT RANGE (b_1)	11.75	12.62	.87
LONG RANGE (b_2)	9.33	8.77	-.56
DIFFERENCE BETWEEN RANGES	-2.42	-3.85	

INTERACTION TABLE: RANGE X RADAR CREW

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)
	SHORT RANGE (b_1)	LONG RANGE (b_2)	
RADAR-CREW 1	11.88	9.97	1.92
RADAR-CREW 2	12.49	10.10	2.39
DIFFERENCE BETWEEN CREWS	.61	.14	

ANALYSIS OF VARIANCE TABLE

FACTORS	SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c = MSS/R.E.MSS
MODIFICATION (A)	.33		1	.33	.12
RANGE (B)	137.34		1	137.34	49.58 **
RADAR CREW (C)	25.65		1	25.65	9.26 **
MOD X RANGE (AB)	7.07		1	7.07	2.55
MOD X RADAR CREW (AC)	.15		1	.15	.05
RANGE X RADAR CREW (BC)	7.80		1	7.80	2.82
MOD X RANGE X RADAR CREW (ABC)	.00		1	.00	.00
TOTAL TREATMENT		178.34	7	25.48	9.20 **
REPLICATE EFFECT		34.67	6	5.78	2.09
RESIDUAL ERROR		116.27	42	2.77	1.00
CORRECTED SS = GSS - CT		329.28	55	5.99	

* SIGNIFICANT AT 5% LEVEL
** SIGNIFICANT AT 1% LEVEL

AVERAGE ACQUISITION TIME (SEC)

	WITH HT COMP. (a_1)		STANDARD MODE (a_2)	
	SHORT RANGE (b_1)	LONG RANGE (b_2)	SHORT RANGE (b_1)	LONG RANGE (b_2)
RADAR-CREW 1 (c_1)	11.50±2.90	8.33±.39	12.26±2.29	7.67±1.19
	S = 3.14	S = .42	S = 2.48	S = 1.29
RADAR-CREW 2 (c_2)	12.00±1.09	10.33±1.54	12.97±1.56	9.87±.49
	S = 1.18	S = 1.66	S = 1.69	S = .53

INTERACTION TABLE: MOD X CREW

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)
	WITH HT. COMP. (a_1)	STANDARD MODE (a_2)	
RADAR-CREW 1	9.91	9.96	.05
RADAR-CREW 2	11.16	11.42	.26
DIFFERENCE BETWEEN CREWS	1.25	1.46	

DATA SET 3

	TABLE OF ACQUISITION TIME DATA — HEIGHT COMPARATOR: a_1							
	SHORT RANGE: b_1				LONG RANGE: b_2			
	RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2
TREATMENT	(1)	d	c	cd	b	bd	bc	bcd
REPLICATE 1	7.7	7.5	9.2	16.0	6.5	10.8	6.2	9.9
2	8.4	9.2	8.7	8.0	6.1	6.7	5.7	10.5
3	9.0	7.8	9.0	12.0	7.3	8.2	5.2	9.5
4	8.5	7.2	6.1	15.4	6.7	6.5	6.6	7.5
5	.1	7.9	7.2	9.7	5.8	8.5	6.8	8.0
6	7.4	16.5	8.2	10.2	6.5	6.3	6.5	9.2
7	5.8	7.8	7.4	11.9	5.2	7.8	7.2	7.4
8	5.9	9.5	8.1	9.9	4.3	6.2	7.0	6.8
9	6.2	7.5	7.0	9.2	6.3	7.4	6.9	7.0
10	8.3	7.7	6.3	12.5	5.5	7.5	6.9	12.7
11	7.7	10.0	8.5	12.0	6.8	5.7	7.3	8.0
12	7.8	7.5	6.3	10.0	5.8	6.8	9.3	9.0
13	10.1	8.8	8.9	8.9	6.1	5.2	5.4	6.5
14	9.2	14.0	8.7	10.9	4.5	6.5	8.0	7.0

	ACQUISITION TIME (sec) — STANDARD MODE: a_2							
	SHORT RANGE: b_1				LONG RANGE: b_2			
	RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2	DEC 56: d_1	APR 57: d_2
TREATMENT	a	ad	ac	acd	ab	abd	abc	abcd
REPLICATE 1	12.5	10.8	15.6	8.2	8.6	10.8	7.2	8.5
2	11.8	11.0	13.3	10.5	7.5	9.8	6.9	9.2
3	9.8	11.8	8.6	9.4	17.8	7.8	6.3	9.7
4	9.0	11.0	7.8	9.5	10.2	10.0	5.9	11.0
5	7.5	9.5	11.8	9.2	7.3	10.0	5.5	8.2
6	9.5	10.0	7.5	12.2	7.2	7.7	5.8	8.2
7	9.0	11.0	7.6	11.0	7.0	12.8	10.6	7.2
8	7.8	10.0	8.3	8.4	6.5	8.2	6.3	9.4
9	10.2	12.0	7.6	11.4	7.2	9.1	5.6	7.5
10	10.0	13.0	8.8	10.9	7.5	11.5	6.4	7.0
11	10.2	9.8	10.1	8.6	8.9	13.0	7.0	7.9
12	9.2	11.0	14.6	8.3	8.2	8.2	5.6	7.2
13	9.7	9.1	9.4	8.3	10.5	6.8	5.1	7.4
14	8.5	7.5	7.5	10.4	7.5	7.1	5.2	10.9

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC

HT. COMP. : $\bar{a}_1$	8.06 ±.42
STANDARD MODE : $\bar{a}_2$	9.09 ±.40
SHORT RANGE : $\bar{b}_1$	9.45 ±.40
LONG RANGE : $\bar{b}_2$	7.68 ±.37
RADAR-CREW 1 : $\bar{c}_1$	8.54 ±.42
RADAR-CREW 2 : $\bar{c}_2$	8.58 ±.42
DEC 56 : $\bar{d}_1$	7.88 ±.60
APR 57 : $\bar{d}_2$	9.28 ±.40

DATA & RESULTS OF DEC 56 & APR 57 L-19 FLIGHTS

ANALYSIS OF VARIANCE TABLE

FACTORS	S S COMPONENTS	TOTALS	D.F.	MSS= SS/D.F.	F_c= MSS/R.E.MSS
MODIFICATION (A)	61.216		1	61.216	20.25 **
RANGE (B)	176.258		1	176.258	58.31 **
RADAR-CREW (C)	0.213		1	0.213	0.07
TIME LAPSE (D)	112.720		1	112.720	37.29 **
MOD. X RANGE (AB)	0.766		1	0.766	.25
MOD. X RADAR-CREW (AC)	56.100		1	56.100	18.56 **
MOD. X TIME LAPSE (AD)	13.260		1	13.260	4.39 *
RANGE X RADAR-CREW (BC)	5.883		1	5.883	1.95
RANGE X TIME LAPSE (BD)	0.145		1	0.145	.05
RADAR-CREW X TIME LAPSE (CD)	7.179		1	7.179	2.37
MOD. X RANGE X RADAR-CREW (ABC)	6.895		1	6.895	2.28
MOD. X RANGE X TIME LAPSE (ABD)	14.150		1	14.150	4.68 *
MOD. X RADAR-CR. X TIME LAPSE (ACD)	4.153		1	4.153	1.37
RANGE X RADAR-CREW X TIME LAPSE (BCD)	0.530		1	0.530	.18
MOD. X RANGE X RADAR CREW X TIME (ABCD)	12.211		1	12.211	4.04 *
TOTAL TREATMENT		471.679	15	31.445	10.40 **
REPLICATE EFFECT		66.811	13	5.147	1.70 *
RESIDUAL ERROR		589.489	195	3.023	1.00
CORRECTED SS = GSS - CT		1128.079	223	5.050	

* SIGNIFICANT AT 5% LEVEL ** SIGNIFICANT AT 1% LEVEL

TABLE OF AVERAGES

	HEIGHT COMP. a_1				STANDARD MODE a_2			
	SHORT RANGE b_1		LONG RANGE b_2		SHORT RANGE b_1		LONG RANGE b_2	
	RADAR-CREW 1 c_1	RADAR-CREW 2 c_2	RADAR-CREW 1 c_1	RADAR-CREW 2 c_2	RADAR-CREW 1 c_1	RADAR-CREW 2 c_2	RADAR-CREW 1 c_1	RADAR-CREW 2 c_2
DEC 56 d_1	7.79±.78 σ=1.26	7.83±.63 σ=1.00	5.86±.49 σ=.85	6.79±.64 σ=1.05	9.82±.78 σ=1.36	9.08±6.60 σ=2.77	8.65±1.56 σ=2.68	6.30±.80 σ=1.38
APR 57 d_2	9.15±1.65 σ=2.88	11.18±1.34 σ=2.32	7.15±.81 σ=1.40	8.50±1.01 σ=1.74	10.18±.78 σ=1.32	9.78±.77 σ=1.33	9.08±1.16 σ=2.00	8.51±.77 σ=1.33

AVERAGE ACQUISITION TIME (sec)

	HEIGHT COMP. a_1		STANDARD MODE a_2	
	DEC 56 d_1	APR 57 d_2	DEC 56 d_1	APR 57 d_2
RDR-CREW 1 c_1	6.88	8.15	9.14	9.98
RDR-CREW 2 c_2	7.31	9.85	8.12	9.13

INTERACTION TABLE MOD X RADAR CREW

	HT. COMP. a_1	STANDARD MODE a_2	AVG. IMPROVEMENT (SEC)
RADAR-CREW 1 c_1	7.51	9.56	2.01
RADAR-CREW 2 c_2	8.58	8.62	.04
DIFFERENCE	-1.07	0.94	

DATA SET 4

TABLE OF ACQUISITION TIME DATA

		ACQUISITION TIME (sec) — HEIGHT COMPARATOR: a_1							
		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2
TREATMENT		(1)	d	c	cd	b	bd	bc	bcd
REPLICATE	1	8.5	12.1	12.0	13.0	5.8	8.8	9.9	12.0
	2	8.6	11.0	15.4	11.6	8.5	7.8	9.2	10.6
	3	9.2	8.5	11.9	11.0	7.8	8.8	11.5	11.2
	4	16.5	8.5	12.5	10.9	7.2	9.5	7.0	8.3
	5	7.8	8.5	12.0	10.6	10.0	8.5	12.7	9.3
	6	7.5	9.5	10.9	12.4	8.0	8.0	10.0	10.2

		ACQUISITION TIME (Sec) — STANDARD: a_2							
		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2	NO-MANEUVER: d_1	ALTITUDE MANEUVER: d_2
TREATMENT		a	ad	ac	acd	ab	abd	abc	abcd
REPLICATE	1	11.0	11.2	8.2	12.8	11.0	8.8	10.0	9.8
	2	11.0	8.6	8.5	11.4	10.0	5.2	10.7	9.7
	3	11.0	10.0	10.5	13.4	8.2	8.5	7.2	9.2
	4	13.0	11.8	9.2	9.9	11.5	8.6	9.4	9.4
	5	9.8	9.0	11.4	13.0	6.8	7.5	7.4	10.1
	6	7.5	15.8	10.4	11.5	7.1	6.8	8.3	10.1

AVERAGE ACQUISITION TIME (sec)

	WITH HT. COMP.		STANDARD MODE	
	NO MANEUVER	ALTITUDE MANEUVER	NO MANEUVER	ALTITUDE MANEUVER
RADAR-CREW 1	8.78	9.13	9.83	8.32
RADAR-CREW 2	11.26	10.93	9.27	10.89

AVERAGE ACQUISITION TIME (sec)

	WITH HT. COMP. a_1				STANDARD MODE a_2			
	SHORT RANGE b_1		LONG RANGE b_2		SHORT RANGE b_1		LONG RANGE b_2	
	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2
NO MANEUVER d_1	9.60 ± 3.56 s=3.39	12.45 ± 1.62 s=1.54	7.88 ± 1.46 s=1.39	10.05 ± 2.06 s=1.96	10.56 ± 1.90 s=1.81	9.70 ± 1.32 s=1.26	9.10 ± 2.11 s=2.01	8.83 ± 1.49 s=1.42
ALTITUDE MANEUVERING d_2	9.60 ± 1.62 s=1.54	11.58 ± .98 s=.94	8.57 ± .65 s=.62	10.27 ± 1.40 s=1.33	11.07 ± 2.75 s=2.62	12.00 ± 1.38 s=1.31	7.57 ± 1.46 s=1.38	9.77 ± .38 s=3.64

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
WITH HT. COMP.: a_1	10.02±0.66
STANDARD MODE: a_2	9.82±0.61
SHORT RANGE: b_1	10.84±.69
LONG RANGE: b_2	9.01±0.49
CREW 1: c_1	9.26±0.67
CREW 2: c_2	10.58±0.53
NO-MANEUVER: d_1	9.78±2.94
ALTITUDE MANEUVER: d_2	10.06±3.00

DATA & RESULTS OF APR 57 & MAY 57 L-19 FLIGHTS

INTERACTION TABLE: MOD. X CREW

	AVERAGE ACQUISITION TIME (sec)		AVERAGE IMPROVEMENT (sec)
	WITH HT. COMP. a_1	STANDARD MODE a_2	
RADAR-CREW 1 c_1	8.95	9.57	0.62
RADAR-CREW 2 c_2	11.09	10.08	-1.01
DIFFEDFNCE	-2.14	-0.51	

ANALYSIS OF VARIANCE TABLE

FACTORS		SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c= MSS/R.E.MSS
MODIFICATION	(A)	1.00		1	1.00	.33
RANGE	(B)	81.40		1	81.40	26.86**
RADAR-CREW	(C)	41.34		1	41.34	13.64**
MANEUVER	(D)	1.81		1	1.81	.60
MOD. X RANGE	(AB)	.81		1	.81	.27
MOD. X RADAR-CREW	(AC)	16.17		1	16.17	5.34*
MOD. X MANEUVER	(AD)	1.71		1	1.71	.56
RANGE X RADAR-CREW	(BC)	.38		1	.38	.12
RANGE X MANEUVER	(BD)	1.08		1	1.08	.36
RADAR CREW X MANEUVER	(CD)	3.08		1	3.08	1.02
MOD. X RANGE X RADAR-CR	(ABC)	2.53		1	2.53	.83
MOD. X RANGE X MANEUVER	(ABD)	10.27		1	10.27	3.39
MOD. X RADAR-CREW X MANEUVER	(ACD)	11.48		1	11.48	3.79
RANGE X RAD. CR. X MANEU.	(BCD)	.40		1	.40	.13
MOD. X RANGE X RAD. CR. X MANEUVER	(ABCD)	.02		1	.02	.01
TOTAL TREATMENT			173.47	15	11.53	3.81 **
REPLICATE EFFECT			6.36	5	1.27	.42
RESIDUAL ERROR			227.22	75	3.03	1.00
CORRECTED SS = 6SS - CT			407.05	95	4.28	

* SIGNIFICANT AT 5% LEVEL
** SIGNIFICANT AT 1% LEVEL

DATA SET 5

		ACQUISITION TIME (SEC)							
		HEIGHT COMPARATOR: a_1				STANDARD MODE: a_2			
		SHORT RANGE: b_1		LONG RANGE: b_2		SHORT RANGE: b_1		LONG RANGE: b_2	
		L-19: c_1	P 2 : c_2	L-19: c_1	P 2 : c_2	L-19: c_1	P 2: c_2	L-19: c_1	P.2: c_2
TREATMENT		(1)	c	b	bc	a	ac	ab	abc
REPLICATE	1	8.5	12.2	6.8	12.5	14.0	6.5	7.8	6.0
	2	7.7	8.2	6.5	6.8	12.5	6.0	8.6	11.8
	3	8.4	12.2	6.1	8.5	11.8	5.8	7.5	9.2
	4	9.0	8.3	7.3	5.2	9.8	8.5	17.0	27.5
	5	6.5	7.5	6.2	7.1	10.8	12.2	10.0	11.3
	6	8.5	10.8	6.7	16.2	9.0	11.8	10.2	9.6
	7	7.1	8.3	5.8	5.8	7.5	9.5	7.3	18.8
	8	7.4	6.7	6.5	5.1	9.5	16.8	7.2	23.5
	9	5.8	13.0	5.2	11.9	9.0	10.5	7.0	10.2
	10	5.9	11.2	4.3	7.5	7.8	13.0	6.5	9.5
	11	6.2	9.2	6.3	11.0	10.2	8.0	7.2	10.9
	12	8.3	14.1	5.5	10.7	10.0	19.5	7.5	11.1
	13	7.7	8.0	6.8	10.9	10.2	24.8	8.9	9.2
	14	8.3	7.8	7.5	10.9	10.5	24.2	14.2	10.0
	15	7.8	8.3	5.8	18.7	9.2	11.3	8.2	8.5
	16	10.1	5.2	6.1	11.2	9.7	22.0	10.5	12.2
	17	6.8	8.0	4.4	5.8	9.5	22.8	6.5	13.1
	18	7.8	7.2	4.5	4.0	10.8	15.8	6.8	21.5
	19	8.2	7.2	4.4	7.3	12.2	14.6	8.7	12.8
	20	9.2	7.1	4.5	6.4	8.5	28.1	7.5	14.1

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
WITH HT. COMP. : $\bar{a}_1$	7.96±.60
STANDARD MODE : $\bar{a}_2$	11.63±1.15
SHORT RANGE: $\bar{b}_1$	10.38±1.01
LONG RANGE: $\bar{b}_2$	9.21±.99
L-19 : $\bar{c}_1$	8.13±.52
PLANE #2 $\bar{c}_2$	11.46±1.22

DATA & RESULTS
OF
DEC 56 L-19 &
OCT 56 FLIGHTS
OF PLANE #2

ANALYSIS OF VARIANCE TABLE

FACTORS	SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c= MSS/R.E.MSS
MODIFICATION (A)	539.12		1	539.12	37.23 **
RANGE (B)	54.41		1	54.41	3.76
PLANE (C)	443.89		1	443.89	30.65 **
MOD X RANGE (AB)	3.39		1	3.39	0.23
MOD X PLANE (AC)	43.37		1	43.37	2.99
RANGE X PLANE (BC)	8.79		1	8.79	0.61
MOD X RANGE X PLANE (ABC)	12.38		1	12.38	0.86
TOTAL TREATMENT		1106.35	7	157.91	10.91*
REPLICATE EFFECT		166.75	19	8.78	0.61
RESIDUAL ERROR		1926.10	133	14.48	1.00
CORRECTED SS = GSS - CT		3199.20	159	20.01	

* SIGNIFICANT AT 5% LEVEL
** SIGNIFICANT AT 1% LEVEL

TABLE OF AVERAGES (SEC)

	WITH HT. COMP. (a_1)		STANDARD MODE (a_2)	
	SHORT RANGE (b_1)	LONG RANGE (b_2)	SHORT RANGE (b_1)	LONG RANGE (b_2)
L-19 (C_1)	7.76±.53	5.86±.47	10.13±.74	8.76±1.24
	S=1.13	S=1.01	S=1.58	S=2.66
PLANE #2 (C_2)	8.03±1.12	9.18±1.80	14.50±3.22	13.04±2.58
	S=2.39	S=3.84	S=6.88	S=5.52

INTERACTION TABLE MOD X PLANE

	AVERAGE ACQUISITION TIME (SEC)		AVERAGE IMPROVEMENT (SEC)
	WITH HT. COMP. (a_1)	STANDARD MODE (a_2)	
L-19 (C_1)	6.81 ±.46	9.45 ±.73	2.64
PLANE #2 (C_2)	9.11 ±1.01	13.82 ±1.98	4.71
DIFFERENCE	2.30	4.37	

DATA SET 6

TABLE OF ACQUISITION TIME DATA

STANDARD MODE: a_2

		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2
TREATMENT		a	ad	ac	acd	ab	abd	abc	abcd
REPLICATE	1	10.8	28.0	8.2	31.0	10.8	21.0	8.5	21.0
	2	11.0	19.0	10.5	27.0	9.8	28.2	9.2	27.0
	3	11.0	15.0	9.4	15.0	7.8	11.5	8.7	13.0
	4	11.0	15.5	9.5	30.0	10.0	14.5	11.0	11.0
	5	8.5	14.0	9.2	17.0	10.0	14.2	8.2	14.2
	6	10.0	15.2	12.2	17.0	7.7	15.1	8.2	14.0
	7	11.0	16.5	11.0	27.0	12.0	13.8	7.2	11.0
	8	10.8	20.5	8.4	15.0	8.2	13.2	9.4	20.3
	9	12.0	20.5	11.4	20.0	8.1	17.7	7.5	19.9
	10	12.8	10.5	10.9	13.0	11.5	18.2	7.0	17.0
	11	9.8	26.0	8.6	22.0	13.0	26.0	7.9	20.9
	12	11.0	15.0	8.3	20.9	8.2	19.8	7.2	19.0
	13	9.1	27.2	8.3	13.0	6.8	19.0	7.4	18.5
	14	7.5	17.2	10.5	14.5	7.1	24.2	10.9	15.9

ACQUISITION TIME (sec)

HEIGHT COMPARATOR: a_1

		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2	L-19: d_1	FL.3: d_2
TREATMENT		1	d	c	cd	b	bd	bc	bcd
REPLICATE	1	7.5	17.2	16.0	15.0	10.8	15.0	8.9	12.0
	2	9.2	18.6	8.0	12.0	6.7	16.0	10.5	12.0
	3	7.8	24.2	12.0	14.0	8.2	13.2	9.5	10.0
	4	7.2	21.0	15.4	18.0	6.5	11.0	7.5	14.2
	5	7.9	21.0	9.7	11.0	8.5	12.0	8.0	9.5
	6	16.5	16.0	18.2	15.0	8.3	15.1	9.2	18.8
	7	7.8	15.2	11.9	17.8	7.0	12.5	7.4	10.0
	8	9.5	15.0	9.0	17.0	6.2	16.2	6.8	9.4
	9	7.5	14.7	8.2	11.2	7.0	9.6	7.0	12.8
	10	7.7	12.5	12.5	11.5	7.5	10.7	12.7	10.0
	11	10.0	23.1	12.0	14.8	5.7	12.9	8.0	10.5
	12	7.5	10.2	10.0	12.0	6.8	18.2	9.0	17.0
	13	8.0	9.0	8.9	14.0	5.2	12.8	6.5	9.0
	14	14.0	12.8	10.0	20.0	6.5	8.2	7.0	12.7

DATA & RESULTS OF APR 56 L-19 & FL.3 FLIGHTS

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC

WITH HT. COMP. : $\bar{a}_1$	[illegible]
STANDARD MODE : $\bar{a}_2$	[illegible]
SHORT RANGE : $\bar{b}_1$	[illegible]
LONG RANGE : $\bar{b}_2$	[illegible]
RADAR CREW 1 : $\bar{c}_1$	[illegible]
RADAR CREW 2 : $\bar{c}_2$	[illegible]
L-19 : $\bar{d}_1$	[illegible]
FL.3 : $\bar{d}_2$	[illegible]

INTERACTION TABLE MOD X PLANE
a. FOR CREW 1

	HT. COMP. a_1 (sec)	STANDARD MODE a_2 (sec)	IMPROVEMENT (sec)
L-19 d_1	8.15 (1.92)	9.99 (1.66)	1.84
FL.3 d_2	14.80 (21.50)	16.79 (21.88)	3.99
DIFFERENCE	-6.65	-6.80	2.15

INTERACTION TABLE MOD X PLANE
b. BOTH CREWS

	HT. COMP. a_1 (sec)	STANDARD MODE a_2 (sec)	IMPROVEMENT (sec)
L-19 d_1	9.00 (1.67)	9.56 (2.44)	0.56
FL.3 d_2	14.04 (2.98)	18.75 (21.40)	4.71
DIFFERENCE	5.04	9.19	X

ANALYSIS OF VARIANCE TABLE

FACTORS		SS COMPONENTS	TOTALS	D.F.	MS = SS/D.F.	F_c = MS/M.S. RES
MODIFICATION	(A)	300.25		1	300.25	30.60**
RANGE	(B)	291.70		1	291.70	22.70**
RADAR-CREW	(C)	2.06		1	2.06	0.19
PLANE	(D)	2006.05		1	2035.05	187.20**
MOD. X RANGE	(AB)	21.07		1	21.07	1.90
MOD. X RADAR-CREW	(AC)	4.27		1	4.27	0.41
MOD. X PLANE	(AD)	250.01		1	250.01	22.71**
RANGE X RADAR-CREW	(BC)	3.20		1	3.20	0.31
RANGE X PLANE	(BD)	7.05		1	7.05	0.66
PLANE X RADAR-CREW	(CD)	20.05		1	20.05	1.88
MOD. X RANGE X RADAR-CR.	(ABC)	5.25		1	5.25	.50
MOD. X RANGE X PLANE	(ABD)	.00		1	.00	.00
MOD X RADAR-CREW X PLANE	(ACD)	56.30		1	56.30	5.31*
RANGE X RADAR-CREW X PLANE	(BCD)	0.01		1	0.01	.00
MOD X RANGE X RADAR-CREW X PLANE	(ABCD)	0.00		1	0.00	0.00
TOTAL EFFECT			3826.31	15	255.09	24.15**
REPLICATE EFFECT			300.25	13	23.13	2.18
RESIDUAL EFFECT			2065.21	195	10.61	1.00
CORRECTED SUM OF SQUARES - CT			6192.06	223	27.77	X

* SIGNIFICANT AT 5% LEVEL   ** SIGNIFICANT AT 1% LEVEL

AVERAGE ACQUISITION TIME (sec)

	WITH HT. COMP. a_1		STANDARD MODE a_2	
	L-19 d_1	PLANE 3 d_2	L-19 d_1	PLANE 3 d_2
RADAR CREW 1 c_1	8.16	14.00	9.06	9.13
RADAR CREW 2 c_2	8.08	15.20	10.79	18.70

AVERAGE ACQUISITION TIME (sec)

		HEIGHT COMP. a_1		STANDARD MODE a_2	
		SHORT RANGE b_1	LONG RANGE b_2	SHORT RANGE b_1	LONG RANGE b_2
RADAR-CREW 1 c_1	L-19 d_1	[illegible]	[illegible]	[illegible]	[illegible]
	d_2	[illegible]	[illegible]	[illegible]	[illegible]
RADAR-CREW 2 c_2	L-19 d_1	[illegible]	[illegible]	[illegible]	[illegible]
	d_2	[illegible]	[illegible]	[illegible]	[illegible]

DATA SET 7

	ACQUISITION TIME (SEC)							
	PLANE #2 : a_1				PLANE #3 : a_2			
	SHORT RANGE : b_1		LONG RANGE : b_2		SHORT RANGE : b_1		LONG RANGE : b_2	
	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2	RADAR-CREW 1: c_1	RADAR-CREW 2: c_2
TREATMENT	(1)	c	b	bc	a	ac	ab	abc
1	6.5	19.6	6.0	16.7	28.0	17.0	15.1	12.0
2	6.0	11.1	11.8	10.6	19.0	27.0	15.8	11.0
3	5.8	17.3	9.2	14.0	15.0	28.0	28.0	12.0
4	8.5	22.5	27.5	30.5	15.0	15.0	12.0	14.2
5	12.2	28.4	11.3	14.7	25.8	30.0	10.8	15.0
6	11.8	26.5	9.6	13.8	15.5	17.0	28.2	11.0
7	9.5	20.1	23.5	12.2	14.0	17.0	18.0	14.0
8	16.8	27.7	10.2	16.2	15.2	27.0	11.5	11.0
9	10.5	13.5	9.5	14.4	15.0	15.0	14.5	20.3
10	13.0	28.8	10.9	24.2	16.5	17.0	22.8	22.6
11	8.0	15.9	11.1	7.8	20.5	20.0	15.1	25.0
12	14.0	16.8	9.2	20.9	20.5	13.0	13.8	16.4
13	24.8	9.0	10.0	10.5	18.5	22.0	13.2	17.0
14	11.3	10.9	8.5	14.0	13.0	20.0	17.7	20.9
15	25.8	14.6	13.1	21.0	26.8	13.0	26.8	20.2
16	22.8	16.9	21.5	10.8	15.0	16.2	15.5	18.4
17	14.6	10.0	12.8	11.2	27.2	15.2	19.0	17.0
18	28.1	14.3	14.1	14.7	17.2	14.5	16.8	15.4

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
PLANE 2 : a_1	15.02±1.51
PLANE 3 : a_2	17.91±1.20
SHORT RANGE : b_1	17.44±1.48
LONG RANGE : b_2	15.50±1.21
RADAR CREW 1 : c_1	15.72±1.48
RADAR CREW 2 : c_2	17.21±1.32

DATA & RESULTS OF PLANE #2 & #3 FLIGHTS (STANDARD MODE)

ANALYSIS OF VARIANCE TABLE

FACTORS	SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c= MSS/R.E.MSS
PLANE (A)	300.16		1	300.16	9.00 **
RANGE (B)	135.53		1	135.53	4.06 *
RADAR-CREW (C)	70.06		1	70.06	2.10
PLANE X RANGE (AB)	0.39		1	0.39	0.01
PLANE X RADAR-CREW (AC)	130.53		1	130.53	3.91
RANGE X RADAR-CREW (BC)	19.58		1	19.58	0.59
PLANE X RANGE X RADAR-CREW (ABC)	0.02		1	0.02	0.001
TOTAL TREATMENT		665.87	7	95.12	2.85 **
REPLICATE EFFECT		408.05	17	24.00	0.72
RESIDUAL ERROR		3969.02	119	33.36	1.00
CORRECTED SS = GSS - CT		5045.14	143	35.28	1.06

*SIGNIFICANT AT 5% LEVEL ** SIGNIFICANT AT 1% LEVEL

TABLE OF AVERAGES (SEC)

	PLANE #2 (a_1)		PLANE #3 (a_2)	
	SHORT RANGE (b_1)	LONG RANGE (b_2)	SHORT RANGE (b_1)	LONG RANGE (b_2)
RADAR CREW 1 (c_1)	13.89±1.56	12.77±2.61	18.76±2.47	17.48±2.73
	S =7.04	S =5.06	S=4.87	S =5.40
RADAR CREW 2 (c_2)	17.99±3.22	15.46±2.79	19.11±2.71	16.30±3.81
	S =6.46	S =5.61	S =5.04	S=7.67

INTERACTION TABLE MOD X CREW

	AVERAGE ACQUISITION TIME (SEC) (a_1)	AVERAGE ACQUISITION TIME (SEC) (a_2)	AVERAGE IMPROVEMENT (SEC)
RADAR CREW 1	13.33	18.12	4.79
RADAR CREW 2	16.72	17.71	0.99
DIFFERENCE	3.39	-0.41	

DATA SET 8

		ACQUISITION TIME (SEC)							
		HEIGHT COMPARATOR: a_1				STANDARD MODE: a_2			
		SHORT RANGE: b_1		LONG RANGE: b_2		SHORT RANGE: b_1		LONG RANGE: b_2	
		PL.2: c_1	PL.3: c_2	PL.2: c_1	PL.3: c_2	PL.2: c_1	PL.3: c_2	PL.2: c_1	PL.3: c_2
TREATMENT		(1)	c	b	bc	a	ac	ab	abc
REPLICATE	1	5.7	21.0	10.7	11.0	12.2	28.0	9.6	15.1
	2	11.2	12.5	8.4	9.8	28.1	19.0	18.8	15.1
	3	8.5	22.5	5.2	13.0	11.8	15.0	12.2	18.2
	4	10.2	24.2	6.4	12.5	15.8	15.0	23.5	23.5
	5	13.0	19.0	10.9	11.0	16.8	25.8	21.5	14.2
	6	11.8	9.8	5.4	13.2	11.3	15.5	13.1	13.2
	7	14.1	12.8	12.8	16.0	19.5	14.0	9.2	23.5
	8	7.2	17.2	4.8	15.1	14.6	15.2	11.8	15.8
	9	9.0	10.2	5.9	9.8	8.0	15.0	10.2	18.0
	10	6.0	14.8	21.0	10.7	10.5	16.5	9.2	22.8
	11	6.2	17.0	7.4	14.2	5.8	20.5	10.9	10.8
	12	5.2	23.1	4.8	13.7	22.0	20.5	14.1	24.2
	13	7.5	21.0	4.8	10.2	24.2	18.5	8.5	21.0
	14	8.3	18.6	11.2	12.9	14.0	13.0	6.0	19.8
	15	12.5	23.8	4.6	26.2	24.8	26.8	12.8	28.2
	16	8.5	14.7	5.2	11.8	22.8	15.0	27.5	12.0
	17	6.7	13.8	16.2	15.0	6.0	27.2	11.1	23.5
	18	12.2	18.6	6.5	14.2	8.5	17.2	11.3	15.8

TABLE OF ACQUISITION TIME DATA

AVERAGE ACQUISITION TIMES BY LEVELS IN SEC	
WITH HT COMP. : $\bar{a}_1$	12.10±1.20
STANDARD MODE: $\bar{a}_2$	16.54±1.42
SHORT RANGE: $\bar{b}_1$	15.18±1.46
LONG RANGE: $\bar{b}_2$	13.45±1.40
PL.2 : $\bar{c}_1$	11.59±1.39
PL.3 : $\bar{c}_2$	17.05±1.18

DATA & RESULTS
OF
PLANE #2 & #3 FLIGHTS
(RADAR-CREW 1)

ANALYSIS OF VARIANCE TABLE

FACTORS		SS COMPONENTS	TOTALS	DEGREES OF FREEDOM	MSS= SS/D.F.	F_c= MSS/R.E.MSS
MODIFICATION	(A)	708.89		1	708.89	32.80 **
RANGE	(B)	107.30		1	107.30	4.97 *
PLANE	(C)	1074.20		1	1074.20	49.71 **
MOD X RANGE	(AB)	15.67		1	15.67	.73
MOD. X PLANE	(AC)	49.59		1	49.59	2.29
RANGE X PLANE	(BC)	6.38		1	6.38	.30
MOD. X RANGE X PLANE	(ABC)	62.81		1	62.81	2.91
TOTAL TREATMENT			2024.84	7	289.26	13.39
REPLICATE EFFECT			643.72	17	37.87	1.75
RESIDUAL ERROR			2622.98	121	21.68	1.00
CORRECTED SS = GSS - CT			5291.54	145	36.49	1.69

* SIGNIFICANT AT 5% LEVEL ** SIGNIFICANT AT 1% LEVEL

AVERAGE ACQUISITION TIME (SEC)

	WITH HT COMP. (a_1)		STANDARD MODE (a_2)	
	SHORT RANGE (b_1)	LONG RANGE (b_2)	SHORT RANGE (b_1)	LONG RANGE (b_2)
PLANE #2 (c_1)	9.10±1.39	8.46±2.28	15.37±3.40	13.41±2.84
	s=2.79	s=4.58	s=6.83	s=5.71
PLANE #3 (c_2)	17.48±2.28	13.35±1.85	18.76±2.47	18.59±2.45
	s=4.59	s=3.73	s=4.97	s=4.93

NOTE: SHORT RANGE < 25,000 YDS
LONG RANGE > 25,000 YDS

INTERACTION TABLE: MOD X PLANE

	ACQUISITION TIME (SEC)		
	WITH HT COMP. (a_1)	STANDARD (a_2)	IMPROVEMENT (SEC)
PLANE #2 (c_1)	8.78 ±1.27	14.39 ±2.13	5.61
PLANE #3 (c_2)	15.42 ±1.54	18.68 ±1.65	3.26
DIFFERENCE	6.64	4.29	-2.35

DATA SET 9

		ACQUISITION TIME (sec)							
		HEIGHT COMPARATOR: a_1							
		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 1: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2
TREATMENT		(1)	d	c	cd	b	bd	bc	bcd
REPLICATE	1	12.8	15.2	17.6	12.7	26.2	10.5	13.6	16.2
	2	16.4	10.1	17.0	13.7	14.2	9.2	24.0	10.9
	3	24.2	20.0	11.0	10.2	10.7	8.6	9.4	17.7
	4	12.5	10.5	15.0	14.3	11.2	8.1	9.4	10.2
	5	13.8	10.0	15.0	13.7	13.7	18.5	10.5	13.3
	6	18.6	11.0	12.8	11.8	11.2	7.5	11.0	12.8

TABLE OF ACQUISITION TIME DATA

		ACQUISITION TIME (sec)							
		STANDARD MODE: a_2							
		SHORT RANGE: b_1				LONG RANGE: b_2			
		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2		RADAR-CREW 1: c_1		RADAR-CREW 2: c_2	
		TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2	TANGENTIAL COURSE: d_1	RADIAL COURSE: d_2
TREATMENT		a	ad	ac	acd	ab	abd	abc	abcd
REPLICATE	1	14.0	16.2	16.2	13.2	15.8	21.0	27.0	13.3
	2	19.0	9.0	15.2	10.2	11.5	13.5	11.0	14.2
	3	18.5	25.2	13.0	23.7	18.2	13.1	17.0	12.1
	4	15.0	6.9	17.0	16.3	15.8	12.5	19.0	10.4
	5	28.0	22.5	15.0	26.4	19.0	28.2	14.2	11.2
	6	16.5	10.0	17.0	23.1	28.0	22.5	15.0	22.6

AVERAGE ACQUISITION TIME BY LEVELS IN SEC	
WITH HT. COMP. : $\bar{a}_1$	13.53±1.20
STANDARD MODE : $\bar{a}_2$	16.94±1.57
SHORT RANGE : $\bar{b}_1$	15.56±1.30
LONG RANGE : $\bar{b}_2$	14.91±1.57
RADAR-CREW 1 : $\bar{c}_1$	15.53±1.66
RADAR-CREW 2 : $\bar{c}_2$	14.94±1.27
TANGENTIAL COURSE : $\bar{d}_1$	15.99±1.37
RADIAL COURSE : $\bar{d}_2$	14.48±1.56

DATA & RESULTS OF RADIAL & TANGENTIAL FLIGHTS PLANE #3

	CREW 1 c_1	CREW 2 c_2
STANDARD MODE : a_1	17.50	16.39
HEIGHT COMP. : a_2	13.57	13.49
IMPROVEMENT	3.93	2.90

INTERACTION TABLE MOD X CREW

	AVERAGE ACQUISITION TIME (sec)		AVERAGE IMPROVEMENT (sec)
	WITH HT. COMP a_1	STANDARD MODE a_2	
RADAR-CREW 1 c_1	$13.57^{\pm 2.00}$	$17.50^{\pm 2.53}$	3.93
RADAR-CREW 2 c_2	$13.49^{\pm 1.41}$	$16.39^{\pm 2.07}$	2.90
DIFFERENCE IN CREWS	.08	1.11	X

ANALYSIS OF VARIANCE TABLE

FACTORS		S S COMPONENTS	TOTALS	D.F.	MSS = SS/D.F.	F_c = MSS/R.E. MSS
MODIFICATION	(A)	278.00		1	278.00	12.20 **
RANGE	(B)	10.14		1	10.14	0.44
RADAR CREW	(C)	8.40		1	8.40	0.37
COURSE TYPE	(D)	54.00		1	54.00	2.40
MOD X RANGE	(AB)	8.00		1	8.00	0.39
MOD X RADAR CREW	(AC)	6.41		1	6.41	0.28
MOD X COURSE TYPE	(AD)	13.06		1	13.06	0.57
RANGE X RADAR CREW	(BC)	3.60		1	3.60	0.16
RANGE X COURSE TYPE	(BD)	0.05		1	0.05	0.00
RADAR CREW X COURSE TYPE	(CD)	31.97		1	31.97	1.40
MOD X RANGE X RADAR CREW	(ABC)	33.14		1	33.14	1.45
MOD X RANGE X COURSE TYPE	(ABD)	8.28		1	8.28	0.36
MOD X RADAR CREW X COURSE TYPE	(ACD)	3.30		1	3.30	0.14
RANGE X RADAR CREW X COURSE TYPE	(BCD)	21.00		1	21.00	0.92
MOD X RANGE X RADAR CREW X COURSE TYPE	(ABCD)	67.00		1	67.00	2.93
TOTAL TREATMENT			549.00	15	36.61	1.60
REPLICATE EFFECT			213.15	5	42.63	1.86
RESIDUAL ERROR			1714.74	75	22.86	F.00
CORRECTED SS = SSS - CT			2476.90	95	26.07	1.14

*SIGNIFICANT AT 5% LEVEL
**SIGNIFICANT AT 1% LEVEL

DATA SET 10

INTERACTION TABLE RADAR CREW X MANEUVER

	AVERAGE ACQUISITION TIME (sec)		AVERAGE IMPROVEMENT
	RADAR CREW 1 c_1	RADAR CREW 1 c_2	
TANGENTIAL COURSE d_1	16.87	15.12	1.75
RADIAL COURSE d_2	14.20	14.76	.56
	2.67	.36	X

TABLE OF AVERAGES (sec)

	WITH HT. COMP. a_1				STANDARD MODE a_2			
	SHORT RANGE b_1		LONG RANGE b_2		SHORT RANGE b_1		LONG RANGE b_2	
	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2	RADAR CREW 1 c_1	RADAR CREW 2 c_2
TANGENTIAL COURSE d_1	16.30±4.71 s=4.00	14.73±2.63 s=2.50	14.53±6.20 s=6.00	12.00±5.80 s=6.04	18.50±5.20 s=5.00	15.37±1.60 s=1.52	16.06±5.81 s=6.53	17.20±5.79 s=6.81
RADIAL COURSE d_2	12.08±4.23 s=4.00	12.73±1.60 s=1.50	10.57±4.31 s=4.01	13.33±8.08 s=4.80	14.97±7.87 s=7.50	16.02±8.04 s=6.54	16.47±5.75 s=5.53	13.77±4.67 s=4.66

RESULTS AND CONCLUSIONS:

a. Table 1 indicates the acquisition times and failure rates for the various radar targets or aircraft flown. The following should especially be noted:

(1) The use of the Height Comparator resulted in a significant reduction in the transfer failure rate for the aircraft types used in the tests.

(2) For radar target or plane #3 the transfer failure rate of weapon batteries without a Height Comparator was inordinately high. In fact 18.3% of the designations resulted in failures. (Acquisition times over 32 seconds were considered failures.

(3) The use of the Height Comparator also resulted in a significant reduction in average acquisition time of the aircraft types used in the tests.

b. It was apparent during the test series that the requirement for multiple operator coordination adversely affects target acquisition.

c. It was apparent that after a brief layoff period, the target track operators performed below par for the first few target acquisitions. This initial failure rate is extremely important when defense systems are operated tactically and subjected to surprise raids. This was evident even for periods as short as 18 hours.

d. The data indicate that the radar-crew combination exerted a statistically significant effect upon acquisition time, and must be considered an important parameter in any analysis of this type. An interaction was present between this and other parameters undergoing analysis. For example, the effectiveness of the Height Comparator in acquiring slow aircraft often depended upon the radar-crew.

e. This analysis also shows the importance of a statistical control, (such as the L-19) when sets of data are separated by large time intervals. The five months time lapse between the November 1956 and the April 1957 flights resulted in a statistically significant increase of 1.4 seconds in average acquisition time. This time lapse played an important role in the comparison between the radar targets, planes #2 and #3.

f. The experimental design required a large number of replications to guard against the possibility of large variances for each treatment combination. This fact was suggested by preliminary tests and the resulting analysis showed this fact to be true. It was therefore felt desirable to replicate the experiment as many times as possible, e.g. at least a dozen times. This feature is one of the main differences between this experimental design and others described in the literature. In spite of the large variances, the analysis was able to proceed to a successful conclusion and meaningful results because of the large number of replications.

RECOMMENDATIONS

In the light of the findings of this analysis, the following recommendations were made to improve target acquisition:

a. Install the automatic Height Comparator (Height Null Meter), or an equivalent device.

b. Initiate a program of daily intensive "on the site" realistic training for operators.

c. Due to the importance of operator training, procedures should be checked and revised when necessary to insure optimum operator performance. An independent team of radar experts using realistic test procedures should select a system and test the performance of the operator personnel.

d. Where large time lapses occur during the tests on defense systems, a statistical control should be utilized. In the test series described herein, the low-speed L-19 army aircraft was used for this purpose.

e. In the design of future tests, and in the statistical reduction of the data, the effect of the radar crew should be clearly differentiated from the effects of the other variables undergoing study.

f. A study should be made of the acquisition procedure and associated equipment with a view toward simplifying and reducing the multi-operator coordination requirements.

ACKNOWLEDGMENTS

The authors wish to express their indebtedness to the many individuals who assisted in the preparation of this report, and in particular to the Specialist Walter LaMotte, Pfc. F. Seltzer, Martin Orr, to the Signal Corps and Air Force pilots, to the engineers who designed the modification, and to the many radar operators.

SHORT GLOSSARY OF TERMS

Altitude Maneuver. Flying the L-19 aircraft on radial and tangential courses in such manner as to render the current acquisition independent of the previous acquisition.

Crew, Radar-Crew, Radar and Crew. These terms are used in their widest meaning to denote the entire man-machine combination.

Height Comparator (Ht. Comp., or Modification). A null-type meter that compares two voltages or currents; one voltage represents the target height as seen by a remote source such as an operations center, and the second voltage is related to the elevation of the radar track antenna.

Plan Position Indicator (PPI). A display that gives azimuth and slant range of targets on the circular face of a cathode ray tube. It is sometimes called a polar coordinate or time base display.

Range Slewing. At any instant of time, the track radar is examining a particular point in range. If the tracking radar is to acquire the target, the above point must be moved inward or outward. This process is called slewing.

EFFECTS OF BALLISTICS AND METEOROLOGICAL VARIABLES ON ACCURACY OF ARTILLERY FIRE

O. P. Bruno
Ballistic Research Laboratories
Aberdeen Proving Ground, Mayland
U. S. Army Ordnance

1. This paper presents preliminary information on an exploratory study which has been undertaken to investigate some of the ballistic and meteorological parameters which affect the accuracy of fire with artillery weapons systems. It is a study which has been in progress at the Ballistic Research Laboratories, at Aberdeen Proving Ground with the cooperation of the Continental Army Command, the U. S. Army Artillery and Missile School at Ft. Sill and the Evans Signal Laboratories of the U. S. Army Signal Corps.

2. With the development of atomic artillery and related tactical concepts requiring relatively small and highly mobile combat units, renewed emphasis has been placed on the development of new doctrine and capability for accurate delivery of both atomic and conventional artillery fire. With atomic artillery particularly, it would be highly desirable to develop a capability for hitting, with a high probability, a target with the first round fired. Development of this capability is important to fully exploit the element of surprise. The effectiveness of the element of surprise is considerably reduced under the customary techniques of adjustment of fire and registration preliminary to firing-for-effect on a target. However, this capability is difficult to achieve because of the many parameters which contribute toward inaccuracy of artillery fire. Some of these parameters are: (1) interior ballistic variations in muzzle velocity caused by gun tube condition; differences in propellant weight, temperature, moisture and other characteristics; shell differences in weight and banding; and (2) exterior ballistic effects including variation in shell weight, surface finish, shape and stability (sometimes expressed as variation in ballistic coefficient). Other factors include meteorological effects, such as wind velocity and direction, temperature, and density; and still other factors, such as accurate determination of distance and azimuth to the target.

3. Let us assume that the Firing Battery has one lot of ammunition and one gun. Further, that they have calibrated (1) their gun tube and ammunition (i.e., they have a good estimate of the velocity level of the gun tube-ammunition combination), and (2) the ballistic coefficient is calibrated (i.e., the difference between the ballistic coefficient of the shell lot and the ballistic coefficient assumed in the Firing Table is known or negligible).

4. Also let us assume that the Firing Battery has meteorological information which can be used to estimate the effect of non-standard meteorological condition on range.

5. Let us also assume that the distance between the gun position and the target is known accurately.

6. Then the error involved in hitting the target with a round of ammunition may be represented by:

$$\sigma_R^2 = \sigma_{V_i}{}^2 \left(\frac{\Delta R}{\Delta V}\right)^2 + \sigma_{C_i}{}^2 \left(\frac{\Delta R}{\Delta C}\right)^2 + \sigma_\theta{}^2 \left(\frac{\Delta R}{\Delta \theta}\right)^2 + \sigma_W{}^2 \left(\frac{\Delta R}{\Delta W}\right)^2 + \sigma_{R_{M_D}}{}^2$$

where σ_R^2 is the variance in range in yards

$\sigma_{V_i}^2$ is the variance in velocity round to round independent of shell weight variations.

$(\frac{\Delta R}{\Delta V})$ is the differential effect in range for a unit change in velocity.

$\sigma_{C_i}^2$ is the variance in ballistic coefficient round to round independent of shell weight variations.

$(\frac{\Delta R}{\Delta C})$ is the differential effect in range for a unit change in ballistic coefficient.

$\sigma_{R_{MD}}^2$ is the variance in metro error measurement among occasions. In this study this will take on several values as we consider various degrees of metro staleness.

σ_θ^2 is the variance in gun tube angle of departure upon firing.

$(\frac{\Delta R}{\Delta \theta})$ is the differential effect in range for a unit change in angle departure.

σ_W^2 is the variance in shell weight.

$(\frac{\Delta R}{\Delta W})$ is the differential effect in range for a unit change in weight.

Through past studies of various calibers of artillery excellent estimates of all of the coefficients and differential effects for ballistic parameters are available. However, reliable estimates for $\sigma_{R_{MD}}$ are not available. The purpose of this study is to obtain estimates of these parameters.

7. It was considered probable that $\sigma_{R_{MD}}$ may vary systematically depending upon the type of meteorological conditions encountered on a day, the distance between the metro station and the firing battery and the staleness of the metro data (i.e., the change in true metro conditions between the time the metro data was taken and the time that the metro information was used). It was also considered advisable to study the variation attributable among metro batteries.

8. The types of experimental design adopted for this study were somewhat dictated by the nature of the parameters which required study and also by economic and logistic considerations. These considerations frequently influence the types of designs which can be adopted where relatively large scale experimentation is involved. In this particular case only four metro batteries were available and it was estimated that they could be conveniently located at distances of approximately 1, 5, 10, and 20 miles from the selected firing position. While metro batteries and their equipment were located in these positions it would be practical for them to take metro observations at 0600, 0800, 1000 and 1200 hours in a full day's work. A two hour interval would be reasonable for them to digest the data and develop the metro message. Consideration of the above factors precluded the random selection of metro staleness. It was also considered desirable to study the factor of staleness independently in order that $\sigma_{R_{MD}}$ could take on various values depending on the degree of staleness. Hence, the statistical design was a Latin Square with two replications where the three factors, days, distances, and metro batteries were studied in the designs for zero hours, 2 hours, 4 hours, and 6 hours staleness independently. In other words, the analysis involved 4 x 4 latin squares with two replications for each of the conditions of staleness under study (0, 2, 4, 6 hours) for each of the two weapons. Since, metro data was developed on each day at 0600, 0800, 1000 and 1200 hours and firings were conducted at 0800 and 1200 hours it was possible to get two sets of latin squares for each of zero hours staleness and two hours staleness, one square for each of four hours and six hours staleness.

Distance or Location

Days \ L	L_1 (1 mi)	L_2 (5 mi)	L_3 (10 mi)	L_4 (20 mi)
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Distance or Location

Days \ L	L_1 (1 mi)	L_2 (5 mi)	L_3 (10 mi)	L_4 (20 mi)
5	D	A	B	C
6	B	D	C	A
7	A	C	D	B
8	C	B	A	D

9. The firing program to develop the necessary information to serve as input data for the design was as follows: The four meteorological batteries were scheduled for occupation of the four different positions on each of the eight days in accordance with the Latin Square Designs indicated previously. The days for firing were selected at random. The Metro Batteries were instructed to take meteorological observations with the Radiosonde GMD-1 equipment at the hours of 0600, 0800, 1000, and 1200 each day. Field Artillery firing batteries were instructed to fire two artillery weapons (different calibers) at 0800 and 1200 on each day. These firings were carried out with rounds from two selected lots of ammunition representing the two calibers. The sample of n rounds fired on each occasion was drawn at random from the lot. For each caliber the charge and the quadrant elevation was fixed for all firings. Three range observation posts were used to measure the range of each round fired. Two doppler chronograph units were used to measure the velocity of each round fired. With this information, it was possible to compute rather accurately the range to each center of impact(corrected for velocity). For each center of impact it was possible to determine the actual effect of the existing non-standard meteorological conditions as opposed to the estimated effect of the non-standard meteorological conditions as computed from the meteorological data and Firing Tables. This latter value represents the input data for each cell in the Latin Square Design; therefore, it was possible to obtain input data for each of several conditions of meteorological staleness; namely, 0 hours, 2 hours, 4 hours, and 6 hours of meteorological data staleness. For example, it was possible to compute the estimated meteorological effect for non-standard conditions existing at 0600 and compare this with the actual effect on range of firings performed at 0800. This represents a condition of 2 hours staleness. The difference represents the error in estimating the effect of non-standard metro conditions and is used as input data in this analysis. Similarly, the firings at 0800 and 1000 were used in conjunction with the meteorological data for 0600, 0800, 1000 and 1200 hours to provide the input information for the study.

The results of the analysis for this study indicated the following:

a. The among Metro Battery differences were not significant, although the training and experience of the personnel of the meteorological batteries varied considerably.

b. That for the conditions existent in this study at Ft. Sill the distances between the field position and the location of the meteorological units at 1, 5, 10, and 20 miles were not significant.

c. That the experimental error σ_{Mi} was fairly constant for all of the Latin Square Designs and that for one weapon it was about 38 yards, while for the other weapon it was 31 yards at 9800 yards range.

d. The component of variance day to day for 0 hours, 2 hours, 4 hours, and 6 hours of staleness was significant and increased accordingly. Charts No. 1 and No. 2 for the two weapons under study show the relationship between the standard deviation in range σ_R as a function of range for each of the ballistic parameters under study; namely, the variation due to velocity

$\sigma_V\ (\frac{\Delta R}{\Delta V})$, ballistic coefficient $\sigma_{c_i}\ (\frac{\Delta R}{\Delta C})$, angle departure $\sigma_\theta(\frac{\Delta R}{\Delta \theta})$, and shell weight $\sigma_W\ (\frac{\Delta R}{\Delta W})$. There are also plotted the estimated values for the components of variation attributable to errors of estimation of the effects of non-standard meteorological conditions where 0, 2, 4, and 6 hours of staleness are involved σ_{D_0}, σ_{D_2}, σ_{D_4}, σ_{D_6}, respectively. Also plotted are the combined estimates for both the ballistic and meteorological sources of variation σ_{R_0}, σ_{R_2}, σ_{R_4}, σ_{R_6} for 0, 2, 4, and 6 hours of staleness respectively.

While this study is relatively limited with regard to the number of types of weapons, the topographical location and to the one range at which firings were performed with each of the two weapons, it is believed that some valuable conclusions can be drawn within the framework of this experiment.

a. That the equipment for measurement of meteorological parameters such as wind velocity, wind direction, temperature, and density does not contribute appreciably toward meteorological errors.

b. That the training and capability of Meteorological Battery personnel is not a particularly limiting factor in developing sufficiently accurate meteorolotical information.

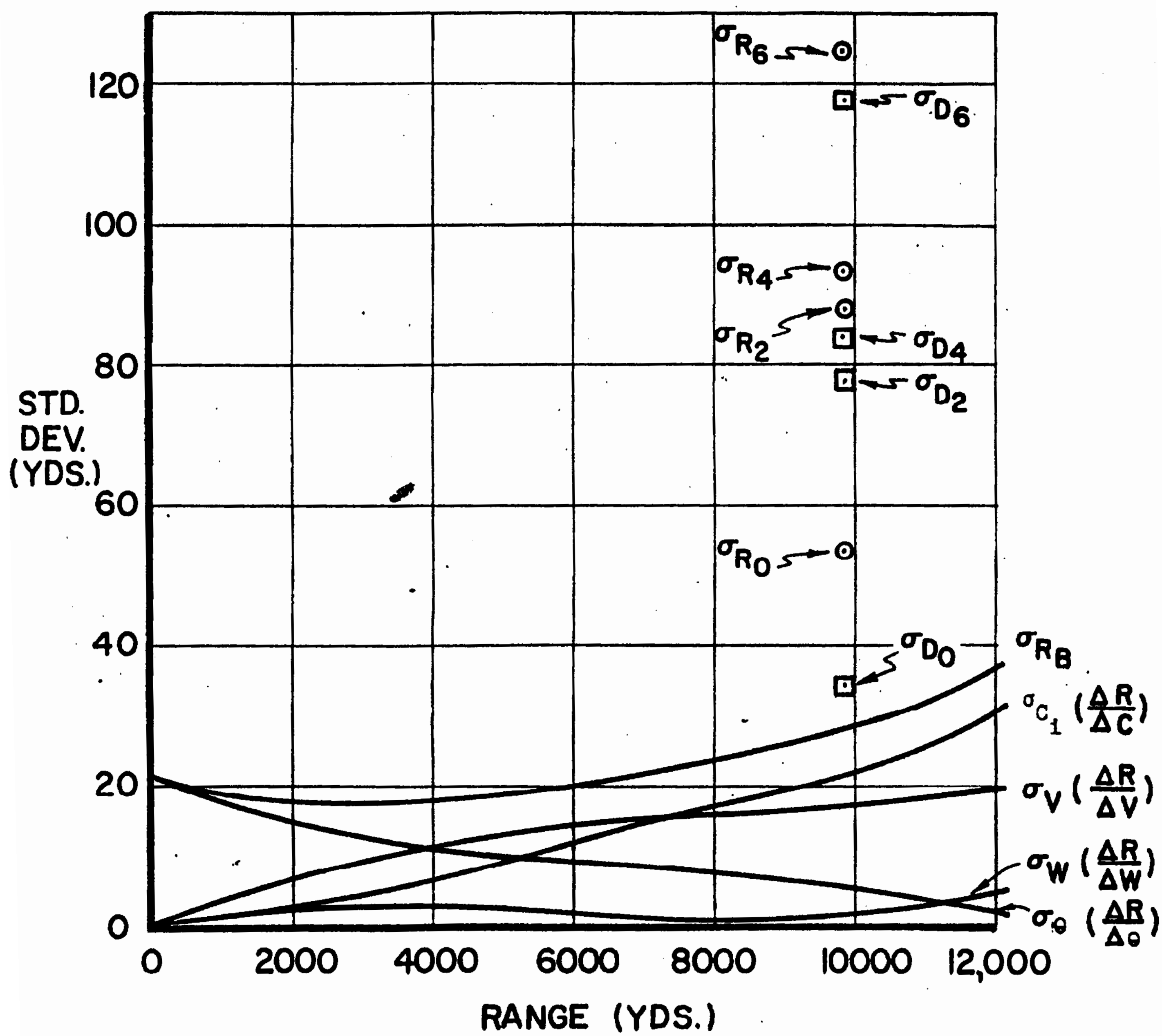
c. That for topographical and meteorological areas similar to Ft. Sill the distances of up to 20 miles between the firing point and location of meteorological batteries is not particularly significant or important.

d. That the most important factor is meteorological staleness. When meteorological data of 0 staleness is used, the error is approximately equivalent to the ballistic errors inherent in the ammunition-gun systems. (Meteorological errors and the ballistic errors contribute approximately equally to the total range error). However, it is recognized that it is not physically possible under the current system to have available meteorological data for 0 hours staleness since an appreciable amount of time is required for the reduction, dissemination, and use of the meteorological information.

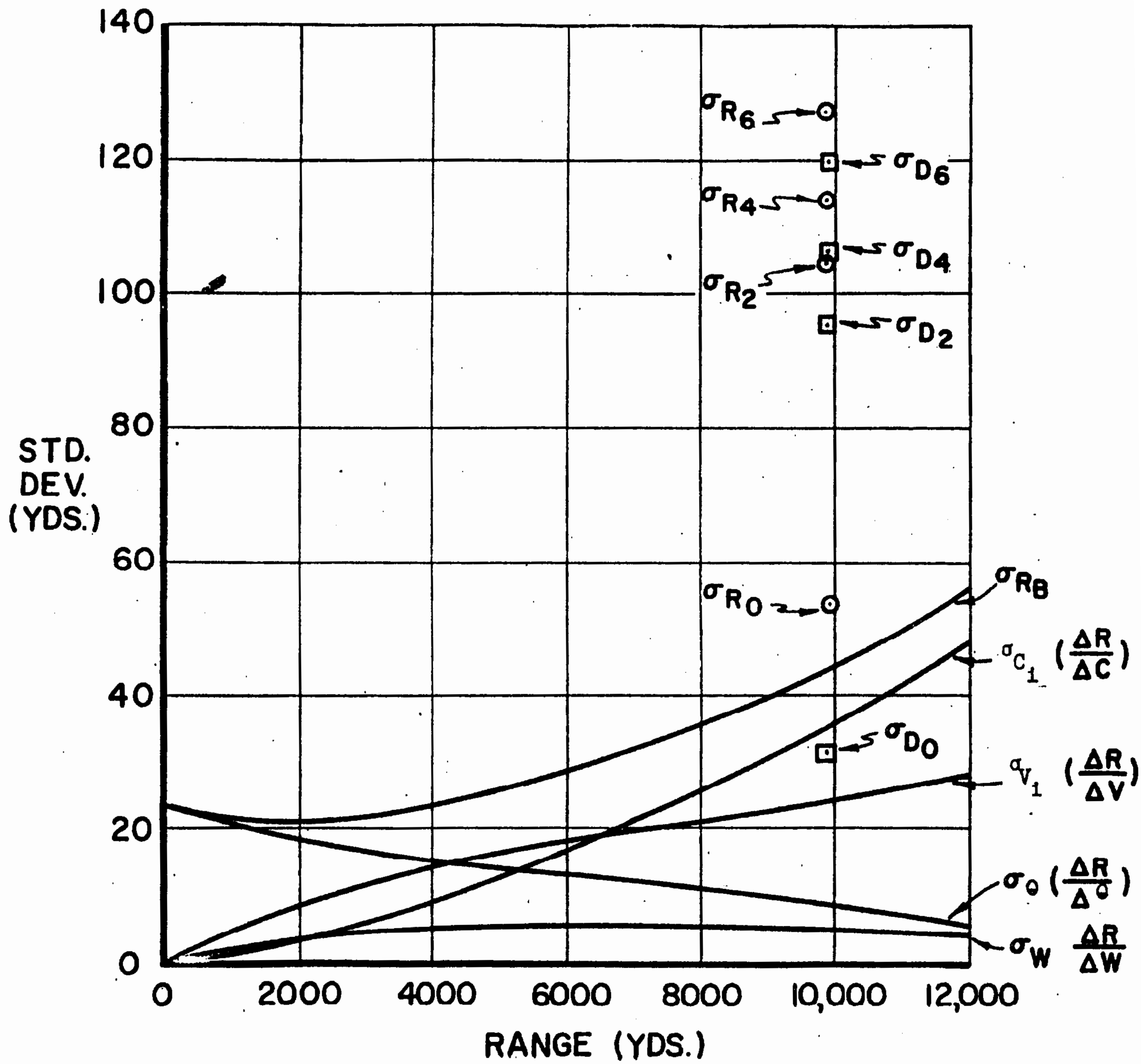
e. That the use of meteorological data which is 2 hours, 4 hours, or 6 hours old contributes appreciably more error than the ballistic errors. It was apparent that the round to round ballistic errors are relatively negligible in comparison to the errors in adjustment for the effect of non-standard metro conditions when 2, 4 and 6 hour stale metorological data is used.

f. It is apparent that the development of capabilities to obtain, reduce, disseminate, and use meteorological information immediately before firings may improve the accuracy of artillery fire and contribute appreciably toward the objective of increasing the probability of hitting the target with the first round fired.

SHELL, H. E. MI (DUALGRAN)
105 MM HOW. CHG VII



SHELL, H.E. M106
8" HOW., CHG V (G.B.)



CHARACTERISTICS OF VARIOUS METHODS OF COLLECTING DATA IN TESTS OF INCREASED SEVERITY

A. Bulfinch
Picatinny Arsenal

SUMMARY. The need for a better understanding of the characteristics of methods for collecting data in tests of increased severity is described. A number of problem areas in Ordnance research are listed in which tests of this kind are required.

Characteristics of various standard methods and some of their modifications have been studied by Monte Carlo techniques. The results of sampling known normal and skewed distributions are evaluated.

The relation of tests of increased severity to reliability testing is pointed out.

CONCLUSIONS. 1. Of the methods studied only two are of general interest:

a. The up-and-down method is most useful as an exploratory method in new situations where nothing is known about the possible outcome. The original version of the method will converge upon the region of the 50% point with the least possible effort regardless of where on the stimulus scale the test is started. The modifications of this method described will converge on other percentage points with the same efficiency. However, the up-and-down method has a number of shortcomings.

b. The run-down method is the most versatile. The original version can accurately determine the location and form of the parent population distribution. Modifications described are completely distribution free and can be used in the extreme tails of the curves for determining such things as safety and reliability.

2. The remaining methods are of little value except in highly specialized cases. Taken alone these methods tell us nothing about the parent population sampled.

INTRODUCTION. To appreciate the need for studying the characteristics of methods for collecting data in tests of increased severity one must know something about the following:

1. The nature of the tests in which these methods are used.

2. The kind of problem in which these tests are useful.

3. The frequency with which problems requiring these tests occur in Ordnance research.

We all know that if we strike an explosive hard enough, it will detonate. If we are careful we can strike it lightly without detonating it. Something like this is also true of delicate instruments. If we strike the instrument hard enough we will destroy it. But if we treat it carefully it will operate as intended. Finding out what happens in between these two extremes (of mechanical shock) is the objective of tests of increased severity. These tests are intended to determine how much stimulus (in various forms such as mechanical shock) is required to cause a given response frequency such as the detonation frequency of an explosive or failure frequency of an instrument. This concept of "how much stimulus is required to cause a response"* has come to be referred to as "sensitivity" to certain stimuli such as mechanical shock, electric energy, temperature, acceleration, etc. The methods used to collect data in these tests are now called sensitivity methods and the data collected is called sensitivity data for explosives and reliability data for missile components and other instruments.

Many problems requiring tests of increased severity for their solution are of long standing but are not recognized as such because of their statistical nature. Further difficulty lies in the fact that sensitivity and reliability data differ from other data in some respects. First, sensitivity and reliability data are binomial in nature. That is, there are only two possible outcomes, success or failure. Secondly, the observed data usually form a cumulative frequency. That is, the frequency of successes (or failures) obtained at any given level of stimulus is an estimate of the sum of all the success (or failure) frequencies up to that stimulus level. As a consequence an understanding of frequency distributions and probabilities (relative frequencies) is required to interpret the data (Ref 9).

Tests of increased severity are usually required when the following question arises: "At what level of stimulus should the test be conducted?" From a statistical point of view the answer is "At the 50% point". Then the question immediately arises "How can the 50% point be found?" This is a problem for tests of increased severity using methods such as the up-and-down method, the run-down method, and the two-stimuli method described later in this report.

However, the Ordnance research engineer is not always interested in the 50% point. He is not interested in explosives that detonate 50% of the time or instruments that function 50% of the time. From a safety standpoint he is interested in determining the maximum stimulus that can be used without causing a single detonation from the explosive. From a reliability standpoint he is interested in the maximum stimulus that can be used without causing a single failure in the instrument. As a result sensitivity methods have been developed for estimating points on the cumulative frequency curve other than the 50% point. The Picatinny Arsenal method and the first-fire-point method described later are two of these.

In this regard it would be interesting to apply the theory of extreme-value distributions to interpret sensitivity and reliability data. This approach has not been used in this study.

Methods for collecting data in tests of increased severity are required in a wide variety of Ordnance research problems. For example, some type of test of increased severity is required to collect useable data in each of the following problem areas:

---

*"Response" can be defined as an explosive detonation or a missile component failure.

1. Mechanical shock sensitivity.

    a. Impact tests of high explosives.

    b. Impact tests of artillery fuzes.

    c. Missile components.

    d. Izod impact test of metals.

    e. Izod impact test of plastics.

    f. Impact or drop test of packing cases.

2. Sensitivity to setback pressures of high explosives.

3. Acceleration sensitivity of missile components.

4. Friction sensitivity of explosives.

5. Velocity sensitifity of fuzes and explosives.

6. Voltage sensitivity of fuzes and missile components.

7. Spark sensitivity of pyrotechnic materials.

8. Temperature sensitivity of explosives and missile components.

In each one of these areas if observations are taken over the full range of responses from zero to 100% failures (or successes), it will be found that the data form a sigmoid cumulative frequency curve. Even when testing to failure in reliability work (Ref 10, the frequency of failures will form a sigmoid curve. As a result accurate reliability statements can only be made when the cumulative frequency percentage point associated with the stimulus level used is known. This percentage point can be determined only by using sensitivity methods such as those described below.

A search of the literature shows that much has been written on methods for tests of increased severity. But most of the work has been done with pure mathematics or with actual materials and equipment. Many of the mathematical treatments have been found to be impractical. Experiments using actual materials contain so many uncontrolled variables that the true characteristics of the data-collecting methods are distorted.

The object of the work reported here is to determine the true chararacteristics of several of the available sensitivity methods and some of their modifications through the use of the Monte Carlo procedure of sampling. It is expected that this procedure will reveal the true characteristics of these methods better than previous approaches, and suggest the need for new techniques for solving present day problems. Methods are required which will give unbiased estimates of the true population means and variances. It is believed that the theory of extreme-value distributions will be useful in this effort.

In Monte Carlo procedures it is assumed that all controlled experiments have the following two characteristics in common:

1. A set of experimental conditions (in the physical sense) is specified. This defines the underlying distribution and its parameters that would be formed if an infinite number of observations were taken under that set of conditions.

2. The order of occurrence of the observed data is always random if no effort is made to bias the data.

From these assumptions, simulated experiments can be conducted as follows:

1. Choose a known distribution which can be considered as representing the distribution defined by the particular set of experimental conditions under investigation.

2. Sample this distribution using a set of random numbers.

Simulated experiments of this kind have the following advantages:

1. They are cheap to conduct.

2. They are more practical than many mathematical approaches.

3. They are free of the usual errors encountered in handling materials and equipment.

4. Reliable estimates of parameters (true values) are economically obtained.

5. Known distributions can be sampled.

6. They can be used to confirm the validity of mathematical models.

Characteristics of the following methods have been studied to date:

1. Original Picatinny Arsenal methods (Ref 5).

2. First modification of Picatinny Arsenal method.

3. Second modification of Picatinny Arsenal method.

4. First-fire point (Ref 2).

5. First-failure point.

6. Up-and-down method (Refs 1 and 6).

    a. Recommended grouped data calculation (Refs 1 and 6).

    b. Usual grouped data calculation.

c. Two-failure modification.

d. Three-failure modification.

e. Ten-failure modification.

f. Fifteen-failure modification.

g. Two-success modification.

h. Three-success modification.

i. Ten-success modification.

7. Run-down method (Ref 4).

8. Two-stimuli method (Ref 3).

EVALUATION OF METHODS:

Original Picatinny Method. This method (Ref 5) starts high on the stimulus scale. If a success* is obtained the next lower (one increment lower) stimulus level is used for the next trial; if a failure is obtained the same stimulus level is used for the next trial. New specimens are used for each trial. This procedure is repeated until a stimulus level is found at which 10 successive failures are obtained. The next higher stimulus level is taken as the result. Why this level is taken as the sensitivity value is not known. The precision of this procedure is very poor. As shown in Table II**, single determinations (from one series of trials) must differ by at least 6.5 units before the difference can be declared significant.

Modified Picatinny Method. The first modification of the Picatinny method used was to take the height at which 10 successive failures were obtained as the result. The precision of this modification is exactly the same as that of the original method. However, the percentage point measured comes a little closer to the 10% point, which is the point the method has been assumed to determine (Table II). This is a very unfortunate percentage point at which to make comparisons of explosives. Former work (Ref 7) has shown that Comp B, RDX, tetryl, and TNT all have the same 7% point. This makes the very small differences among these standard explosives at the 10% point practically indistinguishable.

---

*"Success" is defined as an explosive detonation or a missile component failure.

**The Tables have been placed at the end of this article.

The second modification of the Picatinny method is the same as the first modification except that the stimulus at which 15 successive failures are obtained is taken as the result. This improves the precision somewhat but not to an acceptable extent. Table II shows that two single sets of determinations must differ by at least 5.1 units before the difference can be declared significant. This means that at least 26(5.1 squared) sets of trials must be conducted and averaged before a difference of one unit can be declared significant.

First-Fire Point. The first-fire point (Ref 2) starts low on the stimulus scale. If a failure is obtained the next higher stimulus level is used. The stimulus at which the first fire (explosive detonation or missile component failure) is obtained is taken as the result. This procedure has even poorer precision (Table II) than the Picatinny method and its modifications. This would be expected since the sample size used is smaller. However, the first-fire point method may be useful in reliability testing in situations where reasonable stimulus increments can be established.

First-Failure Point. The first-failure point starts high on the stimulus scale. If a success (detonation) is obtained the next lower stimulus level is used. The stimulus at which the first failure is obtained is taken as the result. The precision of this method is also unacceptable for explosives work. It is similar to the first-fire point in this respect (Table II).

Repeated determinations by any one of the above methods tell us nothing about the magnitude of the parameters of the parent population which we are striving to measure. This can be seen from Table II by comparing the averages and standard deviations obtained with the known parameters ($\mu = 20; \sigma = 5$) of the normal distribution sampled. However, any two of these methods used together will give two points on the cumulative frequency curve of the parent population. The further apart these points are, the more accurate the determination of the curve. If these points are plotted on probability paper the average (50%) point and standard deviation (slope of the line) of the parent population can be obtained by graphical methods within the precision of the method and sample size used. This approach to estimating the parameters is valid for normal distributions and distributions of known form only.

The converse of the Picatinny methods could be used to estimate a point on the curve in the region of the 90% point. But the precision would be expected to be similar to that of the Picatinny methods described above. This modification of the Picatinny method was not included in the present study since modifications of the up-and-down method (described below) can measure both the 10% and 90% regions of the curve. These modified up-and-down methods appear to be superior to the Picatinny methods for the reasons stated below under the discussion of the modified up-and-down methods.

The Up-and-Down Method. The unmodified up-and-down method (Refs 1 and 6) starts any place on the stimulus scale. If a success* is obtained, the next lower stimulus level, one increment below the first, is used for the next trial. If a failure is obtained, the next higher stimulus level is used for the next trial. The stimulus levels must be equally spaced at intervals equal to the standard deviation. This procedure is repeated throughout the test.

*"Success" is defined as an explosion detonation or a missile component failure.

This method is a good exploratory procedure for finding the region of the 50% point. Even when nothing is known about the possible location of the 50% poir the up-and-down method will converge on this region with a minimum number of trials regardless of where on the stimulus scale the testing is started.

The exploratory nature of the up-and-down method makes it very valuable in new situaitons where nothing is known of the possible outcomes. Whereas the original method will converge on the region of the 50% point, modified up-and-down methods can be made to seek out the region of other percentage points. Seven modifications of this method are listed in Table V.

Modified Up-and-Down Methods. The two-failure modification means that two successive failures are required before going to the next higher stimulus level. Only one success is required before going to the next lower stimulus level. The other "failure" modifications were conducted in a similar manner using the indicated number of successive failures before going to the next higher stimulus level. These modifications force the observations to converge on stimulus levels somewhat lower than the average of the parent population.

In the "success" modifications two or more consecutive successes are required before going to the next lower level. Only one failure is required before going to the next higher stimulus level. These modifications force the observations to converge on stimulus levels somewhat higher than the average of the parent population.

The percentage points shown in Table V were obtained by using the equatic of the normal cumulative frequency curve as follows:

$$\bar{X}_1 = \mu \pm t\sigma$$

where: $\bar{X}_1$ = observed mean.

μ = population mean.

t = normal deviate.

σ = population standard deviation.

This equation was solved for "t" and the area under the normal curve associated with the calculated T-value was found from a table of areas under the normal curve. These areas are the percentage points listed in Table V.

In a new situation a combination of one "success" and one "failure" modification of the up-and-down method can be used to determine two points on the cumulative frequency curve of the parent population. From this, reasonable estimates of the population mean and standard deviation can be obtained, if the form of the distribution sampled is known. The nature of these modifications (like the unmodified method) is such that the required percentage point regions can be found with a minimum of effort. The effect of not knowing the magnitude of the standard deviation in new situations may require some repetition to refine the measurements, since for best results the increments used in the up-and-down method should be of the order of the population standard deviation.

These modifications of the up-and-down method are preferred to the Picatinny method or first-fire and first failure methods because the up-and-down methods are more efficient. The method of conducting the up-and-down procedures is better defined and easier to follow consistently without wasted effort.

However, the up-and-down method has its limitations. For example:

1. We have the incongruous situation in which sampling a normal cumulative frequency with the up-and-down procedure forms a frequency distribution that is neither cumulative nor symmetrical (Table VII). Fifty-three per cent of the area under the curve of this distribution is below the mean. Therefore it can be said to have a slight positive skewness, which tends to give slightly low estimates of the mean (Table III). Taking the log of the stimulus units over-compensates for this bias. The fact that the observed frequency is not cumulative raises the question of whether the stimulus level used should be considered the midpoint of the grouped data cell. It is clear that the form of the parent distribution being sampled is the cumulative frequency since increasing the stimulus level can only increase (or decrease) the frequency of a response. The frequency cannot rise to a maximum and then decrease. If the expected distribution is in the form of a cumulative frequency, then the stimulus levels used must be the grouped data cell maxima. In spite of the fact that the observed frequency obtained by means of the up-and-down method is not cumulative, the stimulus levels used must be considered the cell maxima in order to obtain reasonable estimates of the true mean when positive responses are used. When negative responses are used the stimulus levels must be considered the cell minima. The data recorded for the up-and-down method in Tables III through VII have been obtained in this manner. That is, one-half the cell width has been arbitrarily subtracted from the stimulus levels when using positive responses and added to the stimulus levels when using negative responses.

2. Slightly biased estimates of the mean are obtained when the population distribution sampled is skewed (Table VI). These biases are in the direction of the median.

3. The population standard deviation is poorly estimated even when the population sampled is normally distributed. The observed sample standard deviation is significantly less than the population standard deviation (Tables III, V and VI). The usual formula (Refs 1 and 6) for the standard deviation associated with the up-and-down method gives a modified mean square rather than a standard deviation.

4. This method requires the stimulus level to be changed after each trial, which may be impractical in situations where changing the level is complicated or where the responses are not immediately available.

5. The stimulus used must be accurately controllable at predetermined levels so that the stimulus levels are equally spaced at intervals equal to the standard deviation.

A further difficulty with the up-and-down method is the restrictions placed on the observations by the sampling procedure. The conditions under which each observation (except the first) is taken are dependent upon the outcome of the previous observation. As a result all of the observations are concentrated in the central region of the curve (Table VII for normal distribution). The probability of an observation being as far as two standard deviations from the mean in either tail of a normal distribution is 2.28 times per hundred. From Table VII (for normal distribution) it can be seen that at two standard deviations from the mean, 1379 observations with the up-and-down method gave no observation in the upper tail and only one observation in the lower tail. If the observations were random, they would be in proportion to the frequency distribution of the population sampled. Since the condition of collecting data permits observations around the 50% point only but not in the tails, the samples obtained cannot be considered representative of the population. This is reflected in the biased standard deviation obtained. It can therefore be concluded that the data obtained with the up-and-down method is neither independent nor random and is not representative of the parent population sampled.

Run-Down Method. In this method (Ref 4) a given number of successive trials are made at each stimulus level used. The stimulus levels are arranged to cover the entire range of responses (from zero to 100%) in a convenient number of increments. The size of the increments and the number of trials used at each stimulus level can be varied to accomplish the intended purpose. To obtain reasonable precision in the tails, the increments used in the tails of the curve should be smaller and the number of trials used at each stimulus level should be larger than those used in the region of the 50% point.

The sampling procedure of the run-down method does not require a knowledge of the outcome of the previous observation in order to determine the condition under which the next observation will be taken. That is, the outcome of one observation does not affect any other. Therefore it can be said that the observations are independent. There is no restriction in this method as to where the observations are taken. The stimulus levels used can be picked at random anywhere on the curve. Once a level is chosen the observations are completely unrestricted, and therefore occur at random. Since observations are taken over the entire response range (zero to 100%) and occur at random, their relative frequencies will be proportional to those of the parent population and will therefore be representative of the parent population (Table VIII).

Because the data obtained with the run-down method are independent and random and represent the population sampled, this method has the following advantages:

1. The form of the distribution sampled can be determined (Table VIII).

2. Unbiased estimates of thetrue mean are obtained even when the distribution sampled is skewed (Table VI).

3. Unbiased estimates of the true variance are obtained when the distribution sampled is normal (Table III).

4. The validity with which statistical techniques requiring the assumption of normality are used can be evaluated.

5. The acceptability of the new products or new treatments of old products can be based upon the form of the distribution as well as the mean and variance

Additional favorable characteristics of the run-down method are as follows:

1. Basic rules of statistical theory are followed.

2. Observed data form a cumulative frequency as expected (Table VIII).

3. The method is useful in a variety of practical situations since once a stimulus level has been established a number of observations are taken at that level.

4. Comparison of two or more materials or items can be made at any given stimulus level using chi-square tests of significance without any assumption concerning the form of the distributions. This modification of the method is especially useful when the comparison of interest occurs in the extreme tails of the curves, such as when measuring the safety of an explosive or the reliability of a missile component.

5. Prior knowledge of the magnitude of the population standard deviation is not required.

The major disadvantage of the run-down method is the fact that a relatively large sample size (total number of trials) is required to determine the cumulative frequency curve over its entire length. However this disadvantage is tempered by the following:

1. This is the only method which can accurately estimate the cumulative frequency curve over its entire length.

2. If the exact character of the entire cumulative frequency curve is not required, the one-stimulus (described above) or two-stimuli modification (described below) can be used.

An additional disadvantage of the run-down method is the fact that biased estimates of the standard deviations are obtained when the distributions sampled are skewed (Table VI).

Two-Stimuli Method. The two-stimuli method (Ref 3) is a modification of the run-down method. Instead of using several stimulus levels to cover the range of responses, only two stimulus levels are used--thus the name. This method is useable only when the assumption of normality is valid (or the form of the distribution is known) and when the response frequencies obtained from the two stimulus levels differ by as much as 20%.

The advantages of this method under the restrictions mentioned above are as follows:

1. It is simple to conduct and simple to calculate.

2. It uses relatively small sample sizes.

3. It gives unbiased estimates of the mean and standard deviation.

The disadvantages of this method are as follows:

1. It is sensitive to deviations from the assumed form of the distribution.

2. It requires some previous knowledge of the location of the cumulative frequency curve to be sampled.

TABLE I

Known Cumulative Distributions Used in the Monte Carlo Sampling Experiments

True Mean = 20.0

True Standard Deviation = 5.0

Stimulus Levels[a]	Std Dev Units	Distributions Sampled (Area Under Curve, %) Normal Curve	Positively Skewed	Negatively Skewed
35.0	3.00		99	
32.5	2.50	99	98	
30.0	2.00	98	96	
27.5	1.50	93	92	99
25.0	1.00	84	85	86
22.5	0.50	69	74	64
20.0	0.00	50	57	43
17.5	-0.50	31	36	26
15.0	-1.00	16	14	15
12.5	-1.50	7	1	8
10.0	-2.00	2		4
7.5	-2.50	1		2
5.0	-3.00			1

[a] The cell maxima.

TABLE II

Characteristics of Various Methods for Collecting
Sensitivity Data Sampling A Normal Distribution

True Mean = 20.0

True Standard Deviation= 5.0

Method	Incre-ment	Total No. of Trials	Sample Size Used	Percentage Point Measured	Average	Standard Deviation	LSD [a]
PA (original) [b]	σ/5	2240	80	15.5	14.5	2.29	6.5
PA Modified 1 [c]	σ/5	2240	80	11.8	13.5	2.29	6.5
PA Modified 2 [d]	σ/5	2548	52	5.6	11.8	1.78	5.1
First-Fire Point	σ/5	1053	117	25.9	16.2	3.11	8.7
First-Failure Point	σ/5	1044	116	70.7	23.0	2.55	7.1

[a] Least significant difference (for single determinations) to compare two results.

[b] The stimulus level one inch above the height at which 10 successive failures are obtained is used as the average.

[c] The stimulus level at which 10 successive failures are obtained is used as the average.

[d] The stimulus level at which 15 successive failures are obtained is used as the average.

TABLE III

Comparisons Using Large Sample Sizes
Sampling A Normal Distribution with True Mean = 20.0
True Standard Deviation = 5.0, Increment = $\sigma/2$

Methods	Total No. of Trials	Sample Size	Average	Standard Deviation
Up-and-Down				
1. Calc as in Ref 1				
a. Successes	2700	1350	19.8	4.88[a]
b. Failures	2700	1350	19.8	4.92[a]
2. Standard Grouped data calculated				
a. Successes	2700	1350	19.8	2.71
b. Failures	2700	1350	19.8	2.68
Run-Down				
1. Standard Grouped data caloulated				
a. Successes	7200	800	19.9	4.75
b. Failures	7200	800	19.9	4.80
Two-Stimuli				
1. Calc as in Ref 3				
a. Successes	1600	800	20.0	5.03
b. Failures	1600	800	20.0	5.03

[a] These values are actually modified mean squares rather than standard deviations.

TABLE IV

Reproducibility of Method Results
Sampling A Normal Distribution

True Mean = 20.0

True Standard Deviation = 5.0

	Methods					
	Up-and-Down[a]		Run-Down[a]		Two-Stimuli[b]	
Replicate	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
1	19.6	2.46	19.7	5.01	20.2	4.82
2	19.5	3.15	20.0	4.60	20.1	4.34
3	18.4	3.17	19.8	5.02	19.7	3.80
4	18.8	2.80	20.0	4.86	20.1	5.35
5	20.1	2.86	20.0	4.93	19.8	5.24
6	18.9	2.70	20.0	4.83	19.9	5.58
7	20.5	2.60	20.1	4.95	19.6	5.22
8	19.3	2.32	19.9	4.84	19.6	4.42
Ave	19.4	2.77	19.8	4.88	19.8	4.88
Range	2.1	0.85	0.4	0.42	0.6	1.78
	Sample Size = 100 No. of Trials = 200		Sample Size = 100 No. of Trials = 900		Sample Size = 100 No. of Trials = 200	

[a] Increments used equal one-half the true standard division. Standard grouped data calculations were used.

[b] Increments and calculations used are described in Reference 3.

TABLE V

Characteristics of Various Modifications of the Up-and-Down Method Sampling a Normal Distribution

True Mean = 20.0

True Standard Deviation = 5.0

Modification	Increment	Total No. of Trials	Sample Size	Percentage Point Measured	Average	Standard Deviation[a]
None	σ/2	2700	1350	50	19.8	2.5
Two Failure	σ/2	250	74	26	16.8	2.5
Three Failure	σ/2	498	111	18	15.5	2.5
Ten Failure	σ/5	1150	115	8	12.9	2.5
Fifteen Failure	σ/5	1845	123	5.5	11.9	1.2
Two Success	σ/2	405	116	54	20.5	2.0
Three Success	σ/2	720	181	69	22.5	2.0
Ten Success	σ/5	1170	117	90	26.5	1.5

[a] Actual standard deviation of the method, not the modified mean square.

TABLE VI

Effect of the Form of the Distribution on the Characteristics of the Up-and-Down and Run-Down Methods

True Mean = 20.0

True Standard Deviation = 5.0

Increment = $\sigma/2$

Up-and-Down				Run-Down			
Positively Skewed		Negatively Skewed		Positively Skewed		Negatively Skewed	
$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
19.5	2.35	20.6	2.83	19.8	4.27	19.9	6.04
19.4	3.13	20.8	2.52	19.7	4.65	19.8	6.13

$$LSD^{b} = \frac{2.83 \times 5}{\sqrt{500}} = 0.63$$

$$LSD = \frac{2.83 \times 5}{\sqrt{100}} = 1.41$$

[a] The total number of trials used in each case is 1000. This is equivalent to a sample size of 500 in the up-and-down method and a sample size of 100 in the run-down method.

[b] Least significant difference (Ref 1).

TABLE VII

Observed Frequencies for the Up-and-Down Method

Stimulus Levels [a]	Std Dev Units	Normal Curve Area% [b]	Normal Curve Freq [c]	Positively Skewed Area% [b]	Positively Skewed Freq [c]	Negatively Skewed Area% [b]	Negatively Skewed Freq [c]
35.0	3.00			99			
32.5	2.50	99		98			
30.0	2.00	98	0	96	1		
27.5	1.50	93	27	92	10	99	49
25.0	1.00	84	182	85	94	86	183
22.5	0.50	69	431	74	288	64	378
20.0	0.00	50	478	57	381	43	284
17.5	-0.50	31	215	36	201	26	90
15.0	-1.00	16	41	14	25	15	15
12.5	-1.50	7	4	1	1	8	1
10.0	-2.00	2	1			4	
7.5	-2.50	1				2	
5.0	-3.00						
			1379		1001		1000

Distributions Sampled: Normal Curve, Positively Skewed, Negatively Skewed.

[a] The cell maxima.

[b] Taken from Table I.

[c] Observed success frequencies using the following total number of trials:
Normal Curve - 2750 Positively Skewed - 2000 Negatively Skewed - 2000

TABLE VIII

Observed Frequencies for the Run-Down Method

		Distributions Sampled					
		Normal Curve		Positively Skewed		Negatively Skewed	
Stimulus Levels [a]	Std Dev Units	Area% [b]	Freq [c]	Area% [b]	Freq [c]	Area% [b]	Freq [c]
35.0	3.00			99			
32.5	2.50	99	797	98			
30.0	2.00	98	743	96	196		
27.5	1.50	93	673	92	187	99	197
25.0	1.00	84	562	85	176	86	166
22.5	0.50	69	411	74	146	64	125
20.0	0.00	50	254	57	118	43	87
17.5	-0.50	31	122	36	67	26	51
15.0	-1.00	16	53	14	20	15	30
12.5	-1.50	7	19	1	0	8	17
10.0	-2.00	2				4	8
7.5	-2.50	1				2	3
5.0	-3.00					1	0

[a] The cell maxima.

[b] Taken from Table I.

[c] Observed success frequencies using the following number of trials at each stimulus level: Normal Curve - 800 Positively Skewed - 200 Negatively Skewed - 200

REFERENCES

1. Dixon & Massey, Introduction to Statistical Analysis, New York, Mcgraw-Hill Book Co., Inc., second edition, 1957.

2. Anderson, T. W., Staircase Methods of Sensitivity Testing, NAVORD Report 65-46, Statistical Research Group, Princeton University March 1946.

3. Churchman, C. W., Tables for Sensitivity Tests Conducted at Two-Stimuli, Statistical Memo No. 6, Frankford Arsenal Memo Report MR-540, May 1953.

4. Churchman, C. W., Manual for Proposed Acceptance Test for Sensitivity of Percussion Primers, Frankford Arsenal Report R-259A, January 1943.

5. Rinkenbach, W. H. and Clear, A. J., Standard Laboratory Procedures for Sensitivity, Brisance and Stability of Explosives, Picatinny Arsenal Technical Report 1401, Revision 1, February 1950.

6. OSRD report 4040 (NDRC AMP Report No. 101 IR), Statistical Analysis for a New Procedure in Sensitivity Experiments, Statistical Research Group, Princeton University.

7. Bulfinch, Alonzo, Improved Methods and Techniques for Testing Impact Sensitivity of Explosives, Picatinny Arsenal Technical Report 2282, July 1956.

8. Hartvigsen & Vandeback, Sensitivity Tests for Fuzes, NAVORD Report 3496, US Naval Ordnance Test Station, China Lake, California, March 1955.

9. Pieruschka, Erich, Mathematical Foundation of Reliability Theory, Research & Development Division, Ordnance Missile Laboratories, Redstone Arsenal, January 1958.

10. Lusser, Robert, Unreliability of Electronics--Cause and Cure, Research & Development Division, Ordnance Missile Laboratories, Redstone Arsenal, November 1957.

STATING AND TESTING CRITERIA FOR SMOOTHING DATA

Paul C. Cox
White Sands Missile Range

Maurice Kendall, in his second volume of the Advanced Theory of Statistics (page 378) states, "There is voluminous literature on trend fitting which appears to me out of proportion to the importance of the subject." This comment was undoubtedly correct at the time the volume was prepared (i.e. 1948), and as applied to data in economics with which Kendall was primarily concerned, but since that time the requirements for tracking weapons, targets, and many other moving objects, and the recording of many types of signals; the necessity for reducing the error (or noise) in the data, and the development of many new types of equipment to secure and record the data, has made the techniques for smoothing data extremely important in our scientific, engineering, and defense effort. Furthermore, the advances in the development of high speed digital computers has made it possible to use procedures which would have been impractical a few years ago. Thus, developments during the past decade are requiring new and better techniques for data smoothing. It is the purpose of this talk to present:

1. Desirable criteria for choosing a certain smoothing technique.

2. Areas where study and research may contribute toward better procedures.

3. Statistical tests which may be developed that might be used to test these criteria.

The conventional techniques for smoothing data usually consist of the following steps: (a) From the entire set of data select the first N_1 points, where N_1 is usually an odd number; (b) Fit a polynomial of degree r_1 to these points; (c) Choose k_1 points from the center of the N_1 points ($k_1 \ll N_1$). At these k_1 points compute the polynomial values, which will be accepted as the smoothed values corresponding to these k_1 values; (d) If velocity and acceleration data is desired, the polynomial may be differentiated successively; (e) Select a set of N_2 points the first of which is point $\left\{\frac{2 + N_1 - N_2 + k_1 + k_2}{2}\right\}$ and fit k_2 additional values from a second polynomial. Velocity and acceleration data will again be obtained by differentiation. (N_2 and k_2 will probably be equal N_1 and k_1 most of the time, but not necessarily all of the time); (f) This process will be continued with N_3, k_3, and r_3, N_4, k_4, and r_4, etc. until the data has all been smoothed.

The first problem is related to the selection of r. It is desirable that the degree of the polynomial be selected such that the error in the data (the noise) may be eliminated as much as possible, and yet it is not desirable to choose r so large that the smoothed data is over fitted. In such case the smoothed data may be following a noise pattern rather than the desired signal. The use of the F test to determine an optimum value for r is well known, and there is little more that need be said about this technique here. However, it is suggested that the choice of r may also be influenced by a knowledge of the physical characteristics of the data.

That is to say, if the equations of motion are reasonably well known and from this it can be stated that the data should be following a cubic equation (as an example), it then appears that the use of a cubic polynomial to smooth the data would be more desirable than using the F test to determine the value for r. Furthermore, if the F test is not used to determine r, it might be available to test the desirable magnitude of N, k, or something else.

The second problem is related to the choice of N. As a rule, this is selected on an arbitrary basis, but it seems logical that definite criteria may be established for this selection. A few years ago, N was rarely chosen larger than 25 or 30 because the labor would have been prohibitive, but today with the recent developments in high speed computers it is not unreasonable to choose a value for N as large as one or two hundred (possibly even more). There exists a tremendous latitude for the choice of values for N and it appears some criterion may be set up which will indicate a most desirable choice for N, and then a suitable test should be devised for testing this desirability. To mention some concepts related to the choice of N, one might mention that increasing the size of N will definitely decrease the variance of the deviation between the smoothed values and the true values, providing a good fit is retained after taking a larger value for N. This may be illustrated by the well known formula for the linear case in which:

$\sigma_s^2 = \frac{1}{n} + \frac{(t' - \bar{t})^2}{\Sigma(t - \bar{t})^2} \sigma_e^2$; where σ_s^2 is the variance of the smoothed value at $t = t'$ and σ_e^2 is the variance of deviations which exist between observations and the true regression curve. It can easily be seen that if σ_e^2 remains constant that σ_s^2 becomes smaller as N becomes larger. Unfortunately we cannot necessarily make σ_s^2 as small as we please by simply increasing the value of N. If N is to be increased, it must be done either by increasing the lengt of the record or decreasing the size of Δt. In the first case, it may be found that a polynomial will not fit as well for a long record as for a short one. In the latter case, many practical difficulties are involved when Δt gets below a certain value.

Three suggestions are given regarding the selection of N. The first is that if computer programs have been prepared for certain values of N, then it follows that one of these programs will be selected. The second is that in general, N may be increased until σ_s^2 (as well as corresponding variances for the smoothed velocities and accelerations) is made as small as we please, although we will sooner or later reach the point of diminishing returns. The third suggestion is that some test such as the F test might be developed to indicate a most desirable value for N.

The third problem is related to the choice of k. Referring to formula (1) it is clear that the precision is greatly improved as k is chosen to be one and this single value is the central value. However, this is not necessarily the case when using higher degree equations, and when using velocity or acceleration data rather than position data. Furthermore, even in those instances in which the minimum error is obtained when the smoothed value is the central value, it is possible one could choose k larger than one without any appreciable loss in precision.

It then appears that there is need for study to determine how many and which of the N values should be used in the smoothing process. Since starting work on this paper, it has come to my attention that some work has been done on this by a few companies which are interested in data reduction problems. In particular, the Jet Propulsion Laboratory has made some excellent contributions. It appears, however, that nearly everything which has been written on the subject can be found only in internal company reports.

The fourth problem is one related to the computation of velocity and acceleration. The problem is simply this, what is the most efficient method for computing velocity? Is it best to smooth the data, then compute the velocity; compute the velocity from the raw data and then smooth; or smooth the raw data, compute the velocity, and then smooth the raw velocity data? Some work has been done by Mr. Charles Bodwell, formerly of Holloman Air Force Base, and his results are available in "Data Reduction Report Nr. M.T.H.T. 293," White Sands Missile Range.

The final comment I wish to make is that the concept of fitting an orthoganal polynomial of degree r to a set of data is an implication that the true data can be approximated nicely by means of a polynomial of low degree, and all deviations from this are simply noise. This suggests that one should study the possible types of mathematical filters which might filter out the noise, and it is possible that something may work better than the well established orthoganal polynomial. For example, it may be best to fit an exponential function, a sinusoidal function, or simply use a filter designed to eliminate all frequencies above a certain value.

In conclusion, it appears to me there is considerable opportunity to bring the techniques for smoothing data up to the present day needs. There appears to be considerable work in this field at the present time, but to my knowledge, most of the work is found to be in internal company reports and is not readily available to the general public. This immediately points to the need for those agencies which have valuable procedures to make an effort to publish these in scientific journals.

ESTIMATION BY INDIRECT MEANS OF EFFECT OF BACTERIA ON AN UNCHALLENGED HOST

Morris A. Rhian
Biological Warfare Laboratories
U. S. Army Chemical Corps

The data presented for this discussion show two things: (1) the effects on three species of animals of doses of B. anthracis spores and (2) variations in response in the presence or absence of a virulence enhancing factor. For our discussion these data may be regarded as examples of host-causative agent relationships or interactions under several conditions. It is relatively easy to obtain these host-agent interaction data with certain animals, but it is vitrually impossible to obtain comparable data for other animals. If a certain host may not be challenged with a certain disease agent, then one is forced to seek an estimate of the host-agent interaction by indirect means. So the question to be considered in this meeting is: May data from agent-host interactions of the type presented be used to predict the interaction of this agent and an unchallenged host?

B. anthracis spores, prepared in both standard and experimental suspensions, were given by intraperitoneal injection. The effect on the host was measured primarily as the mean time to death of groups of 8 or 10 animals given relatively large doses of spores. A few LD_{50} determinations were made for comparison. In the calculations of mean time to death, the reciprocal transformation was used with values of infinity assigned for time to death of animals that survived the 10-day observation period.

Changes in the effect of the agent on the host were produced by adding egg yolk medium to suspensions of spores. In comparisons of suspensions with and without egg yolk, diluent was used to equalize the concentration of spores in the control suspension. The usual dose by injection was $1/2 \times 10^9$ spores.

In slide 1* the mean times to death for groups of mice given doses of spores from 24 standard and 15 experimental cultures are shown. There was no difference between the two types of culture, but significantly decreased time to death was observed when egg yolk was added to the spore suspensions.

Comparisons of LD_{50} values for mice when spores were injected with and without egg yolk medium are shown in slide 2. Overall, the LD_{50} for mice was reduced approximately 38 fold by the egg yolk. Also, with egg yolk all deaths occurred in 1 day; without egg yolk deaths occurred in 3 to 7 days.

Mean times to death for groups of mice, guinea pigs, and rats, given standard and experimental cultures are shown in slides 3 and 4. In these comparisons of the 3 species, there are instances in which the effects of the treatments were alike in all species and other instances in which they were different. The mean times to death for rats were always less when the spore suspension contained egg yolk, and the effect of the egg yolk was generally greater in the rat than in the other 2 species.

---

* Slides can be found at the end of this paper.

With mice and guinea pigs the mean times to death from spores plus egg yolk were either the same as or less than the comparative times from spore suspensions. However, the responses of these two species were not always the same, and an effect of the egg treatment was shown in one species but not in the other in 4 of the 9 examples.

Questions have occurred to me regarding the type of work illustrated. Do the specific data presented show defects in design or execution of the studies? May data of this kind on the agent-host relationships of several species of animals be used to predict an untried relationship? If the approach illustrated is not satisfactory, is there a way that one may estimate by indirect means the effects of bacteria on an unchallenged host?

SLIDE 1 : Mean Time to Death (MTD) in Hours of Mice Given

B. anthracis Spores a/

	Spores + Diluent	Spores + Egg	Range of Difference, Hr.
$\bar{x}$ of 24 Std. Preps.	10.8	9.2	0 to >6.5
$\bar{x}$ of 15 Exp. Preps.	9.9	7.9	0 to 7.3
Overall $\bar{x}$ of 39	10.5	8.7	0 to 7.3

a/ Culture was mixed with an equal volume of diluent or egg yolk medium, and 1/2 ml was injected intraperitoneally. Dose was approximately 1/2 x 10^9 spores.

SLIDE 2 : LD_{50} of B. anthracis Spores Without and With

Egg Yolk Medium

		LD_{50}	
MICE		Diluent	Egg Medium
Standard Fermentor Culture		218	1
Egg Medium - Shake Flask		383	6.6
Normal Shake Flask		82	5.4
8:1 Ratio Liquid: Agar		3710	65
Normal Shake Flask		320	60
	MEAN	384	10
GUINEA PIG			
8:1 Ratio Liquid: Agar		8500	1000

SLIDE 3 : Mean Time to Death in Hours After Injection of

B. anthracis Spores a/ With Diluent or Egg Yolk

Medium

Culture	Mice Dil.	Mice Egg	Guinea Pig Dil.	Guinea Pig Egg	Rat Dil.	Rat Egg
Std. Ferm.	13.2	11.0	31	22	1980	16
" "	10.6	10.0	36	19	118	19
" "	7.5	7.0	32	16	66	10
" "	10.1	7.5	32	19	57	20

a/ Approximate Doses: Mice - 1/2 x 10^9 Spores; Guinea Pigs - 1 x 10^9 Spores; Rats - 1 x 10^9 Spores

SLIDE 4 : Mean Time to Death in Hours After Injection of

B. anthracis Spores a/ With Diluent or Egg Yolk

Medium

	Mice Dil.	Mice Egg	Guinea Pig Dil.	Guinea Pig Egg	Rat Dil.	Rat Egg
Egg Medium - Firm	7.2	5.8	25	15	6767	7
Normal - Sh. Fl.	14.5	11.5	33	33	300	49
8:1 Liquid/Agar	12.5	11.2	34	22	47	24
Freeze Dried	9.0	7.5	31	22	31	11
Freeze Dried	11.1	9.2	30	17	92	10

a/ Approximate Doses: Mice - 1/2 x 10^9 Spores; Guinea Pigs - 1 x 10^9 Spores; Rats - 1 x 10^9 Spores.

DETERMINATION OF PERFORMANCE CRITERIA FOR QUARTERMASTER CORPS FUNCTIONS

John K. Sterrett
Research and Engineering Division
Office of the Quartermaster General

In the Quartermaster Corps ... the electric accounting machines are being replaced by the high-speed data processing machines at a very accelerated rate. As a result ... the accounting function of the Quartermaster Corps has benefitted tremendously. However, high-speed electronic data processing machines are capable of much more than just counting or keeping records.

It appears logical ... pertinent ... and timely ... to investigate the additionally useful application of these machines in the control and management categories. They are perfectly feasible of application in areas whereby they can aid in the planning and control decisions necessary in logistics enterprises.

As new and more sophisticated data processing machines come into being ... the Quartermaster Corps can be expected to utilize them ... however ... it is folly and wasteful to continue the present practice of giving primary emphasis in these machines to those processes, techniques, and manipulations which formerly were associated with the use of electric accounting machines.

The machines now installed and those programmed for the future are capable of much more than their present contribution to the overall Quartermaster Corps supply mission. Their real potential lies in the fields of supply control ... logistic category specification ... positioning and reporting ... and general stock management. In fact, the major benefits yet to be derived from the present machines will result from the potentials of automation now feasible through the proper utilization of these new devices.

There is every reason to believe that advances in machine technology and computer logic being developed and introduced into the machines of the future ... coincidental with a corresponding integration of logistics and supply operations ... will bring about improvements far greater than as yet has been envisioned. An awareness of ... and alertness to this situation preceded the initiation of the "Study of Future Scientific Quartermaster Corps Control of Inventories" ... a phase of which brings me here before you at this time in the Clinical Sessions.

The overall purpose of the effort in this study is to devise and pursue new and unique approaches to Quartermaster Corps supply management and inventory techniques. It is our intention that the approach shall adequately reflect and be suitably oriented to the potentials inherent in the very latest methods of systems analysis ... further ... we intend that it shall encompass the proper utilization of the most recent advances in the more sophisticated types of high-speed automatic data processing machines.

Significantly important to the successful prosecution of this study is an intimate understanding of the processes necessary for the integrating of the various functions of the supply mission ... namely ... requirements ... procurements ... distribution ... warehousing ... inventory control ... and so forth. No longer can these various phases remain independent.

This integration can only be accomplished through provision and utilization of technological and functional improvements far in excess of anything now in being ... and most likely foreign to most of those things which some people might hold as the proper way to get things accomplished. ... And these people are the ones who must be convinced that this new thing being thrust upon them is really an improvement ... they have to be shown ... and doing so must not interrupt daily operations ... now how do we do it ????? That is one of our many problems which I hope you can shed some light on .....

It is emphasized here that we feel that technological advances are not in themselves sufficient ... the very best mechanization can fall far short of a desired goal if functional relationships also involved in automation activities are not fully understood, appreciated, and made an essential part of the automation flow. Machines and functions must both be integrated if the best in each is to result.

Carried to its logical conclusion in this study ... this could mean the integration of a large number of the now separated functions of the Quartermaster Corps ... integrating of such functions as requirements, distribution, storage, issue, and disposal actions ... and all these being served by data generated by a single centrally controlled data processing organization ... properly manned ... adequately equipped ... and functioning as an integrated whole to the best advantage of all concerned.

Great strides have recently been made in the construction of mathematical models to be used in describing and studying involved business systems and other types of extremely complex management operations. All of these seem to involve a desirable detachment from biased conslusions when too close adherence is held to intuitive and qualitative judgements ... also involved are gaming types of procedures involving the high-speed computers themselves. We are prepared to provide such analysis techniques and high-speed computer applications to our problem too. The trouble as we view the problem is ... where do the standards of comparison come from to determine whether or not one system of approach is better than another ... in fact ... is even better than the present one in existence ... and by how much ?????

This ... then ... is the essence of the problem I bring to you today ...

Before any comparison can be made of replacement techniques ... if any are forthcoming ... an agreement must be reached as to what measures are to be considered pertinent and what combinations of these measures constitute "acceptable" or "superior" performance. No decision on any given technique or method can be reached prior to the identification, acknowledgement and establishment of a set of critieria on which judgements and evaluations are to be based.

Having set such a criteria ... a means must be provided by which various approaches can be tried and the results recorded for comparison. This is what some might call a "controlled experimentation" ... conceivably it could involve complex mathematical models and their manipulations under simulated conditions to those encountered within Quartermaster Corps operations and functions.

Most of the presently employed methods of approach of this sort involve rather extensive utilization of high-speed computing machines because of their easy adaptability to gaming theory which these machines permit ... as well as their large capacity for rapid manipulation which they have as internal functions.

We have, then, three major problems for which we are seeking assistance:

1. The means to obtain the substantiating data upon which the desired set of standards is to be based.

2. The weighting to be assigned to each source of this substantiating data.

3. The manner in which the standards resulting can be applied to arrive at appropriate ratings.

PROGRAM FOR THE INTERLABORATORY DETERMINATION OF COMPRESSION SET OF ELASTOMERS AT LOW TEMPERATURES

S. L. Eisler
Rock Island Arsenal

The purpose of this planned program is to compare the reproducibility of low temperature measurements of the compression set of vulcanized elastomers. This test measures the ability of elastomers to recover, at subzero temperatures, from compressive deformation applied at room temperature.

The Ordnance Materials Research Office and the Elastomers Unit of the Rock Island Arsenal Laboratory have been assigned the responsibility for designing the program, preparing the test specimens for all participants and analyzing the results. This assignment was made by Working Group 4 of Technical Committee 45 of the International Organization for Standardization.

The program, as submitted to the various participants for review and possible suggested changes, contains the following variables:

1. Specimen Sizes (2)
2. Test Temperatures (2)
3. Rubber Compositions (3)
4. Laboratories (7)
5. Replications (duplicates to be run on each of 2 days)

We would like to be able to determine if there is a significant difference in reproducibility between:

1. Laboratories
2. Specimen sizes
3. Test temperatures

and (4) within laboratories between the two test days.

Similar compression set programs have been conducted by individual laboratories involving a smaller number of variables. In these cases, the results have usually been analyzed by means of a series of "F" tests. This, of course, involves testing each level of the second variable separately and often it is not possible to arrive at a definite overall conclusion. For example, compression set measurements were made in one laboratory on nine different elastomers using two specimen sizes. The resultant "F" tests showed a significant difference between specimen sizes in the case of three elastomers but not for the other six elastomers.

The results of another program involving three laboratories, three compounds and two methods were analyzed in a different manner. In this program, each laboratory ran duplicates on each of four days for each compound-method combination. The results were analyzed by preparing an Analysis of Variable Table for each of the six compound-method combinations. A typical table is shown below:

Source	Σ of Sq.	d. f.	m. s.	"F"
Bet. Labs.	34.21	2	17.10	
Bet. Days	24.00	3	8.00	
D X L int.	41.12	6	6.85	3.42*
Within Days	24.02	12	2.00	
Total	123.35	23		

In this example, due to the significant interaction, it was reported that the residual (within days) variance could not be used as a measure of experimental error. Therefore, it was necessary to consider the means of the pairs of duplicates instead of the original readings. This error variance was obtained from the following equation:

$$NS_1^2 + S_2^2 = S_3^2$$

wherein:

S_1^2 = error variance required

S_2^2 = residual variance = 2.00

S_3^2 = pooled main effects and interaction mean squares = 9.03

N = no. of replications = 2

The error variances thus calculated for the two methods were compared by means of an "F" test to determine whether a significant difference existed between the precision of the two methods. This, however, involved three separate "F" tests one for each compound.

The question I should like to present to the panel at this time is what is the most efficient method for analyzing the results of the proposed program in order to compare the reproducibilities between laboratories, specimen sizes and test temperatures as well as within laboratories.

AN APPRAISAL OF SEQUENTIAL ANALYSIS UNDER CONDITIONS RESTRICTED BY THE REQUIREMENT FOR ADVANCED SCHEDULING AND PROGRAMMING

Edgar W. Larson and Walter D. Foster
Biological Warfare Laboratories
U. S. Army Chemical Corps

1. INTRODUCTION. The design of experiments may be broadly defined as the vehicle used to provide answers to questions posed by its partner and teammate, the subject matter field. More and more widely the answers are being accepted on the underlying basis of probability. To narrow the scope of this paper immediately, we have selected from the many current designs that which is known as Wald's Sequential Analysis. Our thinking and limited experience in its use with respect to testing devices designed to aerosolize bacterial suspensions are reported here. There is reason to believe that the principle of sequential analysis may be useful in increasing the efficiency of our testing efforts which are restricted by the requirement for advanced scheduling and programming.

2. TECHNICAL CHARACTERISTICS OF THE PROBLEM. The aerosolizing devices are essentially mechanical, ordnance type, and may use compressed gas, electricity, burning propellants, pyrotechnic fuels, high explosives, or combinations of these as energy sources for the dissemination of bacteria in small airborne particles, starting from concentrates of the organisms either in liquid suspension or as dry powder. The primary responsibility for development of a particular device rests with a design engineer. Several devices undergo research and development concurrently. When a device is in the concept stage it is possible and necessary to delineate the design variables which can conceivably affect performance, disseminating efficiency, and make decisions concerning the practical range of test levels for each variable within which aerosolization performance must be measured. The object of the research and development is to determine the treatment combination or combinations which can be expected to render airborne in small particles the greatest number of viable bacteria from the initial suspension, hereinafter referred to as fill. Further, it is desired that such treatments be expected to produce bacterial aerosols which decay with time after dissemination at a minimum rate. Hence, at least two parameters are required to summarize the results of a single aerosol test. One of these reflects the degree of aerosol stability with time, i.e., a measure of the decay rate, while the second reflects the level of recovery, either the regression intercept or mean.

In conducting aerosol tests, closed, aerosol-tight, testing chambers are employed. The chamber atmosphere is conditioned with respect to temperature and relative humidity; the disseminator is positioned centrally within the chamber and charged with a measured amount of fill, the bacterial density of which has been previously determined; the device is energized and dissemination takes place. The resulting aerosol, which is allowed to age as long as an hour, is sampled periodically for concentration. The basic datum from the trial is per cent recovery computed simply as the percentage of the numbers of airborne bacteria to the numbers of bacteria contained in the initial fill charge. The data are subsequently subjected to a logarithmic transformation because the transformation tends to stabilizevariances between sampling periods and from treatment to treatment and (very conveniently) because the plot of log per cent recovery versus cloud age tends to describe a straight line.

There are several reasons for scheduling and programming in advance. Firstly, the needs of several development programs must be satisfied. Secondly, for the most part the treatment combinations for test are determined by the results from prior experimentation. When decisions are reached concerning the treatments which will be subjected to test, time must be allotted for preparation of drawings, for procurement of fabrication materials and for scheduling machine shop time. Further, the procurement, preparation and characterization of fill materials is time consuming and must be accomplished immediately prior to use because of the instability of the bacteria in suspension. Also, testing technology requirements change from experiment to experiment, laboratory glassware and equipment requirements are changed and there are changes in bacterial growth media and suspending fluid requirements depending upon the experimental objectives. Hence, complex scheduling problems are imposed.

3. EXPERIMENTAL DESIGN CONSIDERATIONS. For the most part, heretofore, we have capitalized on the use of experimental designs with sample sizes fixed in advance. Because of the nature of our problem and the characteristics of the testing system, balanced factorial designs, randomized in either complete or incomplete blocks, have been employed to the greatest extent. It is not uncommon that the design engineer may have a need to investigate the effects of as many as ten variables. If one subjects each variable to test at only three levels, it is obvious that many thousands of treatment combinations are made available. Of course, every treatment combination is not examined, but because the problems are subject to interacting variables, high order (3^4 designs, for example) factorial experiments are executed. The choice of variables for inclusion in a single experiment is generally made from engineering consideations; the engineer has the option, of course, to go back and combine variables from test to test in additional experiments. Our natural experimental block is limited by the number of trials it is possible to complete in a single working day in one aerosol chamber, i.e., from six to nine per day. The aerosol data, reduced to regression intercepts and slopes, are commonly subjected to classical analyses of variance according to the selected design. This procedure possesses certain shortcomings which we would like to overcome. Firstly, the analysis of variance computations are numberous and often involved. Further, there is a tendency to ignore type II errors, i.e., the error of accepting the null hypothesis when it is false. Finally, when an inherently variable biological system is involved whose variance is neither well established nor consistent, there are risks of either under-testing or over-testing. Under-testing fails to yield information permitting a decision while over-testing is expensive. Thus, the approach using fixed sample size is desirable from some standpoints and unsatisfactory in others.

4. SEQUENTIAL ANALYSIS. In the interest of reaching decisions in shorter testing time, we are exploring the possibility of using sequential designs, starting with Wald's designs. Briefly, we want here to review what these designs are and what questions they can answer for us. Primarily, the concept involves testing a null hypothesis against a specific alternative hypothesis with respect to a population mean or variance, offering either one or two sided tests. Knowledge of the population variance is required. When the rates of error, α and β, are specified, together with the alternative hypothesis, the

design is complete. Analysis is achieved by computing a simple statistic which is either tabled or plotted. A decision is reached when the statistic exceeds either of the two bounds, which graphically are shown as the familiar pair of parallel lines. As originally derived, the design was applicable only to a single mean or variance. By a simple modification (undoubtedly discovered and rediscovered by countless users), two means can be accommodated by writing the hypothesis with respect to the difference, remembering, of course, to use the variance of the difference in constructing the design. This modification especially lends itself to the conduct of paired trials.

Naturally with respect to our own requirements, this kind of analysis has certain advantages and disadvantages. Among the disadvantages is this restriction to only two treatments when there are many which need testing. There is no opportunity to estimate interaction, a most important consideration in development work. A third, and again a most serious disadvantage, is the inability to know precisely the termination date in the sequential testing. Finally, using as we do a biological response to evaluate a candidate treatment, we are not always sure we know the variance. Sometimes we can say we know it with confidence; other times not at all. Of course, it is possible in the doubtful cases to resort to the sequential "t" test; however, the only way to relate the scale of standard deviations needed in the "t" test to the scale of measurement such that the alternative hypothesis would then have meaning is through knowledge of the variance or coefficient of variation - a self-contradicting situation.

But on the brighter side, the advantages include the most highly prized desideratum, namely, reduced testing time, which, of course, means reduced testing expense if the difficulties in programming can be obviated. Another advantage comes in requiring the experimenter to consider an alternative hypothesis together with Type I and Type II errors - a concept still relatively unknown outside of statistical circles, especially considering how widely accepted the term "significant difference" is. Finally, nothing appeals to an experimenter more than an analysis which is completed on the same day as the last trial.

5. AN EXAMPLE IN SCREENING. In one of our development projects, dissemination of relatively large quantities of dry fill in an aerosol chamber was required. However, for reasons of both safety and technical feasibility, we could not tolerate the large numbers of bacteria involved if undiluted fill were to be employed. Therefore, we programmed an experiment to search for a diluent which in aerosol would yield results percentage-wise similar to those expected in an undiluted material. Five treatments were included in the experiment: one-to-ten dilutions of the dry bacteria in Microcele, Estercil, cornstarch and talcum - all commercial products - and a one-to-ten dilution of the dry bacteria in the same material previously sterilized. Four trials in a day were completed with the same experimental treatment and two trials were completed with the reference, the undiluted material. We computed the mean for the two reference trials, then developed four differences, one for each trial, from the results with the experimental treatment. These differences were obtained for each of four aerosol parameters including: the intercept and slope from results with a sampler collecting only small aerosol particles and the intercept and slope from results with a sampler collecting only large aerosol particles. The experiment was conducted

in five-day cycles, testing a different treatment each day until on the sixth day, treatment one came up again less our analysis from the first day had indicated we could reach a decision. By choice we specified that when rejection was indicated by any one parameter we would discontinue testing. By the same token final acceptance required all four parameters to be acceptable. For each treatment with respect to the reference we set up the null hypothesis of zero difference against the alternative of .1761 which is the equivalent of a 50% difference in log scale. Type I and II errors were controlled at 5% and 10%, respectively. Analysis during the course of the trials consisted of computing the statistic, ΣD, where D was the difference between treatments.

According to this design, all of the candidate diluents were rejected. Starch and Estercil were rejected after four trials; talcum and Microcele and the sterilized material were all rejected after eight trials. Since the minimum number of trials per day was four, we actually over-tested by two trials for the Microcele and by three trials for the talcum. Executing the experiment as we did, one candidate each day, the scheduling problem was simplified to some extent. At the end of this first cycle of five we knew that part but not all of a second round would be necessary. While we could predict roughly the end of testing, sufficient uncertainty remained to require twelve trials for the sterilized diluent before testing was stopped. All in all, though this constituted a screening type of experiment and therefore was a departure from the usualy type of study, it is considered that the sequential analysis approach served to answer the experimental objective in this case efficiently. Further, the order in which we chose to subject the treatments to test minimized the scheduling problem.

6. SUMMARY AND CONCLUSIONS. Summarizing this discussion, we have considered the sequential analysis approach with respect to its possible shortcomings and advantages when applied to the problem if testing for a research and development program, unique from the standpoint that engineered devices are evaluated by a biological response. Shortcomings of its use are: (a) it is restricted to only two treatments, (b) it provides no opportunity to estimate interaction, (c) if further complicates already complex work scheduling, and (d) it depends upon the hazardous assumption in biological response situations that the population variance is known. Advantages of the approach are listed as follows: (a) it minimizes the amount of testing required, (b) it avoids the problem of serious under-testing and over-testing, (c) it requires the experimenter to consider an alternative hypothesis together with Type I and Type II errors, and (d) it provides immediate answers. Our experience has been limited to the use of sequential analysis in screening type experiments. As applied, the approach appeared to answer the experimental objectives efficiently. Over-all it is concluded that sequential analysis possesses characteristics which limit its value for our purposes. However, under certain conditions this design may constitute a desirable choice among current methods and further study of the concept may produce information broadening its application.

SIMPLIFIED PROCEDURES FOR ESTIMATING PARAMETERS OF A NORMAL DISTRIBUTION FROM RESTRICTED SAMPLES

A. Clifford Cohen, Jr.
The University of Georgia

1. INTRODUCTION. In life testing, analysis of inspection data, dosage-response studies, biological assays, target analyses, and in other related investigations, it is frequently necessary to estimate distribution parameters from restricted samples, in particular from truncated and from censored samples. Truncated samples are those from which certain of the population values are entirely excluded. Censored samples are those in which sample specimens whose measurements fall in restricted intervals of the random variable may be identified and thus counted, but not otherwise observed. Samples of these types are further classified as singly or doubly restricted, depending on whether sample observation is restricted in only one or in both tails of the distribution. Depending on which tail of the distribution is involved, singly restricted samples are still further classified as left or right restricted.

Unfortunately, calculating estimates from samples of these types often involves the soltuion of complicated non-linear equations, a task which is likely to be tedious and time consuming even when appropriate tables are available. Here, we are concerned with reducing this computational labor to a reasonable level for the practical calculation of maximum likelihood estimates of the mean and variance of a normal distribution with probability density function

$$(1) \quad f(x) = (\sigma\sqrt{2\pi})^{-1} \exp -\left[(x-\mu)^2/2\sigma^2\right] \quad -\infty \leq x \leq \infty .$$

The present paper represents a consolidation of results given in [4] for doubly truncated samples and in [6] for singly restricted samples.

For singly restricted samples, the required estimates are obtained by adding simple easily computed corrections which involve only a single auxiliary function of the sample terminus to the sample mean and variance respectively. Calculation of estimates accordingly involves interpolation in only one table. With the exception of estimators given by Gupta [8] , who considered singly censored samples only, previous applicable maximum likelihood estimators have involved two or more auxiliary functions and therefore interpolation in two or more separate tables. Estimators derived by Ipsen [11] for singly censored samples from a normal distribution also involve only a single auxiliary estimating function and thus interpolation in only one table. However, his estimators, which are based on certain moment functions of the restricted distribution, differ slightly from applicable maximum likelihood estimators. Furthermore, his tabular intervals are too wide and his entries contain too few significant digits for accurate interpolation. Gupta's maximum likelihood estimators employ an auxiliary function which unfortunately lacks linearity even over short intervals of his argument. Consequently, his tabular intervals also are in many instances too wide for easy interpolation. Auxiliary functions employed here are approximately linear over moderately wide intervals of the arguments for both truncated and censored samples, so that accurate interpolation between table entries is relatively easy in both cases. Tables and graphs of these auxiliary functions are appended.

In the case of doubly truncated samples, a chart is provided which permit a graphic reading of estimates of the standardized terminals to one or perhaps two decimals, and thus the immediate calculation of estimates of the mean and standard deviation to two or perhaps three significant digits. When greater precision is required, iterative procedures described in [4] may be employed to improve initial approximations obtained from the chart.

Since estimators of this paper were derived by the method of maximum like lihood, for a given sample they lead to identical estimates except for possibl errors of calculation that might be obtained from applicable maximum likelihoo estimators previously obtained by Fisher [7] , Hald [9] , Halperin [10] , Gup the author [1] , and possibly by others. The computing routine given here, ho ever, is believed to be much simpler and easier to carry out. As with maximum likelihood estimators in general, those for truncated and censored samples are consistent and asymptotically efficient. They are to be recommended when samp sizes are at least moderately large. When estimates must be based on samples of size 10 or less - perhaps even on slightly larger samples, it might be preferable to employ linear unbiassed estimators based on order statistics as giv by Gupta [8] in the latter part of his paper and by Sarhan and Greenberg [13]

For the benefit of readers who may wish to delve further into the subject of restricted sampling, a list of some of the pertinent references is appended

2. SINGLY TRUNCATED SAMPLES. Let x_o be a known fixed value of the rando variable, x, which we designate as a terminus or truncation point. Now consid a sample consisting of n observations (values) of this random variable, such that for each observation (i.e. for each sample value), either

(a) $x \geq x_o$, in which case truncation is on the left,

or

(b) $x \leq x_o$, in which case truncation is on the right.

The number of otherwise possible sample values excluded from observation as a consequence of this restriction is not known.

Throughout this paper, we limit our consideration to a random variable with probability density function (1). Since this function is symmetrical about μ, truncation of f(x) on the right at x_o is equivalent to truncation of f(-x) on the left at $-x_o$. Consequently, it is necessary to examine only one of these cases in detail, and for this role, truncation on the left has been selected.

Let F(x) designate the distribution function of x and the probability that a selected value of this random variable meets the requirements for inclusion in a sample that is singly truncated on the left at x_o is given as $1 - F(x_o)$ or in standard units as $1 - f(\xi)$, where

(2) $F(\xi) = \int_{-\infty}^{\xi} \phi(t)\, dt$, with $\xi = (x_o - \mu)/\sigma$, and $\phi(t) = (\sqrt{2\pi})^{-1} \exp - t^2/2$.

The likelihood function for a sample of the type under consideration is

(3) $P(x_o, x_1, \ldots x_n; \mu, \sigma) =$

$$\left[1 - F(\xi)\right]^{-n} (\sigma\sqrt{2\pi})^{-n} \exp\left[-\Sigma_1^n (x_i - \mu)^2/2\sigma^2\right] .$$

Maximum likelihood estimating equations follow as

$$x_o - \mu = \sigma \xi,$$

(4) $$\bar{x} - \mu = \sigma Z,$$

$$s^2 + (\bar{x} - \mu)^2 = \sigma^2 \left[1 + \xi Z\right],$$

where x and s^2 are the sample mean and variance respectively

($\bar{x} = \Sigma_1^n x_i/n$ and $s^2 = \Sigma_1^n (x_i - \bar{x})^2/n$), and where

(5) $$Z(\xi) = \phi(\xi) / \left[1 - F(\xi)\right] .$$

The first equation of (4) follows from the second equation of (2). The last two result from taking logarithms of (3), differentiating with respect to μ and σ^2 in turn, and equating resulting derivatives to zero. The required estimators, $\hat{\mu}$, $\hat{\sigma}^2$ and the auxiliary estimator $\hat{\xi}$ are to be found as simultaneous solutions of (4) in terms of the sample statistics, $\bar{x}$, x_o, and s. Throughout this paper, the symbol ($\hat{}$) serves to distinguish maximum likelihood estimators from the parameters being estimated.

On eliminating $(x - \mu)$ between the last two equations of (4), we have

(6) $s^2 = \sigma^2 \left[1 - Z(Z - \xi)\right]$ or $\sigma^2 = s^2 + \sigma^2 Z(Z - \xi)$.

Eliminating μ between the first two equations of (4) leads to

(7) $\bar{x} - x_o = \sigma(Z - \xi)$ or $\sigma = (\bar{x} - x_o) / (Z - \xi)$.

Combining (6) and (7) given

$$\sigma^2 = s^2 + \left[Z/(Z - \xi)\right] . (\bar{x} - x_o).$$

Now let

(8) $$\theta(\xi) = \frac{Z(\xi)}{Z(\xi) - \xi} ,$$

and the estimating equation for σ^2 assumes the form

(9) $\sigma^2 = s^2 + \theta(\bar{x} - x_o)^2$.

To derive a corresponding equation for estimating μ which does not involve any auxiliary function other than θ, we eliminate $(Z - \xi)$ between (6) and (7) to obtain $\sigma Z = (\sigma^2 - s^2) / (\bar{x} - x_o)$.

On combining this result with (9), we have

(10) $\sigma Z = \theta(\bar{x} - x_o)$.

When (10) is substituted into the second equation of (4), we write the desired estimating equation as

(11) $\mu = \bar{x} - \theta(\bar{x} - x_o)$.

By eliminating σ between (6) and (7), we obtain the more familiar result*

(12) $\left[1 - Z(Z - \xi)\right] / (Z - \xi)^2 = s^2 / (\bar{x} - x_o)^2$.

The system of estimating equations (4) may now be replaced by the equivalent system consisting of (9), (11), and (12). Let $\hat{\xi}$ designate the solution of (12), let $\hat{\theta} = \theta(\hat{\xi})$, and the desired estimators become

$\hat{\sigma}^2 = s^2 + \hat{\theta}(\bar{x} - x_o)^2$,

(13) $\hat{\mu} = \bar{x} - \hat{\theta}(\bar{x} - x_o)$.

As computational aids, tables, and a graph of θ as a function not of ξ, but of $\left[1 - Z(Z - \xi)\right] / (Z - \xi)^2$ have been provided. Since $\hat{\xi}$ is that value of ξ for which $\left[1 - Z(Z - \xi)\right] / (Z - \xi)^2 = s^2 / (\bar{x} - x_o)^2$, we can thereby determine $\hat{\theta}$ directly for any given sample as that value of θ which corresponds to the sample statistic $s^2 / (\bar{x} - x_o)^2$. Accordingly, the necessity for determining $\hat{\xi}$ explicitly prior to calculating $\hat{\theta}$ is eliminated, and since θ is the only auxiliary function appearing in the estimators (13), only the single table of that function is needed in contrast to the two or more tables necessary when employing estimators previously proposed.

Entries of θ in Table 1 were computed from existing tables of normal curve areas and ordinates at equal intervals of ξ . This, of course, resulted in unequal intervals of the argument $s^2 / (\bar{x} - x_o)^2$. Although equal intervals of

---

* cf. for example reference [1] .

this argument might be desirable, a degree of accuracy adequate for most practical applications can be achieved through simple linear interpolation, and the table has proven easy enough to use even with unequal intervals. In view of this fact, and since any graduation for the purpose of equalizing intervals would either result in a loss of significant digits or require complete recomputation of the table, it is offered in its present form.

With x_o, $\bar{x}$, s^2, and accordingly $s^2/(\bar{x} - x_o)^2$ available from the sample data, it is necessary only that we read $\hat{\theta}$ from the table or graph as required and calculate $\hat{\sigma}^2$ and $\hat{\mu}$ from (13). In many applications, $\hat{\theta}$ may be read with sufficient accuracy from the graph of Figure 1. When more accurate values are required, they can be obtained by direct reading or by linear interpolation from Table 1. Only in rare cases should it be necessary to resort to more complicated non-linear interpolative procedures.

Once $\hat{\sigma}^2$ and $\hat{\mu}$ have been computed, $\hat{\xi}$ follows from (2) without the need of additional tables as

(14) $$\hat{\xi} = (x_o - \hat{\mu})/\hat{\sigma} \quad ,$$

where $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$

Although estimators (13) have been derived for samples that are <u>singly truncated on the left</u>, they are equally applicable when samples are <u>singly truncated on the right</u>, as a consequence of the symmetry of the normal probability density function (1). In both cases $\theta > 0$, and as shown in the sketch below, when

<u>truncation is on the left</u>,

$(\bar{x} - x_o) > 0$, and

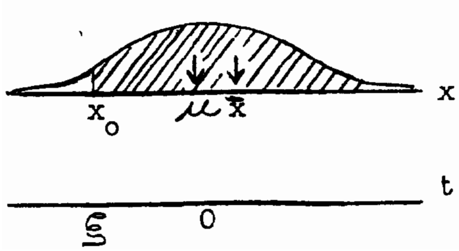
$\mu < \bar{x}$, whereas when

<u>truncation is on the right</u>,

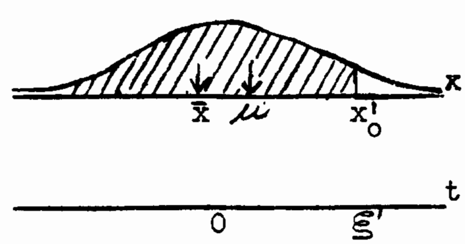
$(\bar{x} - x_o) < 0$, and

$\mu > x$,

TRUNCATION ON THE LEFT AT x_o



TRUNCATION ON THE RIGHT AT x_o'



On the lower scale, t is the standardized normal deviate, $t = (x - \mu)/\sigma$. Thus, when x is normal (μ, σ), t is normal (0,1), and if $x_o = -x'_o$, it follows that $\xi = -\xi'$.

3. SINGLY CENSORED SAMPLES. We consider two types of censored samples. A Type 1 Censored Sample is one in which the terminus or point of censoring is fixed, while a Type II Censored Sample is one in which the number of censored observations is fixed. Within each of these categories, we may have censoring either on the right or left. Here as in the case of truncated samples, the symmetry of f(x) makes it unnecessary to consider both left and right censoring in detail, and again our derivations are confined to left censored samples.

Type 1 Singly Censored Samples

In this category, we consider samples consisting of a total of N observations subject to the restriction that full measurement (i.e. unrestricted observation) of the random variable x is possible if and only if

(a) $x \geq x_o$, in which case censoring is on the left,

or

(b) $x \leq x_o$, in which case censoring is on the right,

where x_o is a known fixed terminus. Let n designate the number of fully measured observations and n_1 the number of censored observations for which it is known only that $x < x_o$ ($x > x_o$, in the case of censoring on the right). Since x_o and N are fixed, both n and n_1 are random variables subject to the condition that $n_1 + n = N$.

The likelihood function for a sample of this type is

$$(15) \quad p = \left[N!/n_1!\, n! \right] \left[F(\xi) \right]^{n_1} \cdot (\sigma \sqrt{2\pi})^{-n} \cdot \exp\left(-\Sigma_1^n (x_i - \mu)^2 / 2\sigma^2\right),$$

where ξ and $F(\xi)$ are given by (2).

In this case the maximum likelihood estimating equations are

$$(16) \quad \begin{aligned} x_o - \mu &= \sigma \xi, \\ \bar{x}_2 - \mu &= \sigma Y, \\ s^2 + (\bar{x} - \mu)^2 &= \sigma^2 \left[1 + \xi Y \right], \end{aligned}$$

where

$$(17)\quad Y(h, \xi) = \left[h/(1-h)\right] Z(-\xi), \text{ with } h = n_1/N.$$

The first equation of (16) comes from (2) and is identical with the first equation of (4) for the truncated case. The last two equations of (16) result from taking logarithms of (15), differentiating with respect to μ and σ^2 in turn and equating to zero. Here as in the truncated case, $\bar{x}$ and s^2 are the sample mean and variance respectively.

Estimating equations (16) which apply in the censored case differ from equations (4) which apply in the truncated case only in that $Z(\xi)$ appearing in (4) has been replaced in (16) by $Y(h, \xi)$ which is defined by (17). Procedures analogous to those employed in the truncated case enable us to replace the system of equations (16) with the equivalent system.

$$\sigma^2 = s^2 + \lambda(\bar{x} - x_0)^2,$$

$$(18)\quad \mu = \bar{x} - \lambda(\bar{x} - x_0) ,$$

$$\left[1 - Y(Y - \xi)\right] / (Y - \xi)^2 = s^2/(\bar{x} - x_0)^2 ,$$

where

$$(19)\quad \lambda(h, \xi) = Y(h, \xi) / \left[Y(h, \xi) - \xi\right].$$

Let $\hat{\xi}$ designate the solution of the third equation of (18), let $\hat{\lambda} = \lambda(h, \hat{\xi})$ and the desired estimators become

$$(20)\quad \hat{\sigma}^2 = s^2 + \hat{\lambda}(\bar{x} - x_0)^2 ,$$

$$\hat{\mu} = \bar{x} - \hat{\lambda}(\bar{x} - x_0) .$$

As computational aids in this case, tables and graphs of λ as a function of h and $\left[1 - Y(Y - \xi)\right] / (Y - \xi)^2$ have been prepared. Since $\hat{\xi}$ is the solution of the third equation of (18), we determine $\hat{\lambda}$ directly for any given sample as that value of λ which corresponds to the sample statistics h and $s^2 / (\bar{x} - x_0)^2$. As in the truncated case, only one table is required.

With h, x_0, $\bar{x}$, s^2 and therefore $s^2/(\bar{x} - x_0)^2$ available from the sample data, it is necessary only that $\hat{\lambda}$ be read from Table 2 (using two-way linear interpolation) or from the graphs of Figures 2 or 3 as required, and that $\hat{\sigma}^2$ and $\hat{\mu}$ be calculated using estimators of (20). In Figure 2, λ is graphed for h = 0(.01).27, while in Figure 3, it is graphed for h = 0(.05).75. In both figures 2 and 3, the range of $s^2/(\bar{x} - x_0)^2$ is (0,1.3). In Table 2, λ is given to 4D for h = .01(.01).05(.05).50 and for $s^2/(\bar{x} - x_0)^2$ = 0(.05)1.00.

As in the truncated case $\hat{\xi}$ can, when required, be obtained from (14). Estimators (20) are equally applicable to both left and right censored samples for the same reasons that estimators (13) apply to both left and right truncated samples.

Type II Singly Censored Samples

In this category, we can consider samples consisting of N observations of a random variable with probability density function (1) such that

(a) the smallest N - n observations are counted but not otherwise measured (in which case censoring is on the left),

or

(b) the largest N - n observations are counted but not otherwise measured (in which case censoring is on the right).

Let x_n designate the smallest (or largest) completely measured observation, and the sample thus consists of n completely measured observations each of which is equal to or greater than x_n (or equal to or less than x_n) plus N - n unmeasured observations about which it is known only that $x < x_n$ (or $x > x_n$).

Estimators for this case turn out to be identical with those of (20) for Type I Singly Censored Samples, when we let

$$x_n = x_o ,$$

$$N - n = n_1 . \tag{21}$$

Although there are no essential differences between estimators for Type I Singly Censored Samples and those for Type II Singly Censored Samples, variances of these estimators differ in the two cases as may be noted in Section 5.

4. DOUBLY TRUNCATED SAMPLES. In this section we consider a sample consisting of n observations of random variable x which has probability density function (1) such that each observation is subject to the restriction $x_o \leq x \leq x_o + w$, where sample terminals x_o and $x_o + w$ are fixed. The logarithm of the likelihood function for a sample of this type is

$$L = -n \ln \left[F(\xi_2) - F(\xi_1) \right]^{-n} - n \cdot \ln\sigma - \left(\Sigma_1^n (x_i - \mu)^2 / 2\sigma^2\right) + \text{constant},$$

where $\xi_1 = (x_o - \mu)/\sigma$ and $\xi_2 = (x_o + w - \mu)/\sigma$.

As derived in [4], maximum likelihood estimating equations may be reduced to

$$\left[\bar{z}_1 - \bar{z}_2 - \xi_1 \right] / (\xi_2 - \xi_1) = (\bar{x} - x_o)/w,$$

$$\left[1 + \xi_1 \bar{z}_1 - \xi_2 \bar{z}_2 - (\bar{z}_1 - \bar{z}_2)^2\right] / (\xi_2 - \xi_1)^2 = s^2/w^2 ,$$

where

$$\bar{z}_i = \phi(\xi_i) / \left[F(\xi_2) - F(\xi_1)\right] , \; i = 1, 2.$$

With $(\bar{x} - x_o)/w$ and s^2/w^2 computed for any given sample, coordinates of the intersection of the corresponding pair of curves in Figure 4 are the required values of $\hat{\xi}_1$ and $\hat{\xi}_2$. With care, these values can be read to within three to five units in the second decimal. The desired estimates then follow as

$$\hat{\sigma} = w/(\hat{\xi}_2 - \hat{\xi}_1), \text{ and } \hat{\mu} = x_o - \hat{\sigma}\hat{\xi}_1 .$$

5. SAMPLING ERRORS OF ESTIMATES. The asympotitic variance-covariance matrix of $(\hat{\mu}, \hat{\sigma})$ is obtained by inverting the matrix whose elements are negatives of expected values of the second order derivatives of logarithms of the likelihood functions. Accordingly we obtain

$$\begin{aligned} V(\hat{\mu}) &\sim \left[\sigma^2/E(n)\right] \left[\hat{\phi}_{22}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)\right], \\ V(\hat{\sigma}) &\sim \left[\sigma^2/E(n)\right] \left[\hat{\phi}_{11}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)\right], \\ \mathrm{Cov}(\hat{\mu}, \hat{\sigma}) &\sim \left[\sigma^2/E(n)\right] \left[-\hat{\phi}_{12}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)\right], \end{aligned} \tag{22}$$

where $E(n)$ is the expected values of the number of completely measured observations, and ϕ_{11}, ϕ_{12}, ϕ_{22} are respectively $-\left[\sigma^2/E(n)\right] E(\partial^2 L/\partial\mu^2)$,

$$-\left[\sigma^2/E(n)\right] E(\partial^2 L/\partial\mu\partial\sigma), \text{ and } -\left[\sigma^2/E(n)\right] E(\partial^2 L/\partial\sigma^2).$$

To simplify the notation, L has been written for ln P, and $\hat{\phi}_{ij}$ for $\phi_{ij}(\hat{\xi})$ or $\phi_{ij}(h, \hat{\xi})$ as applicable.

In truncated samples and in censored samples of Type II, n is fixed and therefore $E(n) = n$. In samples that are Type I singly censored on the left, $E(n) = N\left[1 - F(\hat{\xi})\right]$. The ϕ_{ij} for singly truncated and for singly censored samples of types I and II are

Truncated Samples

$$\phi_{11}(\xi) = 1 - z(\xi)\left[z(\xi) - \xi\right],$$

$$\phi_{12}(\xi) = z(\xi)\left\{1 - \xi\left[z(\xi) - \xi\right]\right\},$$

$$\phi_{22}(\xi) = 2 + \xi\phi_{12}(\xi),$$

Type I Censored Samples

$$\phi_{11}(\xi) = 1 + z(\xi)\left[z(-\xi) + \xi\right],$$

$$\phi_{12}(\xi) = z(\xi)\left\{1 + \xi\left[z(-\xi) + \xi\right]\right\}$$

$$\phi_{22}(\xi) = 2 + \xi\phi_{12}(\xi),$$

(23) Type II Censored Samples

$$\phi_{11}(h, \xi) = 1 + Y(h, \xi) \left[Z(-\xi) + \xi\right],$$

$$\phi_{12}(h, \xi) = Y(h, \xi) \left\{1 + \xi\left[Z(-\xi) + \xi\right]\right\},$$

$$\phi_{22}(h, \xi) = 2 + \xi\phi_{12}(h, \xi)$$

It is to be noted that as $N \rightarrow \infty$, the ϕ_{ij} for type II censored samples approach the ϕ_{ij} for type I censored samples. Likewise as $N \rightarrow \infty$, with the ratio n/N held fixed, then $\left[1 - F(\hat{\xi})\right] \rightarrow n/N$. In this sense, limiting values of variance and covariance of estimates become equal in the two cases.

The variance and covariance of estimators based on samples that are restricted on the left, may be calculated by substituting appropriate values of $\phi_{ij}(\hat{\xi})$ from (23) into (22), where $\hat{\xi} = (x_0 - \hat{\mu})/\hat{\sigma}$ or $\hat{\xi} = (x_n - \hat{\mu})/\hat{\sigma}$ as applicable. For samples that are restricted on the right, calculations are the same except that $\phi_{ij}(-\hat{\xi})$ from (23) are substituted into (22).

In the case of doubly truncated samples, variance and covariance of estimates may be computed as described in [4].

The assistance of Mr. Walt G. Herstman, who performed the calculations necessary for the compilation of Tables 1 and 2, and who rendered material aid in the preparation of the charts of Figures 1, 2, and 3 is gratefully acknowledged.

6. ILLUSTRATIVE EXAMPLES. To illustrate the practical application of estimators derived in the preceding sections of this paper, the following examples have been selected.

Example 1. Left truncated. To insure meeting a lower specification limit of 0.1215 in. on the thickness of a certain insulating washer, all production of this component is sorted through go, no-go gages, and all of thickness less than this value are discarded. For a random sample of 100 washers selected from the screened (i.e. the retained) production, it is found that $\bar{x} = 0.124624$ and $s^2 = 2.1106 \times 10^{-6}$. Since n = 100 and $x_0 = 0.1215$, then $(\bar{x} - x_0) = 0.003124$ and $s^2/(\bar{x} - x_0)^2 = 0.21627$. By linear interpolation in Table 1, we obtain $\hat{\theta} = 0.02012$. Even without Table 1, this value might have been read from Figure 1 to three decimals as 0.030. Under the assumption that x is normally distributed, we employ estimators (13) and calculate

$$\hat{\sigma}^2 = 2.1106 \times 10^{-6} + 0.03012(.003124)^2 = 2.405 \times 10^{-6}, \text{ and}$$

$$\hat{\sigma} = 0.00155,$$

$$\hat{\mu} = 0.124624 - 0.03012(.003124) = 0.1245.$$

From (14) $\hat{\xi} = (0.1215 - 0.1245)/.00155 = -1.94$.

In determining the asymptotic variances and covariance of $\hat{\mu}$ and $\hat{\sigma}$, $Z(-1.94) = 0.062399$ is calculated from the defining relation of (5) with the aid of ordinary tables of normal curve areas and ordinates. This value might have been obtained from "The normal probability function: Tables of certain area-ordinate ratios and their recriprocals", published as an editorial in Biometrika, Vol. (42), (1955), pp. 217-22. Using the truncated sample formulas of (23), we calculate $\phi_{11}(-1.94) = 0.8751$, $\phi_{12}(-1.94) = 0.3048$, and $\phi_{22}(-1.94) = 1.4087$. Using these values with $E(n) = n = 100$, and with $\hat{\sigma}^2$ as calculated above, we employ (22) to calculate $V(\hat{\mu}) \sim 2.98 \times 10^{-8}$, $V(\hat{\sigma}) \sim 1.85 \times 10^{-8}$ and $Cov(\hat{\mu}, \hat{\sigma}) \sim -0.65 \times 10^{-8}$. It then follows that $\sigma_{\hat{\mu}} = \sqrt{V(\hat{\mu})} \sim 1.7 \times 10^{-4}$, $\sigma_{\hat{\sigma}} = \sqrt{V(\hat{\sigma})} \sim 1.4 \times 10^{-4}$, and $\rho_{\hat{\mu}, \hat{\sigma}} = Cov(\hat{\mu}, \hat{\sigma}) / \sqrt{V(\hat{\mu}) \; V(\hat{\sigma})} \sim -0.28$.

In the case of truncated samples and type I censored samples, these calculations may be somewhat simplified with the aid of tables of elements of the variance-covariance matrices given by *Hald [9]. Similar tables for type I censored samples were given earlier by Stevens [15]. Gupta [8] tabled corresponding matrix elements for type II censored samples while the author and Woodward [2] tables the matrix element necessary for calculating $V(\hat{\theta})$ in the truncated case.

Example 2. Right Censored Type I. A reaction time test is terminated at the end of ten hours in order to eliminate the effects of certain contaminants which are troublesome when the test is continued over a longer period. For specific sample of this type, $s = 10$, $n = 62$, $n_1 = 38$, $\bar{x} = 8.75$, $s^2 = 1.1043$, $(\bar{x} - x_o) = -1.25$, $s^2/(\bar{x} - x_o)^2 = 0.70675$, and $h = 0.38$. Two-way linear interpolation in Table 2 gives $\hat{\lambda} = 0.71$. This same value might have been read from the graphs of Figure 3. Accordingly, using estimators (20) we calculate

$$\hat{\mu} = 8.75 - 0.71(-1.25) = 9.64,$$

$$\hat{\sigma}^2 = 1.1043 + 0.71(1.5625) = 2.2137,$$

$$\hat{\sigma} = 1.49, \text{ and } \hat{\xi} = 0.244.$$

Since this sample is censored on the right, we need the $\phi_{ij}(-\hat{\xi})$ in order to determine the variances and covariance of $\hat{\mu}$ and $\hat{\sigma}$. Accordingly, we calculate values of $Z(-0.244)$ and $Z(0.244)$ as defined by (5). From the type I censored formulas of (23) we evaluate $\phi_{11}(-0.244)$, $\phi_{12}(-0.244)$, and $\phi_{22}(-0.244)$. With $E(n) = 100 \left[1 - F(-0.244)\right] = 100 F(0.244)$, and $\hat{\sigma}^2$ as calculated above, we employ (22) to calculate $V(\hat{\mu}) \sim 0.070$, $V(\hat{\sigma}) \sim 0.301$, and $Cov(\hat{\mu}, \hat{\sigma}) \sim -0.132$.

---

* Hald's tables are also available in his "Statistical tables and formulas", published by John Wiley and Sons (1952).

Example 3. Right Censored Type II. A sample of $N = 300$ electric light bulbs were left tested until $n = 119$ has burned out with the result that $\bar{x} = 1304.832$ hrs., $s^2 = 12128.250$, and $x_n = 1450.000$ hrs. Accordingly $s^2/(\bar{x}-x_n)^2 = 0.575515$, $n_1 = 300 - 119 = 181$, and $h = 181/300 = 0.6033$. Visual interpolation from the graph of Figure 3, gives $\hat{\lambda} = 1.36$, and using estimators (20) we now calculate

$$\hat{\mu} = 1304.832 - 1.36(1304.832 - 1450.000) = 1502 \text{ hrs.},$$

$$\hat{\sigma}^2 = 12128.250 + 1.36(1304.832 - 1450.000)^2 = 40789, \text{ and}$$

$$\hat{\sigma} = 202 \text{ hrs. From (14) } \hat{\xi} = (1450 - 1502)/202 = -0.257.$$

This example was originally given by Gupta [8] , and to the number of singificant digits given, the above estimates are in agreement with those which he calculated. A more accurate determination of $\hat{\lambda}$ and correspondingly more accurate determinations of $\hat{\mu}$ and $\hat{\sigma}$ are possible by calculating additional values of λ, Y and related functions directly from tables of the normal curve areas and ordinates or from the Biometrika editorial tables (loc. cit.) and then interpolating as summarized below.

$s^2/(\bar{x} - x_n)^2$	$-\xi$	λ
0.575304	0.25690	1.35712
0.575515	0.25693	1.35719
0.576081	0.25700	1.35735

With $\hat{\lambda} = 1.35719$ as determined above, a recalculation using estimators (20) gives more accurate values as $\hat{\mu} = 1501.853$, $\hat{\sigma} = 201.815$, and of course $\hat{\xi} = -0.25693$.

This is a right censored type II sample, and in order to determine variances and covariances of the sample estimates, we must evaluate the $\phi_{ij}(h, -\hat{\xi})$; that is, $\phi_{11}(h, 0.257)$, $\phi_{12}(h, 0.257)$ and $\phi_{22}(h, 0.257)$, where $h = 0.6033$. Calculating these values using the type II formulas of (23) then with $E(n) = n = 119$, and with $\hat{\sigma}^2$ as determined above, we substitute into (23) and subsequently calculate

$\sigma_{\hat{\mu}} = \sqrt{V(\hat{\mu})} \sim 16.6$, $\sigma_{\hat{\sigma}} = \sqrt{V(\hat{\sigma})} \sim 14.9$, and $\rho_{\hat{\mu},\hat{\sigma}} \sim -0.57$. The rather high correlation between estimates reflects the high degree of censoring in this example.

Example 4. Doubly Truncated. To illustrate estimation in the doubly truncated case, we consider an example in which the entire production of a certain bushing is sorted through go, no-go gauges, with the result that items of diameter in excess of 0.6015 in. and those less than 0.5985 in. are discarded. For a random sample of 75 bushing selected from the screened production,

$\bar{x} = 0.600\ 149\ 133$ in., $s^2 = 0.000\ 000\ 371\ 187$, $x_o = 0.5985$ and $w = 0.0030$.

Thus $\bar{x} - x_o = 0.001\ 649\ 31$, $(\bar{x} - x_o)/w = 0.54978$, $s^2/w^2 = 0.041\ 242$, and visual interpolation between the curves of Figure 4 gives:

$$\hat{\xi}_1 = -2.52 \text{ and } \hat{\xi}_2 = 2.00.$$

Accordingly,

$$\hat{\sigma} = w/(\hat{\xi}_2 - \hat{\xi}_1) = 0.0030/\left[2.00 - (-2.52)\right] = 0.00066, \text{ and}$$

$\hat{\mu} = x_o - \hat{\sigma}\,\hat{\xi}_1 = 0.5985 - (.00066)\,(-2.52) = .60016$. Employing iterative procedures as described in [4], the above initial values may be improved upon to yield $\hat{\mu} = 0.60017511$ and $\hat{\sigma} = 0.00066302$.

Table 1.

Auxiliary Estimating Function θ for Singly Truncated Samples

$s^2/(\bar{x} - x_o)^2$	θ	$s^2/(\bar{x} - x_o)^2$	θ
0.062 46	0.04 335	0.155 82	0.008 09
.064 05	.04 413	.156 86	.008 32
.065 69	.04 490	.157 90	.008 56
.067 39	.04 626	.158 95	.008 81
.069 16	.04 768	.160 01	.009 06
0.071 00	0.04 940	0.161 07	0.009 32
.072 91	.03 115	.162 14	.009 59
.074 90	.03 140	.163 22	.009 86
.076 96	.03 170	.164 31	.010 14
.079 11	.03 206	.165 40	.010 42
0.081 34	0.03 249	0.166 50	0.010 72
.083 66	.03 301	.167 61	.011 02
.086 08	.03 362	.168 73	.011 33
.088 59	.03 435	.169 85	.011 64
.091 21	.03 522	.170 98	.011 96
0.094 21	0.03 624	0.172 12	0.012 30
.096 77	.03 745	.173 27	.012 64
.099 72	.03 887	.174 42	.012 98
.102 79	.001 05	.175 58	.013 34
.105 98	.001 25	.176 75	.013 71
0.109 31	0.001 48	0.177 92	0.014 08
.112 77	.001 74	.179 11	.014 46
.116 37	.002 05	.180 30	.014 86
.120 11	.002 41	.181 50	.015 26
.124 00	.002 83	.182 71	.015 67
0.128 05	0.003 31	0.183 93	0.016 09
.132 26	.003 86	.185 15	.016 52
.136 63	.004 49	.186 38	.016 96
.141 17	.005 22	.187 62	.017 41
.145 88	.006 05	.188 87	.017 87
0.150 76	0.007 01	0.190 12	0.018 35
.151 76	.007 21	.191 38	.018 83
.152 76	.007 42	.192 65	.019 33
.153 78	.007 64	.193 93	.019 83
.154 80	.007 86	.195 21	.020 35

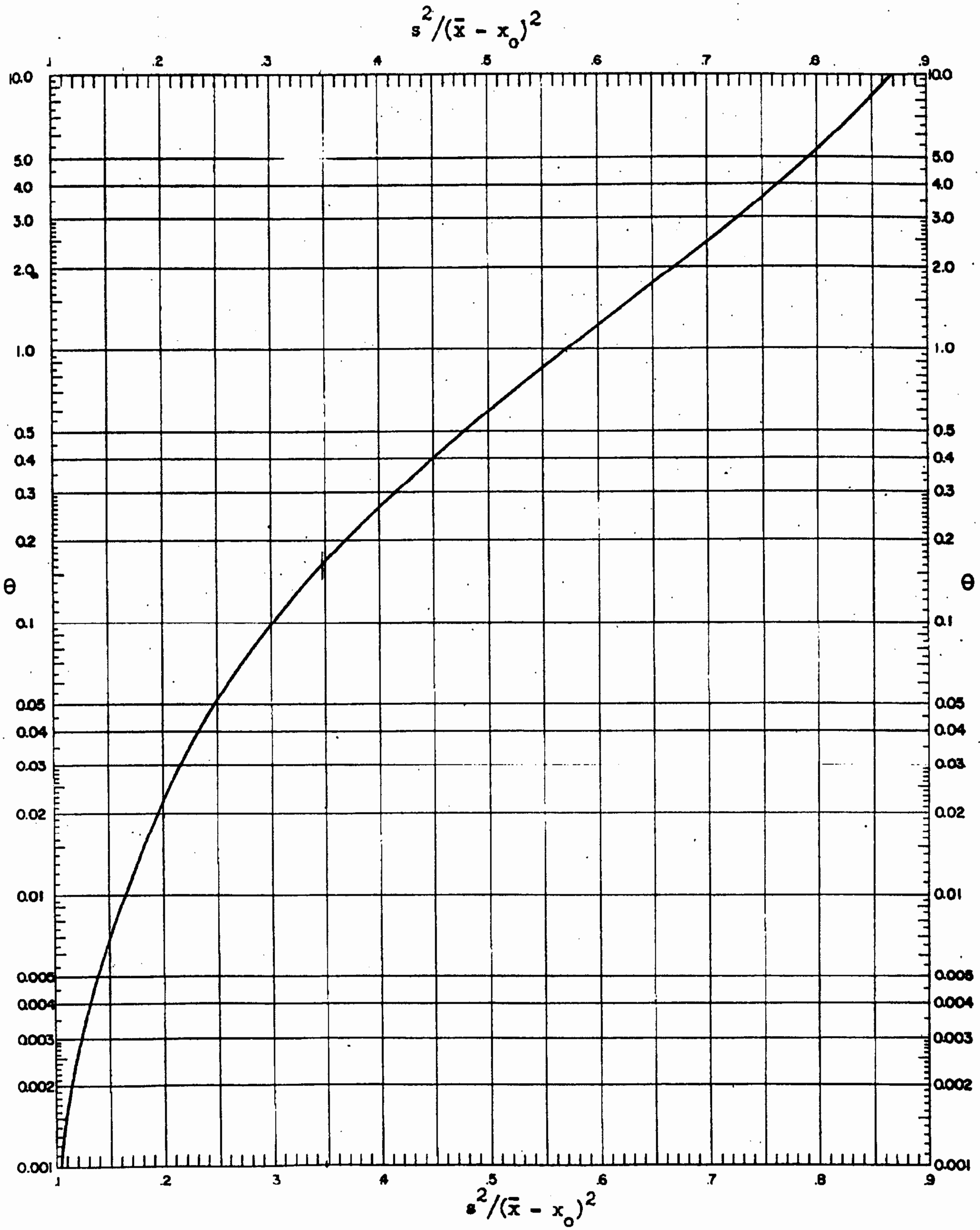
$s^2/(\bar{x} - x_o)^2$	θ	$s^2/(\bar{x} - x_o)$	θ
0.196 51	0.020 88	0.254 44	0.054 73
.197 81	.021 42	.256 04	.055 98
.199 12	.021 98	.257 65	.057 25
.200 43	.022 54	.259 26	.058 55
.201 75	.023 12	.260 88	.059 87
0.203 09	0.023 72	0.262 50	0.061 21
.204 43	.024 32	.264 14	.062 58
.205 77	.024 94	.265 78	.063 98
.207 13	.025 57	.267 42	.065 41
.208 49	.026 22	.269 07	.066 86
0.209 86	0.026 88	0.270 73	0.068 33
.211 24	.027 55	.272 40	.069 84
.212 62	.028 25	.274 08	.071 37
.214 01	.028 95	.275 74	.072 93
.215 41	.029 67	.277 43	.074 52
0.216 82	0.030 41	0.279 12	0.076 14
.218 24	.031 16	.280 82	.077 79
.219 66	.031 93	.282 52	.079 47
.221 02	.032 72	.284 23	.081 18
.222 53	.033 52	.285 94	.082 92
0.223 98	0.034 33	0.287 66	0.084 69
.225 43	.035 17	.289 39	.086 49
.226 89	.036 02	.291 12	.088 33
.228 36	.036 89	.292 86	.090 20
.229 84	.037 78	.294 60	.092 10
0.231 32	0.038 69	0.296 35	0.094 03
.232 81	.039 62	.298 11	.096 00
.234 31	.040 56	.299 70	.097 99
.235 82	.041 53	.301 63	.100 0
.237 33	.042 51	.303 40	.102 1
0.238 85	0.043 52	0.305 18	0.104 2
.240 38	.044 54	.306 98	.106 4
.241 91	.045 59	.308 75	.108 5
.243 45	.046 65	.310 54	.110 8
.245 00	.047 74	.312 34	.113 0
0.246 56	0.048 85	0.314 14	0.115 3
.248 12	.049 98	.315 95	.117 6
.249 69	.051 14	.317 76	.120 0
.251 27	.052 31	.319 57	.122 4
.252 85	.053 51	.321 40	.124 9

$s^2/(\bar{x} - x_o)^2$	θ	$s^2/(\bar{x} - x_o)^2$	θ
0.323 23	0.127 4	0.399 21	0.266 0
.325 06	.129 9	.401 16	.270 5
.326 90	.132 5	.403 11	.275 2
.328 73	.135 1	.405 07	.279 9
.330 57	.137 7	.407 02	.284 7
0.332 43	0.140 4	0.408 97	0.289 6
.334 28	.143 2	.410 90	.294 5
.336 13	.146 0	.412 88	.299 5
.338 00	.148 8	.414 83	.304 5
.339 86	.151 7	.416 80	.309 6
0.341 73	0.154 6	0.418 76	0.314 8
.343 61	.157 6	.420 72	.320 1
.345 48	.160 6	.422 67	.325 4
.347 36	.163 6	.424 63	.330 8
.349 25	.166 7	.426 59	.336 2
0.351 13	0.169 9	0.428 53	0.341 7
.353 02	.173 1	.430 51	.347 3
.354 92	.176 4	.432 47	.353 0
.356 82	.179 7	.434 43	.358 8
.358 72	.183 0	.436 39	.364 6
0.360 62	0.186 4	0.438 35	0.370 5
.362 53	.189 9	.440 31	.376 4
.364 43	.193 4	.442 27	.382 5
.366 35	.196 9	.444 23	.388 6
.368 26	.200 6	.446 19	.394 8
0.370 18	0.204 2	0.448 15	0.401 0
.372 10	.207 9	.450 10	.407 4
.374 02	.211 7	.452 06	.413 8
.375 95	.215 5	.454 02	.420 3
.377 88	.219 4	.455 97	.426 9
0.379 81	0.223 4	0.457 92	0.433 5
.381 74	.227 4	.459 88	.440 2
.383 67	.231 4	.461 83	.447 1
.385 61	.235 5	.463 78	.454 0
.387 55	.239 7	.465 73	.460 9
0.389 47	0.243 9	0.467 67	0.468 0
.391 43	.248 2	.469 62	.475 1
.393 37	.252 6	.471 57	.482 4
.395 32	.256 9	.473 51	.489 7
.397 27	.261 4	.475 45	.497 1

A more extensive table listing larger entries of $s^2/(\bar{x} - x_o)^2$ is available in reference [b] .

$s^2/(\bar{x} - x_o)^2$	θ	$s^2/(\bar{x} - x_o)^2$	θ
0.477 39	0.504 5	0.552 83	0.880 3
.479 32	.512 1	.554 65	.891 8
.481 26	.519 5	.556 46	.903 3
.483 20	.527 5	.558 27	.915 0
.485 13	.535 3	.560 07	.926 8
0.487 06	0.543 2	0.561 84	0.938 8
.488 99	.551 2	.563 66	.950 8
.490 91	.559 3	.565 46	.962 9
.492 84	.567 5	.567 24	.975 2
.494 76	.575 8	.569 02	.987 5
0.496 68	0.584 1	0.570 80	1.000 0
.498 63	.592 6	.572 63	1.012 6
.500 51	.601 2	.574 34	1.025 3
.502 42	.609 8	.576 10	1.038 1
.504 33	.618 5	.577 86	1.051 1
0.506 28	0.627 3	0.579 65	1.064 1
.508 14	.636 3	.581 36	1.077 3
.510 04	.645 3	.583 11	1.090 6
.511 93	.654 4	.584 85	1.104 0
.513 85	.663 6	.586 58	1.117 5
0.515 72	0.672 9	0.588 31	1.131 1
.517 61	.682 3	.590 04	1.144 9
.519 49	.691 8	.591 76	1.158 8
.521 38	.701 4	.593 47	1.172 8
.523 25	.711 1	.595 18	1.186 9
0.525 13	0.720 9	0.596 89	1.201 1
.526 97	.730 8	.598 59	1.215 5
.528 87	.740 8	.600 28	1.230 0
.530 74	.750 9	.601 97	1.244 6
.532 60	.761 1	.603 66	1.259 3
0.534 46	0.771 4	0.605 34	1.274 2
.536 31	.781 9	.607 01	1.289 2
.538 16	.792 4	.608 68	1.304 3
.540 01	.803 0	.610 35	1.319 5
.541 85	.813 7	.612 01	1.334 9
0.543 69	0.824 5	0.613 66	1.350 4
.545 53	.835 5	.615 31	1.366 0
.547 36	.846 5	.616 96	1.381 7
.549 19	.857 7	.618 59	1.397 6
.551 01	.868 9	.620 23	1.413 6

$s^2/(\bar{x} - x_o)^2$	θ	$s^2/(\bar{x} - x_o)^2$	θ
0.621 86	1.429 7	0.777 12	4.42
.623 48	1.446 0	.781 95	4.59
.625 09	1.462 3	.786 66	4.77
.626 71	1.478 8	.791 26	4.96
.628 31	1.495 5	.795 74	5.14
0.629 91	1.512 3	0.800 12	5.33
.631 51	1.529 2	.804 39	5.52
.633 10	1.546 2	.808 55	5.73
.634 68	1.563 4	.812 62	5.94
.636 26	1.580 7	.816 58	6.14
0.637 84	1.598 1	0.820 44	6.36
.639 40	1.615 7	.824 21	6.58
.640 97	1.633 4	.827 88	6.80
.642 52	1.651 2	.831 47	7.03
.644 08	1.669 2	.834 96	7.26
0.645 62	1.687 3	0.838 37	7.50
.647 35	1.705 7	.841 69	7.74
.648 70	1.724 0	.844 93	7.98
.650 23	1.742 5	.848 09	8.23
.651 75	1.761 1	.851 17	8.49
0.653 27	1.779 9	0.854 17	8.75
.660 76	1.88	.857 10	9.01
.668 14	1.98	.859 95	9.28
.675 36	2.08	.862 74	9.55
.682 44	2.19	.865 45	9.83
0.689 38	2.30	0.868	10.11
.696 18	2.41	.871	10.40
.702 84	2.53	.873	10.69
.709 36	2.65	.876	10.99
.715 74	2.77	.878	11.29
0.721 98	2.90	0.880	11.60
.728 08	3.04		
.734 05	3.17		
.739 88	3.32		
.745 58	3.46		
0.751 16	3.61		
.756 60	3.76		
.761 91	3.92		
.767 10	4.08		
.772 17	4.25		



ESTIMATION CURVE FOR SINGLY TRUNCATED SAMPLES

Figure 1

Table 2.

Auxiliary Estimating Function λ' for Singly Censored Samples

$s^2/(\bar{x} - x_0)^2$ \ h	.01	.02	.03	.04	.05	.10	.15
.00	.0101	.0204	.0309	.0416	.05245	.1102	.1734
.05	.01055	.02129	.03222	.04334	.05467	.1143	.1793
.10	.01095	.02208	.03340	.04490	.05659	.1180	.1848
.15	.01131	.02280	.03446	.04632	.05836	.1215	.1898
.20	.01164	.02346	.03545	.04763	.05999	.1247	.1946
.25	.01195	.02408	.03638	.04886	.06152	.1277	.1991
.30	.01224	.02466	.03725	.05002	.06297	.1306	.2034
.35	.01252	.02521	.03808	.05112	.06434	.1333	.2075
.40	.01278	.02574	.03887	.05217	.06566	.1360	.2114
.45	.01304	.02624	.03962	.05318	.06692	.1385	.2152
.50	.01328	.02673	.04035	.05415	.06813	.1409	.2188
.55	.01351	.02720	.04105	.05509	.06930	.1432	.2223
.60	.01374	.02765	.04173	.05600	.07044	.1455	.2558
.65	.01396	.02809	.04239	.05687	.07154	.1477	.2291
.70	.01417	.02851	.04303	.05773	.07260	.1499	.2323
.75	.01438	.02893	.04365	.05855	.07364	.1520	.2355
.80	.01458	.02933	.04426	.05936	.07465	.1540	.2386
.85	.01478	.02972	.04485	.06015	.07564	.1560	.2416
.90	.01497	.03011	.04542	.06092	.07660	.1580	.2445
.95	.01515	.03048	.04599	.06167	.07755	.1599	.2474
1.00	.01534	.03085	.04654	.06241	.07847	.1617	.2502

In type II censored samples x_0 is replaced by x_n.

$s^2/(\bar{x} - x_o)^2$ \ h	.20	.25	.30	.35	.40	.45	.50
.00	.2427	.3185	.4021	.4941	.5961	.7096	.8368
.05	.2503	.3279	.4130	.5066	.6101	.7251	.8539
.10	.2574	.3366	.4233	.5184	.6234	.7400	.8703
.15	.2640	.3448	.4329	.5296	.6361	.7542	.8860
.20	.2703	.3525	.4422	.5403	.6483	.7678	.9012
.25	.2763	.3599	.4510	.5506	.6600	.7810	.9158
.30	.2819	.3670	.4595	.5604	.6712	.7937	.9299
.35	.2874	.3738	.4676	.5699	.6821	.8060	.9437
.40	.2926	.3803	.4755	.5791	.6927	.8179	.9570
.45	.2976	.3866	.4831	.5880	.7029	.8295	.9700
.50	.3025	.3928	.4904	.5967	.7129	.8407	.9826
.55	.3073	.3987	.4976	.6051	.7225	.8517	.9949
.60	.3118	.4045	.5046	.6133	.7320	.8625	1.0070
.65	.3163	.4101	.5114	.6213	.7412	.8730	1.0188
.70	.3206	.4156	.5180	.6291	.7502	.8832	1.0303
.75	.3249	.4209	.5244	.6367	.7590	.8932	1.0416
.80	.3290	.4261	.5308	.6441	.7676	.9031	1.0527
.85	.3331	.4312	.5370	.6514	.7761	.9127	1.0636
.90	.3370	.4362	.5430	.6586	.7844	.9222	1.0742
.95	.3409	.4411	.5490	.6656	.7925	.9314	1.0847
1.00	.3447	.4459	.5548	.6725	.8005	.9406	1.0951

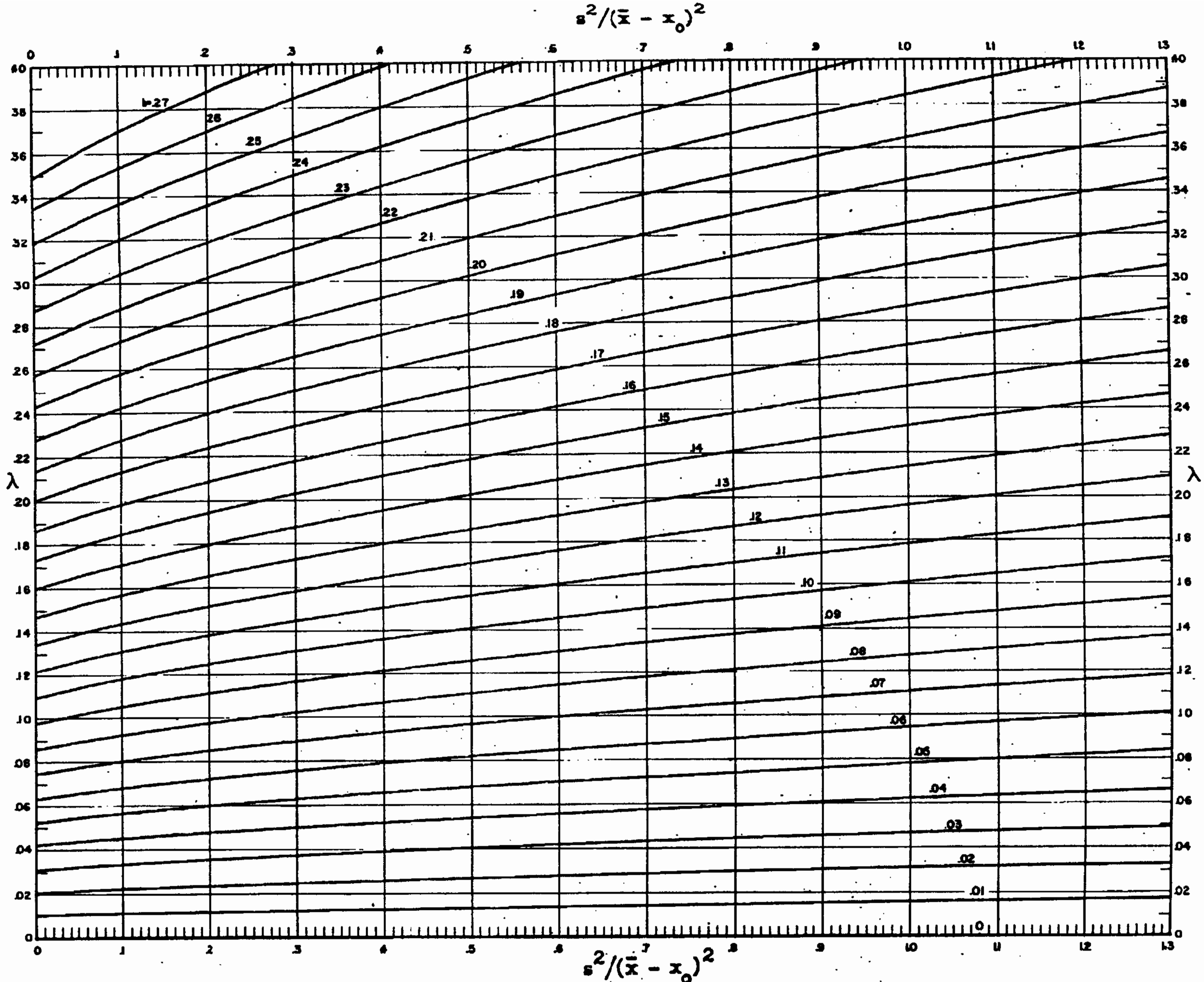


Figure 2

Estimation Curves For Singly Censored Samples
h = 0(.01) .27

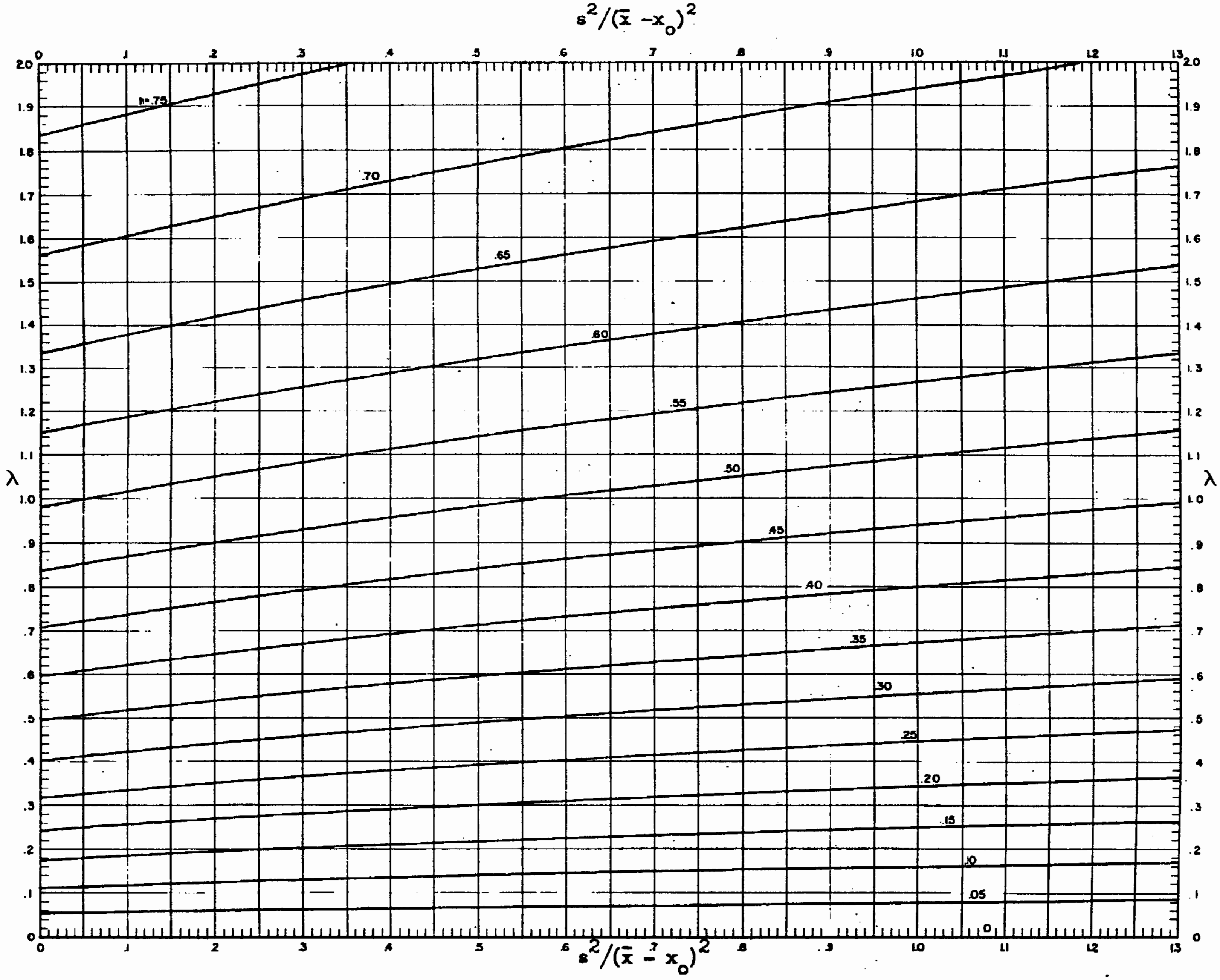


Figure 3 Estimation Curves For Singly Censored Samples
h = 0(.05) .75

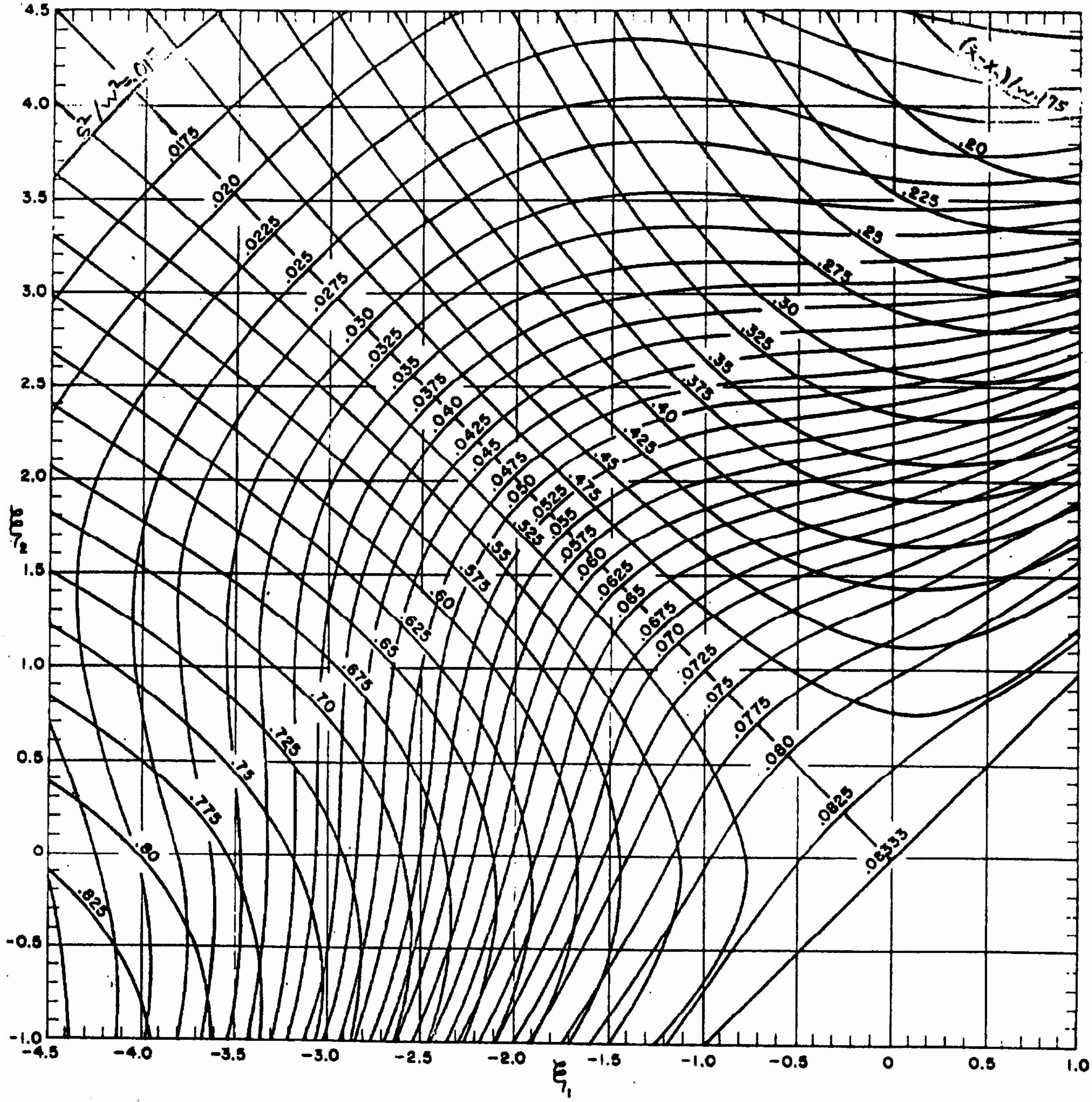


Figure 4

INSTRUCTIONS

1. Locate $(\bar{x} - x_o)/w$ curve corresponding to sample value of this quantity. Interpolated if necessary.
2. Follow curve located in (1) to point where it intersects with s^2/w^2 curve for corresponding sample value. If necessary, interpolate here also.
3. Coordinates of intersection determined in (2), which may be read on scales along the base and the left edge of chart, are the required value of ξ_1, and ξ_2.

REFERENCES

1. COHEN, A.C., JR., "Estimating the mean and variance of normal populations from singly truncated and double truncated samples", Ann. Math. Statist., Vol. 21 (1950), pp. 557-69.

2. COHEN, A.C., JR., and WOODWARD, JOHN, "Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions", Biometrics, Vol. 9 (1953), pp. 489-97.

3. COHEN, A.C., JR., 'Restriction and selection in samples from bivariate normal distributions", J. Amer. Stat. Assn., Vol. 50 (1955), pp.884-93.

4. COHEN, A.C., JR., "On the solution of estimating equations for truncated and censored samples from normal populations", Biometrika, Vol. 44, (1957), pp. 225-36.

5. COHEN, A.C., JR., "Restriction and selection in multinormal distributions", Ann. Math. Statist., Vol. 28 (1957), pp. 731-41.

6. COHEN, A.C., JR., "Simplified Estimators for the normal distribution when samples are singly censored or truncated", Technical Report No. 14, Contract No. DA-01-009-ORD-463, Dept. of Math., University of Georgia, (1958). This paper is to appear in Issue No.3 of Technometrics.

7. FISHER, R.A., "Properties of H_n functions", Math. Tables, Vol. 1, British Assn. for Advancement of Sciences (1931), pp. xxvi-xxxv.

8. GUPTA, A.K., "Estimation of the mean and standard deviation of a normal population from a censored sample", Biometrika, Vol. 39,(1952) pp. 260-73.

9. HALD, A., "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point", Skandinavisk Aktuarietidskrift, Vol. 32 (1949), pp. 119-34.

10. HALPERIN, M., "Estimation in the truncated normal distribution", J. Amer. Statist. Assn. Vol. 47, (1952) pp. 457-65.

11. IPSEN, J., "A practical method of estimating the mean and standard deviation of truncated normal distributions", Human Biology, Vol. 21, (1949) pp. 1-16.

12. PEARSON, KARL and LEE, ALICE, "On the generalized probable error in multiple normal correlation", Biometrika, Vol. 6 (1908) pp. 59-68.

13. SARHAN, A.E. and GREENBERG,B.G., "Estimation of location and scale parameters by order statistics from singly and doubly censored samples", Ann. Math. Statist.,Vol. 27 (1956) pp. 427-51.

14. STAMPFORD, M.R., "The estimation of response-time distributions, Part II", Biometrics, Vol. 9 (1952) pp. 307-69.

15. STEVENS, W.L., "The truncated normal distribution", (Appendix to paper by C. I. Bliss on: The calculation of the time mortality curve.) Ann. Appl. Biol., Vol. 24 (1937) pp. 815-52.

STATISTICAL PROBLEMS ASSOCIATED WITH MISSILE TESTING

Dr. Charles L. Carroll, Jr.
RCA Service Company
Missile Test Project. Patrick Air Force Base

There are a large number of statistical problems that arise in connection with the testing of missiles at the Atlantic Missile Range. These problems are somewhat different from the usual statistical problems and do not appear to have captured the fancy of statisticians who are not actively working at a missile range. In spite of this, there are some significant problems and areas for future statistical research that are suggested by missile testing. If this country's leading statisticians are made aware of these problems and encouraged to pursue them, it is believed that contribution to the national defense effort plus stimulation to the individual statisticaan will be accomplished. Good work on these and related problems is being carried on at the various missile test ranges but in some cases, due to the pressure of getting a job done on time, it is not possible to formulate and solve the problems in their most general and abstract form. It is felt that more work of this type is needed and it is the purpose of this paper to bring to your attention a few of these statistical problems.

Before going into details concerning these problems, it would seem proper to say something about the nature of the work that is done at the Atlantic Missile Range and by RCA Service Company. In general, research and development type missile programs are the type that use the range. In this connection information and data useful in developing and evaluating guidance systems, propulsion systems, aerodynamic characteristics and the weapon system itself are required. RCA is under contract to assist in developing an adequate range from the instrumentation point of view, operate the range instrumentation on all tests and reduce the data obtained from the instrumentation for all missile contractors using the Atlantic Missile Range.

In order to cover an integrated part of the statistical problems in the allotted time, attention is to be focused on those problems arising from trajectory measuring systems. These are systems giving measurements which can be used to reconstruct the trajectory or the path of the missile. It might be said in passing that the accuracy requirements for some of these systems are extremely rigid, for example, errors of less than one part in 1,000,000 where the measurements are made from a point 500 miles from the place where the event is happening.

There are at present, two basic types of trajectory measuring systems: (1) optical and (2) electronic. In the optical systems, there are:

(a) Cine-Theodolite

This is an optical instrument installed in astrodome towers. The Askania Kth 53 is the standard theodolite in use and requires two operators, one to track in azimuth and one in elevation. Angular information from precision glass dials are photographed on each frame together with the missile image. Dial photography is by means of strobic lamps with all theodolites synchronized to "read out" at the same time. Position data is obtained using least square techniques from data from two or more instruments.

(b) Fixed Metric Camera

The standard fixed metric camera system at AMR consists of CZR and RC-5 cameras mounted on three-axes gimbal mounts capable of being oriented to cover the desired field of view. Each individual camera gives the direction of a ray in space from the camera to the missile. Least square methods lead to position data. In most cases, the cameras are controlled remotely from the firing sequencer.

(c) Ballistic Camera

The ballistic camera system at Atlantic Missile Range includes BC-4 cameras and K-37 cameras. These cameras photograph flashing lights or flares at night. The positions of the stars are used to orient the camerss. Because of the high inherent accuracy and reliability of the system, the ballistic cameras are used for evaluation and in-flight calibration of electronic trajectory systems. At the present time, the BC-4 system is the most accurate instrumentation on the range.

In addition there are other optical systems: engineering sequential optical systems, intermediate focal length tracking telescopes, large tracking telescopes (IGOR), ROTI.

In the ecectronic area there are several tracking systems:

(a) Radar

There are several types of radar on the range:

Mod II, a modified SCR-584 radar,

Mod IV, an X-band radar-modification NIKE missile tracking radar,

FPS-8, L-band, AF early-warning air surveillance radar,

FPS-16, a high precision radar developed by RCA, Defense Electronic Products.

In general, the radar gives azimuth, elevation and range of the missile and boresight corrections are applied to get position data.

(b) AZUSA

AZUSA is a high precision electronic tracking device using a crossed base line which gives at a sampling rate of 10 samples per second, two direction cosines and the range.

(c) DOVAP

Using the Doppler principle, an increase (over the previous reading) in range sum from the transmitter to the missile and back to the receiver is obtained. Given three or more simultaneous readings, least square methods can be used to obtain position data.

There are several other electronic tracking systems under development, for example, EXTRADOP, COTAR, SECOR.

In connection with the trajectory systems there are two types of problems of particular interest.

(1) Real Time Problems

In these problems, the computation must be accomplished essentially simultaneously with the event. Examples are:

(a) Impact prediction

(b) Apogee prediction

(c) Nose cone location

(d) Quick look data

(2) Data Reduction Problems

In data reduction problems time is not the primary consideration, but maximum effort is exerted to obtain optimum amount of information from the data collected.

In performing these functions many interesting statistical problems present themselves.

Problem (1) The Accuracy Problem

Basically the question here is: How accurate are the data obtained from each instrument and what can be done to improve the accuracy.

We are interested in the accuracy problem from several points of view:

(i) From a test-by-test point of view. In order for the data to be useful in the evaluation of guidance systems, propulsion systems, etc. on an indiviaual test, the data must be known to be sufficiently accurate. In addition, this information is required to evaluate and improve the performance of the system.

(ii) From the long range point of view. In order to develop a range with the required capabilities, it is necessary that for each instrumentation system the following be known.

(a) The inherent or theoretical accuracy capabilities of the system as it exists on the range.

(b) The accuracy that is being achieved on the range operationally.

(c) Methods for improving the operational accuracy and making it approach the inherent accuracy of the system.

(d) Modifications in the hardware of the system required to improve the inherent or theoretical accuracy of the system.

(iii) Error Studies for the various systems. This involves an understanding of: (a) systematic errors and (b) random errors.

(a) Present methods for getting information about the systematic errors are:

(1) Comparison of measured data from the system with data computed from ballistic camera data, which is considered an order of magnitude better than most tracking data.

(2) Comparison of data from the given system with a best estimate of the trajectory obtained from all instrumentation.

(3) Construction of a mathematical model of the system and the analysis of the systematic errors.

(4) Study of residuals when least square methods are used.

It should be kept in mind that in general the systematic errors for the various systems are from one to one hundred times as large as as the random errors. The systematic errors are not constant and appear to behave as stochastic variables.

(b) Present methods for getting information concerning the random errors:

(1) Variate difference methods using 20-50 consecutive data points.

(2) Polynomial curve fitting using the F-test.

(3) Considering the residuals of a moving are technique.

It is particularly desirable to have a good estimate of the random errors for a particular system on a particular test as functions of time. This information is needed to settle such important questions as whether the successive errors in a measured quantity are correlated or not, what are optimum smoothing functions and what are the best methods of estimating velocity and acceleration data from position data.

Problem (2) Determine Optimum Methods of Determining Velocity and Acceleration Data from Position Data.

Almost all present range instrumentation measures quantities which lead most directly to position data. There is a great need for obtaining accurate velocity and acceleration data which tax the state of the art in both range instrumentation and data reduction. This problem is extremely important in both real time problems and in data reduction problems.

In data reduction, most of the present methods assume the errors in successive measurements are uncorrelated and use moving arc techniques which essentially fit a second or third degree polynomial by least square methods to from two to three seconds of data, evaluating the first and second derivative of this polynomial at the mid point.

For the real time problems, methods have been developed called "almost least square" techniques which do the equivalent of this in a much shorter time.

Problem (2) is closely related to Problem (1) in that to settle it, the nature of the random errors must be known explicitly. Related problems are:

Problem (2.1) Given that the errors in the measured position data $x_i, x_{i+1}, \ldots, x_{i+n}$ obtained at intervals of time t are correlated with known autocorrelation function R_v, determine methods of obtaining $\frac{dx_{i+j}}{dt}$ which are efficient from a computational point of view (i.e., could be used in real time).

Problem (2.2) Given that the errors in the measured position data $x_i, x_{i+1}, \ldots, x_{i+n}$ obtained at intervals of time t satisfy the relation

$$l_{i+j} = k_1 \cos k_2 t$$

determine methods of obtaining $\frac{dx_{i+j}}{dt}$ which are efficient from a computational point of view.

Problem (3) Determine Methods Which Can Be Programmed For An Electronic Computer For Editing Data.

Since enormous quantities of data are processed, attempts are made to automate its handling as much as possible through the use of electronic computers. But no matter how it is done, it is the old problem of rejection of outlying data and no completely satisfactory solution is available.

It is desired to remove the data that is in error but under no circumstances should useful information be removed from the data.

There is interest in this problem in:

(a) Both real time problems and data reduction problems.

(b) Editing both input and output data for the computer.

(c) Editing large discrepancies in the data.

(d) Fine grain editing of data.

In some applications, an attempt is made to remove only the very large errors. In other applications, there are certain peculiarities which are to be edited, for example, in a certain record the error will either be a 1 or a 0 and it is desired to isolate the error and remove it.

Problem (4) Determine Optimum Methods of Smoothing of Data.

This problem is closely related to No. 3 and much has been said and written about it but it remains an important area for additional work.

Problem (5) Extend the Variate Difference Method to Unequispaced Intervals.

The variate difference method that has been developed for equispaced time intervals has been very useful at AFMTC. At the present time a Monte Carlo evaluation of this method is underway. There are applications in which it would be desirable to have the variate difference method extended to unequispaced variables.

Problem (6) Design of Experiments.

An example will suffice - a single piece of complicated electronic tracking equipment has been developed to meet certain specifications:

(1) Accuracy

(2) Reliability

(3) Maintainability

Design a test to determine whether the specifications have been met or not.

Problem (7) Given an Instrumentation System Consisting of n Instruments with Known Accuracy, Considering Geographic Limitations, Determine the Location of the Instrument Sites to Give Optimum Accuracy with Respect to Trajectory Data for a Specified Intended Trajectory.

Problem (8) Determine the Reliability of a Specified Tracking System, Range Safety System or Communication System.

Problem (9) Given a Pencil of n Lines in Space, Possibly Specified by a Set of n Azimuth and Elevation Angles as Measured from a Single Point P, Where all Lines Are Subject to Errors in Measurement, Determine the Conical Surface Which Best Fits this Data.

Problem (10) Given m Sets of Pencils, Each Pencil Consisting of n_i Rays in Space from a Point P_i , Where the Rays Are Subject to Errors in Measurement but Known to Intersect in a Curve, Determine the Best Estimate of the Curve.

APPLICATIONS OF SEQUENTIAL TYPE DESIGNS AND ANALYSES TO FIELD TESTS

Harold R. Rush
Quartermaster Field Evaluation Agency
Headquarters Quartermaster Research and Engineering Command, US Army

INTRODUCTION. It is a well-known and often pointed out fact that the conditions under which an experiment is conducted may differ widely from conditions of actual usage. With this in mind a program for testing Quartermaster items of food, clothing, and equipment is conducted under controlled field conditions by the Quartermaster Research and Engineering Field Evaluation Agency. This testing is accomplished prior to any standardization procedures. A segment of this program is implemented by eight accelerated Wear Courses on which test items are subjected to the equivalent of months of simulated normal wear in a comparatively short time.

It has been found, however, that even an accelerated wear course can require the use of considerable numbers of test subjects for extended periods of time. Clearly the number of test subjects cannot be reduced without a corresponding loss of precision in the test. If the number of subjects cannot be reduced, an alternative approach is to seek some method for reducing the manhour requirements per subject. Considerable savings in time have been realized in quality control work and other industrial situations by use of sequential type procedures. Thus it seemed logical to investigate the possibility of applying such procedures to these field experiments where savings in time was a desired goal.

The present investigation concerns an attempt to adapt the concept of sequential type analysis for use with accelerated wear field tests of fabric durability. The term sequential analysis is used here in its broader sense and does not refer to the sampling procedure developed by Wald (15) and applied by many investigators in various fields (2,4,5,6,8,13). The important distinction between the field test situation and those circumstances which have thus far proved amenable to sequential analysis concerns the sampling or observation procedure. In the more typical sequential analyses a series of independent observations are made on items subjected to a given test, and the hypothesis is accepted, rejected, or a decision made to sample additional independent items subjected to the same test. In field tests, the sample size is usually established prior to the test and testing is cyclical in nature producing cumulative wear or degradation. In these instances, additional observations represent one more test cycle for the entire fixed sample rather than one more sample subjected to a predetermined amount of testing.

RESEARCH PROCEDURES. The procedures followed in devising an explicit method for application of sequential type analysis to certain kinds of field test data can be described as a logical extension or representation of what many investigators have done intuitively. When an experiment has progressed to some logical stopping point, an experimenter frequently will apply appropriate statistical tests to determine to what extent observed differences may be attributable to chance. If differences are not statistically significant at the desired level he may examine the data carefully for trends, inconsistencies, etc., and then decide that the experiment should be continued because a few more cases, assuming comparable results, will provide the desired probability level or that the experiment should be stopped because the erratic results thus far make it improbable that significant differences would be obtained even with a large increase in the number of observations. This intuitive approach can have obvious advantages over rigid adherence to a predetermined test length. However, as pointed out by Anscombe (1) it is still quite susceptible to error.

Some decision criteria other than intuition is obviously needed to formalize the reasoning behind such thinking. Three requisites can be used to enable a more definitive and formal basis for making such judgments: a knowledge of the magnitude of the true differences one wishes to detect, a valid estimate of the population variance, and a reliable estimate of the probability of detecting a given difference at the specified significance level. The exact calculation of this probability, given by Neyman et al (12) is rather complicated and unwieldy. However, an adequate approximate solution for this probability, denoted as P, can be obtained simply by solving for t_2 in the following equation: (The notation, in general, follows that used in (7)).

$$r = \frac{2\sigma^2}{\delta^2} (t_1 + t_2)^2$$

where

δ = specified true difference desired to detect

σ^2 = population variance (estimated from the experiment)

t_1 = value of t, with degrees of freedom for σ^2, from Students' "t" table corresponding to a fixed risk, α, of rejecting the null hypothesis when it is true (one-tailed test)

r = number of replications per treatment

Having obtained t_2, the desired probability P of obtaining a significant result is the probability that a value of t, with appropriate degrees of freedom, should exceed t_2. Since the ordinary t-table gives probabilities, which we designate as P', that a value lies outside the limits $\pm t_2$, the required probability, P, is 1 - (1/2)P'.

The value of P can be calculated at regular discrete intervals during the test and used as the basis for a decision to stop testing or to continue testing until the next interval. The criterion value of P for such a decision can be set arbitrarily depending on the purpose of the investigator and the nature of the particular testing situation. Figure I* has been prepared to facilitate the determination of the P values associated with various values of r, δ and σ.

The utility of such a procedure was evaluated by applying it to data obtained from several tests previously conducted on the Field Evaluation Agency's Fabric Courses. The data from these field tests become available

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

in time sequence and a large number of test subjects are required for protracted periods of time. Thus, such tests are typical of those which appear amenable to sequential procedures and from which maximum benefits in efficiency might be expected.

The Field Evaluation Agency's Fabric Courses are designed to enable comparative evaluation of fabrics in terms of durability under conditions approximating accelerated normal wear. The courses consist of a series of obstacles which when traversed by subjects wearing test garments produce types and amounts of wear similar to that observed in garments salvaged from normal field use but at a faster rate. The complete course is used for testing cotton fabrics, Figure 2, and a modified version is used for wool fabrics, Figure 3. One pair of trousers made from each test fabric is assigned each man participating in the test. The garments are worn through a pre-planned number of cycles with a cycle consisting of two traversals of the fabric course and one laundering. The trousers are examined after each cycle and failures are charted and scored by means of a weighted scoring system developed by the Field Evaluation Agency. A statistical analysis is made at the end of the number of cycles designated in the test plan to determine if differences in average wear scores of fabric types are significant. Statistically speaking, the experimental design used is that of randomized blocks, where each test subject is a "block". The order of wear of the fabric types is randomized with each fabric type being represented during each wear cycle.

Since it would be unwise to terminate a test prior to the end-point in durability for which an item is designed or before a reliable trend has become established, the selection of a test interval at which the test termination criterion will be first applied must be based on experience with the items and test methods involved. For the Fabric Course much background information was available for use in determining the earliest point at which it would be desirable to examine the data. A typical graph of the average cumulative wear score per cycle for a fabric tested on the fabric course climbs steeply during the earlier cycles, then reaches a point at which the curve tends to flatten out. This flattening of the curve indicates that the fabric has been worn beyond a point where the wear score will faithfully reflect further fabric deterioration. The point for the initial analysis of the data should be after reliable wear trends have developed and prior to this flattening of the wear curves. Data available from a large number of past tests suggests that reliable wear trends have developed when wear scores approximate values of 40 for cotton fabrics and 25 for wool serge and similar wool blend fabrics.

Since there is usually considerable variation in the number of cycles required to obtain these critical scores for different fabrics within a test, testing continues until the lowest average wear score approximates 40 for cotton and 25 for wool and wool blends. It should be emphasized that these values are applicable to the fabrics investigated in this study and that they can change as the fabrics investigated change. For example, critical scores for wool shirtings are higher than those of wool serge. A careful check of cumulative wear plots should be maintained across successive tests to determine these critical scores for the different types of fabrics.

Having established the point at which an analysis of the data would start, the familiar analysis of variance was performed and an appropriate test for comparing means, such as Duncan's multiple range test (10) was employed. The investigation would end in the unlikely event, that all differences between adjacent ranked means exceeded $\sqrt{2}\, s_{\bar{x}} t_{0.05}$, that is, all fabrics differed significantly from each other. Normally, though some of the means may be found to differ significantly, other means will be grouped in such a manner that the decision to accept the null hypothesis for these means or to continue the test is surrounded by uncertainty. These are the means of concern and to which the criteria for termination are applied. It is at this point that a value of P is computed, utilizing the sample variance as an estimate of σ^2 and setting δ at 20 per cent as experience has shown that detection of differences in fabric durability of less than 20 per cent is unlikely. This value of P is then evaluated against some predetermined level. For the fabric course a P of .80 was selected as giving reasonable protection levels against committing a Type II error. To facilitate computations, the relationship between selected values of r, associated standard errors, and true mean differences for P = .80 are given in Figure 4.

RESULTS. The results of the application of this type of sequential analysis to a test of cotton fabrics (13) can be seen from Table I (Slide #5). In that test four different cotton fabrics were run for 10 cycles on the fabric course. Using the usual procedure, only that analysis shown in the last row of Table I would be made. However, examination of the mean wear scores by cycle indicated that sequential statistical analysis should start after cycle 7 where the lowest average wear score, 38.5, was an adequate approximation of the minimum wear score required for establishment of reliable trends. Analysis of the data at this cycle allowed definitive statements to be made with respect to types K and CW. The maximum difference between adjacent ranked means for the remaining fabrics, in this instance CS and KR, expressed as a per cent of the general mean was 14 per cent. With a standard error of 38 per cent and a sample size of 37, it is seen from Figure I that the probability of detecting a difference of as much as 20 per cent is .73. Applying the stipulated decision criteria the test would be terminated at this point. Comparing these results with those obtained at the end of the full ten cycles showed that substantially the same conclusions would be drawn with the same level of confidence. The seemingly desirable inverse relationship between the number of cycles and the magnitude of the standard error as seen in Table I can be misleading, since it is accompanied by decreasing proportional differences between fabric wear scores as maximum wear is approached thereby suggesting loss in sensitivity.

In a similar manner, the foregoing procedure was applied to an additional 2 cotton fabric and a 3 wool fabric tests. It was found that the proposed procedure worked quite well for cotton tests, allowing a reduction in the test period of from 2 to 3 cycles. The same conclusions with respect to fabric comparisons would be made after the proposed shortened test period as were made at the original end of the test. Applying the same criterion, that of conformance of results, to the three woolen tests, it was found that only one behaved in such exemplary fashion. On the other two woolen tests, some discrepancies in conclusions were noted. In these instances, however, it was felt that the shortened test period gave a truer evaluation of fabric differences.

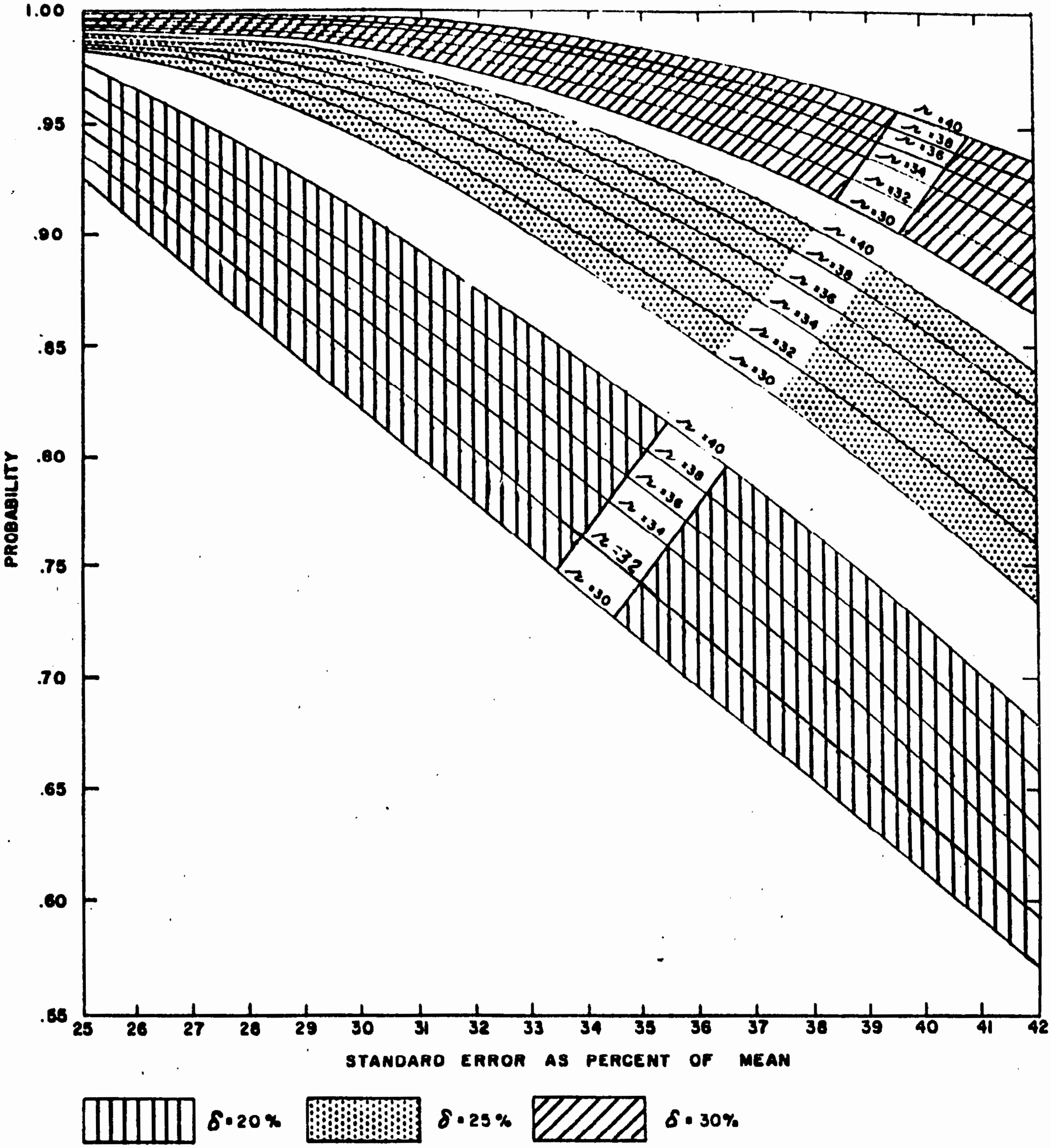
DISCUSSION. This method is not only of immediate value in enabling more efficient fabric course testing while sustaining essentially the same results as the longer, less efficient procedure, but is equally important for its potential application to other comparable testing situations. Speaking only of the realm of field testing, it might profitably be applied to durability testing of many types of footwear, socks, gloves, and other clothing items. Tests of shipping containers and gasoline drums are other items which might be susceptible to this sort of analysis if the tests are designed with that in mind. Recognizing the present achievement, the desirability of investigating the possibility of similar applications in all instances where data become available for analysis in time-sequence and the other requisites are approximated is obvious.

In other such testing situations at the Agency which will be investigated, a more general solution formulated by Tang (14) may be applicable. He has investigated the sensitivity of the analysis of variance test for the general case of t treatments, and prepared tables from which the size of the Type II error can be determined, given the number of replicates, the treatment effects, the size of the Type I error and a reliable estimate of σ^2. Tang's procedure is concerned with testing for differences among all treatments in a group, whereas the question posed in this study is whether, after any particular cycle of wear, any two adjacent ranked treatments (fabrics), differ significantly from each other.

A quite recent article by Bechofer (3) presents a multiple decision procedure for selecting the best one of several normal population means with a common unknown variance, and, as pointed out in the article, the problem of selecting and ordering the t populations with the largest population means also can be treated within the same general theoretical framework. However, the sampling procedure which is the orthodox sequential one, could be unworkable for the type of experiment discussed here. Consider a test of five treatments (fabrics). It would require each test subject approximately 4 weeks to complete 10 traversals of the fabric course for this number of treatments. In other words four weeks would be required to obtain a single observation, and an additional four weeks for each observation thereafter. It is apparent that the tests would take several months to run under this sampling procedure.

SUMMARY. A sequential type approach to analysis of data obtained from accelerated wear field tests is devised through adaptation of the statistical concept of the power of a test. The method requires that the experiment be cyclical in nature and that the data become available in time sequence. From a knowledge of the magnitude of the differences it is desired to detect in the experiment and an estimate of the population variance, the probability of detecting such a difference at a given significance level is determined. This probability is used after each cyclical analysis in making a decision to continue or to terminate testing. Application of this procedure to a number of accelerated wear fabric tests conducted on the Fabric Course demonstrated a potential savings of 20 per cent to 33 per cent in test personnel manhours without loss of meaningful information. The adaptation of such methods to other field tests appears highly feasible.

Figure 1.



Probability of detecting true differences for specified values of δ, r and σ

COTTON FABRIC COURSE

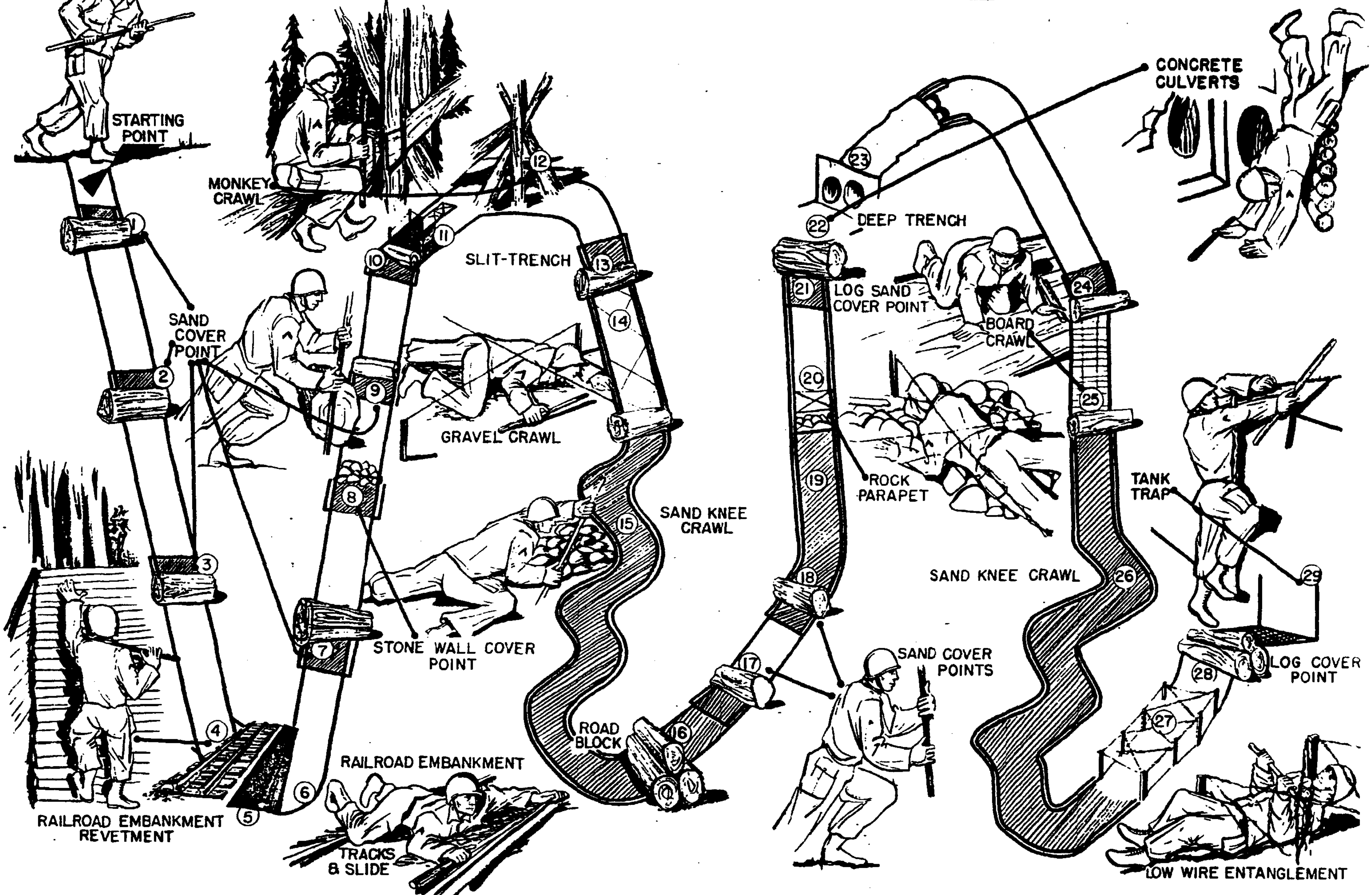


Figure 2.

WOOL FABRIC COURSE

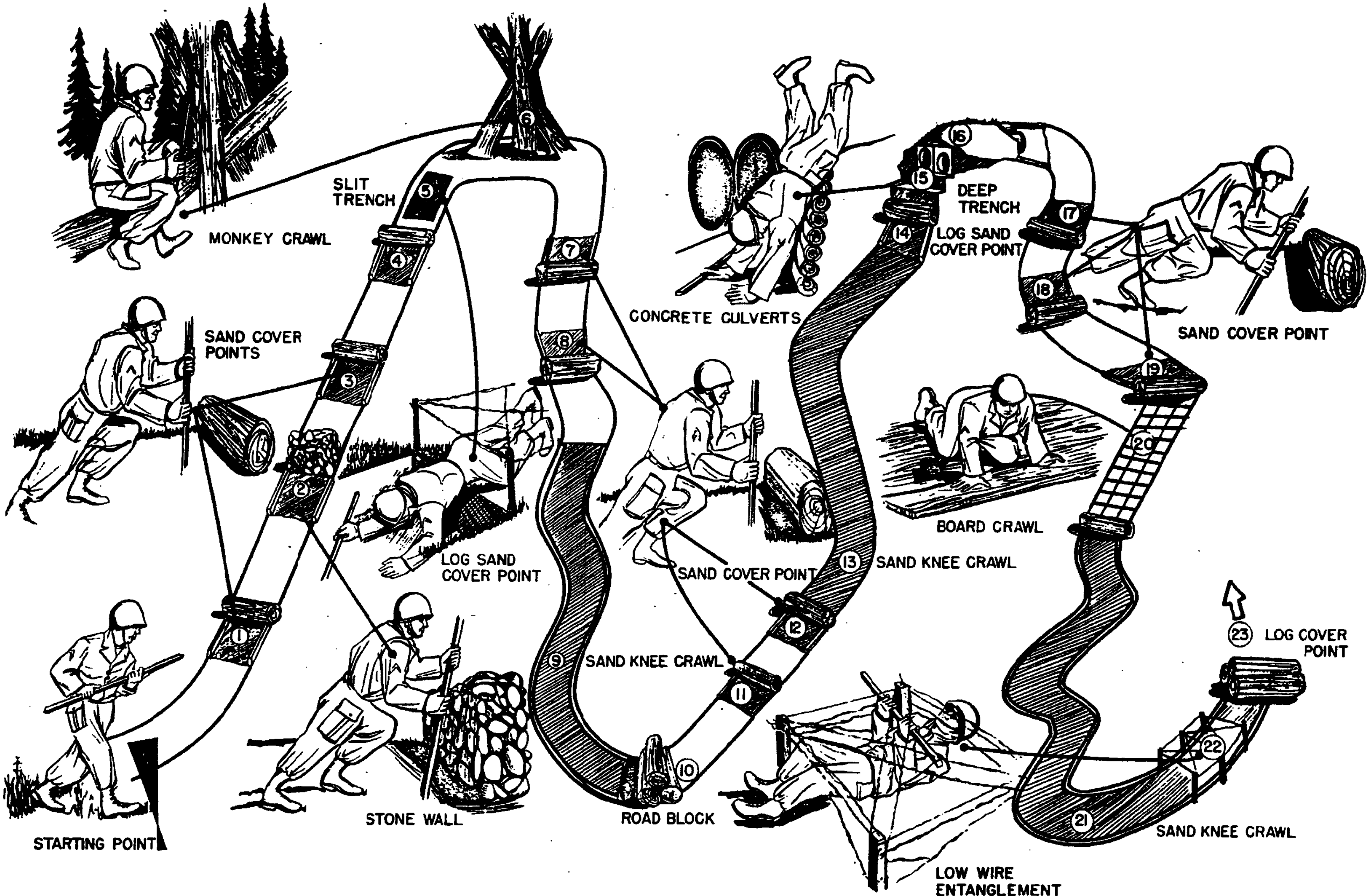


Figure 3.

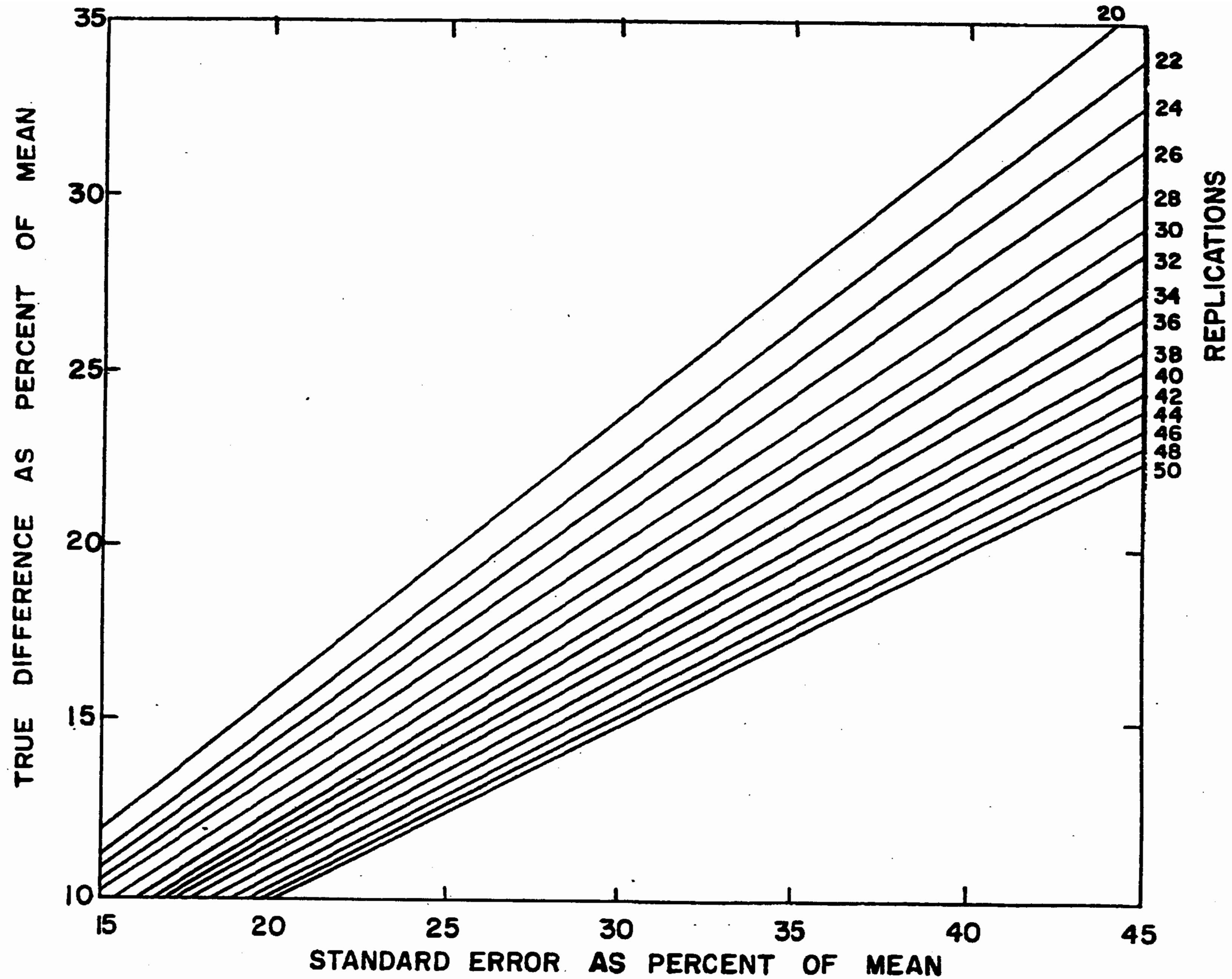


Figure 4. Possible detectable differences for 20 to 50 replications. Test of significance at the 5% level, probability = 80%

TABLE I

ANALYSIS OF DATA FROM FEA 53095, COTTON FABRIC TEST

(r = 37)

Cycle	Average Wear Score and Multiple Comparison Test[a]	Standard Error[b]	F Value	Mean Wear Score	Value of P for δ = 20%
6	CS KR K CW 30.3 38.4 48.7 67.2	42	24.82	46.1	0.65
7	CS KR K CW 38.5 46.5 62.8 81.2	38	27.86	57.3	0.73
8	CS KR K CW 45.5 53.4 72.5 91.6	37	26.80	65.7	-----
9	CS KR K CW 55.0 63.0 83.7 101.0	33	25.46	75.7	-----
10	CS KR K CW 65.9 76.7 94.8 112.6	30	22.21	87.5	-----

[a]Duncan's multiple range test (10)
[b]As a percent of the mean.

BIBLIOGRAPHY

1. Anscombe, F. J., Fixed Sample Size Analysis of Sequential Observations, Biometrics 10, 89-100, 1954.

2. Armitage, P., Sequential Tests in Prophylactic and Theurapeutic Trials, Quart. Journ. Med. 23, 255-274, 1954.

3. Bechofer, Robert E., A Sequential Multiple Decision Procedure for Selecting The Best One of Several Normal Populations With a Common Unknown Variance, And Its Use With Various Experimental Designs, Biometrics 14, 408-429, 1958.

4. Bradley, Ralph A., Some Statistical Methods in Taste Testing and Quality Evaluation, Biometrics 9, 22-28, 1953.

5. Bross, I. D. J., Sequential Medical Plans, Biometrics 8, 183-187, 1952.

6. Brownlee, K. A. et al., The Up-and-Down Method with Small Samples, Jour. Amer. Stat. Assoc. 48, 262-277, 1953.

7. Cochran, W. G. & Cox, Gertrude, Experimental Design, (2nd Ed.) New York: Wiley, 1957.

8. Dixon, W. J., and Mood, A. M., A Method for Obtaining and Analyzing Sensitivity Data, Jour. Amer. Stat. Assoc. 43, 109-126, 1948.

9. Duncan, D. B., Multiple Range and Multiple F Tests, Biometrics 11, 1-42, 1955.

10. Fiske, D. W. & Jones, L. V., Sequential Analysis in Psychological Research, Psychol. Bull. 51, 264-276, 1954.

11. Matthews, J. M. FEA 53095, Wear Resistance of 9 oz. Cotton Sateen Blended with Rayon, QM R & E Field Evaluation Agency, Ft. Lee, Va.: 1954.

12. Neyman, J., Iwaszkiewicz, K. and Koldziejczyk, St., Statistical Problems in Agricultural Experimentation, Suppl. Jour. Roy. Stat. Soc., 2, 107-154, 1935.

13. Radkins, Andrew P., Sequential Analysis in Organoleptic Research Triangle, Paired, Duo-Trio Tests, Food Research 23, 225-234, 1958.

14. Tang, P. C., The Power Function of the Analysis of Variance Tests With Tables and Illustrations of Their Use, Stat. Res. Mem. 2, 126-149, 1938.

15. Wald, A., Sequential Analysis, New York: Wiley, 1947.

A SEQUENTIAL OBSERVATIONAL PROGRAM USED IN A STUDY OF A RESPONSE SURFACE FOR A COMPLEX WEAPONS SYSTEM*

William J. Wrobleski
The University of Michigan

The family of probability distributions associated with the number of aircraft destroyed by a MISSILE MASTER Anti-Aircraft Defense System in raids which belong to a relevant raid space is a response surface of considerable importance to numerous different agencies. Each of these agencies must make a variety of decisions about the MISSILE MASTER System, decisions which involve huge costs and must reflect the essential characteristics of this family of probability distributions.

In order to study this response surface a MISSILE MASTER System was viewed as a stochastic structural relation between:

(1) the primary random variable (i.e., the number of aircraft destroyed)
(2) a set of secondary random variables (i.e., certain service times).

A truncated sequential observational program was conducted in stages. It was designed to study the secondary random variables and some of their interrelationships which were imposed because the system acted as a stochastic structure between primary and secondary variables.

This experimental design included specification of:

(1) the random variables observed during the program,
(2) the sampling plan for each stage of the observational program,
(3) the terminal decisions made at the end of each stage, and
(4) the statistical techniques used to make terminal decisions.

Except for the initial stage, the sampling plan for each successive stage of the observational program was based upon terminal decisions made at the end of each of the preceding stages. It was specified prior to the beginning of the observational program that if experimentation continued through a fifth stage it was to be terminated at the end of that stage regardless of the terminal decisions obtained from the five experimental stages.

From these studies a representation or model of a MISSILE MASTER System was constructed using a digital computer, and a second observation program for estimating the response surface from this representation was designed.

Before proceeding to a more particularized account of the observational program whose general characteristics have been sketched, I shall outline abstractly the central estimation problem which motivated and spanned the entire investigation. This abstract formulation will establish a common

---

* This work was conducted under contract to the United States Army Signal Engineering Laboratories Contract DA-36-039 SC-64627.

frame of reference for the subsequent accounts of the observational program without necessitating a detailed description of a MISSILE MASTER System. It will also provide motivation for the selection of those random variables that were studied in the observational program.

To begin let us consider the measurable space (Z, Ω) where Z is the collection of non-negative integers and Ω is the family of all subsets of Z. In addition we consider a set R, called the <u>raid space</u>. Let <u>a</u> be the generic symbol for an aircraft and <u>r</u> the generic symbol for a raid contained in R. The relation <u>$a \epsilon r$</u> means "a is an aircraft in r"; and we let <u>n(r)</u> denote the cardinality of the set $\{a \mid a \epsilon r\}$. For each $r \epsilon R$, the equivalence relation $i \underset{r}{\sim} j$ in Z defined by $i \equiv j \pmod{n(r) + 1}$ decomposes Z into $n(r) + 1$ disjoint and exhaustive subsets $Z_k(r) = \left\{ z \epsilon Z \,\middle|\, \frac{z-k}{n(r) + 1} \epsilon Z \right\}$, $k = 0, 1, 2, \ldots, n(r)$.

Correspond each $z \epsilon Z_k(r)$ with k and denote this Ω measurable function from Z to Z by $N(z \mid r)$.

A MISSILE MASTER Anti-Aircraft Defense System, hereafter denoted by S, consists of a ring of missile batteries and an automatized coordination center for the missile batteries of the ring. Now interpreting $N(z \mid r)$ as the number of aircraft in the raid $r \epsilon R$ destroyed by S, S may be viewed as a mechanism by which an observation of the random variable $N(z \mid r)$ is generated. In essence S generates a probability distribution $\pi(r, S) = \{\pi_0, \pi_1, \ldots, \pi_{n(r)}\}$ on (Z, Ω) where $\pi_k = \pi_k(r,S) = \pi(Z_k(r) \mid S) = Pr(\{z \mid N(z \mid r) = k\} \mid S)$, $k = 0, 1, \ldots, n(r)$. Denoting $\{\pi(r,S) \mid r \epsilon R\}$ by $\pi(R,S)$, the problem of estimating $\pi(R,S)$ was the central problem which motivated the entire observational program. $\pi(R,S)$ has been variously called the <u>response surface</u> of S relative to R, the <u>performance characteristic space</u> of S relative to R, and the <u>output state space</u> of S relative to R.

Following are some questions which occur in designing an observational program for estimating $\pi(R,S)$:

(1) If J denotes an index set for the number of different raids to be flown against S during the observational program, what should its cardinality n(J) be?

(2) Corresponding to each $j \epsilon J$, what choice should be made for the tuple (r_j, m_j) where $r_j \epsilon R$ and m_j denoted the number of replications of r_j to be made during the observational program?

(3) Should $n(J)$, m_1, $m_2, \ldots$, or $m_{n(J)}$ be random variables?

(4) Is there a natural "estimation topology" for the space $\pi(R,S)$? If there were such a topology O, the r_j's could be selected so that $\{\pi(r_j, S) \mid j \epsilon J\}$ is an O- dense set in $\pi(R,S)$.

Given a tuple, (r, m), m observations of the random variable $N(z \mid r)$ would be made during the observational program, say $N_1, N_2, \ldots, N_m$, and $\pi(r,S)$ estimated by $\hat{\pi}(r,S) = \{\hat{\pi}_0, \hat{\pi}_1, \ldots, \hat{\pi}_{n(r)}\}$ where $\hat{\pi}_k = m_k/m$ and m_k denotes the number of N_j's which are equal to k, $j = 1, 2, \ldots, m$ and $k = 0, 1, \ldots, n(r)$.

However, the number of aircraft destroyed by S in a mock raid can not be observed directly; and so the following question ariese in designing an observational program for estimating $\pi(R,S)$: Does there exist a random variable which is observable and from which the value of $N(z|r)$ can be inferred?

Let b be the generic symbol for a missile battery, b|a the generic symbol for the exent that b destroys a, and P(b|a) the generic symbol for the probability that the exent b|a occurs. Suppose P(b|a) were known as a function $P(\delta)$ of the distance $\delta = \delta(a,b)$ of a from b. Then if at the instants $\eta_1 \leq \eta_2 \leq \ldots \leq \eta_\ell$ when b simulates a missile launch at a, the distances $\delta_1, \delta_2, \ldots, \delta_\ell$ of a from b were known for each $b \epsilon B_a$ (the set of $b \epsilon S$ that simulate a missile launch at a) and for each $a \epsilon r$, an observation of the random variable $N(z|r)$ could be generated.

Therefore, if P(b|a) were known as a function of δ, a reasonable observational program would be the following:

(1) Employ a sampling plan evolved from consideration of questions 1, 2, 3 and 4; and for each replication of a raid specified by the sampling plan observe the distance of b from a at the instants of simulated missile launches for each battery - aircraft assignment combination (b,a) made during the raid.

(2) From these observations calculate P(b|a) from the function $P(\delta)$ and generate an observation of the random variable $N(z|r)$.

(3) Employ the estimate suggested, namely $\hat{\pi}_k = m_k/m$, to make the terminal decisions about the magnitude of $\pi_k \epsilon \pi(r,S)$ for $k = 0, 1, 2, \ldots, n(r)$.

The modifications of this observational program which yield the observational program used for experimentation with the MISSILE MASTER System that was studied can be motivated by considering the nature of system effects on $\pi(R,S)$. For this purpose a useful classification of system effects is the following:

(1) STRUCTURAL EFFECTS produced by the characteristics and configurations of the system's equipment complex. The number of missile batteries belonging to S is an example of a characteristic of the battery configuration of the system's equipment complex which yields a structural effect of $\pi(R,S)$.

(2) PROCEDURAL EFFECTS produced by the system's rules of operation or standing operating procedures which weld together components of the system's equipment complex. For example, the missile batteries and positional information tracking components of the system's equipment complex are connected by the system's assignment doctrine. The assignment doctrine yields a procedural effect on $\pi(R,S)$.

(3) OPERATOR EFFECTS produced by individual differences in operating the system's equipment complex and executing the system's standing operating procedures.

(4) ENVIRONMENTAL EFFECTS produced by the climatic and topographical characteristics of the system's locale.

From this classification of system effects two notable reasons can be derived for modifying the observational program that has been described. The first reason is that operator difference may produce such variability in the system's structural and procedural complexes that additional replications of raids are necessary to obtain significant results. In other words inherent in this observational program are two untenable risks: namely, the risk of having the cost of the program rise to prohibitive heights before significant results are obtained and the complementary risk of having to terminate the program before significant results are obtained in order to avoid astronomical costs. These risks cannot be completely eliminated from any observational program whose objective is to provide experimental information for estimating $\pi(R,S)$. But can a set of random variables be found whose members have a meaningful relation to the problem of estimating $\pi(R,S)$ and are adaptable to an observational program in which these risks are reduced?

The second reason for modifying this observational program stems from the desire to study $\pi(R,S')$ for different systems S' obtained from S through simple modifications of its structural and procedural complexes without having to design and execute another observational program for the system S'. In terms of this reason for modifying the suggested observational program, can a set of random variables be found whose members have a meaningful relation to the problem of estimating $\pi(R,S)$ and are relatively independent of the structural and procedural effects of S?

By considering $[r,S]$ (the raid-system phase space) we will see that affirmative answers can be given to both of these questions. The raid-system phase space although conceptually trivial is difficult to characterize notationally. Essentially it consists of the following points:

(1) The instant $\tau_{\underline{a}}$ of arrival of <u>a</u> for each $a \in r$.

(2) The instant $\chi_{\underline{a}}$ of detection of <u>a</u> for each $a \in r$.

(3) The instant $\kappa_{\underline{a}}$ of entry of <u>a</u> for each $a \in r$.

(4) The instant $\rho_{(a,b)}$ of assignment of <u>a</u> to <u>b</u> for each $a \in r$ and $b \in S$. (If <u>a</u> is not assigned to <u>b</u> put $\rho_{(a,b)} = \infty$).

(5) The instant $\psi_{(a,b)}$ of acquisition of <u>a</u> by <u>b</u> for each $a \in r$ and $b \in S$. (If <u>a</u> is not acquired by <u>b</u> put $\psi_{(a,b)} = \infty$).

(6) The instants $\{\eta_{(a,b)}\}$ of simulated missile launches by <u>b</u> and <u>a</u> for each $b \in S$ and $a \in r$. (If <u>b</u> does not simulate a missile launch at <u>a</u> put $\eta_{(a,b)} = \infty$).

(7) The instant $\nu_{\underline{a}}$ that <u>a</u> reaches its bomb release point.

In addition, for each $a \in r$, let $v_a(t)$ denote the velocity of a at the instant t, and $\varphi_a(\alpha,\beta)$ the path of a between the time instants α and β. Then if $[r,S]$ were known and if, for each $a \in r$, $\varphi(\tau_a,\nu_a)$ and $v_a(t)$ for $t \in (\tau_a,\nu_a)$ were known, the set of distances $\{\delta_{(a,b)}\}$ of a form b at the instants $\{\eta_{(a,b)}\}$ of simulated missile launches could be computed and so an observation of $N(z|r)$ generated using the formula $P(\delta)$ for $P(b|a)$.

Let us look at the following random variables which are system epoch times and can be computed from $[r,S]$:

(1) $\chi - \tau$, the time from arrival to detection.
(2) $\kappa - \chi$, the time from detection to entry.
(3) $\rho - \kappa$, the time from entry to assignment.
(4) $\psi - \rho$, the time from assignment to acquisition.
(5) $\eta - \psi$, the time from acquisition to missile launch.

$\chi - \tau$ and $\eta - \psi$ are not as dependent upon the structural and procedural complexes of S as are $\kappa - \chi$, $\rho - \kappa$, and $\psi - \rho$. These random variables are composed of a waiting time component and a service time component. For the random variables $\kappa - \chi$, $\rho - \kappa$, and $\psi - \rho$ the waiting time component is more significant than the service time component, and the waiting time component is dominantly affected by the structural and procedural complexes of S.

For example, consider the time from detection to entry, $\kappa - \chi$; $\kappa - \chi = W(\kappa - \chi) + S(\kappa - \chi)$ where $W(\kappa - \chi)$ is the time spent in waiting from the instant of detection to the instant system entry service commences and $S(\kappa - \chi)$ is the length of the system entry service period. Among other things $W(\kappa - \chi)$ depends upon the number of system entry service units and upon the system's entry SOP and is affected by the structural and procedural complese of the system. But these system effects assume a different role when $S(\kappa - \chi)$ is considered.

It was through this series of modifications and for the reasons indicated that the observational program previously sketched was moulded into one which had as its primary objectives the study of certain service time distributions and decision processes associated with a MISSILE MASTER System. We are now in a position to give a more particularized account of this observational program: especially of the sequential sampling plan used, the terminal decisions made, and the statistical procedures used to make these terminal decisions. I shall do this only for that part of the observational program aimed at the study of system service time distributions. To do this for the complementary part of the observational program aimed at the study of system decision processes additional description of a MISSILE MASTER System would be required.

Beginning, let T denote one of the system service times and $F(t|r,S)$, $t>0$, the distribution function of T for the raid $r \in R$ and the system S; in other words, $F(t|r,S) = Pr\ (T < t|r,S)$. Translated into technical terms one objective of the observational program was to specify the family of distribution functions $\{F(t|r,S) \mid r \in R\}$.

Prior to experimentation with a MISSILE MASTER System a survey was initiated to uncover possible characteristics of this family of distribution functions from analogous studies carried out on similar service units. These were studies like the typical time and motion investigations performed by industrial engineers. In addition a time and motion study using full scale wooden models of components of a MISSILE MASTER System was conducted for the same purpose. From these efforts, and prior to any experimentation with a MISSILE MASTER System, the following tentative hypotheses were constructed about the family $\{F(t|r,S)\ |r\epsilon R\}$ of distribution function:

(1) $$F(t|r,S) = \int_0^t f(u|r,S)\, du$$

where $$f(u|r,S) = (2\pi\sigma^2)^{-1/2}\, u^{-1} \exp\left[-(2\sigma^2)^{-1}(\ln u-\mu)^2\right]$$

and $\mu=\mu(r,S)$, $\sigma = \sigma(r,S)$.

(2) $$\mu(r,S) = \mu_0 + \mu_1\, n(r) + \mu_2\, v(r) + \mu_3\, h(r)$$

where $$v(r) = (n(r))^{-1} \sum_{a\epsilon r} \int_{\tau a}^{\nu a} (\nu_a - \tau_a)^{-1}\, v_a\,(t)\, dt,$$

$$h(r) = (n(r))^{-1} \sum_{a\epsilon r} \int_{\tau a}^{\nu a} (\nu_a - \tau_a)^{-1}\, h_a\,(t)\, dt,$$

and $h_a(t)$ denotes the height of $a\epsilon r$ at the instant t.

(3) $$\sigma(r,S) = \sigma(r',S) \text{ for } r,\ r' \in R.$$

We call two raids r, r' ϵ R equivalent provided $n(r) = n(r')$, $v(r) = v(r')$ and $h(r) = h(r')$. This equivalence relation decomposes R into disjoint and exhaustive subsets, and one and only one of these subsets corresponds to each triple (n,v,h). The sampling plan for the first stage of experimentation consisted of specifying k - triples $(n_1, v_1, h_1,)$, (n_2, v_2, h_2), ..., (n_k, v_k, h_k), and selecting a raid r from the k equivalence classes of R corresponding to each of these triples for which φ_a was a straight line path between the instants τ_a and ν_a for each a ϵ r, and both $v_a(t)$ and $h_a(t)$ were constant in the interval (τ_a, ν_a) and identical for each a ϵ r. For each r specified in the first stage of the sampling plan the $n(r)$ observations of T (obtained from the experiment by flying r against S) were used to estimate the parameters μ and σ in the log-normal distribution by the maximum likelihood method. The hypothesis that the observed sample can from a log-normal distribution specified by these estimated parameters was tested using the ω^2 - test. For those raids for which this hypothesis was not rejected, the hypothesis that the logarithms of the service times associated with each such raid had the same variance was tested by using Bartlett's homogeneity of variance test. Finally, for the maximal subset of raids for which neither of these two hypotheses was rejected, normal regression theory was applied to examine the means of the logarithms of these service times as functions of the indicated raid parameters.

These terminal decisions were made at the end of each stage of the sampling plan using the statistical procedures mentioned. Based on the terminal decisions that occured, the partition of R into disjoint and exhaustive subsets was left unchanged, or was made finer by adding new raid parameters to be investigated, or was made coarser by dropping parameters that appeared to be inconsequential. The next stage of the sampling plan was then constructed to reflect these changes.

No formal statistical procedures were followed to arrive at the decision to continue or discontinue experimentation with a system service time. This was done on an intuitive basis which reflected the experimental results that obtained from previous stages. However, experimentation with system service times that continued through five successive stages was discontinued regardless of the terminal decisions that occured.

Using this observational program and the stage by stage truncated sampling plan, terminal decisions and statistical decision procedures described, the service time distributions and decision processes alluded to have been studied using a MISSILE MASTER System. They have been synthesized with the aid of a digital computer into representations of systems S and S' (obtained from S through modifications of its structural and procedural complexes) so that $\pi(R,S)$ and $\pi(R,S')$ can be estimated.

I have discussed an approach to the problem of estimating a response surface associated with a complex stochastic structure. This approach involves reducing the primary estimation problem to a number of relevant secondary estimation problems which are partially amenable to classical statistical studies. The results of such studies can then by synthesized into an estimate of the response surface. It is an approach which possesses a certain intuitive appeal; it has evolved and is continuing to develop in almost every area of scientific research which involves the investigation of complex stochastic structures.

This approach, however, does not fit comfortably into any current statistical theory. The more "global" types of statistical procedures demanded by it have not been invented. Nevertheless, the application of this method leads to studies of random variables and their distribution functions which can be subjected to sophisticated measurement and statistical analysis, and thus to precise knowledge about the stochastic structure in the "small".

But it must be kept in mind that because of the complicated structural context by which these random variables are related to each other, a representation of this structural relation will possess a considerable degree of unfaithfulness. This may lead to a lack of precise information about the stochastic structure in the "large" due to distortion of the estimate of the response surface obtained from the representation because of its unfaithfulness.

SOME STATISTICAL ASPECTS OF PREFERENCE AND RELATED TESTS

C. I. Bliss
The Connecticut Agricultural Experiment Station
and Yale University

INTRODUCTION. A cursory review of the recent literature on sensory testing netted more than 90 references, most of them within the last five or ten years. They range from the highly practical and even naive to excursions in mathematical statistics. The restriction of my title to preference testing narrows the field, but not enough. Sensory tests range from the checking of individual preferences in questionnaires to ratings by expert panels, as in the professional judging of tea or milk (Fenton, 1957). I will limit myself to comparative tests in the middle of this range, where the subject bases his verdict on a subjective criterion. In preference testing, our attention is directed as much to the population of which our subjects are a sample as to the materials being compared, so that methods designed for small expert panels may be quite impracticable with the larger numbers that represent supposedly a given population.

EXPERIMENTAL DESIGN. Experimental preference tests should be restricted to what we may call the same sensory dimension, avoiding comparisons between diverse items such as peaches, salmon, sauerkraut and milk in the same series (Peryam and Haynes, 1957). Preferences between different kinds of fruit or between different varieties of the same kind would more nearly fit the pattern I propose discussing. Since they are comparative judgements between two or more similar items, we need not rate them on a "hedonic" scale ranging from "dislike very much" to "like very much" in five or ten or more steps. Despite their widespread use, individual ratings of this type complicate a preference test unnecessarily. Both the mean preference for the several items and their spread over the rating scale will differ from one subject to another, introducing differences in both the average response and the variance. Because of their lesser efficiency per aliquot, I would also avoid triangular and duo-trio tests in preference studies (Gridgeman, 1955).

Within these restrictions, what kinds of comparative tests seem to me best for preference studies? If we have just two items to compare, we might ask, for example, "Between these two varieties of peach, which do you prefer?" If we have three or more varieties, we might present them in pairs, so that the subject could compare each variety separately with every other variety in a so-called paired comparison (Bradley, 1953; Jackson and Fleckenstein, 1957). Given four varieties A to D, for example, each subject would compare them in six pairs: A-B, A-C, A-D, B-C, B-D, and C-D. As each pair is presented, we might ask additionally "Is your preference slight, moderate or strong, or in reality non-existent?" (Scheffé, 1952; Bliss et al, 1956; Carroll, 1958). Although differences between subjects in the strength of their preferences may introduce heterogeneity in the error, the gain in information will often justify this risk.

An alternative design when three or more items are to be compared is to ask the subject to rank them in sequence from the most to the least preferred (Bliss et al, 1943, 1953; Greenwood and Salerno, 1949). To avoid bias from the order of presentation, the order can be randomized for each subject or balanced with a Latin square, the rows representing subjects or sessions, the columns order of presentation, and letters the items to be compared. Since the design should facilitate comparisons within a selected set of stimuli, ranking is most effective when these are qualitatively similar. If the critical stimuli are qualitatively dissimilar, the subject may have less difficulty in choosing between the two members of a pair than in ranking three or more in order. Under these circumstances paired comparisons would be preferred.

Ranking a series of three to five or six varieties or "treatments" may work well, but with longer series sensory fatigue can blunt the subject's ability to discriminate. If the testing of all possible pairs requires too many replicates, ranking in incomplete balanced blocks may be the solution (Hopkins, 1954; Murphy et al, 1957). In one of these known as the Youden square, each row, representing the order of presentation, contains all varieties, and each column the varieties compared in one session by one subject. Within columns every variety occurs equally often with every other variety in the series. The two Youden squares in Table 1* provide the testing of 7 varieties in groups of three and of four, the upper and lower sections together forming a 7x7 Latin square (Youden, 1940). A simplified rank analysis for incomplete block designs has been described by Dykstra (1956).

The scope and efficiency of many preference tests can be enlarged with a factorial design of the treatments. In a 2x2 factorial, for example, American and Dutch process cocoa syrups were prepared by both the "hot" and "cold" methods and the four combinations tasted in a paired comparison by each of 30 subjects (Reid and Becker, 1956). Only the interaction proved significant, leading to the recommendation that the "hot" method be used for American cocoa and the "cold" method for Dutch cocoa. Other factorial taste tests include a 2x2 factorial on pesticide flavors in apples (Bliss et al, 1956), a 2x3 palatability test on kale (Greenwood and Salerno, 1949), and a 4x2x2 taste test on jam (Gridgeman, 1956). As the number of factors is increased, the experiment may be kept within bounds with a fractional factorial, as shown by Carroll (1958) for five formulation variables of a pudding, each at two levels. By selecting a particular set of 16 from the 32 possible paired comparisons and a Scheffé rating for the degree of preference, she measured the effect of each variable with a marked gain in efficiency.

---

* Tables are to be found at the end of this article.

A preference test, of course, should also be controlled in its environment, in preventing collusion between subjects, and in uniformity of the samples, all of which are noted in the literature. Even time of day may be important (Harries, 1953).

Ideally, the subjects in a preference test are a random sample from the population whose preferences we are trying to measure. In some experiments this ideal can be approximated, as in tests reported by Pangborn et al (1957), but more often, we may be limited to personnel available to the testing laboratory, who in time may become seasoned as subjects. Training or practice sessions have so increased the consistency and sensitivity of the ratings in some sensory tests (Bennett, 1956), that a practice session may be warranted even in large scale preference testing. Since we wish to infer the preferences of a theoretically unlimited population from our sample of subjects, larger panels may be justified than in other types of tests. Among the studies on panel selection are comparisons with consumer surveys (Murphy et al, 1958), the sequential testing of prospective subjects (Bradley, 1953; Armitage, 1957), and the relation of screening tests with water solutions to judging ability for foods (Mackey and Jones, 1954). If the test material or the total number of responses is limited, we can broaden our sampling by testing each subject only once, but for precise comparisons, a design giving two or more complete replicates from each individual would be preferred.

If the experimental procedure is satisfactorily controlled, the test materials are uniform, and the motivation of the subjects is adequate, inconsistency in a subject's preferences in a sensory test may have two different explanations. All concentrations of a critical component in the test materials may fall below the subject's threshold of perception, so that his choices are essentially random. Alternatively, the actual difference between two concentrations of a critical component where both are above his threshold may not exceed the just perceptible difference between them, so again his choice is random. Although both factors are important, I shall consider here only the threshold of perception and its distribution, where I have been struck by some analogies with experiments on insecticides and drugs.

<u>CONTINUOUS DISTRIBUTIONS OF SENSORY THRESHOLDS</u>. Let us suppose that we are determining preferences for samples of orange juice, which differ primarily in their content of sugar or of acid. We will further assume that the differences between samples are greater than the just-perceptible sensory difference, if the subject can detect the sugar or acid at all. A subject with a threshold for the critical component that falls below that of the sample with the smallest concentration will be able to distinguish one juice from another. Whatever may be the level of sugar or acid which he prefers, whether high, low or in the middle, his preferences in replicate tests can be internally consistent because he can recognize the differences between all of them. By contrast, a subject with a sensory threshold above the concentration of sugar or acid in all or most of the sample juices will be unable to separate those falling below his threshold, and will rank them at random or by some secondary factor. His replicate ratings will be internally inconsistent. Hence, it would be useful to know how the sensory thresholds for primary taste sensations such as sweet, sour, salt and bitter, are distributed in a sample of tasters.

Two experiments in this area have concerned water solutions of sucrose (Baker et al, 1954) and of tartaric acid (Baker et al, 1958). Each series of five and ten test solutions respectively represents a geometric progression in steps of two, so large a log interval that only the median threshold can be estimated profitably for each subject. Each was asked to identify the tube containing the test solution when it was paired against water, and he sampled each concentration in his critical range 15 or more times over a number of sessions. Concentrations below his sensory threshold have an expectation of 50% correct answers, which increases to 100% as the concentrations reach and exceed the subject's threshold. The statistical problem is two-fold: (1) how can we compute the median threshold concentration for each subject, and (2) how are these thresholds distributed in the population from which these subjects may be considered a random sample?

In animal tests, many drugs and toxicants produce no perceptible change in the organism until they reach a critical concentration, when an all-or-none reaction occurs. The critical dose which just produces a response is a measure of its threshold level at the time of the test, although in repeated trials with the same dose the animal may react on some occasions and not on others, as has been reported, for example, in the convulsive response of individual rats to the drug thujone (Sampson and Fernandez, 1939). If the animal is tested repeatedly in, say, 20 trials, with doses ranging from one to which it does not respond to one to which it responds invariably, the percentage of positive reactions plots commonly against the log-dose of drug as a symmetrical sigmoid curve. A similar relation might be expected with the thresholds for a taste stimulus, as has been shown for sucrose by Gridgeman (1958).

If the variation in the threshold results from a number of factors acting concurrently, some raising and others lowering the level, a suitable model would interpret the sigmoid relation as an integrated normal curve. The stimulus interpolated from this curve at a net response of 50% would estimate the individual's threshold as the median of a normal distribution. Our simplest procedure is to convert each sigmoid curve to a linear form by plotting the standardized normal deviate or probit for each observed net percentage against the log-concentration of the stimulus. The normal curve is paralleled so closely by the logistic over most of its range that substantially the same result can be obtained with either function (Gridgeman, 1958).

Taste tests have one complication, which is implied by the term "net percentage response". When the subject is unable to discriminate, we start with an expectation of 50 instead of 0 percent of correct answers. This is analogous to an insecticide experiment with 50% natural mortality, where the natural mortality and that attributable to the toxicant act independently of each other. In a taste test each response between 50 and 100% of correct identifications can be corrected for a base line of 50% with the entomologist's adjustment for natural mortality by computing

$$\text{Net \% response} = \frac{\text{observed \%} - 50}{1 - 0.50}$$

When percentages of more than 50% are observed at the smaller concentrations in a graded series and are not succeeded by larger values, they may exceed 50% by chance and can be omitted as not relevant to the experiment.

For analysis, each net response in the intermediate zone between 0 and 100 percent is transformed to its empirical probit by a suitable table, such as that given by Fisher and Yates (1957). Since responses of 0 and 100 percent have empirical probits of minus and plus infinity respectively, we may adapt Berkson's (1953) useful dodge for our preliminary estimate and replace the first net zero percent below the intermediate zone and the first 100% above the intermediate zone by the percentages 100/N and 100 (N-1)/N respectively, where N is the total number of pairs sampled by a single subject at a given concentration. These are then transformed to empirical probits.

Given the coded log-concentrations (x) and the corresponding empirical probits (y), we may compute a trial straight line for each individual by simple least squares without weighting. If the slopes (b) of these lines for subgroups of two or more subjects agree sufficiently, they can be combined into improved, more stable estimates b_c. The subjects in Baker's experiments were grouped primarily by the number of responses in the intermediate zone, those tested with sucrose into two homogeneous groups and those with tartaric acid into three groups, with significantly different slopes. Each individual's threshold in coded log-units may be estimated provisionally from the unweighted means, $\bar{x}$ and $\bar{y}$, and b_c as

$$X_5 - \bar{x} + (5-\bar{y})/b_c$$

where X_5 is the coded mean of a log-normal sample with an estimated standard deviation $s = 1/b_c$. The calculation of the provisional estimate is illustrated for subjects "C" and "N" in Table 2.

For a definitive result, these initial estimates are improved iteratively by maximum likelihood. This involves computing the expected probit Y at each x, replacing the empirical probits with their estimated working probits y, and weighting each y by $w = N\{Z^2/Q(P+1)\}$. The term in brackets is the weighting coefficient for 50% natural mortality, which has been tabeled by Finney (1952). The first weighted regression will often answer the experimenter's requirements. Its calculation is illustrated in Table 3 with the data for the two subjects in Table 2. Additional iterations have been computed from the present data, omitting from b_c any individuals with less than two responses in the intermediate zone.

Agreement with our model can be checked by χ^2. The separate slopes for the subjects tested with both sucrose and tartaric acid showed better than average agreement with the combined slopes for their respective groups but the vatiation of y about several of the curves with lesser slopes was significantly heterogeneous. A composite χ^2 over all tests, however, showed adequate agreement with the underlying hypothesis ($\chi^2 = 82.77$, n = 75).

From the weighted means $\bar{y}$ and $\bar{x}$, the coded threshold concentration X_5 for sugar or acid was recomputed for each subject with the relevant combined slope b_c. To test whether these log-thresholds X_5 were distributed normally, each series has been arranged in increasing order and plotted in Figure 1, where the ordinate is the corresponding rankit or expected average deviate for a sample of N (=15 and 24) from a normal population with zero mean and unit standard deviation (Fisher and Yates 1957, Table XX). Since the trend of each series of plotted points is substantially linear, we may consider the distribution of these two taste thresholds as essentially, log-normal.

The distribution of thresholds could play an important role in testing preferences between similar foods in a series. If the concentration of a critical component in the samples offered to a panel were less than some of their thresholds, it could influence the choice of only those individuals with a small enough threshold to taste this component. Hence, inconsistencies in a subject's response may depend upon the relation of his sensory threshold for the critical factor to that of the population whose preferences are being sampled.

DISCONTINUOUS DISTRIBUTIONS OF THE SENSORY THRESHOLD. The variations of sensory thresholds between individuals may not be continuous and substantially normal or Gaussian, as in the above tests with sucrose and tartaric acid. Instead a clear-cut discontinuity may divide the population into two categories of tasters and nontasters. Perhaps the best known case is that of solutions of phenyl-thio-carbamide (PTC), which to some people are exceedingly bitter and to others tasteless. Tasters can be separated from non-tasters by a solution of about 1/5 molar. In a study of some 3700 individuals, the proportion of tasters in the population was about 71% (Cotterman and Snyder, 1939). Geneticists have traced the dichotomy to a single pair of alleles. Individuals homozygous for the recessive gene find PTC tasteless and individuals with one or both of its dominant allele find it extremely bitter. Although in saturated solutions (4/5 molar) all individuals can detect some bitterness (Blakeslee, 1932), the frequency distribution of the taste threshold is sharply bimodal, with a fair spread between tasters and little or no spread between the non-tasters.

Blakeslee and his associates (1935, 1948) have extended their studies to olfactory as well as to taste reactions. Individuals varied widely not only in their thresholds but also in their preferences as to whether a given odor was pleasant, indifferent or unpleasant. If the preference for one of two varieties of a given food, for example, were to depend upon a well-marked bimodality in the taste or olfactory threshold, the situation would parallel that separating the placebo reactors from the placebo non-reactors in experiments with drugs.

In a clinical experiment that is especially relevant, four preparations A to D were compared as headache remedies (Jellinek, 1946). Tablets of the four preparations, identical in color, shape, size and taste but with the compositions shown in Table 4, were distributed to headache-prone patients under code designations concealed from both the patients and their cooperating physicians. They were given in successive two-week periods to a total of 199 patients in an order determined for each group of 49 or 50 by the rows in a Latin square. Each patient was to take a tablet every time he developed a headache and to record whether his headache was relieved within half an hour. When the results were analyzed, the mean success rates for the four preparations were A 0.84, B 0.80, C 0.80, and D 0.52, the first three giving relief significantly more often than the placebo (D) but not differing among themselves.

Jellinek then examined the frequency distribution of the number of successes reported for the placebo and found it U-shaped, as in the two series in the Table 5 for patients reporting five headaches in a two-week period. On the hypothesis that patients who never reacted to the placebo had physiological headaches and those who were relieved by the placebo on one or more occasions had psychological headaches, he divided his subjects into two series, numbering 79 and 120 respectively. Their mean success rates and the combined analysis of variance for each series are given in Table 6. The placebo non-reactors discriminated between the three true drugs significantly but the placebo reactors gave all three the same success rate as the placebo itself (0.86). Jellinek concluded that "discrimination among remedies for pain can be made only by subjects who have a pain on which the analgesic can be tested".

SEPARATION OF SENSITIVE AND INSENSITIVE SUBJECTS. How can this principle be applied to the testing of preferences where we have no "placebo" to separate the sheep from the goats? One possible criterion is consistency in the response of each subject. This would require replication within subjects, so that the mean ratings of each individual can be tested against an error term based upon his own inconsistency. The direction of preference or its additivity would not be a criterion, only the requirement that a subject designated as "having preferences" must show some stability in at least one choice in replicated tests. In parallel analyses, all variances between "treatments" for those without preferences should be of the same magnitude as the error, but for those with preferences one or more comparisons should be significant or approach significance.

The relative size of the two groups would itself be an important outcome of the experiment and we could minimize or omit any preliminary screening of the taste panel. Within the sensitive group, of course, individuals may have sufficiently diverse preferences, that no direct comparisons between treatments are significant, but in this case a significant interaction of treatment by subject would testify to a difference in opinion as contrasted with no opinion. Whether the indifferent subjects represent "taste blind" individuals, as in the PTC test, or the high threshold end of a Gaussian distribution would probably require additional experiments. In the latter case, we would expect the proportion of non-tasters to be relatively unstable and more dependent upon the exact concentration of the critical component in the test materials than if the distribution of sensory thresholds were bimodal.

As an example, I will apply this procedure to a paired comparison on the relative palatability of Cortland apples from trees which had been treated with four different spray combinations in a single experimental orchard (Bliss et al, 1956). Each spray mixture contained one of the two fungicides, thiram (Th) or sulphur, and one of the two insecticides, lead arsenate (L) or parathion, in a 2x2 factorial design. Apples from each treatment were chosen at random from the fall harvest, washed in detergent suds, rinsed, cored and quartered. These were then made into sauce and quick frozen, the yield from one-quarter of each of ten apples providing sufficient test material for a given treatment in a single taste session.

Twenty-five subjects, students and faculty at the University of Connecticut, participated in six sessions arranged in three pairs, the second session of each pair following two days after the first. In the first, third and fifth sessions, the six possible pairs of the four treatments, all with concealed identities, were presented to each taster in an order determined by assigning him to one row in a 6x6 Latin square. In the second, fourth and sixth sessions, he tasted the same sequence of pairs as in the preceding session but with the order within each pair reversed. Each subject recorded not only his preference within each pair but also whether the difference was slight, moderate, large, or really non-existent.

The degrees of preference have been transformed to rankits for $N = 7$, and analyzed in terms of Scheffé's extension of the Thurstone-Mosteller model. This model postulates a subjective continuum within each subject on which the sensations developed by the stimuli are arranged on a linear scale, the sensations for each stimulus varying normally about a mid-point. The six rankits in each of the six replicated tests for each subject, one for each pair of samples, varied about zero with 36 degrees of freedom. Six of these represent differences between the means of the paired stimuli and 30 the remaining variability. From the six mean differences we can isolate three factorial comparisons, each with one degree of freedom, representing the preference between the fungicides thiram and sulphur, between the insecticides lead arsenate and parathion, and their interaction. The sum of squares for the remaining three degrees of freedom measures non-additivity on the subjective continuum.

Each factorial effect for each subject, disregarding the direction of the preference, was compared against his residual variability. In 11 of the 25, no effect was significant at $P \leq 0.15$, with $P < 0.20$ for only five of the 33 comparisons and no two of these in the same individual. These subjects apparently were either indifferent or insensitive to any specific flavors which might be associated with the four toxicants. Presumably, their thresholds lay either in the upper end of a normally distributed population of taste thresholds or in the upper portion of a bimodal distribution.

Two analyses of variance have been computed in Table 7 for 11 inconsistent and for the 14 consistent subjects. In agreement with our hypothesis, no comparison in the first group is significant, although the contrast between thiram and sulphur is twice its error variance ($F = 2.22$). By contrast, the other 14 subjects differ significantly in their preferences for the two direct factorial comparisons (rows 5 and 6), but their disagreement is not sufficient to preclude a significant "vote" for thiram in preference to sulphur ($P < 0.025$) and for lead arsenate in preference to parathion ($P < 0.05$). The interaction in row 3, measuring the dependence of the preference for the insecticide upon which fungicide was present (or vice-versa), approaches significance. The assumption of an additive scale or linear subjective continuum is justified by variance ratios in rows 4 and 8 of $F < 1$ or not significant. In comparison with the composite analysis in the original paper, the present subdivision of the 25 subjects into two groups has sharpened our test of the disagreement between those with a consistent preference. Since each direct effect has been compared against its interaction with subjects, its overall significance is somewhat smaller than before.

SUMMARY. By their very name, preference tests measure the comparative response of individual subjects to a series of two or more items, most commonly in taste tests. When these can be presented to each subject in pairs, scoring the direction and degree of preference between the two samples provides a more pertinent criterion than rating each item separately on a hedonic scale, which introduces a needless source of variation. When the samples do not differ enough qualitatively to make ranking difficult, three or more items may be presented in each set and the subject asked to rank them in order of preference. The structure and order of presentation within each set may represent an arrangement in randomized groups, Latin squares, or in balanced incomplete blocks such as the Youden square.

Statistical methods are suggested for estimating the individual sensory thresholds for a given stimulus and then describing their distribution. Of other designs available for this purpose, the experimenter would be well advised to consider sequential procedures. Experiments for measuring thresholds are primarily of value in explaining the results of preference tests. Where our objective is descriptive rather than explanatory, the separation of subjects by the consistency of their replicated responses into two series, one sensitive and the other indifferent, should prevent the individuals without preferences from concealing the critical evidence of those with preferences and thereby increase the efficiency of our experiments. The proportion of individuals without preferences would then constitute one out-come of the experiment.

Table 1. A 7×7 Latin square that divides into two Youden squares. (Youden, 1940)

Order of tasting		Replicate No. 1	2	3	4	5	6	7	Letters occur together in same replicate
1	1	A	B	C	D	E	F	G	once
2	2	B	C	D	E	F	G	A	
3	3	D	E	F	G	A	B	C	
4	1	C	D	E	F	G	A	B	twice
5	2	E	F	G	A	B	C	D	
6	3	F	G	A	B	C	D	E	
7	4	G	A	B	C	D	E	F	

Table 2. Provisional estimate of EC50 for tasters "C" and "N" from paired difference tests against water of 5 concentrations of sucrose increasing by multiples of 2 from 0.034% at x=1 (Baker et al, 1954)

Taster	Sol'n. x	No.(+) / total	Net %	Probit y	Calculation in coded log-concentrations:
"C"	1	8/19	0	3.4	$\bar{x}$ = 2.5, $\bar{y}$ = 4.575
	2	10/18	11	3.8	$[x^2]$ = 5, [xy] = 5.05
	3	12/18	33	4.6	Slope for "C", b = 1.01
	4	15/15	100	6.5	Pooled slope b_c = 1.120
	5	15/15	100	Omit	EC50 = 2.879 (with b_c)
"N"	1	10/19	5	3.4	$\bar{x}$ = 3, $\bar{y}$ = 5.32
	2	13/17	53	5.1	$[x^2]$ = 10, [xy] = 7.2
	3	16/19	68	5.5	Slope for "N", b = 0.72
	4	14/15	87	6.1	Pooled slope b_c = 0.725
	5	15/15	100	6.5	EC50 = 2.559 (with b_c)

Table 3. First weighted estimate of EC50 for subjects "C" and "N" in Table 2, where $Y = \bar{y} + b_c (x-\bar{x})$ from Table 2.

Taster	Sol'n. x	Expected Y	Weight w	Working probit y	wx	wy
"C"	1	2.9	.0			
	2	4.0	1.1	3.8	2.2	4.18
	3	5.1	4.0	4.6	12.0	18.40
	4	6.2	2.6	6.8	10.4	17.68
	5	7.4	.5	7.8	2.5	3.90
"N"	1	3.9	.9	3.5	.9	3.15
	2	4.6	2.6	5.1	5.2	13.26
	3	5.3	4.5	5.5	13.5	24.75
	4	6.0	3.0	6.1	12.0	18.30
	5	6.8	1.3	7.3	6.5	9.49

Statistic	"C"	"N"	Statistic	"C"	"N"
Σw	8.2	12.3	$[wxy]$	7.837	10.993
$\Sigma(wx)$	27.1	38.1	b	1.587	.770
$\Sigma(wy)$	44.16	68.95	b_c	1.670	.840
$\bar{x}$	3.305	3.098	$[wy^2]$	13.350	9.171
$\bar{y}$	5.385	5.606	B^2	12.437	8.461
$[wx^2]$	4.938	14.283	χ^2	.913	.710
			X_5	3.074	2.377

<u>Table 4.</u> Clinical test on the comparative effectiveness of headache remedies in patients assigned in equal numbers to four groups. (Jellinek, 1946).

Experimental design

Group No.	Successive 2-week periods 1	2	3	4
I	A	B	C	D
II	B	A	D	C
III	C	D	A	B
IV	D	C	B	A

Drug composition

A – Ingredients a,b,c
B – " a,c
C – " a,b
D – Placebo (pharmacologically inert)

<u>Table 5.</u> Frequency distribution of "successes" with placebo, as reported by subjects who had taken drug D for 5 attacks of headache.

No. of headaches relieved	0	1	2	3	4	5
No. of subjects, this study	22	1	5	7	8	16
" " " , later study	27	0	1	5	10	19

<u>Table 6.</u> Combined analysis of success rates for the four groups in Table 4.

Subjects	Placebo non-reactors				Placebo reactors			
Mean rates	No. of subjects	Rates for drug A	B	C	No. of subjects	Rates for drug A	B	C
	79	.88	.67	.77	120	.82	.87	.82
Term	DF	MS		F	DF	MS		F
Subjects	78	.181		2.08	119	.119		1.86
Drugs	2	.999		11.48	2	.073		1.14
Subjects×drugs	156	.087			238	.064		

Table 7. Analysis of variance of a 2×2 factorial experiment on off-flavor in apples sprayed with thiram (Th) or sulfur and with lead arsenate (L) or parathion; from paired comparisons with degrees of preference transformed to rankits. (Bliss et al, 1956)

Row No.	Comparison of	Inconsistent subjects DF	MS	F	Consistent subjects DF	MS	F
1	Fungicides (Th)	1	.545	2.22	1	8.417	7.42*
2	Insecticides (L)	1	.239	.97	1	3.740	4.91*
3	Interaction Th×L	1	.004	.02	1	.681	3.07
4	Non-additivity	3	.168	.68	3	.020	.09
5	Tasters × Th	10	.193	.78	13	1.134	5.10**
6	" × L	10	.167	.68	13	.761	3.43**
7	" × Th×L	10	.284	1.16	13	.132	.59
8	" ×non-add.	30	.345	1.40	39	.164	.74
9	Within tasters	330	.246		420	.222	

* $P < 0.05$, ** $P < 0.001$

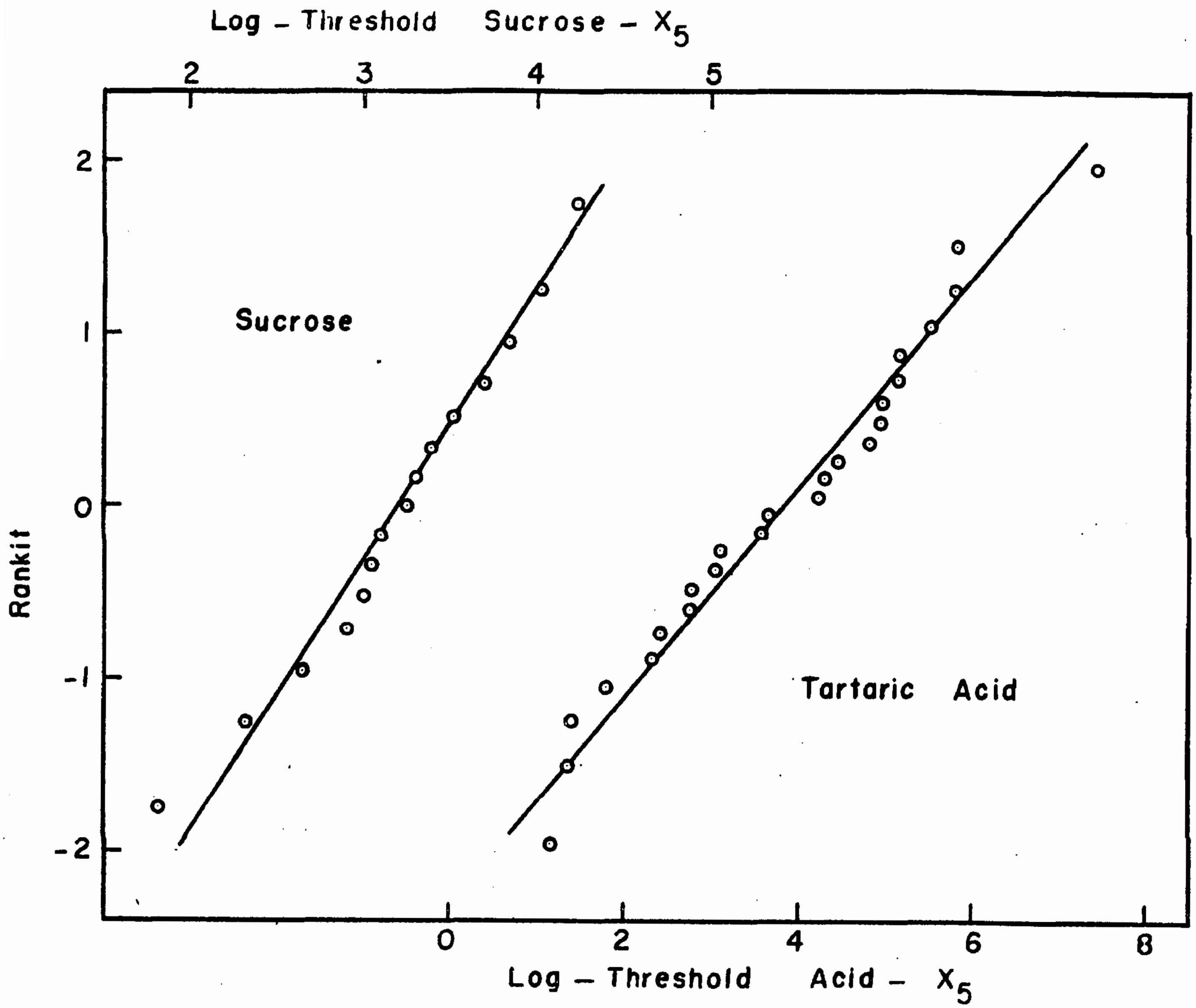


Figure 1. Graphic test for agreement with a normal distribution of the coded threshold concentrations X_5 in 15 subjects for sucrose and in 24 subjects for tartaric acid. Successive concentrations increase by doubling from 1 = 0.034% of sucrose and 1 = 586×10^{-8} moles per liter of tartaric acid; at rankit 0, the straight lines pass through means of $\bar{X}_5$ = 3.210 for sucrose and $\bar{X}_5$ = 3.834 for tartaric acid with slopes of b = 1/s, where the standard deviation s = 0.635 and 1.651 respectively.

LITERATURE CITED

1. Armitage, P. Restricted sequential procedures. Biometrika, 44, 9-26 (1957).

2. Baker, G.A., Amerine, M.A. and Roessler, E.B. Errors of the second kind in organoleptic difference testing. Food Research, 19, 206-210 (1954).

3. Baker, G.A., Mrak, V., and Amerine, M.A. Errors of the second kind in an acid threshold test. Food Research, 23, 150-154 (1958).

4. Bennett, G., Spahr, B.M. and Dodds, M.L. The value of training a sensory test panel. Food Technology, 10, 205-208 (1956).

5. Berkson, J. A statistically precise and relatively simple method of estimating the bio-assay with quantal response based on the logistic function. J. Am. Stat. Assoc., 48, 565-599 (1953).

6. Blakeslee, A.F. Genetics of sensory thresholds: taste for phenyl thio carbamide. Proc. Nat. Acad. Sci., 18, 120-130 (1932).

7. Blakeslee, A.F. and Salmon, T.N. Genetics of sensory thresholds: individual taste reactions for different substances. Proc. Nat. Acad. Sciences, 21, 78-90 (1935).

8. Bliss, C.I., Anderson, E.O. and Marland, R.E. A technique for testing consumer preferences, with special reference to the constituents of ice cream. Storrs Agr. Expt. Sta., Bull. 251, 1-20 (1943).

9. Bliss, C.I., Greenwood, M.L. and McKenrick, M.H. A comparison of scoring methods for taste tests with mealiness of potatoes. Food Technology, 7, 491-495 (1953).

10. Bliss, C.I., Greenwood, M.L. and White, E.S. A rankit analysis of paired comparisons for measuring the effect of sprays on flavor. Biometrics, 12, 381-403 (1956).

11. Bradley, R.A. Some statistical methods in taste testing and quality evaluation. Biometrics, 9, 22-38 (1953).

12. Campbell, W.I.P. and Blakeslee, A.F. Would a rose smell so sweet by any other name? Horticulture, 24, 333 (1948).

13. Carroll, M.B. Consumer product testing statistics. Flavor Research and Food Acceptance, Reinhold Publ. Corp., New York, 162-174 (1958).

14. Cotterman, C.W. and Snyder, L.H. Tests of simple Mendelian inheritance in randomly collected data of one and two generations. J. Am. Stat. Assoc., 34, 511-523 (1939).

15. Dykstra, O. A note on the rank analysis of incomplete block designs-applications beyond the scope of existing tables. Biometrics, 12, 301-306 (1956).

16. Fenton, F.E. Judging and scoring milk. Farmer's Bull. 2111, 1-20 (1957).

17. Finney, D.J. Probit Analysis. 2nd. Ed. Univ. Press, Cambridge, (1952).

18. Fisher, R.A. and Yates, F. Statistical Tables for Bio-logical, Agricultural and Medical Research. Oliver and Boyd, Edinburgh (1957).

19. Greenwood, M.L. and Salerno, R. Palatability of kale in relation to cooking procedure and variety. Food Research, 14, 314-319 (1949).

20. Gridgeman, N.T. Taste comparisons: two samples or three? Food Technology, 9, 148-150 (1955).

21. Gridgeman, N.T. A tasting experiment. Applied Statistics, 5, 106-112 (1956).

22. Gridgeman, N.T. Application of quantal response theory to the cross-comparison of taste-stimuli intensities. Biometrics, 14, 548-557 (1958).

23. Harries, J.M. Sensory tests and consumer acceptance. J. Sci. Food. Agric., 10, 477-482 (1953).

24. Helgren, F.J., Lynch, M.J. and Kirchmeyer, F.J. A taste study of the saccharin "off-taste". J. Am. Pharm. Assoc., 44, 353-355 (1955).

25. Hopkins, J.W. Incomplete block rank analysis: some taste test results. Biometrics, 10, 391-399 (1954).

26. Jackson, J.E. and Fleckenstein, M. An evaluation of some statistical techniques used in the analysis of paired comparison data. Biometrics, 13, 51-64 (1957).

27. Jellinek, E.M. Clinical tests on comparative effectiveness of analgesic drugs. Biometrics Bull., 2, 87-100 (1946).

28. Mackey, A.O. and Jones, P. Discernment of primary tastes in water solution compared with judging ability for foods. Food Technology, 8, 527-530 (1954).

29. Murphy, E.F., Clark, B.S. and Berglund, R.M. A consumer survey versus panel testing for acceptance evaluation of Maine sardines. Food Technology, 12, 222-226 (1958).

30. Murphy, E.F., Covell, M.R. and Dinsmore, J.S., Jr. An examination of three methods for testing palatability as illustrated by strawberry flavor differences. Food Research, 22, 423-439 (1957).

31. Pangborn, R.M., Simone, M. and Nickerson, T.A. The influence of sugar in ice cream. I Consumer preferences for vanilla ice cream. Food Technology, 9, 679-682. (1957).

32. Peryam, D.R. and Haynes, J.G. Prediction of soldiers' food preferences by laboratory methods. J. Appl. Psych., 41, 2-6 (1957).

33. Reid, A.W. and Becker, C.H. A study of cocoa syrups for taste preference. J. Am. Pharm. Assoc., Sci. Ed., 45, 160-162 (1956).

34. Sampson, W.L. and Fernandex, L. Experimental convulsions in the rat. J. Pharmacol. Exptal. Therapeutics, 65, 275-280 (1939).

35. Scheffé, H. An analysis of variance for paired comparisons. J. Am. Stat. Assoc., 47, 381-400 (1952).

36. Youden, W.J. Experimental designs to increase accuracy of greenhouse studies. Contrib. Boyce Thompson Inst., 11, 219-228(1940).

STATISTICS IN THE TEXTILE INDUSTRY

L.H.C. Tippett

British Cotton Industry Research Association
Shirley Institute, Manchester, England.

Statistics is now so widely used in many industries and the usages are so commonly understood, that it seems better for me to "highlight" a few special features of the usage of statistics in the textile industry than to attempt a wearisome, comprehensive list of the applications. In doing this I shall deal with applications that have come under my notice in my work with the British cotton industry, but I acknowledge that people working in the cotton and other branches of the textile industry in several countries could add much of interest to the subject.

My paper is divided into two parts. Part I mentions briefly a few general points and Part II deals more fully with the design of experiments.

Part I. General

I suppose that if one were to say in a word what statistics is about (or the sort of statistics under consideration at this conference) one would say that it deals with variation - with its description and measurement, and with its effects on scientific inference and decision. In many fields - in the manufacture of engineering components to a dimension, for example - variation is little more than a nuisance, but in textiles it is so important an attribute of the quantities of technical interest, that it is studied in its own right. Moreover, not only is the degree of variation important, but its pattern is also.

For example, cotton yarns vary in mass per unit length or thickness along their length and this variation can take the form of a mixture of random fluctuations, of almost-periodic fluctuations (with a period-length phase and amplitude that varies from place to place) and of strictly periodic fluctuations. The almost-periodic fluctuations are due largely to the variation in the length of the fibres and are inherent in cotton spinning. The strictly periodic fluctuations are usually caused by machinery defects such as eccentric rollers or faulty gears, and can be eliminated. When yarn is used, say, as weft or filling in a cloth or knitted into hosiery, periodic variations even of slight degree can form a pattern that is unpleasing to the eye, whereas random or almost-periodic variations of greater degree would in the same circumstances be harmless.

Thus, in the analysis of the causes and effects of yarn thickness variation, account must be taken of the pattern. In the statistical treatment a number of devices are used; periodograms, correlograms, and curves relating the variance of mass per unit length to the specimen-length. But of these, only some form of periodogram analysis leading to the identification of period lengths and amplitudes has, as far as I know, led to conclusions of technical importance. Electronic devices are now available for measuring the degree of variation and for identifying periods, and these are used in mills for appraising yarn quality and diagnosing the causes of defects.

Samples are much used in textiles, and there are two special features to which I call attention. The first is that in taking samples of fibres account must be taken of a bias towards selecting long fibres, either by adopting a technique that eliminates such bias (as is done for cotton) or by calculating the bias for different modes of selection (as is done for wool). The second feature is that hardly any textile appraisals are by attributes (involving classification into defectives and non-defectives); almost all are by measured variables. Most of the statistical literature on industrial sampling and most of the sampling plans apply to attributes, and so have little application to textiles.

Many experiments are done in the textile industry, in research departments and institutes and in mills, sometimes in order to increase technological knowledge for general application and sometimes in order to provide information on the best conditions of processing for some particular situation. There is much uncontrolled, and uncontrollable, variation in textile processing, and so the statistical design of experiments finds important application. Since that is the main subject of this conference, I shall devote the remainder of this paper to it.

Part II. The Design of Experiments

Although most space in textbooks on the design of experiments is devoted to statistical aspects, it is soon discovered by the practitioner and is widely understood that the whole situation has to be taken into account, technical as well as statistical. Dr. G.E.P. Box and his colleagues of the Statistical Techniques Research Group at Princeton have begun to study the wider aspects of experimentation systematically, and I am sure that in the years to come we shall see important progress. At present, however, the experimenter has to rely largely on unsystematized experience and common sense. This is the sort of situation in which case-histories are especially useful. Accordingly, I propose to discuss in some detail a field of experimentation in textiles, and shall try to bring out the general issues lying behind some of the particular considerations involved.

I shall discuss experiments to investigate the sizing of warp yarns for weaving. The warp threads are those that extend lengthways in a long piece of cloth; the process of weaving consists in interlacing with them the cross-threads, known as the "weft" or the "filling". During weaving, the warp is subjected to a good deal of abrasion and to considerable fluctuating tensions, and in order that it may withstand this rough treatment it is "sized" - i.e. the yarn is given a protective coating of some adhesive such as (for cotton) starch mixed with other ingredients. The subject of experimentation is the determination of the most suitable ingredients for the size and of the optimum amount to be put on the warp.

This subject has been investigated for many years, and a good deal is known about the sizing of the older fibres such as cotton and rayon with natural sizes (starches, gums, and gelatine); but the coming of the new synthetic fibres and adhesives has given the subject a new lease of life as one for investigation. Attempts have been made over the years to elucidate the fundamentals of sizing and weaving, and to develop laboratory tests; but the problem has proved to be intractable and, although some progress has been made, practical action requires the information derived from empirical experiments, conducted in a research institute or the mill, in which yarns are sized in different ways and their weaving performance observed. I shall discuss fully the topic of performance, but for the present shall characterise it as the warp breakage rate. From time to time during weaving the warp threads break and have to be mended. The warp breakage rate is important, and a low rate is, of course, to be desired.

What Factors?

In designing an experiment, the first thing to be decided is: what variables, or, in the jargon, what factors shall be investigated. The technologist's first answer will undoubtedly be the type of size and the amount on the warp. The type of size is not a simple thing since there are usually at least two ingredients: an adhesive such as starch or polyvinyl alcohol (PVA) and a lubricant such as tallow. On consideration, however, the technologist will agree that there are other factors that have an effect on weaving performance and that should be considered. The relative humidity of the atmosphere in which the weaving is done, the complex of factors under the heading of loom settings, and the cloth particulars (which may range from those for a fine cambric or poplin or dress fabric to those for a coarse sheeting) are only a few.

According to the classical method of experimentation one would investigate each of these factors, one at a time; but that is not good enough. The optimum amount of size is very different for one based on starch than for one based on (say) carob bean gum; the effect of relative humidity is not the same for all sizes, yarns, and cloth constructions; and so on. In statistical language there are interactions between the factors, and for complete information all relevant factors must be investigated in a so-called factorial experiment. The issue of the factorial versus the classical experiment had once to be argued; now it is decided and factorial experiments are generally regarded as the correct thing.

But in practice difficulties arise. The number of factors can be very large and if they are all included the experiment may become unmanageably large. In the book Design and Analysis of Industrial Experiments, edited by Dr. O.L. Davies, experiments in the chemical field with as many as five factors are described, but such a scale of operation would be impracticable in the field I am dealing with, and a selection of factors has to be made. I suspect that more often than not it is possible to think up more factors than can be dealt with in one experiment.

In sizing, the main ingredient of the size is the adhesive, and the technologist will usually be able from his general knowledge and the results of laboratory work to decide what other ingredients can reasonably be incorporated, and in what proportions. In this way, type of size as a factor can be reduced to the adhesive, although the situation once more becomes more complicated if two adhesives are used. The simultaneous inclusion of type of size (simplified in the way described) and the amount on the warp cannot be avoided since their interaction is very important.

These two factors result in an experiment that is as large as can usually be handled at one time and so the other factors are usually excluded. Relative humidity and loom settings are troublesome to vary and most technologists will be prepared to act on the assumption that their influence on the optimum type and amount of size is of a second order of importance. If the investigation is done in a mill, the management will be interested in one cloth at a time, and so cloth particulars can be excluded as a variable factor. A research institute serving an industry is interested in a wide range of cloths, but will prefer to cover the range by dealing with a limited number of typical cloths, and finds it acceptable as well as convenient to have a separate experiment for each one. Then, as results for each cloth are obtained, manufacturers weaving that cloth or something near it can immediately apply them; and as the results for different cloths accumulate, a pattern begins to emerge so that the whole picture can be filled out without exhaustive investigation.

How many Levels?

A second question to be decided is how many values of each factor there should be in the experiment - in statistical language, how many levels there should be. A related question is how they should be disposed. For example, we might in one experiment have two types of size (i.e. two levels of type) each at four amounts on the warp, (i.e. four levels of amount). This would give eight variations (termed treatments) in all.

When the factor is qualitative, as is the type of size, there is little to say about the chioce of levels except that it is the job of the technologist. When the factor is a measureable variable, two levels are enough provided there are grounds for believing that the relationship between the measured effect of the factor (termed generally the response and exemplified here by the warp breakage rate) and the level is nearly linear over the range of interest, or at least that there is no maximum or minimum in the curve. Many factorial experiments are done with two levels of each factor, and such seem to be very suitable for exploring a relatively unknown field in order to discover which factors are important.

In our sizing-weaving experiment, however, we have to take account of the fact that the breakage rate-amount of size curve usually has a minimum, and it is the breakage rate around this minimum that we require to know. Moreover, it is not quite enough to know exactly the minimum - we need to know the shape of the curve. For example, the curve might be like that in Fig.1*, rising more steeply on one side of the minimum than the other, and it would then be important to know this. In ordinary mill practice the amount of size cannot be controlled precisely, and in routine production one would aim at a percentage of size somewhat in excess of A, so that a small deviation below the aimed-at value would not lead to the large increase in response that would result from a small deviation below the value A.

---

* Figures are to be found at the end of this article.

In order to obtain the information required, at least four levels of amount of size are necessary. For comparing two types of size, one would require at least eight treatments - sometimes a formidable requirement. If the two types of size are similar in general character, say two starches, one might be able to reduce the number of treatments to (say) six by assuming the same general shape of curve for response to amount for each type. Then, I think that most technologists would use four treatments to establish the curve for one type and two to establish the general level of response for the other. If one had enough confidence in the identity of shape of the two curves I suppose that the most efficient arrangement would involve three treatments for each type, but I doubt if any technologist would have such confidence.

If the types of size are very different, I think that most technologists, if they could not have eight treatments in one experiment, would want to conduct three experiments, each with perhaps four treatments. The first experiment would explore the breakage rate-amount curve for one type, the second for the other, and the third would establish the relationship between the types at two amounts chosen after the separate curves have been established. (I am presuming that the average level of response for each experiment is uncontrollably different so that the results of the three cannot be compared directly.)

Someone will be sure to say that such a set of experiments is too expensive or troublesome. What we do then depends on the circumstances. If an adequate experiment really is impracticable, simplifying assumptions (or even guesses) may have to be made, and the resulting information may be better than nothing. Or the information from an inadequate experiment may be so slight as not to be worth the cost of the experiment. The last thing we should do is to allow anyone to be deluded into thinking that adequate information can be derived from an inadequate experiment.

Sometimes, there is no great pressure to reduce the number of treatments, and hence of levels, for each factor; can there be too many? For the purpose of this discussion let us suppose that we wish to explore the breakage rate-amount of size curve for one type of size, that its shape is roughly that of Fig.1, and that eight or ten treatments in the experiment are tolerable. Is it better to have eight or ten different amounts of size or to have four or five different amounts, measuring the response for each amount twice? This kind of question requires more discussion than I can give here, and I will venture an opinion, (for which I would not go to the stake) which is that in most situations it is best to have the minimum number of points necessary to delineate the broad outline of the curve, and correspondingly to have the maximum information for each point that experimental resources permit.

What Levels?

A decision has to be taken as to the range over which the experimental variables, the factors, shall be varied. If there are only two levels for any factor, and if the response curve is linear, the more widely the levels are spaced the more precisely is the slope of the response curve determined; but also, the more likely is the assumption of linearity to be seriously in error. This statement, I think, could be extended. If the mathematical form of the response curve is known, I suspect that for many forms the more widely the levels are spaced the more precisely is the curve determined for a given experimental error. But seldom is the form of the response curve known; seldom is it of a simple mathematical form. Technically we are only interested in the response curve in a certain region. The phenomena behind the curve are complex and I doubt if information for areas far outside the region of interest gives much information for areas within that region. When the curve is like that of Fig.1, for example, very low amounts of size lead to catastrophic results and are to be avoided; quite high amounts of size are tolerable, but even so they may not be advisable. For example, the response curves for two sizes I and II may be as shown in Fig. 2. Practically we would be interested in the region between A and B, anu I doubt if results at C would help.

Of course, in practice we cannot always define the region of interest or practicability. Then we have to make the best guess we can from previous knowledge of the kind of thing that happens, or a preliminary exploratory experiment may be desirable. In sizing-weaving experiments a good deal of prior knowledge is available, and if new fibres and new sizing substances have physical properties not very different from those previously encountered, it is not difficult to suggest a suitable range of variation.

At best, however, the range of interest is not known precisely, and in order to be sure of covering it, the experimental range should extend slightly beyond the presumed range of interest.

It does not require much reflection to decide that the experimental amounts of size need not be the same for the two types, nor need they cover the same range. For example, if the response curves were as shown in Fig.3, as they might easily be, one would explore the region A - B for size I and C - D for size II. There would be no interest in comparing the sizes at the same amount. This is convenient, for in most practice it is not possible to control the amount of size on the warp closely. One aims at a certain amount but only achieves something fairly near it and then, by subsequent analysis, determines the actual amount.

Seldom is there enough information to justify one in spacing the levels for each factor at other than equal, or nearly equal, intervals within the chosen region.

What to do with Excluded Factors?

Factors that are excluded from the experiment can either be controlled at a constant level or they can be allowed to vary and contribute to the random errors. Which we do depends on many circumstances.

For example, two important factors excluded from our sizing-weaving experiment are the relative humidity of the atmosphere in the weave room and the loom settings. An up-to-date mill will have the relative humidity controlled at a certain level, and results applicable to that level will be appropriate; relative humidity there will be controlled. A less up-to-date mill may not have such control, and then the relative humidity should be allowed to vary over the range normally experienced. This may present difficulties since relative humidity may have a seasonal fluctuation and an experiment extending over a substantial part of a year may be unduly burdensome and protracted.

Loom settings, on the other hand, vary somewhat from loom to loom in most mills, variations being associated with loom overseer. These variations must be covered by the experiment and contribute to the random errors.

So far I have discussed these issues from the view-point of the individual mill seeking empirical information for local application. A research institute serving an industry will have a wider interest and, logically, should cover the full range of conditions that occur in many mills. If the experiments are done in mills but are under the control of the institute, it will be practicable only to treat each mill experiment independently as though the work were being done for an individual mill, and to generalise as the results for different mills accumulate. When the experiments are done in experimental workrooms the experimenter will usually prefer to control all the excluded variables each at one level, as far as possible. Then he will have sound results for a defined set of conditions, which can be built into a rising edifice of knowledge. Any generalisations that it may be expedient to make at any time will be the result of speculation guided by such knowledge as is available.

Sometimes it is not easy to say what is meant by constancy of a factor. For example, consider a size mixing containing an adhesive such as starch, and a lubricant such as tallow. The effect of the lubricant, although not unimportant, is secondary to that of the adhesive, and often the amount and type of adhesive is investigated, the lubricant being kept "constant" in type and amount as an excluded factor. What is constancy here? Would one keep constant the absolute weight of lubricant per 100 lb of yarn sized or the weight of lubricant relative to that of the adhesive? The only sure way of answering this question is to do a factorial experiment. In the meantime, the usual view is that the lubricant lubricates the size rather than the yarn, and the weight relative to the amount of size is the basis adopted.

What Responses?

Two general points arise when deciding for an experiment what observations shall be taken and what measurements shall be considered as responses. These are (1) the view must be so broad that all relevant effects are considered, and (2) compromises must be struck when there are conteracting effects. I shall illustrate these.

In our sizing-weaving experiments, the people responsible for production and wages are interested in reducing warp breaks experienced in weaving to a minimum since they add to the weavers' work-load. The quality control department are interested in cloth quality which, other things being equal, is imporved as warp breaks are reduced. But some sizes might reduce warp breaks to a minimum but be deleterious to other aspects of quality such as "cover" or "cannage". Further, some sizes give low warp breakage rates but may be difficult to remove in finishing, and would be unacceptable to the people in the finishing department. Finally some sizing materials are more expensive than others: some prepared starches, for example, cost twice as much per ton as their natural equivalents. The experimenter should take all these considerations into account in deciding what observations to take and how to appraise the results.

If all aspects - warp breaks, cloth quality, and so on - can be evaluated in terms of costs, it is relatively easy to decide on the optimum size. But such evaluation is not possible for all aspects, and the technologist must assess the different results qualitatively and use judgment in striking the best compromise. He will probably choose from the sizes that give nearly the lowest warp breakage rate those that are satisfactorily removed in finishing, and if there is further room for choice of these he will select those that give the best cloth quality.

Sometimes the statistical analysis is facilitated by mathematically transforming the variable in which a response is measured - by analysing the square root of the breakage rate, for example. I have done this sort of thing but am not sure that such action is not sometimes an exercise in statistical virtuosity rather than a good thing to do. In any event we must remember that the final report has to be made to a technologist and all figures must be given in terms that mean something to him. He can interpret a warp breakage rate but not its square root.

What Experimental Plan?

The experimental plans or designs now available are many more than the simple randomised blocks and Latin squares which held the field in the early days of the subject. In the experiments I am discussing, however, experimental treatments are few and simple plans are appropriate.

The natural experimental unit in a sizing-weaving experiment is a warp, containing yarn to make several hundred yards of cloth, sized with one size throughout. This will go independently into one loom in the weave room and will take between two and six weeks to weave. At the sizing process warps are produced one at a time successively from the machine, each taking perhaps an hour to run; and thirty or so warps, requiring three or four shifts to run, form a set. Sizing is an almost continuous process, with only very short stops at the end of each warp and somewhat larger stops at the end of each set. The capacity of the size tank is considerable, so that it is quite a business to change the size in type, although the amount put on the warp can somewhat more easily be changed in a downwards direction by adding water in the "sow box" (which contains the size actually in the process).

The problem is to superimpose on this industrial set-up an experiment so that production is not interfered with unduly. Suppose that there are two types of size I and II, and four amounts of each, leading to eight treatments, say I1, I2, I3, II1, II2, II3, II4, and that there are four warps for each treatment.

The statistician's ideal would probably be an arrangement with four blocks each of eight consecutive warps, and the treatments distributed at random within each block. This would be intolerable to the mill since it would involve changing the size at the end of each warp. The most that the mill is likely to tolerate is a new type of size for each of four shifts, with the amounts of size being successively reduced within each, so that the arrangement would be:

Shift A:	I1	I1	I2	I2	I3	I3	I4	I4
Shift B:	II1	II1	II2	II2	II3	II3	II4	II4
Shift C:	II1	II1	II2	II2	II3	II3	II4	II4
Shift D:	I1	I1	I2	I2	I3	I3	I4	I4

Then each pair of shifts would form an independent sub-experiment with two replicate warps for each treatment. The arrangement violates the canons of sound experimentation since the treatments are not distributed at random (they are in order of decreasing amount of size), and the replicates are consecutive. But this, or something like it, is the best that the mill is likely to tolerate.

When the warps are produced one would like them to go into the weave room according to some pattern, but it is likely that they will have to go into the looms as they become vacant, the warps being chosen at random only if several are available when one is called for. However, this arrangement is likely to be substantially a random one.

An experiment of this sort is not valueless, even though it is not entirely satisfactory. The effect of amount of size cannot be disentangled from that of the order of sizing, but the order effects are unlikely to be the same for the two sub-experiments and, with care, should be small. Moreover, the variance between replicates within the same sub-experiment can be compared with that between sub-experiments to show whether there is a substantial position effect. I think that technologists, with the background knowledge they possess, will easily reach useful conclusions from the results of such an experiment.

I think that when experiments are superimposed on normal factory production, it will usually be advisable to have two or three small independent sub-experiments and to make the arrangement within each sub-experiment simple to operate, introducing such randomisation and subtleties of arrangement as are expedient, but not worrying overmuch if the arrangement is more systematic than a statistican would like. A former colleague, Mr. R.E. Peake, describes an experiment in a spinning mill (Applied Statistics, 2, 1953, pp 184-192) in which a Latin square arrangement would have been appropriate were it not that that would have involved a certain group of machines working continuously for several weeks on the same product. This condition could not be ensured and the experiment had to be divided into independent sub-experiments, each lasting about a week. Within each sub-experiment a random arrangement of treatments was feasible, so that the whole formed a randomised block experiment.

When an experiment is done at a research station or institute, complication in the arrangement is practicable and may be desirable. For example, in a sizing-weaving experiment, we may have four warps each with a different size, and each divided into four sub-warps. These can be woven simultaneously in four looms and at the end of each sub-warp the warps can be interchanged between looms on a Latin square plan. Then in the analysis, loom effects, which can be quite substantial and contribute to the errors in the above plans for factory use, are eliminated from the comparisons between warps. Adequate replication of the sizing can be achieved by having two or more sets of four warps. If there are eight treatments, an 8 x 4 plan may be used.

What Size of Experiment?

In principle two things are required to decide the size of experiment: the precision with which the response is to be determined, and the extent and pattern of error variations likely to be encountered. On the second, a good deal is known for sizing and weaving experiments as conducted in Lancashire. Mr. E. Bradbury and Mr. H. Hacking have dealt with experiments in factories (Journal of the Textile Institute, 40, 1949, pp P532-P551) and Mr. V.R. Main, and I have dealt with experiments as conducted at the Shirley Institute (Journal of the Textile Institute, 32, 1941, pp T209-T220). I do not think that it is necessary to do more than make crude estimates from standard errors calculated on large-sample theory.

In many experiments in textiles we are interested in the rate at which the yarn breaks in processing, or in the incidence in time or space of various defects, and these are chance events distributed more or less at random, i.e. more or less according to the Poisson or negative exponential law. This is convenient because it enables us to calculate in advance how large an experiment needs to be. In practice, there are other uncontrolled variations superimposed on the chance variations so that the size of experiment so determined is too small. Nevertheless the calculations are useful in showing roughly the scale of experimentation required, and in setting a lower limit. It is disconcerting to many experimenters and practical men to find that the necessary scale is much larger than anything that they had previously contemplated, and that carefully controlled, small-scale experiments, perhaps with the warp yarns to be compared woven side by side in strips in the same loom, do not suffice. Such arrangements cannot reduce the purely chance variations.

If no prior data are available it is necessary to proceed in a sequential way, starting with a fairly small experiment, examining the results, and then extending the experiment stage by stage until adequate precision is attained.

Execution of Experiment and Collection of Results

It is axiomatic that after an experiment has been planned, the specified procedures and conditions should be closely adhered to and the data should be correctly recorded. These things are for the technologist or experimenter rather than the statistician, and are apt to be taken too much for granted.

Good statistical design is not a substitute for careful experimental control - it is complementary. In our weaving experiments at the Shirley Institute, we have found that by unremitting attention to detail, the precision of the results has been improved enormously. This is not a matter about which I can find much to say, but I do emphasize its importance.

Analysis of Results

The standard statistical procedure for treating the results of an experiment is to analyse the variance and test the significance of the various effects. This is always a good thing to do in order to restrain the ever-optimistic experimenter from reading into the results more information than is there. And the comparison of the error variance for a particular experiment with error variances commonly experienced provides a check that the control has been good.

But the main scientific or technological interest lies in measuring the response for different values of the variable - in measuring, for example, the relationship between amount and type of size, and the mean warp breakage rate. For this, the plotting of graphs in the usual way provides a great help, and we have found that if a simple experiment is well planned and carried out, the technologist can interpret the results without recourse to recondite statistical methods. Statistical principles find their most important application in the planning stage rather than in the stage of analysis of results.

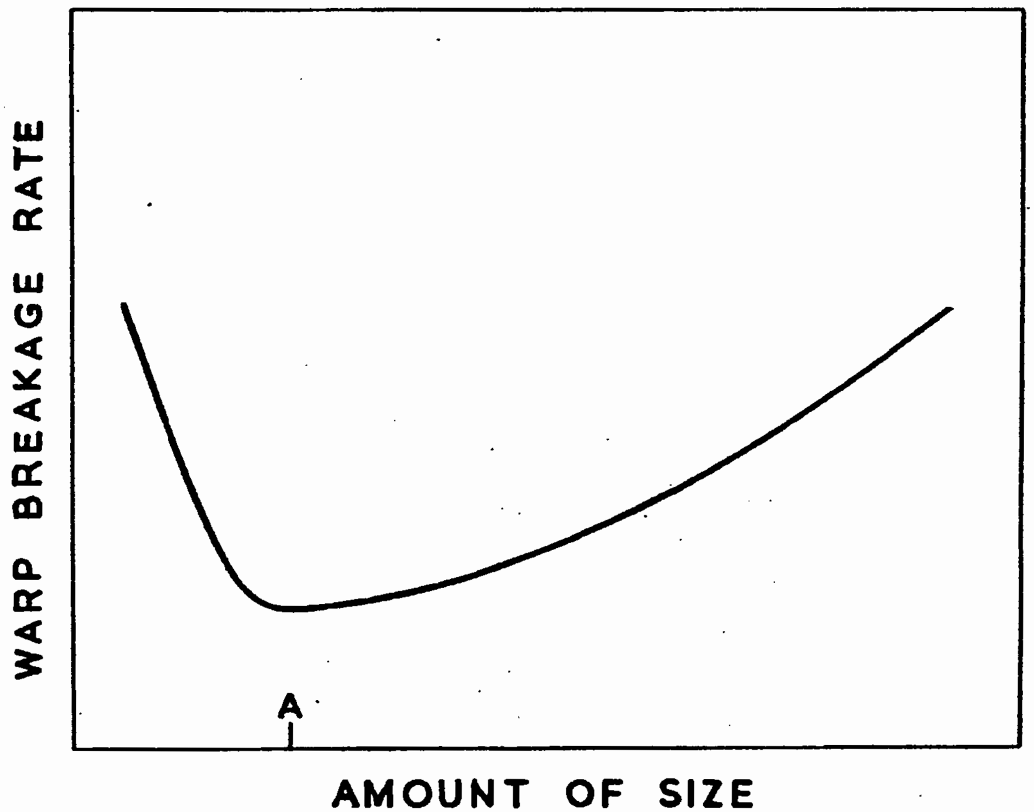


Figure 1

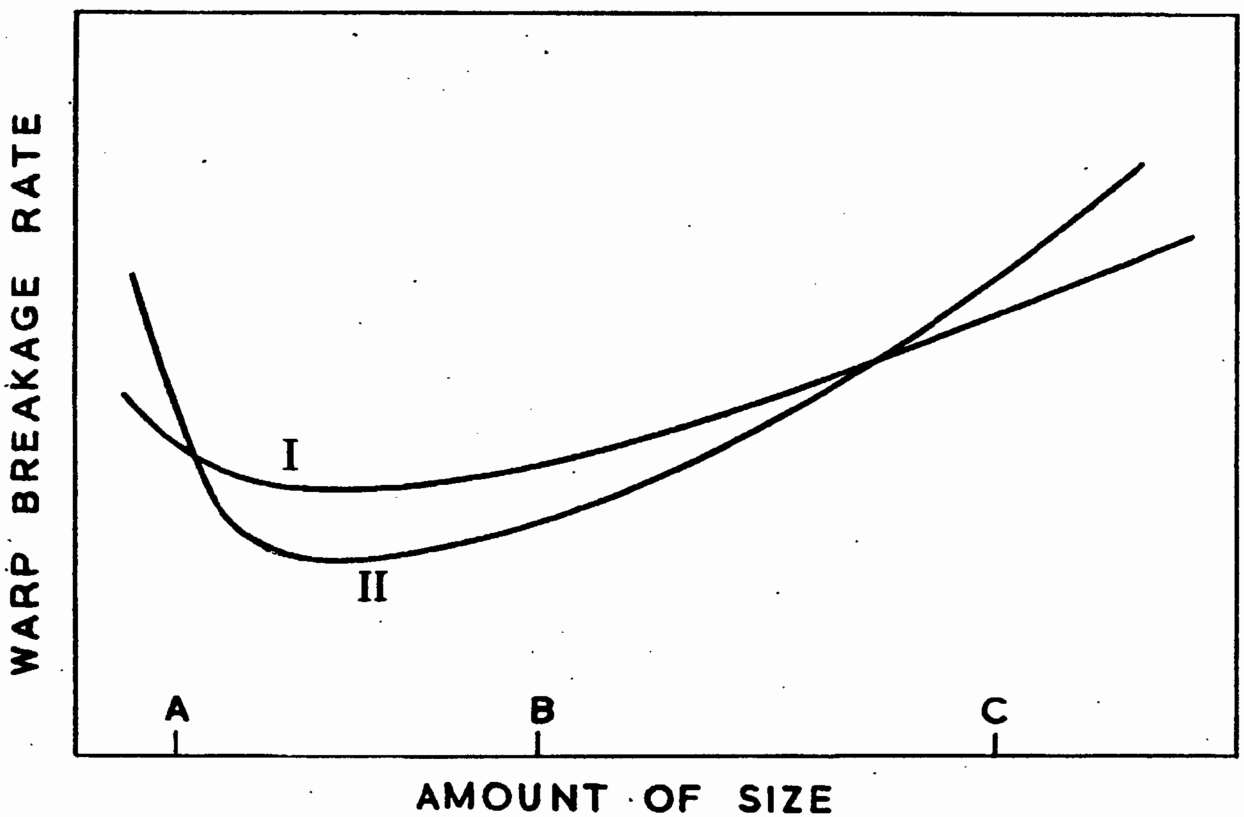


Figure 2

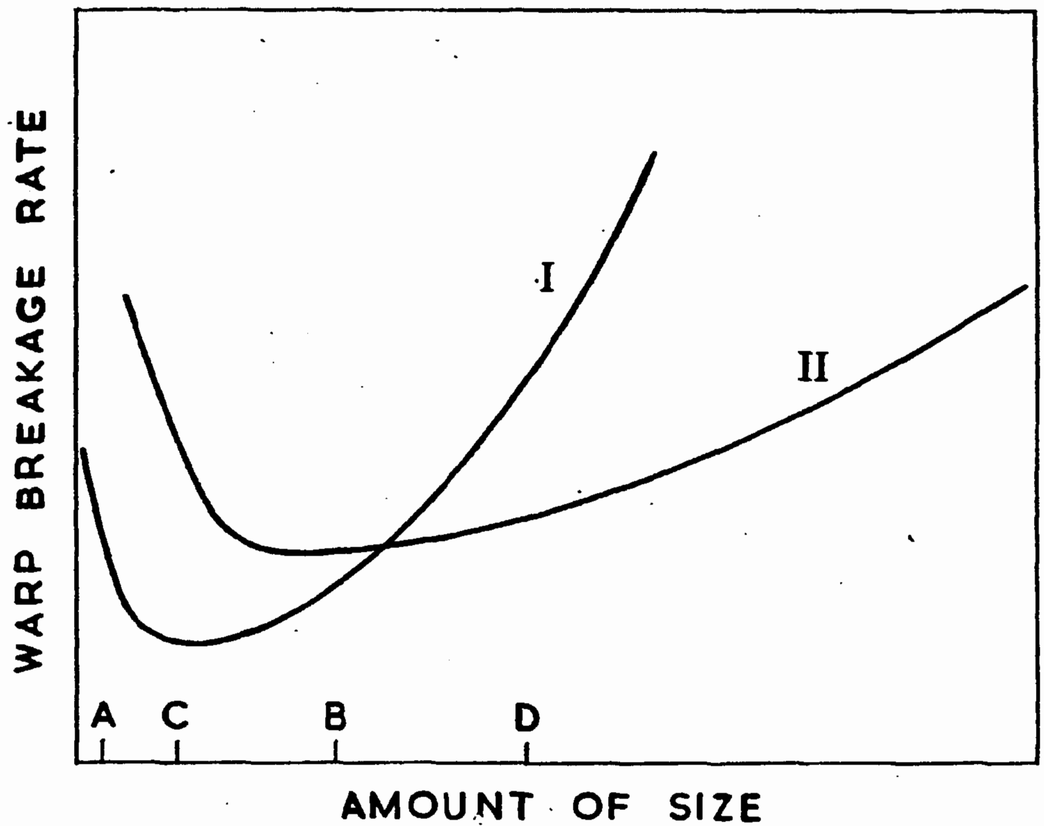


Figure 3