RELIABILITY, PROBABILITY, AND BINOMIAL INFERENCE

W. A. Thompson, Jr.
University of Delaware

INTRODUCTION. This paper attempts to answer two questions: (1) What do we mean by the reliability of a piece of equipment (a radio for example), and (2) in what senses can we assure this reliability? The emphasis will be primarily on the logical and philosophic aspects of the problem rather than on technic. This is not, then, a "how to do it" manual on reliability but rather a preliminary discussion of the foundations of the subject. In order not to obscure the foundations, the examples have been kept very simple; however, these illustrations have been adapted from the author's own practical experience and hence have a basis in fact.

The problem of reliability prediction and control may be looked at from several points of view. 1. Given an existing piece of equipment, how can we analyze and/or test it in order to determine its reliability? 2. What corrective actions may be initiated in the production process which will improve reliability? 3. What testing and sampling methods can we use which will prove to the customer that the equipment has the reliability he asked for? Or from the customers point of view, what methods can be used to monitor the reliability of equipment which is being purchased? 4. What guiding principles (such as redundant components) should be followed in order to design reliability into the equipment? A first answer to the question of "What is reliability assurance?" is that it is the collection of all techniques which throw light on one of the four preceding questions. We hope that the remainder of this paper will show that the first answer is not the whole answer.

I. RELATION TO SCIENTIFIC METHOD IN GENERAL. Before focusing our attention on the details of reliability engineering it would perhaps be advisable to examine the relation of the subject to scientific method in general. It is the point of view of common sense that a piece of equipment is real and that the properties which we perceive are in the equipment. Thus a radio is a real object which has several qualities in it. Examples of these qualities might be 1. being painted green, 2. being hard, or 3. being reliable. Adopting this common sense point of view, we find that the circumstances surrounding the operation of the equipment are usually too complicated to comprehend in their entirety. Thus it is necessary to idealize the real radio and study a

conceptual one instead. This conceptual radio is called a model. The trick is to choose a model which is simple enough to work with and yet complicated enough to describe the operation of the device to a desired degree of approximation.

In short, knowledge about the reliability of a piece of equipment is knowledge about an abstraction rather than knowledge about the real device. Thus it will be important for us to distinguish between the real device which common sense tells us is "out there" and the conceptual or abstract DEVICE whose reliability we may study. As in the previous sentence we make this distinction by using all capital letters whenever the abstract noun is intended.

The job of specifying the nature of the EQUIPMENT is not as easy as it at first appears and careful study of this aspect of the undertaking is frequently very fruitful. Quite aside from the question of representative samples there is usually a difference between the universe of devices which are investigated and the universe to which the conclusions are applied. For example, one procedure in determining the average failure time of a DEVICE is to test a single unit until it fails, replace the failed part and inspect the unit to restore it to its original condition then test this repaired unit until it fails, etc. The average failure time is taken to be the long run ratio of total operating time to total failures. Here the universe of investigation is the collection of all repaired states of a single unit while the universe of application is surely the collection of all units which are produced by a certain process. The validity of the entire study depends on how closely these two universes approximate each other.

Without going into too much detail the DEVICE will be characterized by a process of cross classification. Thus the DEVICE may consist of all devices of a certain design produced in 1950 at a specific factory and to be used in the continental United States. The possibility of classifying a set of objects so that the set has properties other than those used in the classification is a basic assumption and an observed fact of science.

One classification which deserves special attention is the class of failures to be considered. Several possibilities are illustrated by the following questions. Will the radio work when the customer first drops in a battery? If it does work initially then how long will it give satisfactory service? Does satisfactory service mean no failures or does it mean only easily repaired failures? Must the radio work continuously or only for a few hours each day? Are we interested in failures due to design faults, the process being out of control, or only in workmanship failures? Several

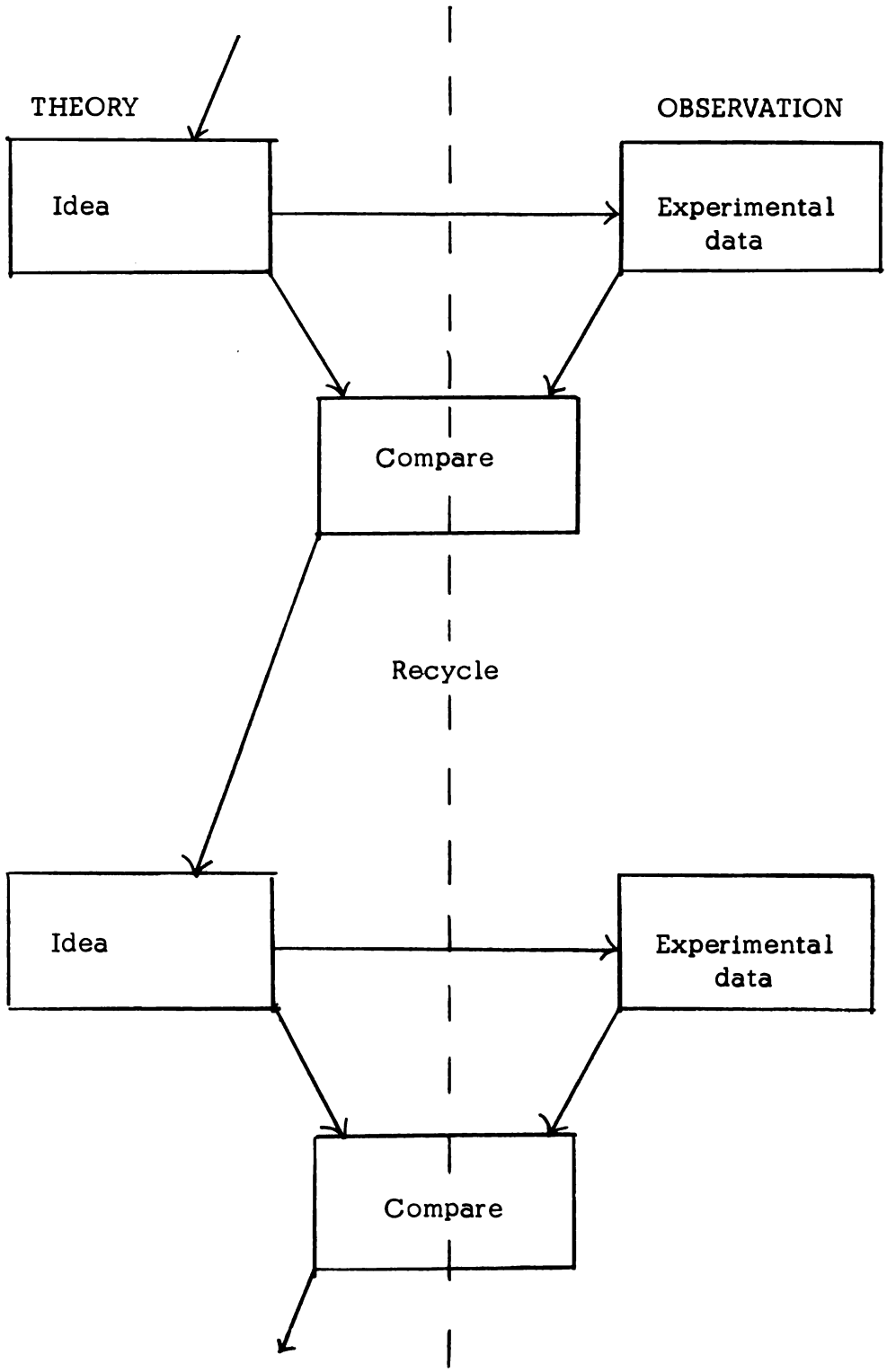
different techniques would be needed in order to treat these different questions but in this paper we will be able to discuss in detail only those situations where the device either fails or does not fail. In particular the important literature connected with life testing is much too extensive to be reviewed at this writing.

Certainly there is no single scientific method which provides a simple set of rules for "doing science"; however, complex activities are frequently better understood by considering a simplified version of them. In this spirit we present the following view of scientific method as a process of successive approximation. Schematically this process may be shown as proceeding as in H. C. Sweeney's diagram which appears here as Figure 1. An example follows of how the diagram could work in a reliability situation. The initial idea might be the reliability design specifications of a complex industrial valve to be used in turning off a rocket motor at a preassigned time. An experimental valve is then built and tested. It is found that with sufficient adjustment this experimental unit can be made to work for short periods of time in the laboratory. It is decided that if the design is altered and care is taken in the construction then the equipment will be feasible. A limited production run is undertaken to provide several units for test and demonstration. All valves work well in the laboratory but when the items are shipped for demonstration it is found that few of them work at all and that considerable training is needed to adjust the valve for initial operation. If the valve is designed more ruggedly and installation training is provided, it is believed that the reliability will be satisfactory. The decision is made to manufacture and promote the item. When the valve is mass produced and user tested, anticipated failures did not materialize for those valves which were carefully handled but very few of the devices receiving rough treatment performed satisfactorily. The reader should not think of the example as stopping at this point but as continuing on through cycle after cycle, each new model of reliability depending on all the previous models and all the previous data gathered. However, the general pattern is now clear. The reliability of the DEVICE is an evolutionary thing with the universe of investigation continuing to change and for the most part expand.

II. RELIABILITY AND PROBABILITY. The more advanced reader will want to skip this and the next section. There are many aspects of reliability which are not probabilistic in nature; for example, we might discuss with relevance the industrial revolution, piece-work and management science. Though these topics are interesting and important they do not fall within the scope of this

FIGURE 1

SCIENTIFIC METHOD AS SUCCESSIVE APPROXIMATION



article. As the title indicates this paper concerns itself with examining that substantial portion of reliability which is related to probability. In this section we take the point of view that the meaning of "the device functions correctly" becomes clear as soon as the abstract DEVICE, whose reliability is being studied, is specified. Remember that the abstract DEVICE is determined by a process of cross classification where the criteria of classification include such things as the conditions of operation, the kinds of failures to be considered, etc. For the purposes of this paper, the reliability of a DEVICE is then defined to be the probability that it functions correctly. This definition is more general than some in that the role of time to failure is not explicitly emphasized. Thus when time is a determining factor it may be included in the definition of "functions correctly" but if time is unessential the definition still applies. It is instructive to discuss this meaning of reliability by recalling a few of the various concepts of probability which have been proposed and specializing these to the reliability situation.

Before discussing probability it is necessary to have in mind the primitive concept of an experiment. For example, in studying the effect of gravitation on falling bodies a ball is rolled down a prearranged inclined plane and the resulting time required for the ball to make the trip is recorded. It is customary to think of this experiment as being deterministic in that the conditions of the experiment completely determine the result. For other experiments there will be various possible outcomes and it will be impossible to predict which of these will occur on any particular performance. This latter kind of experiment is termed probabilistic.

The possible outcomes of a probabilistic experiment are called events. Some of these events are thought of as being simple or indivisible while others are considered to be compound in that they are composed of simple events connected together by means of the words "and", "or", "not". If the occurrence of the event A necessarily implies the occurrence of B then we say that B contains A as a subevent. The compound event which contains all possible outcomes of the experiment is worth naming, we call it the universal event. The simple events are to be thought of as having the following two properties: 1) they are exhaustive in the sense that when the experiment is performed one of the simple events always occurs and 2) they are mutually exclusive, it being impossible that two simple events will occur on the same performance of the experiment. The illustration of a valve which is to shut off the flow of a liquid at a given instant provides an example which is typical of reliability work. Here the simple events are i. the flow of liquid is stopped and ii. the liquid continues to flow.

A more complicated probabilistic experiment is provided by setting an alarm clock (or fuze) and measuring the alarm timing error which equals the actual time when the bell first rings minus the time for which the alarm is set. If the measuring instruments are accurate to the nearest tenth of a second and if we think of the alarm error, in seconds, as being plotted on a line then the simple events will be all multiples of 1/10 within a certain interval, say +300 to -300. Since the length of this interval is long compared to the accuracy of the instrument (600 versus 1/10) it is convenient to idealize this experiment to include all points on the line as simple events. This kind of idealization is essential in many situations in order to obtain a model which is of manageable proportions. The model is particularly appropriate since if the measuring accuracy is improved to say a hundredth of a second then all multiples of 1/100 between +300 and -300 actually are possible outcomes of the experiment. In passing note that if the model is to be an accurate approximation to the true situation it will be necessary to assign a very small amount of probability to that portion of the line which is outside the interval +300 to -300.

The historically first concept of probability is referred to herein as the equally likely definition. This requires that we decide on m equally likely simple events of which c imply the occurrence of a compound event of interest. The probability of the latter event is then taken to be the ratio of c to m. Thus if the six faces of a die are taken to be equally likely and if C denotes the compound event of observing an even number of dots on the up face then the probability of C, written $P(C)$, is $3/6$.

If B and C are two events then a third event, (C and B), may be formed by considering their simultaneous occurrence. Further, if c simple events imply C and of these d imply (C and B) then the elementary formula

$$(d/m) = (c/m) \cdot (d/c)$$

yields the important multiplication rule:

$$P(C \text{ and } B) = P(C) \cdot P(B|C),$$

where $P(B|C)$ stands for the probability of the event B if the occurrence of C is made part of the conditions of the experiment.

The most important objection of defining probabilities in terms of equally likely is that many of the most interesting applications cannot be formulated in this way. While the equally likely idea of probability works quite well for the illustration of rolling a die; in both the valve and the alarm clock examples it would be exceedingly difficult to intuitively fix upon m mutually exclusive and equally likely simple events. The equally likely definition originates from a time when it was thought necessary to base all mathematics on "self evident truths" but there is nothing self evident about equally likely for the reliability type of problem and hence it would seem that we must look elsewhere for an adequate concept of reliability. If we adopt the point of view that it is essential to be able to verify a probability by observation and we ask what is available to check the correctness of a probability then we are immediately led to the frequency definition.

In order to talk about frequencies we must first discuss what is meant by n performances of the same experiment. If n experiments differ only in unimportant conditions then we will say that they are n performances of the same experiment. The apparently necessary vagueness at this point seems to be a drawback to <u>defining</u> probabilities as frequencies. We will see that independence is what is needed here, but this concept will be defined in terms of probability and hence a definition of probability in terms of independence would be circular.

If, in n performances of an experiment, a specified event C occurs s times then the <u>frequency</u> of occurrence of C is s/n. According to the <u>frequency definition</u>, s/n is a measurement of a permanent numerical property of the experiment; this property, if it exists, is called the probability of C. In short, probability is that property of an experiment which is measured by frequency. All measurements are inaccurate to a greater or lesser degree; in the present instance we would expect the accuracy with which s/n measures P(C) to increase with n. If in n trials, C occurs s times and the compound event (C and B) occurs s_1 times then $s_1/n = (s/n) \cdot (s_1/s)$ which again yields the multiplication rule:

$$P(C \text{ and } B) = P(C) \cdot (B|C).$$

There is a difference however, in that the probabilities involved are now probabilities in the frequency rather than the equally likely sense.

We may try to interpret our concept of reliability in terms of frequency probabilities; this yields something like the following: Reliability is that property of a DEVICE which is measured by the frequency of correct functioning. Though it is implied by our convention concerning the use of nouns with all capital letters, it is worth emphasizing that a frequency probability statement refers to an individual item only in so far as it is a member of a class. If we choose a particular valve from an assembly line and mark an * on it, then we should not speak of the reliability of the valve with the *, but only of the reliability of a class of valves of which this particular one is a member.

In the previous two definitions we have taken probability itself to be primitive. In the first instance we disguised this fact by using the words equally likely and in the second we emphasized the measurement of probability and ignored the nature of the concept. It appears that it may be necessary to take probability, and hence reliability, as in intuitive and undefined concept . If this is so then we might as well shift the emphasis from the nature of probability to how we want it to behave. It is this line of reasoning which leads to the axiomatic theory. According to this treatment a number not less than zero and not greater than one is allowed to correspond to each of the events of a probabilistic experiment; the number corresponding to the event C is denoted by P(C) and is called the probability of C. We insist that this correspondence have the following two properties: 1.) the probability of the universal event is one, and 2.) if A, B, C ... are mutually exclusive events then

$$P(A \text{ or } B \text{ or } C \text{ or } \ldots) = P(A) + P(B) + P(C) + \ldots .$$

For axiomatic probabilities the logical status of the multiplication rule is simply that of a definition. By analogy with the frequency and equally likely theories we define conditional probabilities so that the multiplication theorem will hold.

Two experiments are said to independent if $P(C \text{ and } B) = P(C) \cdot P(B)$ whenever C and B are possible results of the first and second experiments respectively. From the multiplication rule we see that this is the same as requiring $P(B) = P(B|C)$ i.e., the probability of every possible result of the second experiment is independent in the grammatical sense of what has happened on the first experiment. If the potential results and their probabilities are identical for a sequence of n independent experiments then the sequence is said to consist of n independent trials. Earlier when we

introduced the frequency definition it was the concept of independent trials which was needed but unavailable. Of course, the idea of independent trials could be developed in the above manner from either the equally likely or the frequency definitions, but independence would be available only after probability is introduced and hence could not be made the basis for defining probability.

The main difficulty with the axiomatic approach is that at this point one is tempted to say, "Yes, but what is probability?". This objection is lessened somewhat by the law of large numbers [8] which shows that there is a close relation between the axiomatic theory and the frequency definition. The axiomatic development does not yield an exact counterpart of the precise frequency definition. In fact, if f_n is the frequency of occurrence of the event C in the first n performances of an experiment and if the axiomatic probability of C is 1/2 then it is conceptually possible that C could occur in each of an infinite sequence of performances. In this eventuality f_n would equal 1 for all n and the sequence would approach 1 which is not equal to 1/2. However, in the axiomatic theory it is demonstrable that in a sequence of independent trials, f_n approaches P(C) with probability one. Thus the cases where f_n does not approach P(C) forms a negligible exception.

At first consideration the axiomatic theory seems to add little to an understanding of the nature of reliability since it requires the introduction of many events which are not clearly pertinent. However, the following consequence of the law of large numbers (which in turn can be developed as a consequence of the axiomatic theory) seems to give some real insight into the nature of reliability: Except for cases having probability zero, the reliability of a DEVICE is the limit of the frequency of correct functioning in a sequence of independent trials.

We now return to the probabilistic experiment of observing whether a valve does or does not shut off the flow of a liquid at a given signal. If this experiment is performed on n identical valves and if the probability of correct functioning at each performance is p then it will often be appropriate to think of the n performances as constituting independent trials. Under these circumstances the probability of exactly s successful valves in some particular order is $p^s (1-p)^{n-s}$.

Thus if $\binom{n}{s}$ denotes the number of ways of ordering s successful and n-s unsuccessful valves, then the probability of exactly s successful

valves in an arbitrary order is

$$\binom{n}{s} \rho^s (1-\rho)^{n-s} .$$

The entire series of n performances is called a binomial sample and s is said to have the binomial distribution.

It is important to point out that 99% reliability does not mean that at least 99 out of each 100 pieces of equipment do what they are supposed to. The law of large numbers hypothesizes a sequence of independent trials and the frequency definition requires n performances of the same experiment. It is conceivable that each of 100 devices might have a 1% probability of failure, but that if one fails then all the rest will fail also. However, even if the 100 devices are independent, the reliability of a single device must be at least .9995 in order to insure that 99 out of 100 items will function properly with an axiomatic probability of 95%[1]; and in fact, no reliability less than 1 will guarantee that 99 out of 100 items do what they are supposed to do.

III. TESTS OF SIGNIFICANCE. The logic of a test of significance is patterned after that of proof by contradiction; the main difference being that in certain places the word "false" is replaced by the phrase "very unlikely". Thus to accept R as being true, provisionally assume that R is not true. Next collect a sample and examine this sample using a probability argument. If the sample is very unlikely under our provisional assumption, then two explanations are possible. We may either believe that R is false and we have observed a very unlikely event or we can believe that R is true. Many people will prefer the second explanation, as the first is akin to "believing in miracles". Of course it is possible that the truth of R may be considered to be a "miracle" of higher order than the above mentioned very unlikely event, but this would seem to indicate that either the conclusion possibilities have been artificially restricted or all explanations of the sample are unlikely and consequently that the sample is not the whole of the available information.

[1] Consider the equation $100\rho^{99}(1-\rho) + \rho^{100} = .95$.

A manufacturer agrees to supply one of his customers with radios which have a 99% initial probability of working correctly. How may the customer check to see whether the manufacturer is abiding by his agreement? The only answer which immediately presents itself is to gather and examine experimental evidence. We take the simplest case where this evidence constitutes a binomial sample. Thus knowing that s out of n radios perform correctly we would like to determine whether ρ could be as large as 99%. It is intuitively clear that large s values are consistent with the belief that ρ exceeds 99% and that small s values are inconsistent with this belief. Thus we may adopt a cut-off value, v, such that if s is less than v we declare that ρ is too small and the manufacturer has violated his agreement.

Applying the general procedure of the second previous paragraph we find, from tables of the binomial distribution, that if an event which occurs with probability .0138 is held to be very unlikely than 48 is an appropriate value for v. More explicitly, if 3 or more radios out of 50 fail to work initially then we declare that the manufacturer has violated his agreement; however, if 2 or fewer failures are observed, then we say merely that the data is consistent with believing that $\rho \geq 99\%$. But the data would be consistent with other beliefs as well. For example, 2 failures is even more consistent with the belief that $\rho = 96\%$. Thus we cannot claim to have proved that the manufacturer is living up to his agreement.

Before leaving tests of significance, it should be pointed out that this theory gives no help in defining the expression "very unlikely" and experience has shown that examining the specific situation only helps a little. The difficulty of precisely defining the meaning of this expression is then a noticeable weak point in the theory of tests of significance.

IV. DECISIONS AND HYPOTHESIS TESTING. Continuing the radio example of the previous section, suppose that instead of desiring to monitor reliability we want to make a "decision" concerning whether the reliability is adequate or not. It should be pointed out that some writers would refuse to recognize a difference in these two objectives. Here, we do not comment on this point but discuss one possible rational basis for making such decisions. Presumably, any rule for making decisions should take account of the experimental evidence, s; hence we speak of the decision rule D(s) which, for example, makes the decision D(n) when all tested radios work satisfactorily. Also, there will be certain monetary or other losses

associated with making any decision. The economist is more optimistic and speaks of the gain in making a certain decision but this is, of course, an equivalent point of view since loss is negative gain. The loss depends on the decision which in turn depends on s so that $L[\rho, D(s)]$, the loss incurred in using the rule $D(s)$ if ρ is true, is a quantity which is subject to chance. For any particular rule the loss may be averaged over the $n+1$ possible values of s to obtain the expected loss, or risk, if $D(s)$ is used and ρ is the true reliability. With this formulation it now seems reasonable to choose that decision rule which minimizes the risk in some manner.

In order to explain one possible solution to this problem we pretend, as a simplifying assumption, that there are only two possible wrong decisions and only two different losses. More precisely we postulate that $D(s)$ can make only two decisions; the rule must decide either D_0, that reliability is adequate, or D_1, that it is inadequate. Further, the reliability will be called satisfactory or unsatisfactory depending on whether $\rho \geq 99\%$ or $\rho \leq 97\%$. The region $97\% < \rho < 99\%$ is a zone of indifference where we do not much care whether the reliability is pronounced adequate or inadequate. As a notational convenience we will sometimes refer to the decision rule D and write $D = D_0$, for example, when we mean that the rule makes the decision D_0. In addition we take the various possible losses to be as in Table 1.

TABLE 1

LOSSES FOR THE SIX POSSIBLE DECISION-TRUE RELIABILITY COMBINATIONS

true reliability \ decision	D_0: adequate	D_1: inadequate
adequate	0	L_1
indifferent	0	0
inadequate	L_0	0

Now write $P(D = D_1 | \rho)$ for the probability that the decision rule, D,

makes the decision D_1 when ρ is the true reliability and define $P(D = D_0 | \rho)$ similarly; then the risk equals $L_1 \cdot P(D = D_1 | \rho)$ or $L_0 \cdot P(D = D_0 | \rho)$ according as $\rho \geq 99\%$ or $\rho \leq 97\%$ and is zero when ρ is in the zone of indifference. Hence we may speak of there being, effectively, only two kinds of risk. The central idea of this method may now be stated, it is that we should choose a decision rule which constrains one kind of risk but minimizes the other. Or what amounts to the same thing, we demand that $P(D = D_1 | \rho)$ should not exceed a given small but positive number α throughout the region $\rho \geq 99\%$ and subject to this restriction we then seek to minimize $P(D = D_0 | \rho)$ in the region $\rho \leq 97\%$. This method of minimizing the risk is an interpretation of the Neyman-Pearson [19] hypothesis testing theory from Wald's [24] decision rule point of view.

A very important technical result, the Neyman-Pearson lemma [17], may be used to show that if α is .0138 then the practical procedure of the previous section accomplishes the Neyman-Pearson minimization. In more detail, let $D^*(s)$ consist of deciding that reliability is adequate when 2 or fewer out of 50 radios fail and otherwise deciding that reliability is inadequate. Then $P(D^* = D_1 | \rho)$ does not exceed .0138 throughout the region $\rho \geq 99\%$ and if D^1 is any other decision rule satisfying this requirement then

$$P(D^* = D_0 | \rho) \leq P(D^1 = D_0 | \rho) \text{ for all } \rho < 97\%.$$

Because of the way in which the Neyman-Pearson risk minimization is carried out, it seems that tests of hypotheses are subject to two different logical interpretations. First, of course, we may look at hypothesis testing as a method of choosing wisely between two possible gambles. This is the minimum risk point of view emphasized in this section. But secondly, since the number α is always taken to be small, the test of significance argument would seem to apply as well.

It can be fairly argued that the previous paragraph fails to fully exploit the power of the decision rule point of view. We may, for instance, be able to proceed as follows. Any decision of consequence will result in a course of action; here, the customers course of action might quite reasonably be to continue or cancel the manufacturer's contract according as

reliability is decided to be adequate or inadequate. The consequence of canceling the contract of a manufacturer who is producing radios of adequate reliability would almost certainly be the permanent loss of a conscientious supplier. On the other hand, if the contract is allowed to remain with a producer of goods having inadequate reliability then the customer will acquire a number of radios of poor quality. However, this error will presumably be rectified at a later date and the damage is only temporary compared with the permanent loss of a conscientious supplier. Accordingly we assign to L_1 and L_0 the somewhat arbitrary values of 10 and 1 respectively. The risk becomes $10\ P(D = D_1|\rho)$ for $\rho \geq 99\%$, $P(D = D_0|\rho)$ for $\rho \leq 97\%$ and 0 in the indifference region. It is very appealing to require that the two kinds of risk be equal in some sense and subject to this restriction minimize their common value. Because of technical difficulties we can't quite accomplish this objective but we can do something very similar; we can choose the minimax rule [24]. As ρ is varied, each decision rule will have a maximum risk; we choose that decision rule which minimizes the maximum risk. Again it is almost obvious that we may restrict outselves to decision rules which have a cut-off value, v, such that reliability will be declared inadequate if and only if $s < v$. Table 2 shows the way in which the binomial distribution may be used to compute the cut-off value yielding the minimax rule. Once again we will decide that reliability is adequate when and only when 2 or fewer out of 50 tested radios fail.

TABLE 2

DETERMINING THE MINIMAX CUT-OFF VALUE

v	Maximum risk for $\rho \geq 99\%$	Maximum risk for $\rho \leq 97\%$
50	3.95	.22
49	.89	.56
48	.14	.81
47	.02	.94
46	.00	.98

The minimax value of v depends strongly on the relative magnitudes of L_0 and L_1; if L_1 and L_0 equal 2 and 1 respectively then v would be 49 instead of 48.

V. CONFIDENCE BOUNDS. Again we consider the probabilistic experiment of observing whether a valve does or does not shut off the flow of a liquid at a given signal and we consider that a binomial sample is available. By a well known technique, see Mood [17, p. 233] we may obtain r_C, a 95 per cent lower confidence bound, for ρ. In the derivation of r_C we contemplate the consequence of the occurrence of an event which has probability .95; r_C is determined so that ρ is greater then r_C whenever this event occurs. But according to the frequency interpretation of probability such an event will occur in 95 per cent of a large number of samples; thus the statement "ρ is greater than r_C" should, in the long run, be correct 95 per cent of the time. This is the logical basis of the lower confidence bound. The figure 95% was, of course, chosen for definiteness and other confidence levels are possible.

At this writing the method of confidence bounds is one of the most important techniques for "assuring reliability" but, as Wilson [26] indicates, in perhaps a majority of the cases it is inapplicable. An indication of this is the large sample size required to give a useful bound. From column 2 of Table 4 we see that even with 95 per cent confidence, which is frequently thought to be inadequate, the confidence bound does not get to be interesting until the sample size is in the neighborhood of sixty. This requirement is frequently at odds with common sense.

For example no one would think it necessary to observe sixty consecutive failures of a frosted light bulb before stating that the bulb is defective. If the bulb does not burn for the first several trials then we reason that the filament is broken and we throw the bulb away. The reader may consider the above example to be unfair since the various trials are not independent and hence the reasoning used in deriving the lower confidence bound does not apply. But this is exactly the point. We are not attempting to criticise confidence bounds when they apply but to indicate that the knowledge of a physical theory can override the statistical considerations in a way which will cause current statistical methods to be inapplicable. In a certain sense the best way to assure the reliability of a device is to build it in conformity with a physical theory which says that it should work.

Following the line of reasoning of the preceding paragraph, it is tempting to classify situations as being either deterministic or empirical and to maintain that statistical methods should be appropriate to the latter but not the former. However, this is not typically the case; many scientific inference situations will have a deterministic as well as a statistical aspect. Even in the lightbulb example it is sometimes possible to "repair" a burned out bulb by striking it sharply with the eraser of a pencil. Thus we might want to try the bulb say three times before throwing it away rather than just once as a completely deterministic approach would indicate.

If we are to make a completely isolated and purely empirical statement about the reliability of a device then the sample sizes of Table 3 are probably in fact essential. However, for most problems the situation is quite different; the device is usually designed according to well tested scientific principles and other similar devices will frequently be available to help in appraising reliability. The difficulty is in formalizing the vague but real evidence given by an analogous but not identical situation. What the light-bulb discussion points out is that, even though they may be difficult to develop, there is a definite need for statistical methods which allow for the role of prior knowledge in reliability and other scientific inference.

TABLE 3

NINETY-FIVE PER CENT LOWER BOUNDS ON RELIABILITY
IF NO FAILURES ARE OBSERVED IN A SAMPLE OF SIZE n

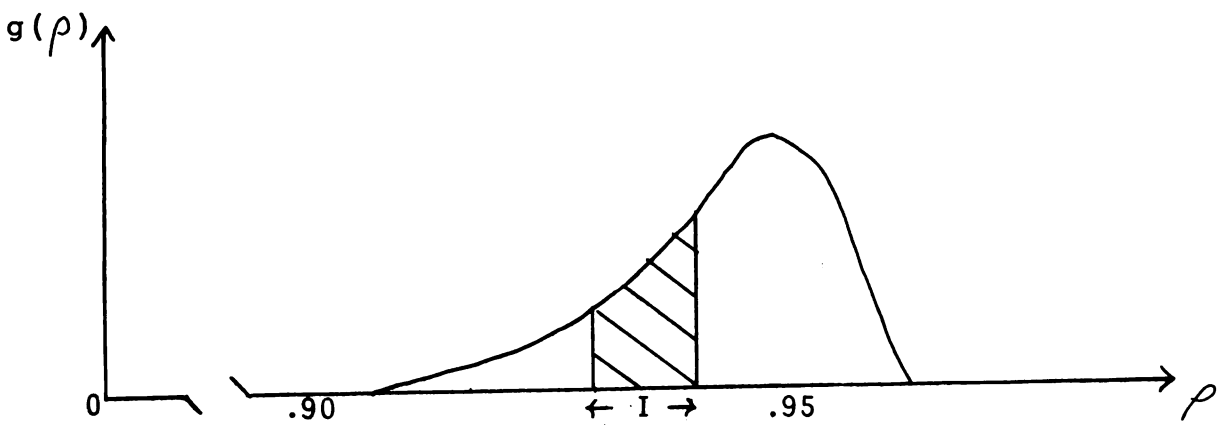
(1) n	(2) Confidence Bound	(3) Bayesian (prior beta distribution a=160, b=8.4)	(4) Stable Estimation
2	.22	.92	.14
5	.55	.92	.54
10	.74	.92	.74
20	.86	.93	.86
30	.90	.93	.90
60	.951	.940	.951
90	.967	.947	.967
200	.985	.963	.985
300	.990	.971	.990
1000	.997	.988	.997
3000	.9990	.9957	.9990

VI. SUBJECTIVE PROBABILITIES. The method of confidence bounds has been called the blind man's approach to reliability assurance in that no use is made of knowledge which may be available from previous results on similar components. As we previously pointed out, the confidence procedure frequently requires large sample sizes and yields uninteresting bounds. Where smaller sample sizes and tighter bounds are justified it is because the investigator has prior knowledge about analogous components. In order to discuss the form which this prior knowledge might take, we return to the example of determining ρ, the reliability of a valve.

From design considerations and a knowledge of the reliability of other similar valves the investigator may believe that ρ is very likely to be near .95 and that it is extremely unlikely that ρ will be smaller than .90. If a complete and explicit statement of this prior knowledge about ρ were possible then it might be representable as in Figure 2. $g(\rho)$ is a function defined for each ρ between 0 and 1 inclusive such that the area under $g(\rho)$ is 1.

FIGURE 2

A REPRESENTATION OF PRIOR KNOWLEDGE



The strength of the investigator's hunch that ρ is in the interval I is taken to be the area (shaded in Figure 2) under $g(\rho)$ and over I. This area is called the investigator's prior probability that ρ is in I. In order to reflect the investigator's prior knowledge the bulk of the area must be near .95 and almost all of it must be over the interval .90 to 1.

One function which is versatile enough to yield a curve with the general characteristics of Figure 2 is

$$g(\rho) = \text{const.}\ \rho^{a-1}(1-\rho)^{b-1},\ a>0,\ b>0.$$

For this particular choice of g, ρ is said to have the beta distribution. The average of the beta distribution is $a/(a+b)$; hence the prior knowledge tells us that it would be desirable to pick a and b to satisfy $a/(a+b) = .95$. In order to satisfy the second hunch we choose a and b so that the area under $g(\rho)$ over the interval 0 to .90 is .01. These two restrictions are enough to completely determine a and b; we find that $a = 160$ and $b = 8.4$. With this choice of a and b the beta distribution reflects the assumed prior knowledge that ρ is near .95 and almost certainly greater than .90.

The prior probability of the previous paragraphs is a special case of subjective probability. In the sense of Savage [21, 22], subjective probability if very much like what we have called axiomatic probability, except that it is completely determined by the betting odds which an idealized rational person would be willing to offer. "Though we are not quite like that person, we wish we were, ...", and Savage emphasizes this by referring to the idealized individual as "thou". Earlier we chose the parameters a and b of our prior beta distribution so that the area under $g(\rho)$ and over the interval 0 to .90 would be .01. From the subjective point of view this means that "thou wouldst barely" be willing to offer odds of 99:1 that $\rho > .90$ against $\rho \leq .90$.

In order to discuss the logical status of the multiplication rule for subjective probabilities write $P'(X) = P(X|C)$ for every event X which implies C. In accordance with the frequency and equally likely theories we should assign the probabilities, $P'(X)$, to the events X so that the multiplication rule will hold. But is it necessary to base the subjective assignment of conditional probabilities on the frequency or equally likely definitions? That the answer is in the negative is shown by an argument of Savage [21, section 3.5]. We give a more elementary intuitive argument due to Kemeny, et al. [11]. The universal event for the conditional probability experiment has been reduced to C but no new information about subsets of C is available. Thus, if X and Y are arbitrary subsets of C then thou wouldst want the betting odds of X against Y to be unchanged by the knowledge that C has occurred; i.e., $P'(X)/P'(Y) = P(X)/P(Y)$, or

$P(X)/P'(X) = P(Y)/P'(Y) = k$, a constant. Hence $P(X) = k\,P'(X)$ for all subsets, X, of C. To determine k we note that $P(C) = k \cdot P'(C) = k \cdot 1 = k$ and finally $P(X) = P(C)\,P'(X)$ for all X contained in C. Now let B be an arbitrary event not necessarily contained in C. The compound event (C and B) is, however, contained in C and hence

$$\begin{aligned} P(C \text{ and } B) &= P(C) \cdot P'(C \text{ and } B) \\ &= P(C) \cdot P(C \text{ and } B \,|\, C) \\ &= P(C) \cdot P(B \,|\, C), \end{aligned}$$

which is the multiplication rule.

VII. MINIMUM EXPECTED RISK. "When subjective probability is taken seriously decision, loss and other economic concepts, though they remain important, become relatively uninteresting because in principle the solution of every decision problem is simply to minimize expected risk with respect to the subjective probability that applies at the moment of making the decision." We illustrate this comment of Savage's [22] by considering a method of estimating reliability which uses prior knowledge. Thus, to continue the previous example, if n valves are tested and s of them are successes then how would this alter our hunch that the reliability of the valve is very near .95? Let r denote such an altered estimate of reliability. If we decide to adopt the beta distribution representation of prior knowledge then r should clearly depend only on a, b, s and n since these four quantities determine the prior knowledge and the additional experimental evidence concerning ρ. In the interest of simplifying the notation we suppress the dependence of r on a, b, n and write $r = r(s)$ to emphasize that, because of chance variation in s, r is also subject to chance.

We now attempt to find a rational basis for choosing among the many possible estimates of reliability. If ρ is the true reliability of the valve then there is a certain loss in acting as though ρ is equal to r(s); denote this loss by $L[\rho, r(s)]$. If r actually equals ρ then the loss would presumably be zero. For any particular estimator, L may be averaged over the n+1 possible values of s to obtain the average loss if r(s) is used and ρ is the true reliability. This average loss may be called the risk in using r(s) if ρ is true. The reader will not fail to notice that this is

another use of the decision rule idea which was first introduced in Section IV. Now, taking account of the prior probability distribution of ρ, the expected risk may be calculated to be a positive multiple of the sum of the $n+1$ terms

$$\binom{n}{i} \int_0^1 L(\rho, r(i))\ \rho^{a+i-1} (1-\rho)^{n+b-i-1} d\rho, \quad i = 0, 1, \ldots, n.$$

It would seem to be reasonable to use that estimator, $r(s)$, which minimizes the expected risk. We may find this estimator by separately minimizing each of the $n+1$ terms whose sum is the average risk. Taking the squared error loss function, $L(\rho, r) = (\rho - r)^2$, then the estimator having minimum average risk is $r(s) = (a+s)/(n+b+a)$. In particular, if all n valves function correctly then our altered estimate of ρ is $r(n) = (n+a)/(n+b+a)$. The squared error loss function is one of many choices which has the general properties desired; it is zero when r and ρ are equal and is an increasing function of the error of estimation. There is a modestly valid power-series reason for being interested in mean squared error [21, p. 233].

VIII. BAYES' THEOREM. In order to illustrate how the next concept of inference might work, we adopt the beta distribution representation of prior knowledge concerning reliability. Again consider that n valves are tested and that s of them work satisfactorily. In the light of the further evidence given by the newly tested valves what now should be the state of the observers mind? If B is the event that s out of n valves function correctly, then $P(B)$ is computed by averaging the probability of B for given reliability, ρ, with respect to the prior distribution of reliability. Thus

$$P(B) = \int_0^1 \binom{n}{s} \rho^s (1-\rho)^{n-s} g(\rho)\, d\rho .$$

Note that we have integrated over the entire range of ρ and hence $P(B)$ does not depend on ρ. Now define C to be the event that the reliability is in the interval $(\rho, \rho + \Delta\rho)$ and use the multiplication rule.

$$P(B) \cdot P(C \mid B) = P(C \text{ and } B) = P(B \mid C) \cdot P(C) .$$

Hence in the limit as $\Delta\rho$ approaches 0

$$P(C \mid B) = [1/P(B)] \binom{n}{s} \rho^s (1-\rho)^{n-s} g(\rho)\Delta\rho .$$

Remembering the meanings of the events C and B we may write the last equation as

$$h(\rho \mid s)\, d\rho = \text{constant} \quad \rho^s (1-\rho)^{n-s} g(\rho)\, d\rho ,$$

where the constant will be independent of ρ . $h(\rho \mid s)$ is called the conditional probability density function of ρ given s and the meaning of h is that if we know the value of s then we may graph h as a function of ρ ; the probabilities of various conditional events are then calculated by measuring areas under h. This is, of course, a special case of the more general result known as Bayes' theorem. Allowing $g(\rho)$ to be represented by the beta distribution we obtain as a special case

$$h(\rho \mid s) = \text{constant} \quad \rho^{a+s-1} (1-\rho)^{n+b-s-1} .$$

The observers subjective probability concerning the true reliability is again given by the beta distribution, but the parameters have been altered by further experimental evidence. This altered distribution is called the posterior distribution of reliability.

A very impressive fact concerning Bayesian inference in the present form, is that it fits nicely into the successive approximation scheme of scientific method. Thus in Figure 1, the initial idea is the prior beta distribution of Section VI. If a = 160 and b = 8.4 in this prior distribution and if the experimental data consist of observing ten additional successful valves then the data and the prior distribution combine to yield a posterior beta distribution

with parameters $a = 170$ and $b = 8.4$. In Figure 1, the posterior distribution of the first cycle becomes the new idea and hence the prior distribution of the second cycle. If fifteen further successes are observed, then the third "idea" of reliability is that it has a beta distribution with parameters $a = 185$ and $b = 8.4$.

The Bayesian concept which is analogous to the 95% lower confidence bound would seem to be to calculate that value of ρ, say r_B, such that 95% of the posterior distribution of ρ is greater than r_B. The practical matter of actually evaluating r_B is not quite as easy as it at first appears but when r_B is near one, as we would expect in the reliability situation, we may write

$$r_B = 1 - \chi^2 (.05)/(2a' + 2b' - 1)$$

where a' and b' are the parameters of the posterior distribution and $\chi^2(.05)$ is such that a chi-square deviate with $2b'$ degrees of freedom will exceed $\chi^2(.05)$ with probability 5%. This method of evaluating r_B is based on an approximation due to Tukey which, however, appears in a paper of Kimball and Leach [13]. It is interesting to compare the Bayesian bounds obtained in this way with the confidence bounds in the second column of Table 3; column three of this same table has been prepared to facilitate such a comparison. Note that if the sample size becomes much larger than about 60, the confidence bound is more strigent than the Bayesian bound; this is because the experimental evidence indicates a greater reliability than the prior opinion would warrent. Finally, the observer's posterior opinion of the exact value of reliability is obtained by integrating the posterior distribution over the interval (0, 1) to obtain $(a+s)/(n+b+a) = r(s)$, the minimum mean square error estimate of Section VII. $r(s)$ when derived by the method of the present section may be called the <u>average posterior estimate</u> of reliability to distinguish it from the numerically equal but conceptually different estimate of Section VII.

IX. <u>STABLE ESTIMATION</u>. The reader will not have failed to wonder whether it is often possible to formulate prior knowledge specifically enough

to describe it by means of a beta distribution or for that matter any specific prior distribution. In many instances prior knowledge will be a hazy, fuzzy, ill-defined thing which an individual feels but is unable to specify. At first consideration this seems to severely limit the cases in which Bayes' theorem can be applied; however, sharp peaks in the prior density function, g, correspond to relatively precise prior opinions and hence if prior knowledge is ill-defined then g must be broad and relatively flat over small ranges of ρ. Savage [22] points out that if the distribution is flat over narrow ranges it matters little how it is defined. For this reason, the method to be explained below is called the theory of stable estimation.

To illustrate for the type of problem under consideration remember that Bayes' theorem gave the result that $h(\rho|s)$ is proportional to $b(s|\rho)\, g(\rho)$ where h, b and g are respectively the posterior density of ρ given s, the binomial distribution and the prior density of ρ. Now, in those instances where the prior knowledge is ill-defined it seems that though we would not know g specifically, we would know some of its properties. It seems that sometimes g should be taken to be a wide flat curve with no marked peaks. In contrast, when considered as a function of ρ, b would usually be sharp and pointed in the neighborhood of s/n and would be quite small for values of ρ which are much removed from s/n. Under these circumstances $b(s|\rho)\, g(\rho)$ is well approximated by $b(s|\rho)\, g(s/n)$. Therefore h is a probability density which is well approximated by a constant multiple of $b(s|\rho)$ considered as a function of ρ, i.e., $h(\rho|s) =$ constant $\rho^{s}(1-\rho)^{n-s}$ to a good approximation. Again the posterior distribution is of beta type but this time the parameters are $s+1$ and $n-s+1$. In fact the results are the same as if we had assumed a prior beta distribution with parameters $a=b=1$ but surely the method of obtaining the result is more satisfactory.

Much of what was said in the previous section would now apply here as well. For example, $h(\rho|s)$ may be taken as the prior probability for the second stage of a successive approximation and we may calculate average posterior estimates and lower Bayesian bounds just as we did before. The results of several sample computations of such lower bounds appear in column four of Table 3. It is interesting to compare columns two and four of this table since the prior knowledge is assumed to be imprecisely known for both the confidence and the stable estimation bounds; the agreement is, in this instance, remarkable. In general for large s the agreement will be good since the stable estimation lower bound is the corresponding confidence bound with sample size and number of successes each increased by one;

see for example Mood [17, p. 235]. Stable estimation seems to describe the situation in which the experimenter finds himself in a large number of practical situations and once accepted the theory has far reaching implications.

We now mention an interesting related idea which may be referred to as Anscombe's prior distribution [1]. We are asked to consider a "unique trial for which there is no clearly relevant past experience" but where "every effort will have been made, in design and production, to prevent failures." What prior density function $g(\rho)$ would represent the initial opinion of an open-minded unprejudiced observer? Such an observer's initial opinion about ρ would presumably be diffuse and hence g would be continuous and would change only slowly with ρ, except possibly in the neighborhood of $\rho=1$. "There is little loss of generality in assuming that (g) is of beta type, . . ." In view of the assumed considerable effort to prevent failures it most unlikely that ρ will be really small. Hence in the prior beta distribution, a will be large relative to b and at least as large as 1. The behavior of $g(\rho)$ near $\rho=1$ will then, depend mainly on b. For $b < 1$, $g(\rho)$ becomes infinite at $\rho=1$ and to choose $b > 1$ implies a definite belief that ρ is bounded away from 1. If we choose b=1 then $g(\rho)$ will be a slowly increasing function for all ρ, so that over any short interval $g(\rho)$ is not far from uniform. For the experimental circumstances envisaged, it is accordingly proposed that the prior density function should be taken to be of beta type with parameters b=1 and a large relative to 1. Though Anscombe has in mind a particular kind of sequential acceptance procedure, his prior distribution could presumably also be used for the purposes explained in our section entitled Bayes' Theorem.

X. CONCLUSIONS. We have seen in the preceding sections that, starting with prior distributions, we can build up a fairly elaborate mathematical theory of inference. We are able to "get answers" and these answers are appealing in that they make use of prior knowledge. But does this mathematical theory describe the physical situation? Should we want to do the job for which this theory provides the foundation?

Some would answer these questions in the negative on the grounds that the reliability of a device is not subject to chance variation but is a fixed although unknown constant. Thus, they continue, it is foolish and misleading to assign a probability distribution to reliability. Further the whole idea of subjective and prior probabilities is incorrect since the probability of an event would differ from person to person and it would be

impossible to empirically check the correctness of a probability. Finally, the investigator will merely have some vague feelings about reliability which he will be unable to formulate in the precise manner indicated in Section VI and even if he were able to formulate his prior knowledge it wouldn't take the form of ascribing a beta distribution to ρ.

The last criticism has been anticipated in the discussion of stable estimation. There it was pointed out that under certain circumstances the prior distribution will be broad and if it is broad then it matters little how the prior distribution is exactly defined. In general this final objection will be met if Bayesian procedures can be found which do not depend too strongly on the precise nature of the prior distribution.

To discuss the first objection intelligently it is necessary to recall the distinction that we made in Section I between the device of common sense and the DEVICE whose reliability is capable of being studied. Remember that the DEVICE is an abstraction; it is the common sense device as seen by the observers measuring instruments and interpreted with respect to his preconceived ideas. Even if the reliability of a concrete device (which we haven't defined) is a fixed but unknown constant it is not completely clear that the same is true of the reliability of the abstract DEVICE. It might be that, in the transition, reliability assumes a distributional character due to the lack of accuracy of the process of observation and interpretation. We hasten to re-emphasize that the reliability of a common sense device cannot be studied empirically since as soon as it is formulated and observed it interacts with our senses and preconceived ideas and becomes a DEVICE.

As regards the second objection it does seem that a situation in which there is no agreement among individuals about basic scientific questions would come close to destroying the character of physical science. Thus prior distributions which are the results of whims or hunches may form a proper subject matter for psychology or sociology but in the physical sciences prior distributions should not be used unless they have some sound basis. Except where we are studying how the mind itself works we should not use prior distributions which are pulled out of thin air, but should insist that they be advanced as the consequence of some more or less valid argument. Such a valid argument could be a physical theory or a development along the lines of stable estimation for example. If a valid prior distribution is not available then the confidence bounds and tests of hypotheses of classical statistics can be used. Thus the second criticism of subjective probabilities leads to a clarification of the practical cases to which Bayesian inference can be applied rather than a refutation of the method.

It is to be expected that many objectivists would maintain that there are no practical situations where a valid prior distribution would be available while the subjectivists would tend to claim that their techniques are almost universally applicable. If the reader is interested in the details of this controversy he may consult the insightful and valuable works of Lindley [15], Savage [21], Edwards [7] and Tukey [23].

ACKNOWLEDGMENTS AND BIOGRAPHICAL NOTES. It is essential that we show how the ideas of this paper have been derived from and are related to the other literature. Where possible, this has been done by a direct reference in the text. However, in many places it has seemed appropriate to acknowledge help received but not to attribute a particular idea to a perhaps unwilling author. These latter, more indirect, references are collected here as biographical notes.

The papers of Costello [3], Cox [4], and Duffett [6] were responsible for arousing the author's interest in the foundations of reliability. Much of the motivational material in the present paper has been adapted from these three sources. The opinions of Section I on scientific method were primarily derived from the works of Crawford [5], Frank [10], Lenzen [14], Nagel [18], Polya [20], Wallis and Roberts [25], and Wilson [26]. The discussions of Feller [8], Kendall [12], Loève [16] and Nagel [18] were found to be very useful in the development of Section II. Sections VI, VIII, and particularly III were influenced by the book of Fisher [9]. This paper was markedly improved by the helpful criticisms of C. C. Braunschweiger and particularly L. J. Savage. Savage's constructive comments were nearly as extensive as the earlier manuscript.

BIBLIOGRAPHY

[1] F. J. Anscombe, "Testing to establish a high degree of safety or reliability", Bulletin de L'Institut International de Statistique, Brussels, 33^{e} session (1961).

[2] R. J. Buehler, "Confidence intervals for the product of binomial parameters", Jour. Am. Stat. Ass'n., vol. 52 (1957), pp. 482-493.

[3] D. L. Costello, "Myths, misconceptions and mean-time-between failures", Transactions of the Middle Atlantic Regional Meeting, Am. Soc. Qual. Control, Milwaukee, 1960.

[4] Paul C. Cox, "Some statistical problems related to missile safety"; Proceedings of the 6th Conference on the Design of Experiments in Army Research, Development and Testing; Aberdeen Proving Ground, Md., 1960.

[5] C. C. Crawford, Integration of the Fields of Knowledge, Texas Western College, unpublished notes, 1955.

[6] James R. Duffett, "Reliability", Proceedings of the 6th Conference on the Design of Experiments in Army Research, Development and Testing; Aberdeen Proving Ground, Md., 1960.

[7] Ward Edwards, "Behavioral Decision Theory", Annual Review of Psychology, Annual Reviews Inc., Palo Alto, California, vol. 12 (1961), pp. 473-498.

[8] William Feller, An Introduction to Probability Theory and Its Applications, Wiley, 1950.

[9] Sir Ronald A. Fisher, Statistical Methods and Scientific Inference, Hafner, 1956.

[10] Philipp Frank, Philosophy of Science, Prentice-Hall, 1957.

[11] Kemeny, Snell and Thompson, Introduction to Finite Mathematics, Prentice-Hall, 1956.

[12] M. G. Kendall, "Reconciliation of theories of probability", Biometrika, vol. 36 (1949).

[13] A. W. Kimball and E. Leach, "Approximate linearization of the incomplete β-function", Biometrika, vol. 46 (1959), pp. 214-218.

[14] Victor F. Lenzen, "Procedures of Empirical Science", International Encyclopedia of Unified Science, U. Chi. Press, 1938.

[15] D. V. Lindley, "Professor Hogben's 'Crisis' - a Survey of the Foundations of Statistics", Applied Statistics, vol. 7 (1958), pp. 186-198.

[16] Michel Loève, Probability Theory, D. Van Nostrant, 2nd Ed., 1960.

[17] Alexander McFarlane Mood, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1950.

[18] Ernest Nagel, "Principles of the theory of probability", International Encyclopedia of Unified Science, U. Chi. Press, 1939.

[19] J. Neyman, Lectures and Conferences on Mathematical Statistics and Probability, U. S. Department of Agriculture , Washington, 1952.

[20] G. Polya, How to Solve It, Princeton University Press, 1945.

[21] L. J. Savage, The Foundations of Statistics, Wiley, 1954.

[22] L. J. Savage, "Subjective probability and statistical practice", The Foundations of Statistical Inference: A symposium, Methuen and Co., London, To be published.

[23] John W. Tukey, "Conclusions vs. Decisions", Technometrics, vol. 2 (1960), pp. 423-433.

[24] Abraham Wald, Statistical Decision Functions, Wiley, 1950.

[25] W. Allen Wallis and Harry V. Roberts, Statistics a New Approach, Free Press, Glencoe, Ill., 1956.

[26] E. Bright Wilson, Jr., An Introduction to Scientific Research, McGraw-Hill, New York, 1952.

AN EXPERIMENT ON AIRCRAFT VULNERABILITY

Bruce Taylor
Garth McCormick
Research Analysis Corporation

The problems involved in the conduct, data reduction, and analysis of data from a large scale field experiment are not always fully recognized by designers and planners of such experiments. It is the intention of this paper to present the history of one such experiment including some of the results, and to draw from it some general conclusions about the practicality of conducting experiments on that scale.

The experiment was conducted during October and November 1958 at the Hunter Liggett Military Reservation, California. It was a US Army Combat Development Experimentation Center (USA CDEC) experiment in which ORO participated. Experimentation was conducted concurrently but independently by both agencies. Data was collected by ORO only during October 1958 and this paper is concerned with the ORO experimentation.

The problem was to determine the effectiveness of some light antiaircraft weapons and small arms against low-flying tactical aircraft in the forward battle area. At the time of the experiment many new aircraft systems were under study by different Army organizations. Because of the missile threat at medium and high altitudes most tactics envisioned low-flying aircraft. In the region below 600 feet only light arms presented a threat to aircraft in the forward areas. There was however very little known about response and detection capabilities of an antiaircraft crew under alert and non-alert conditions. Furthermore aiming error and hit probability data did not exist for effectiveness of small arms against aircraft flying below 600 feet. The experiment was designed to obtain these basic data from the field.

It was considered important to relate these data to measurable variables which were thought to be instrumental in determining weapon effectiveness (Fig. 1). The specific problem then to which the experimental results were addressed was the following. What is the effect of target velocity, altitude, crossing range, terrain mask angle and gun-crew alert status upon single-weapon vs. single-aircraft engagement kill probability and upon related quantities such as gun-crew detections ranges and number of rounds fired?

The experiment consisted of a series of single aircraft passes over sites containing antiaircraft units. Four sites were selected to give a variety

of terrain features (Fig. 2). For any aircraft pass only three sites were occupied. Each of these contained the group of weapons shown in Table 1.

Table 1

WEAPONS USED

1 Twin 40 AA
4 M-1 Rifles
2 BARS
2 50 Cal MG (mounted on APC's)
1 Quad 50 AAMG

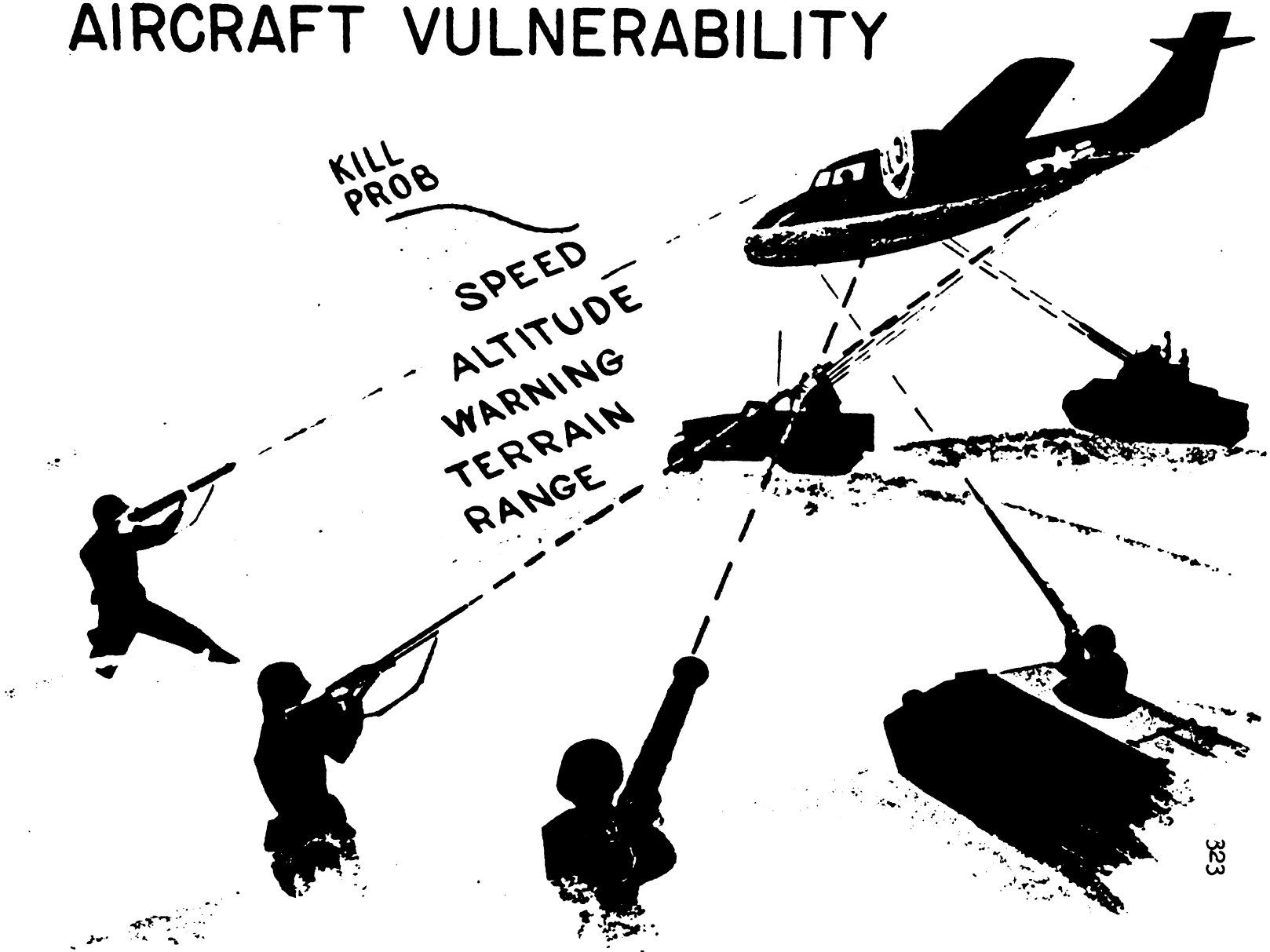
In addition a Redeye simulator was placed in each of two sites.

The targets for all weapons were the Army L19 observation aircraft flying at a velocity of 75 KT, the Air Force T37 jet trainer flying at 200 KT, and the Air Force F100 fighter flying at 325- and 450-KT. During the experiment the aircraft flew straight and level courses with respect to the ground weapons. It was planned that these targets be generalized to other aircraft of the same general size and velocity class by using the average vulnerability characteristics of these and related aircraft.

The operation of all weapons was simulated in the sense that camera film records were obtained from the experiment rather than actual target damage from real rounds fired. Two of the weapons were further simulated. The Redeye was a mock-up, or specially designed dummy to simulate to the Redeye operator some of the more important characteristics of the weapon. The Twin 40mm was also simulated in that the gun crew did not use its computing sight.

DESIGN OF EXPERIMENT. Each aircraft flight was carefully controlled so that with respect to each weapon site pre-specified values of the independent variables were taken on. A factorial design with two replications was originally intended. That is, every combination of the values of

AIRCRAFT VULNERABILITY



the independent variables shown in Table 2 would be taken on during two aircraft-site passes. Because of equipment and time limitations only one replication was possible, and after data reduction the factorial design was incomplete. Where possible these gaps were filled by standard statistical techniques.

Using a regression analysis it was planned to determine the functional relationship between the dependent and independent variables. The basic method described by Milne[1], is that of approximation by orthogonal polynomials. This requires that the independent variables take on values at equally spaced intervals; and this requirement was met by the selection of values as seen in Table 2.

Table 2

Velocity (In knots)	Altitude (In feet)	Mask (In degrees)	Crossing Range	Warning
75	0	No	Overhead	Surprised
200	200	$0-5^{o}$		
325	400	$5-10^{o}$	Crossing (300 meters)	Warned
450	600	$10-15^{o}$		

Having explained the intention and design of the experiment, it is still necessary to describe some of the details of instrumentation, troop training, and data reduction in order to give some notion of the limitations of method and hence applicability of the results of the experiment.

INSTRUMENTATION - SOURCES OF DATA

There were four major sources of data from the experiment, gun, cameras, radars, phototheodolites, and pen records.

Two types of gun cameras were used (Fig. 3). The first, used only for the Ml rifle, was the Robot Star 35mm still camera. This was mounted on

[1] Milne, Numerical Calculus, Sect. 71, Princeton University Press, Princeton, New Jersey.

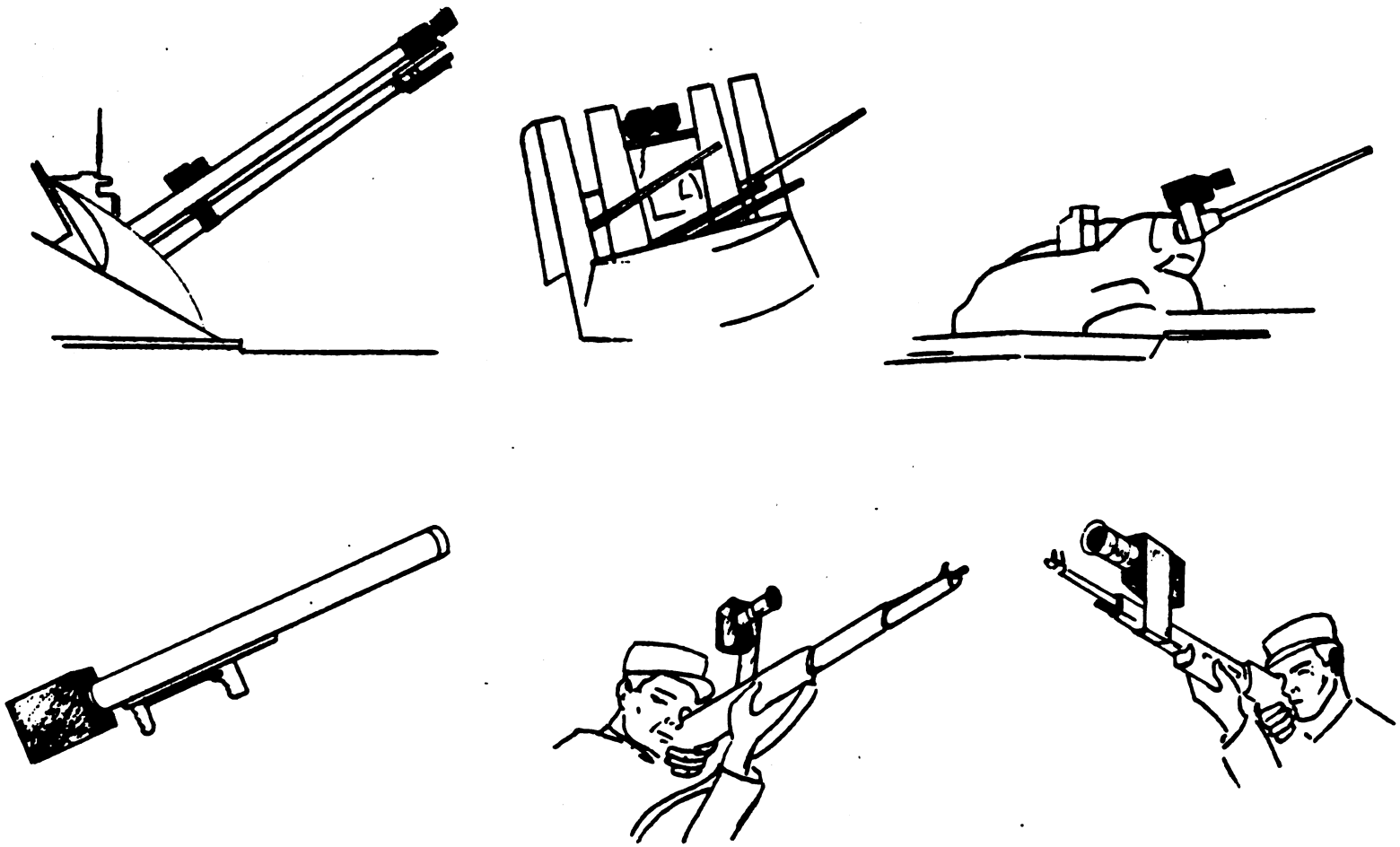


Fig. 3 - ARTIST'S OUTLINE SKETCHES OF CAMERA LOCATIONS ON WEAPONS USED

a. Twin 40 (M-42) - Two AN-N6 cameras, USACDEC mount to rear, ORO mount on muzzle.
b. Quad 50 (M-16) - Two AN-N6 cameras, both on mounts fastened to overhead bar.
c. APC 50 - One AN N6 camera on USACDEC mount near weapon base.
d. Rad Eye Simulator - AN N6 camera within rear housing.
e. M-1 Rifle - Robot Star camera on ORO mount located over weapons center of gravity.
f. BAR - AN N6 camera on USACDEC mount, mounted forward of weapon's center.

the rifle and instrumented via its trigger. When the trigger was pulled a picture would automatically be taken and the film advanced by one frame. All the other weapons were mounted with the AN-N6 movie camera. During tracking of the aircraft the camera took a movie of the aircraft while the trigger of the weapon was depressed.

Two M-33 type radars were used to track the aircraft and record its instanteous positions. The aircraft range, azimuth, and elevation were read directly on dials which were photographed every half second by cameras.

To insure adequate aircraft tracking data at all altitudes, four phototheodolites were stationed about the test site. Cameras recorded at intervals of very half second the azimuth and elevation angles of the aircraft as well as its relative position to the phototheodolite crosslines as seen through the optical sights.

The cameras, radars, and phototheodolites were synchronized by pen recorders. Twenty pens operated for each site during an engagement. Each of 14 of the pens was wired to a weapon and recorded the trigger movements of the weapon. One of the pens was activated by a button pressed by an ORO observer on the site who indicated when the aircraft came into view, when it was first observed by the troops, and when it disappeared from view. Two other pens were controlled by Twin 40 and Quad 50 observers who recorded by pushing a button when the first turret motion occurred, when first tracking started, and when tracking ended. In addition to measures of crew responses which could be gotten directly from the pen records, the pen records made it possible during a later phase of data reduction to assign real times to gun camera pictures and radar observations.

For one month prior to the experiment an extensive training program for the gunners was conducted. The men were given instructions on how to lead moving aircraft and how to estimate range properly.

A firing doctrine was established for each weapon and the gunners were trained in using their weapons in accordance with this doctrine. Effective ranges were set for the different weapons: The effective range for the Twin 40 was 1,500 yards, for the Quad 50 and Single 50 800 yards, for the M1 rifle and BAR - 400 yards, and for the Redeye 2,000 yards. The M1 riflemen were not to fire more than 8 rounds during one engagement. Redeye gunners were instructed to fire continuously as long as the aircraft was in range.

For all other weapons the gunners were instructed to fire in simulated bursts as long as the plane was within range of their weapons.

During and after the experiment the "raw data", that is, the pen records, gun camera films and pictures of radar dials, were read-out on to IBM cards. It was decided not to reduce the phototheodolite data to cards because of the prohibitive cost and because most of those data duplicated the radar data. The result of this reduction was approximately 200,000 cards. The data from these cards were then put on magnetic tapes.

The next stage in the data reduction consisted of a series of computer programs (using the 1103A Univac Scientific Machine) which eliminated a considerable amount of bad data (about 30% of the total) and by sorting and other techniques associated the pen record data with the gun camera and radar data. This provided a means for assigning a real time to each radar and gun camera observation.

The data were then in shape so that the model used in determining measures of weapon effectiveness could be employed.

The most important measure of weapon effectiveness is the probability that during an engagement, fire from the weapon will result in the "kill" of the aircraft. This measure is called "Engagement Kill Probability" (EKP). The bulk of computations were directed toward computing this quantity for each weapon-aircraft engagement.

The first step was to simulate the path of the bullet as indicated for each gun camera frame and to determine the closest approach of the bullet to the aircraft during its flight. The distance to the aircraft at the closest approach of the bullet was called the "miss distance". Necessary to this simulation was the capability of determining the bullet position at any point in time. This was accomplished by a series of curve fits to ballistic data found in firing tables. Orthogonal polynomial methods were used to do this.

The main results of each bullet-aircraft simulation were the co-ordinates of the bullet at point of closest approach in the miss distance plane. This plane is that which has its origin at the center of the aircraft and is perpendicular to the line from the gunner to the aircraft. For each weapon-aircraft engagement the bullet co-ordinates in the miss distance plane provided a dispersion pattern which was found to be approximately normally distributed in each direction. The center of the pattern was always found to be biased about the aircraft. The bias tended to be proportional to the velocity of the aircraft.

Aircraft vulnerability data were combined with the dispersion pattern in order to compute the probability that a single bullet would "kill" the aircraft. The vulnerability data were gotten from BRL publications. The vulnerable areas of typical aircraft were averaged and that area assumed to be one standardized vulnerable square centered at the aircraft. The single shot kill probability (or the probability that a single bullet would strike the vulnerable portion of the aircraft) was simply the double integral of the bivariate normal distribution function determined by the bullet dispersion patterns over the vulnerable square in the miss-distance plane.

The number of rounds fired by each weapon during an engagement was gotten from the known firing rates of the weapons and the experimentally measured time during which the gunners fired at the aircraft. The engagement kill probability then was simply:

$$\text{Ekp} = 1 - (1 - \text{Single Shot Kill Probability})^{(\text{No. of Rounds Fired})}$$

In addition to Engagement Kill Probability other measures of weapon effectiveness and crew response were computed. These were associated with the independent variables. As mentioned previously each engagement was designed to correspond to different values of the independent variables. However, after the experiment the radar data were examined to determine the actual altitude, velocity, and crossing range of the aircraft. This resulted in a reclassification of about 10% of the aircraft-site engagements with respect to the independent variables.

The original method of fitting orthogonal polynomials was not used because of the form in which it was apparent that some of the dependent variables were influenced by the independent variables. For example, the single shot kill probability varied as the inverse square of the velocity. An approximation to this statement using a polynomial whose variables contained positive powers would have been misleading and not very useful.

The relation of the independent to dependent variables where it was found to exist was gotten by straight-forward methods. Most of these results can be tabulated and presented but no explicit meaningful formulas given. Several examples of results will show that more clearly.

RESULTS. DETECTION RANGE. One useful measure of the ability of ground observers to visually detect aircraft is the slant range at which the aircraft is first observed. The variables of warning, crossing range, and velocity appeared to have no measurable influence on the detection range in this experiment. The two most important variables were terrain masking and target altitude.

Figures 4, 5, and 6 show the large difference in detection range cumulative probability as a result of the different mask angles. There is a gradual decrease in the difference between the 0-5 degree masks and no-mask cases as altitude increases. At 600 feet the difference disappears entirely.

Figure 7 shows the detection range probability when the data are averaged over all controlled variables except altitude. Each mask group contributes an unequal amount of data to each altitude and the effect of altitude is not as regular as it is when taken by mask case.

FIRING TIME. The firing time for a weapon is the time between start-fire and end-fire times for a given engagement. It was recorded automatically from a trigger pull switch. Firing time is useful as a rough experimental indication of the number of real rounds that could be fired during an engagement. As was expected firing time was an inverse function of felocity. The variables of mask and altitude were also important as shown in Figures 8, 9, and 10.

Figure 11 contains a plot of the bullet dispersion pattern means for the M1. Pre-experimental calculations had assumed that bullet dispersion patterns would always be centered at the target and that the effect of the different variables would be in the variance of the distributions. However, as indicated in this figure the centers are displaced roughly proportional to the velocity of the aircraft. This statement was true for those weapons which required the gunners to lead the aircraft. In the case of the Twin 40 and Redeye which aimed directly on the target, however, the dispersion means were very close to the center of the aircraft and showed little if any effect of velocity. Figure 12 demonstrates this for the Twin 40. These patterns include range effects. In an effort to mask out all but angular aiming errors a new quantity was introduced called "miss angle aiming error" and was computed as the ratio of the miss distance divided by the slant range to the aircraft. Figure 13 shows the mean of this quantity plotted against velocity for all the weapons. This demonstrates quite clearly

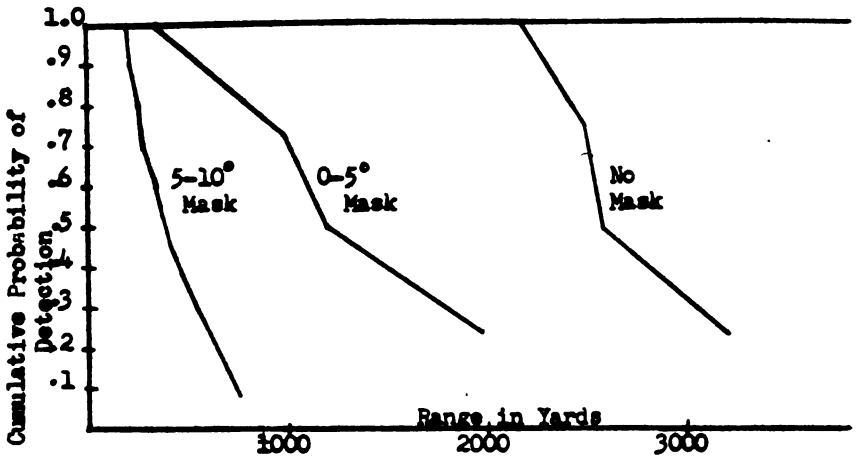


Fig. 4 Detection ranges at 200 feet altitude

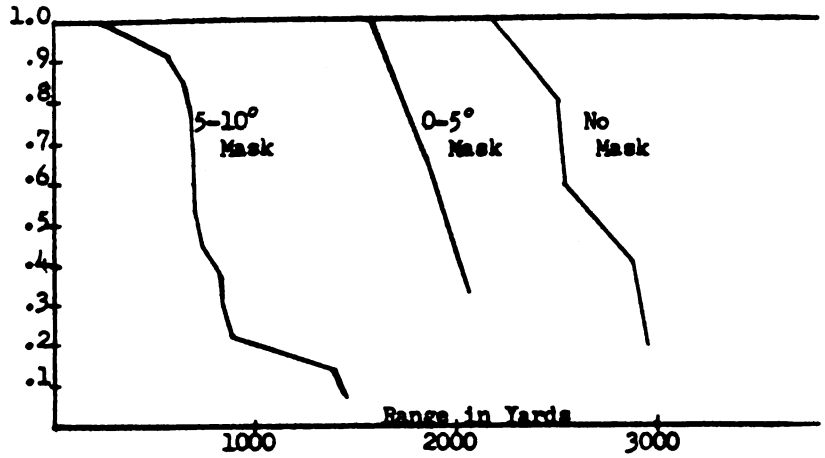


Fig. 5 Detection ranges at 400 feet altitude

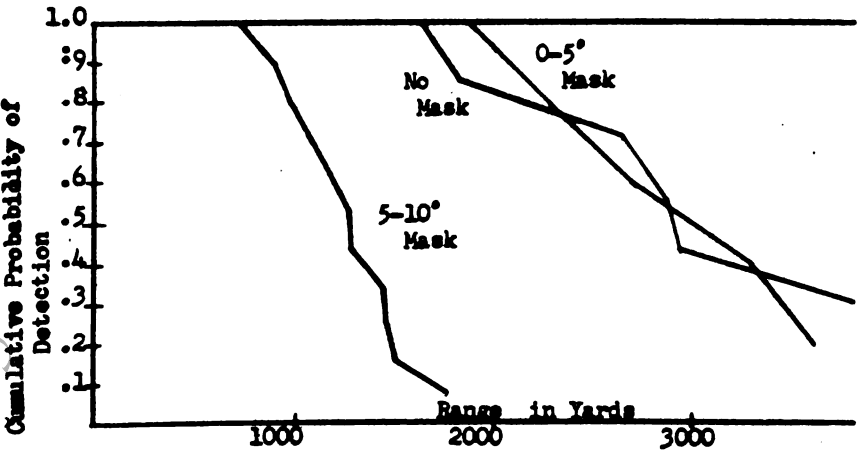


Fig. 6 Detection ranges at 600 feet altitude

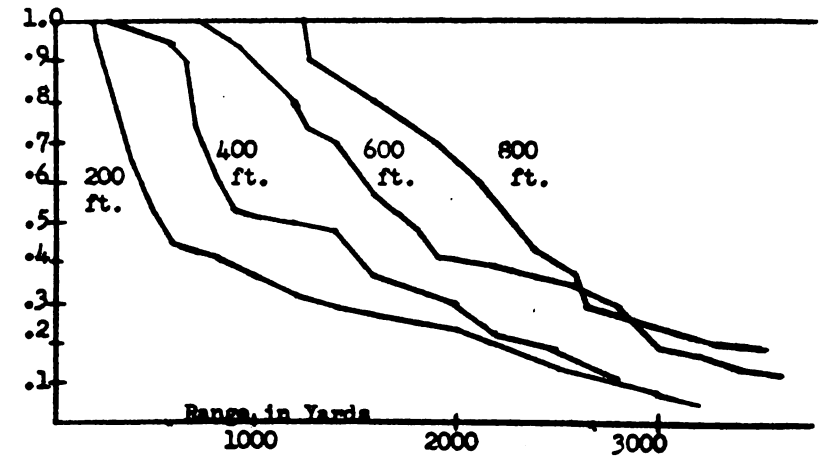


Fig. 7 Detection ranges for all altitudes

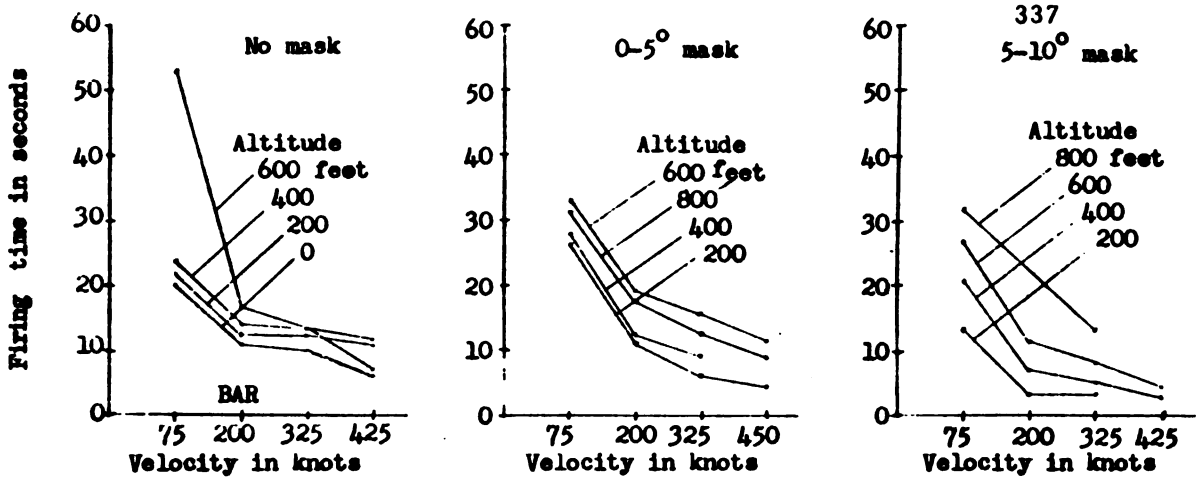


Fig. 8 BAR Firing Times, Averaged Over Warning and Crossing Range

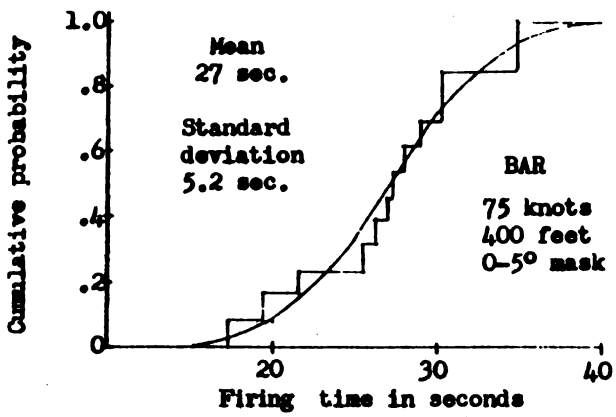


Fig. 9

BAR Firing Time Cumulative Distribution for all Data, at 75 kt, 400 ft, and 0-5° Mask. Smooth Curve is a Normal Distribution with Mean and Standard Deviation as Shown

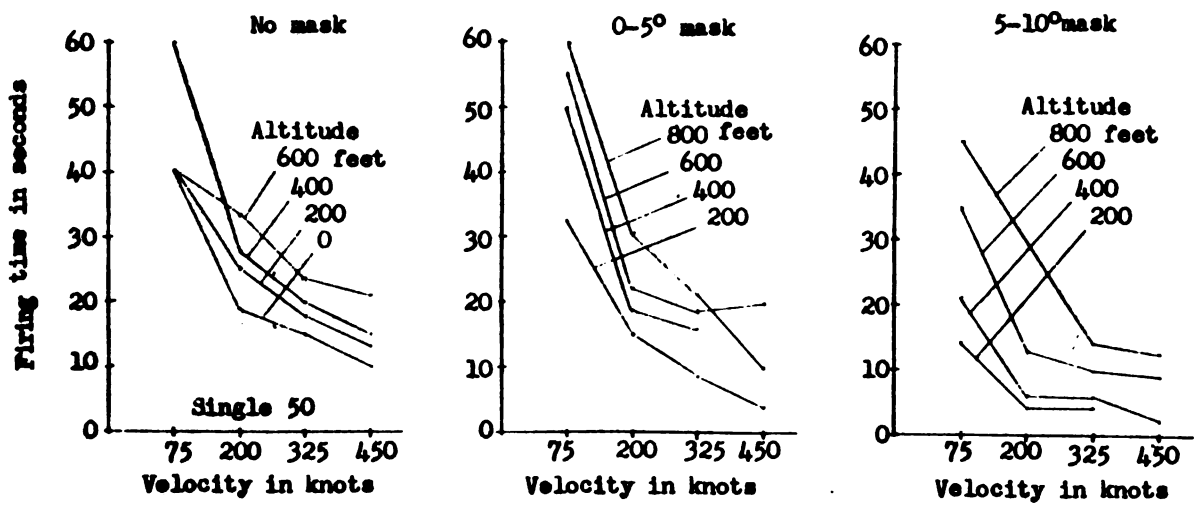


Fig. 10 Single 50 Cal. Machine Gun Firing Times, Averaged Over Warning and Crossing Range.

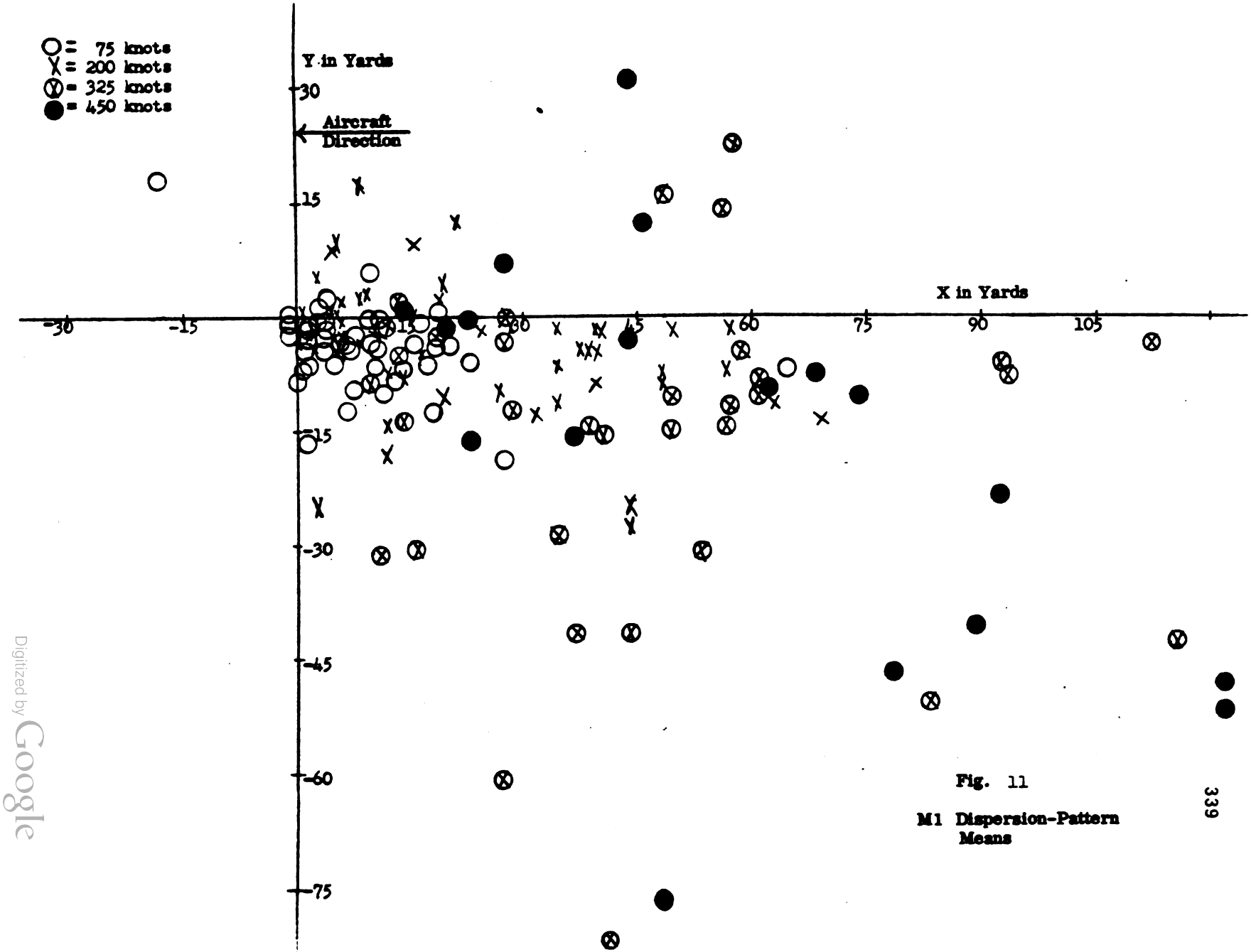


Fig. 11

M1 Dispersion-Pattern Means

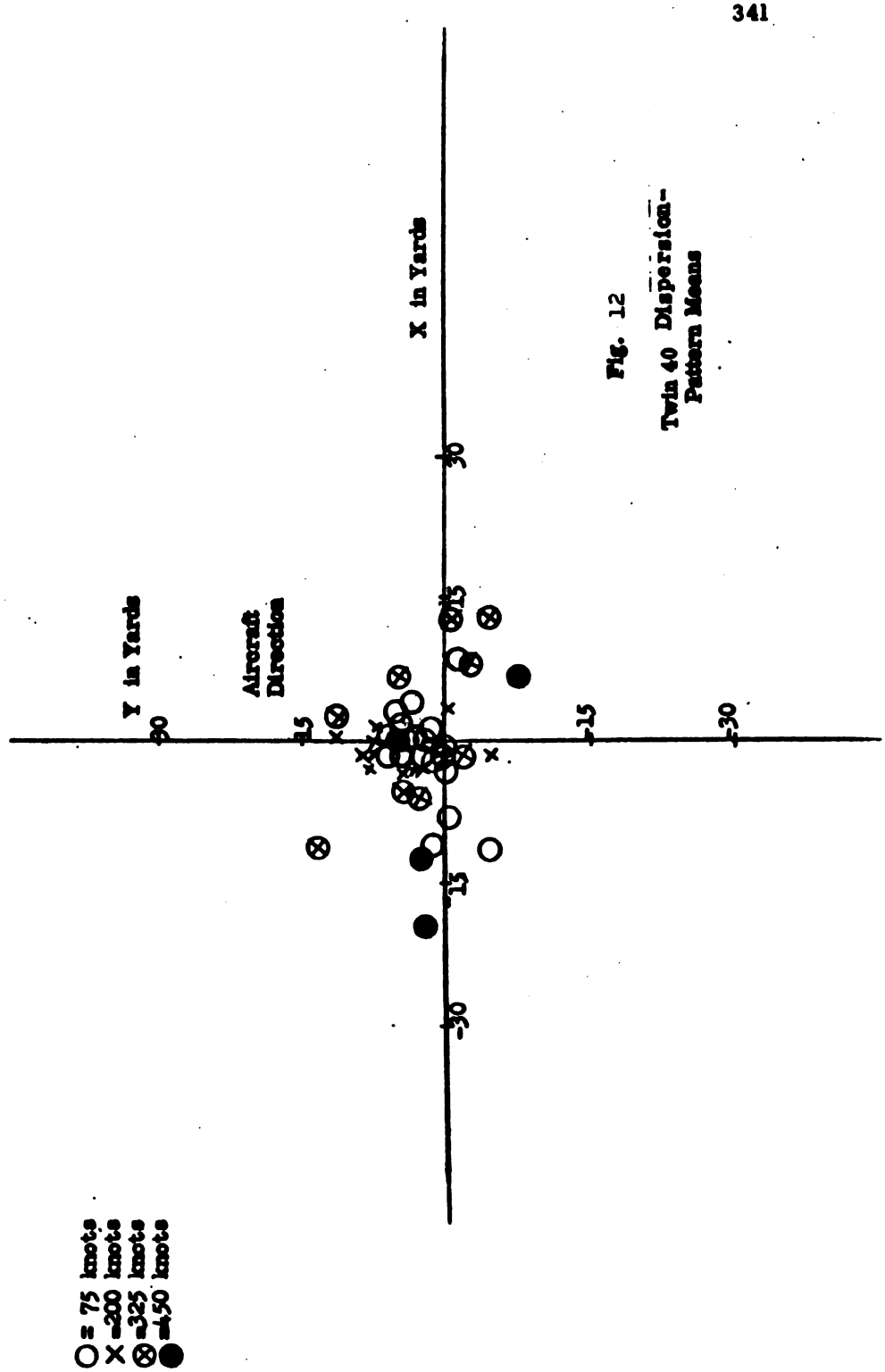


Fig. 12
Twin 40 Dispersion-Pattern Means

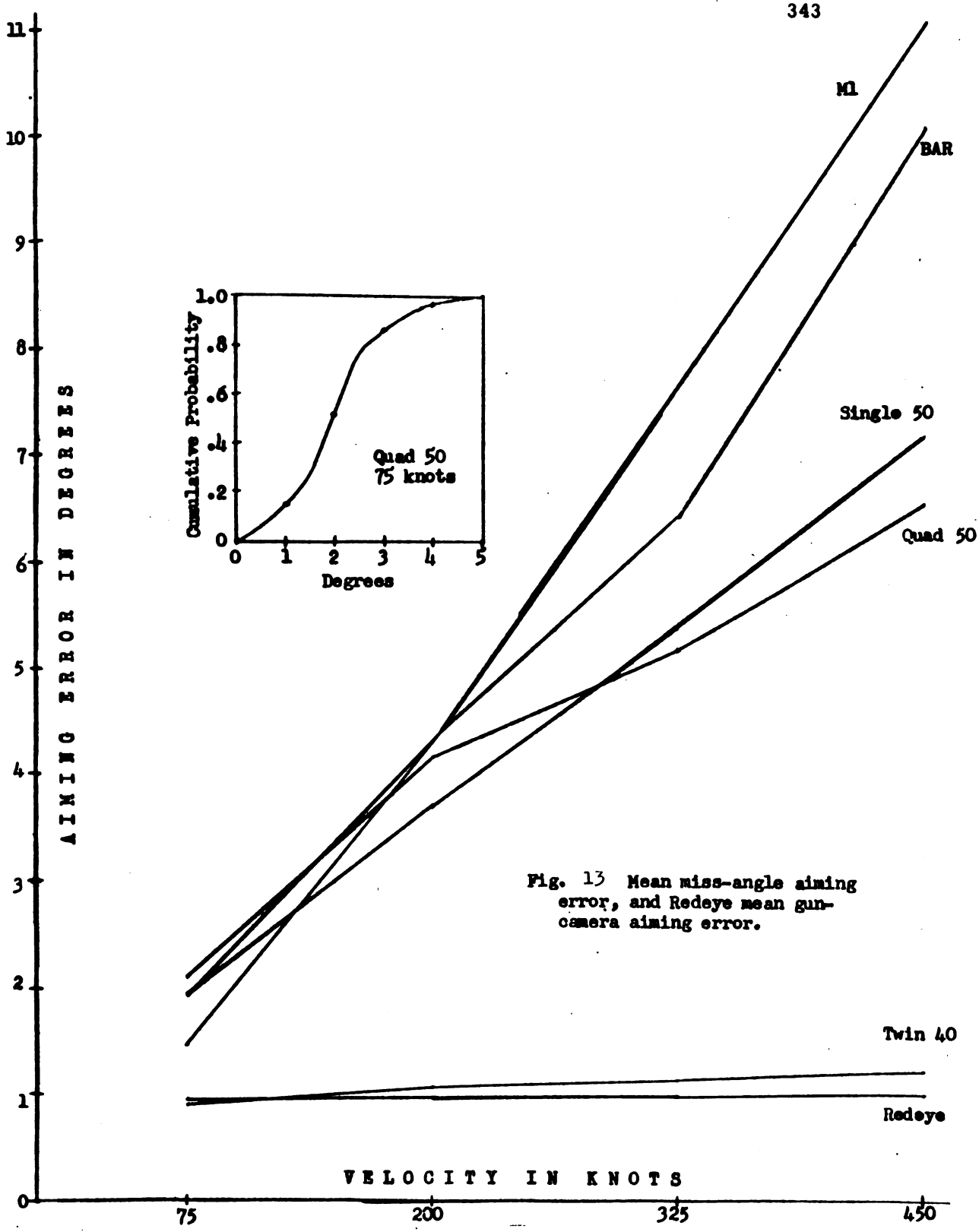


Fig. 13 Mean miss-angle aiming error, and Redeye mean gun-camera aiming error.

the advantage especially at high velocities that weapons not requiring lead have over those that do.

Single-shot kill probabilities seen in Figures 14 and 15 show similar sharp decreases as velocity changes from 75 to 200 knots. The single shot kill probabilities for the M1, BAR, Single 50 and Quad 50 seemed to vary as the inverse square of velocity. Other variables were important in some cases but no regularities were found.

Finally, engagement kill probabilities as shown in Figures 16 and 17 were most strongly influenced by velocity. The data for the cases at higher velocities were scarce but the general trend for the weapons which required some lead by the gunners was for engagement kill probability to vary as the inverse cube of velocity. The Twin 40 Ekp was approximately a slowly decreasing function of the inverse of the velocity although data at higher velocities is incomplete.

One particular assumption which determines the results is the size of the vulnerable area. Different areas other than the standard one were tried but the general shapes of the curves remained the same.

The flow of data from the field was such that at no single point in the flow was there raw data which was recognizable as such. The flow started with undeveloped film and pen records. The information on these media was not digitalized until a later time, and then only in very large amounts. The film had to be developed and then its information transferred to IBM cards and listed before any numerical data could be seen. There was an analogous set of steps for the pen records . . . The data handling was something like real-time data processing of the kind that is used during missile tracking and testing. It was quite different from the classical procedure of experimentation in which an observer records in numerical form what he sees on an instrument dial or number of objects he counts, the small amount of recorded data then being inserted in explicit mathematical formulae for computation of final results. Hence, the report on this experiment does not contain any raw data, nor a set of formulae which when applied by themselves to the raw data will yield the results contained in the report . . . A corollary of this view is that, like missile testing data handling, this experiment cannot efficiently yield results with just one trial. It requires more or less repeated runs in order to obtain an efficient operation. Since the ORO part of the aircraft vulnerability experiment was a one-of-a-kind event, data was lost, time was consumed, and some inefficiency resulted.

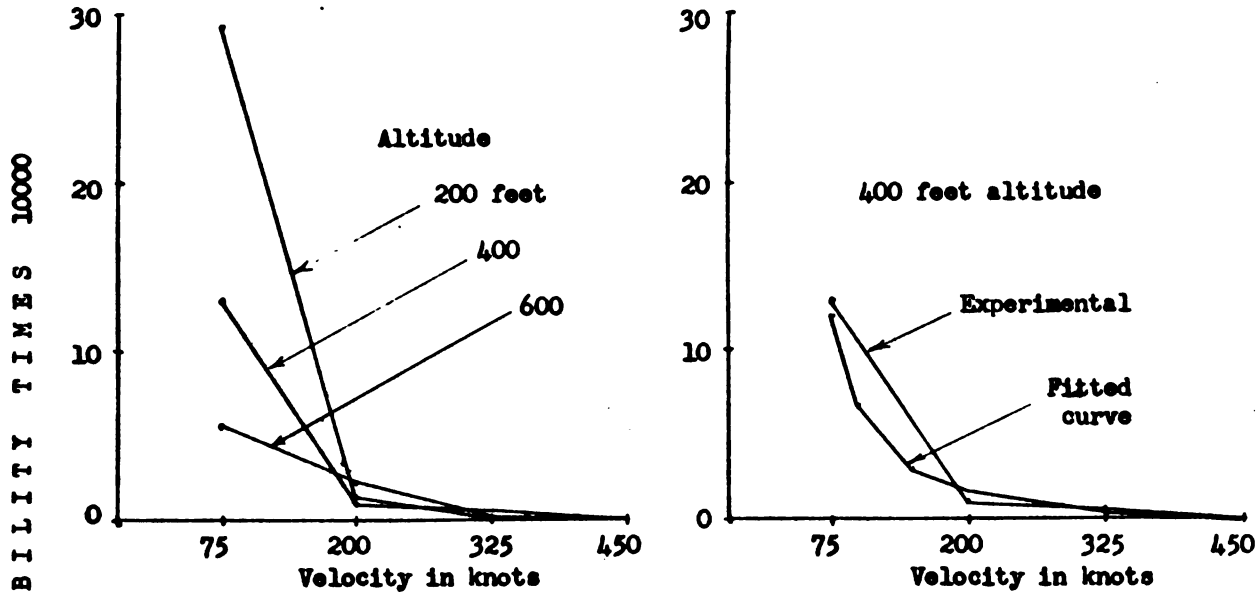


Fig. 14 Single 50 Single-Shot Kill Probability

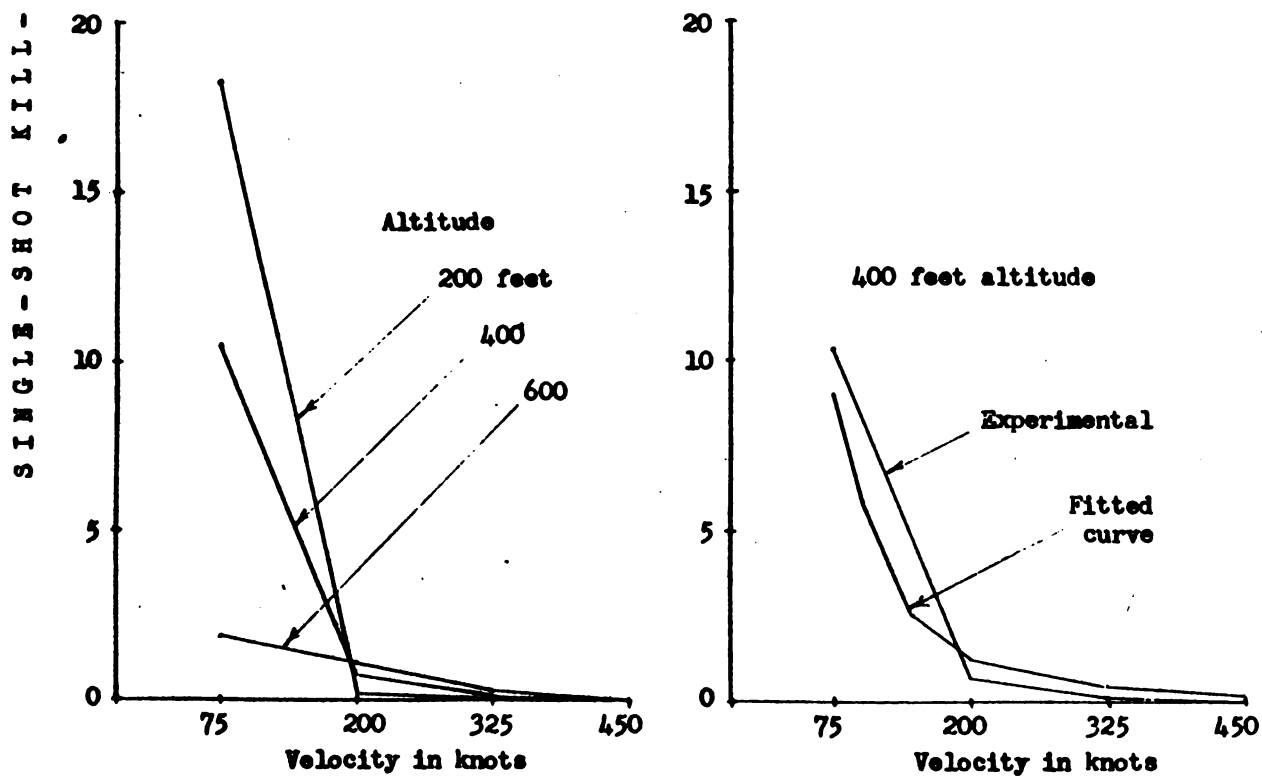


Fig. 15 Quad 50 Single-Shot Kill Probability

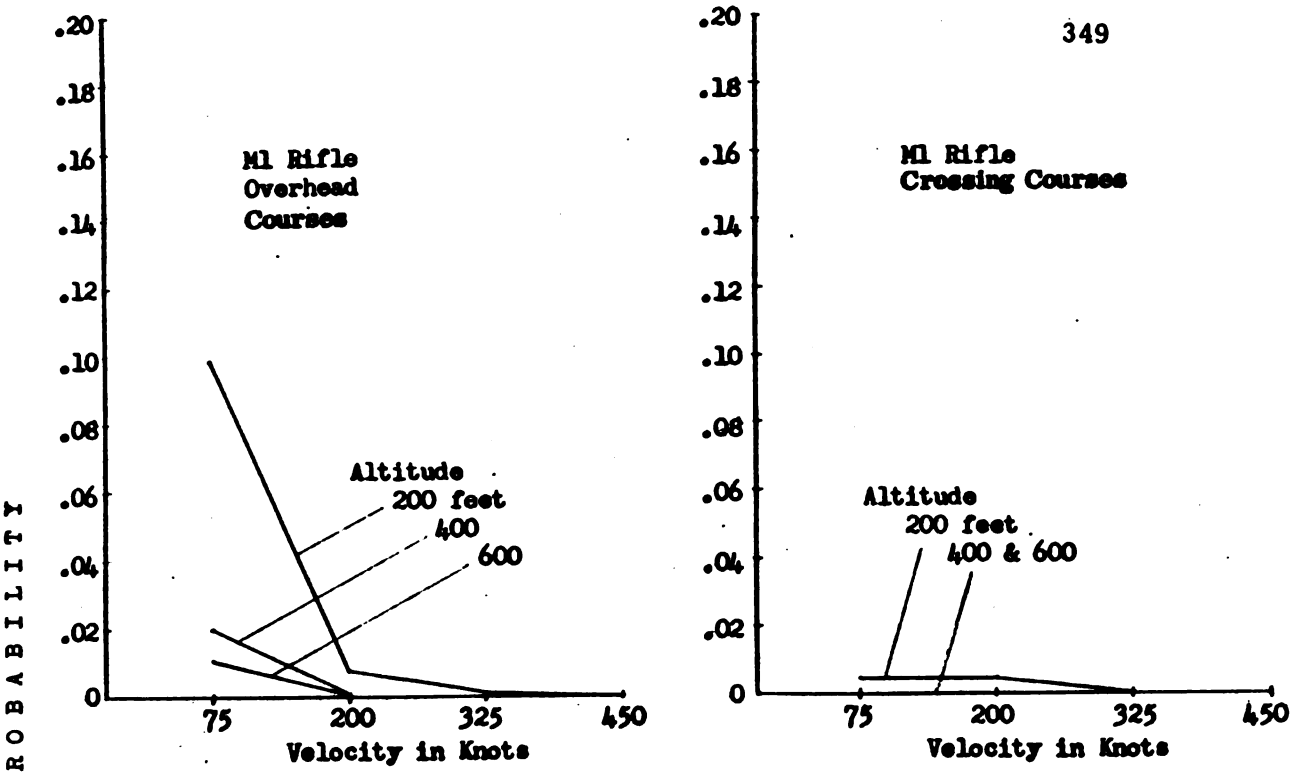


Fig. 16 M1 Engagement Kill probability for a Single Aircraft, Shown for Overhead and Crossing Courses

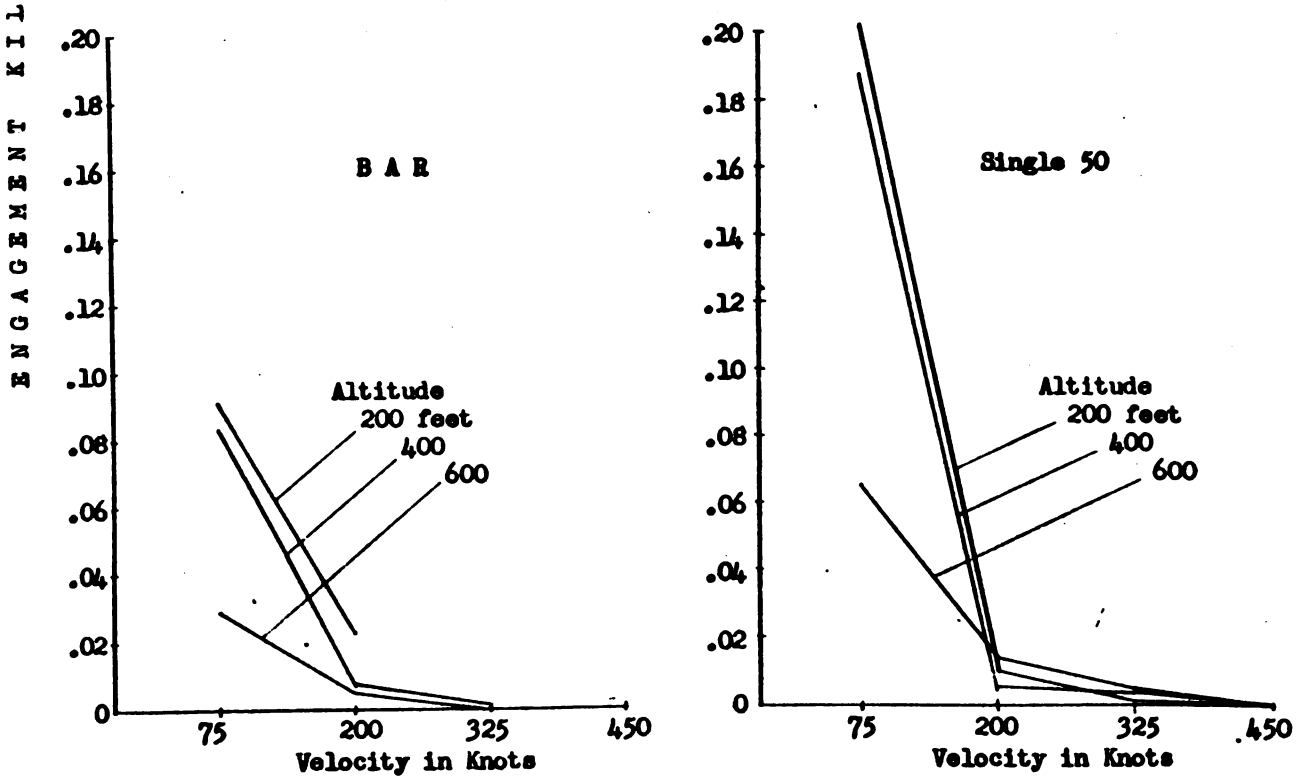


Fig. 17 BAR and Single 50 Engagement Kill Probability for a Single Aircraft

APPRAISAL OF THE EXPERIMENT. It has become apparent during the period of data transfer, reduction, and analysis that this experiment was entirely too large a scale for the results that were desired. The concept of divide and conquer is as applicable here as in war itself. This field test was really a set of 5-10 smaller experiments that could much better have been performed serially so as to yield a set of principles and a body of field testing experience at each stage to be fed into the next experiment. Thus separate experiments on detection ranges, response time, aiming errors, and individual weapons taken alone, would have been appropriate and would have made it clear and feasible that originally planned experimentation should be expanded or reduced in scope. By carrying out a very large scale test it was hoped to conserve resources, especially expensive aircraft operations. This result did occur but only at the price of a 2 to 3 year waiting period for results, loss of significant results from recognition of corrective measures long after the experiment was run, and a large turnover in the needed experimenters during the critical data transfer, reduction, and analysis phases.

Similar experiments in the future should be broken down into components and run separately, and/or incorporated into a continuing program of experimentation and data treatment.

A METHOD OF WEAPON SYSTEM ANALYSIS

H. G. Smith
Picatinny Arsenal

ABSTRACT. A method of determining optimum weapon system parameters using the optimization criterion of minimum weapon system weight is discussed.

Secondly, means of obtaining the probable error in range due to probable errors in ballistic coefficient, muzzle velocity, and angle of elevation is discussed.

Also, the possibility of determining optimum weapon system parameters using the optimization criterion of maximum ballistic accuracy is presented.

The purpose of this paper is to present a method of optimizing weapon system parameters for minimum system weight and a method of approximating the ballistic errors of projectiles. Also, the concept of designing weapon system parameters to achieve maximum ballistic accuracy is presented.

In many applications, a weapon system of minimum possible weight is highly desirable, especially where high maneuverability is required. Since the weapon system weight is approximately proportional to the kinetic energy of the projectile at the muzzle, the criterion of minimum kinetic energy is used in this analysis for simplicity.

In order to determine the optimum point of minimum kinetic energy required to obtain a given range, the following procedure may be used. First, one may select various values of ballistic coefficient. For each ballistic coefficient selected, one may find the muzzle velocity required to obtain the desired range from the appropriate ballistic tables or other suitable solution of the trajectory equations. One may then compute the product of the ballistic coefficient and the square of the muzzle velocity which is proportional to the kinetic energy as long as the shell diameter and form factor remain constant. By plotting a curve of this product, one can readily determine the optimum ballistic coefficient by finding the ballistic coefficient corresponding to the minimum point of the curve.

An alternate method of determining the optimum ballistic coefficient which involves slightly more calculations but may be easier to visualize may also be utilized. As before, one starts by selecting various values of ballistic coefficient and finding the corresponding muzzle velocities required to obtain the desired range. One then can compute the shell weight corresponding to each ballistic coefficient using the appropriate shell diameter and form factor. Using the shell weight and muzzle velocity, one may compute the kinetic energy and plot a curve of the kinetic energy versus the shell weight. The optimum shell weight can be determined from the curve at the point where the kinetic energy is a minimum. The optimum ballistic coefficient can be computed using the optimum shell weight and the appropriate shell diameter and form factor.

Once the optimum ballistic coefficient and corresponding muzzle velocity required to obtain the desired range are known and the projectile acceleration limit and piezometric efficiency are either known or estimated, the length of travel (length of gun tube) can be calculated from the following expression:

$$L = \frac{V^2}{2g} \cdot \frac{1}{(P.E.)\,a}$$

where: L - Length of travel

V - Muzzle velocity

g - Acceleration due to gravity

(P.E.)-Piezometric efficiency

a - Acceleration limit (G's)

This equation is easily derived by equating the work done on the projectile by the propellant gases throughout the travel of the shell in the gun tube to the kinetic energy of the projectile at the muzzle.

$$\text{Work Done} = \bar{P}\, A\, L$$

$$\text{Kinetic Energy} = \frac{1}{2}\frac{W}{g}V^2$$

where: $\bar{P}$ - Average pressure

A - Bore area

L - Length of travel

W - Projectile weight

g - Acceleration due to gravity

V - Muzzle velocity

$$\bar{P}\,A\,L = \frac{1}{2}\frac{W}{g}V^2$$

$$\frac{\bar{P}\,A}{W} \cdot L = \frac{V^2}{2g}$$

$$(P.E.) = \frac{\bar{P}}{P_p} \quad \text{(by definition)}$$

where: P_p - Peak pressure

(P.E.) - Piezometric efficiency

$\bar{P}$ - Average pressure

$$\bar{P} = (P.E.)\,P_p$$

$$(P.E.)\,\frac{P_p A}{W} \cdot L = \frac{V^2}{2g}$$

$$\text{However: } \frac{P_p A}{W} = \text{Peak acceleration in G's} = a$$

$$(P.E.)\, a\, L = \frac{V^2}{2g}$$

Therefore, the length of travel is:

$$L = \frac{V^2}{2g} \cdot \frac{1}{(P.E.)\, a}$$

As an aid to the design engineer, system analysis charts for the projectile types most encountered can be compiled. A separate chart is required for each range of interest. A typical chart would include curves of kinetic energy required to obtain the desired range versus shell weight for several shell diameters with constant form factor, curves of ballistic coefficient versus shell weight for several shell diameters with constant form factor, a curve of muzzle velocity required to obtain the desired range versus ballistic coefficient, and curves of length of travel versus muzzle velocity for several acceleration limits with constant piezometric efficiency.

Figures 1 through 4 (to be found at the end of this article) are typical examples of these charts. The charts were compiled for Type I Projectiles for ranges of 5,000, 10,000, 20,000, and 30,000 meters using a form factor of .5 and a piezometric efficiency of .5. Kinetic energy and ballistic coefficient curves are shown for shell diameters of 4, 6, 8, and 10 inches and length of travel curves are shown for acceleration limits of 2,000, 4,000, 6,000, 8,000, and 10,000 G's. The charts can be used for form factors other than those used in the compilation of the charts by multiplying the shell weight and kinetic energy obtained from the chart by the ratio of the form factor desired to the form factor used in the chart. Similarly, the charts can be used for piezometric efficiencies other than those used in the compilation of the charts by multiplying the length of travel obtained from the chart by the ratio of the piezometric efficiency used in the chart of the piezometric efficiency desired.

These charts can be very useful to the design engineer, not only in the optimization of a weapon system, but also in visualizing the relationships between the parameters required to obtain the desired range.

In the design of any projectile, the design engineer is concerned with the attainment of acceptable accuracy with preliminary designs. The accuracy of a newly designed projectile should be estimated in the preliminary stages of design in order to avoid the situation of discovering that the accuracy of a new design is unacceptable after a great deal of time and effort has been spent in the design of the projectile.

In order to approximate the accuracy that might be expected from the projectile, the design engineer must decide on the variations that might be expected in the parameters that determine the projectile's trajectory and determine the effect of each of the parameter variations on the range dispersion before the total range dispersion can be computed. The effects of the major parameter variations on the range can be obtained from the appropriate exterior ballistic tables by selecting a convenient change in a parameter, finding the corresponding change in range due to the parameter change, and dividing the change in range obtained by the change in the parameter to find the change in range due to a unit change in parameter. This procedure should be followed for each parameter at each range of interest.

As an aid to the design engineer, these differential effects can be computed from ballistic tables and plotted so that the design engineer can pick the differential effects off a graph and thereby simplify the task of estimating the ballistic errors.

The following factors have been found to be convenient in describing the major differential effects. They are the range-velocity sensitivity factor (the percent change in range due to a one percent change in muzzle velocity at at constant ballistic coefficient and angle of elevation), the range-ballistic coefficient sensitivity factor (the percent change in range due to a one percent change in ballistic coefficient at constant muzzle velocity and angle of elevation), and the range-angle of elevation sensitivity factor (the percent change in range due to a one percent change in angle of elevation at constant muzzle velocity and ballistic coefficient). The major parameter variations considered, for convenience since they are independent of each other, are velocity variation at constant weight, weight variation, form factor or drag coefficient variation, and angle of elevation variation. When the parameter variations and the sensitivity factors are known, the probable error in range can be determined from the following equation.

$$P_r^2 = (S_v P_v)^2 + (S_c P_w + nS_v P_w)^2 + (S_c P_i)^2 + (S_\phi P_\phi)^2$$

where: S_v - Range-velocity sensitivity factor

S_c - Range-ballistic coefficient sensitivy factor

S_ϕ - Range-angle of elevation sensitivity factor

P_r - Probable error in range in percent

P_v - Probable error in muzzle velocity in percent (not including any error due to variation of projectile weight)

P_w - Probable error in projectile weight in percent

P_i - Probable error in form factor (or drag coefficient) in percent

P_ϕ - Probable error in angle of elevation in percent

n - Logarithmic rate of change of muzzle velocity with projectile weight.

The Ordnance Engineering Design Handbook, ORDP 20-140, TRAJECTORIES, DIFFERENTIAL EFFECTS, AND DATA FOR PROJECTILES, gives the following approximate values for n.

Rifled gun with multiperforated propellant grains - $n = -.3$

Rifled gun with single-perforated propellant grains - $n = -.4$

Smooth bore mortar with flake propellant - $n = -.47$

Recoilless rifle with multiperforated propellant grains - $n = -.65$

Figures 5 and 6 show typical plots of the three sensitivity factors for Type I Projectiles. Figure 5 consists of curves of range-velocity sensitivity factors versus muzzle velocity for several values of ballistic coefficient at an angle of elevation 30^{o} and curves of range-ballistic coefficient sensitivity factor versus ballistic coefficient for several values of muzzle velocity at an angle of elevation of 30^{o}. Figure 6 consists of curves of range-angle of elevation sensitivity factor versus angle of elevation for various values of ballistic coefficient at a muzzle velocity of 440 meters per second.

Since the curves are functions of the ballistic coefficient, the curves may be used for any form factor as long as the ballistic coefficient used is computed using the applicable form factor. It should be noted, however, that the curves may only be used for projectiles which have drag functions similar to the drag function of the projectile type for which the curves were compiled.

These curves can be useful to the design engineer, not only in estimating the range dispersion, but also in visualizing the parameter changes that may be necessary to improve the accuracy.

Once the various parameter variations are known or assumed, it is possible to optimize the parameters using an optimization criterion of minimum ballistic error or, in other words, maximum ballistic accuracy. This may be done by selecting several values of ballistic coefficient, finding the corresponding muzzle velocities required to obtain the desired range at the desired angle of elevation, computing the probable error in range using the various parameter probable errors and the appropriate sensitivity factors, and plotting a curve of the range probable error versus ballistic coefficient. The optimum ballistic coefficient can be determined from the curve at the point where the range probable error is a minimum.

This procedure is particularly useful in the design of projectiles for existing weapons or any application where the primary design consideration is accuracy.

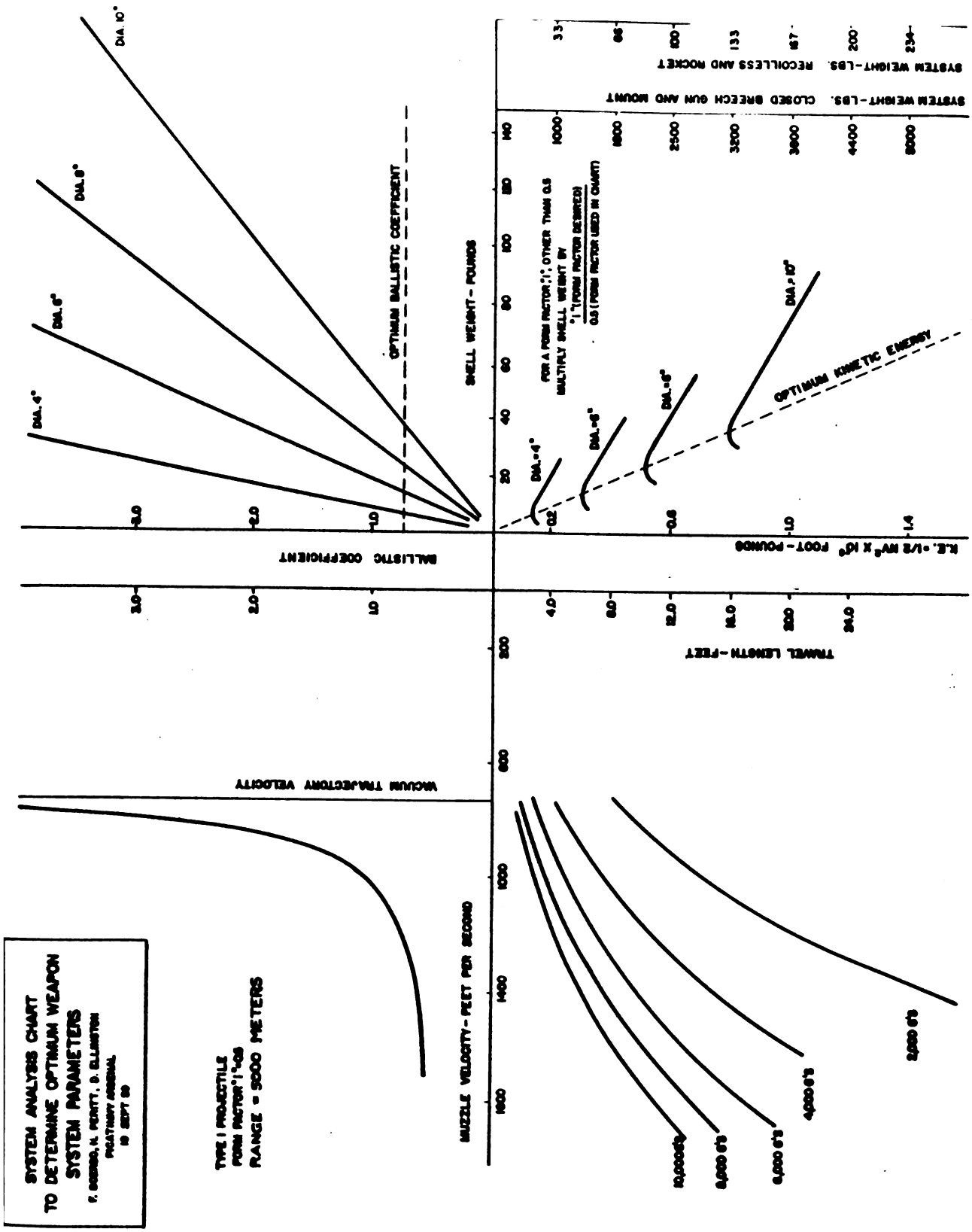


FIGURE 1.

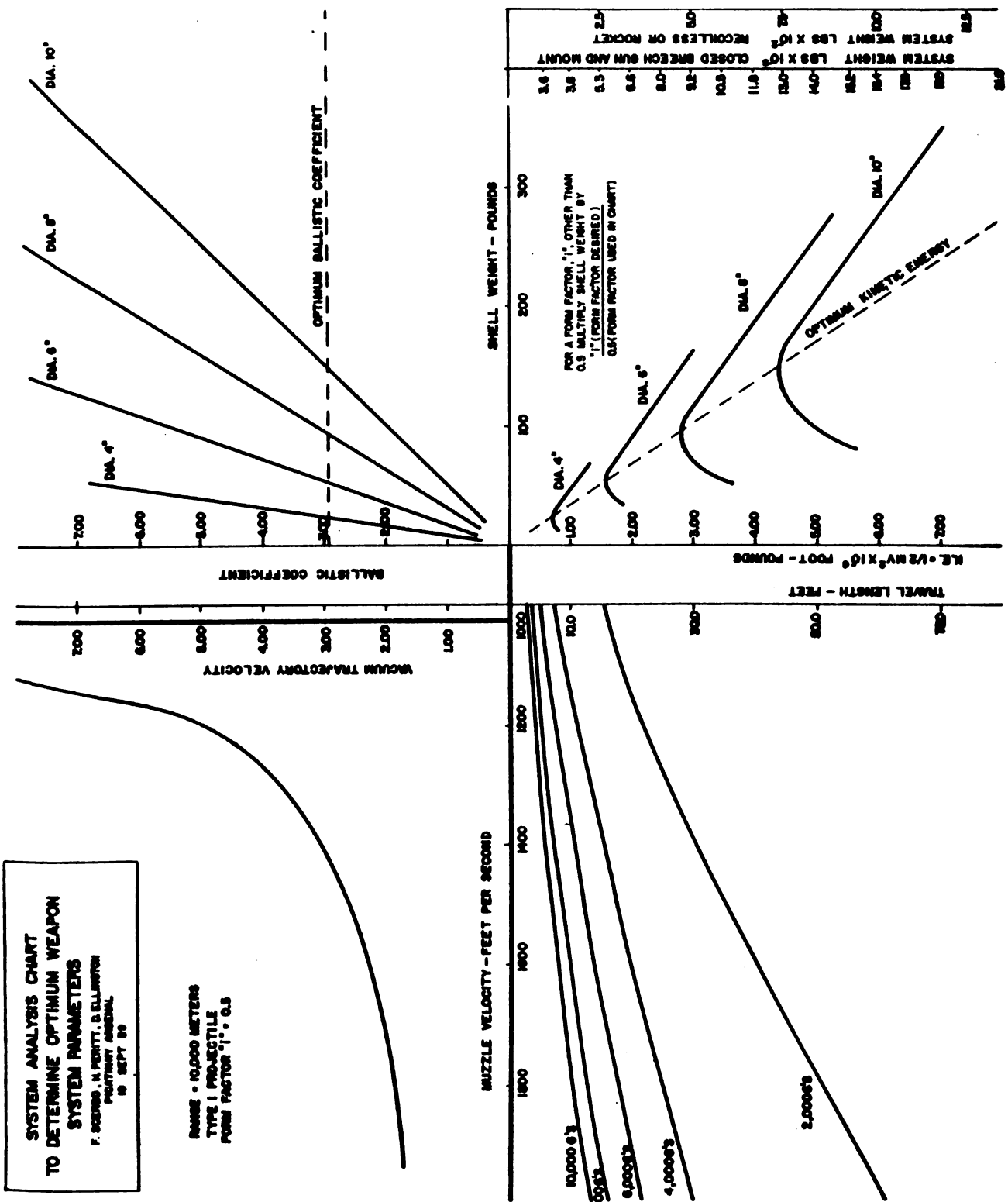


FIGURE 2.

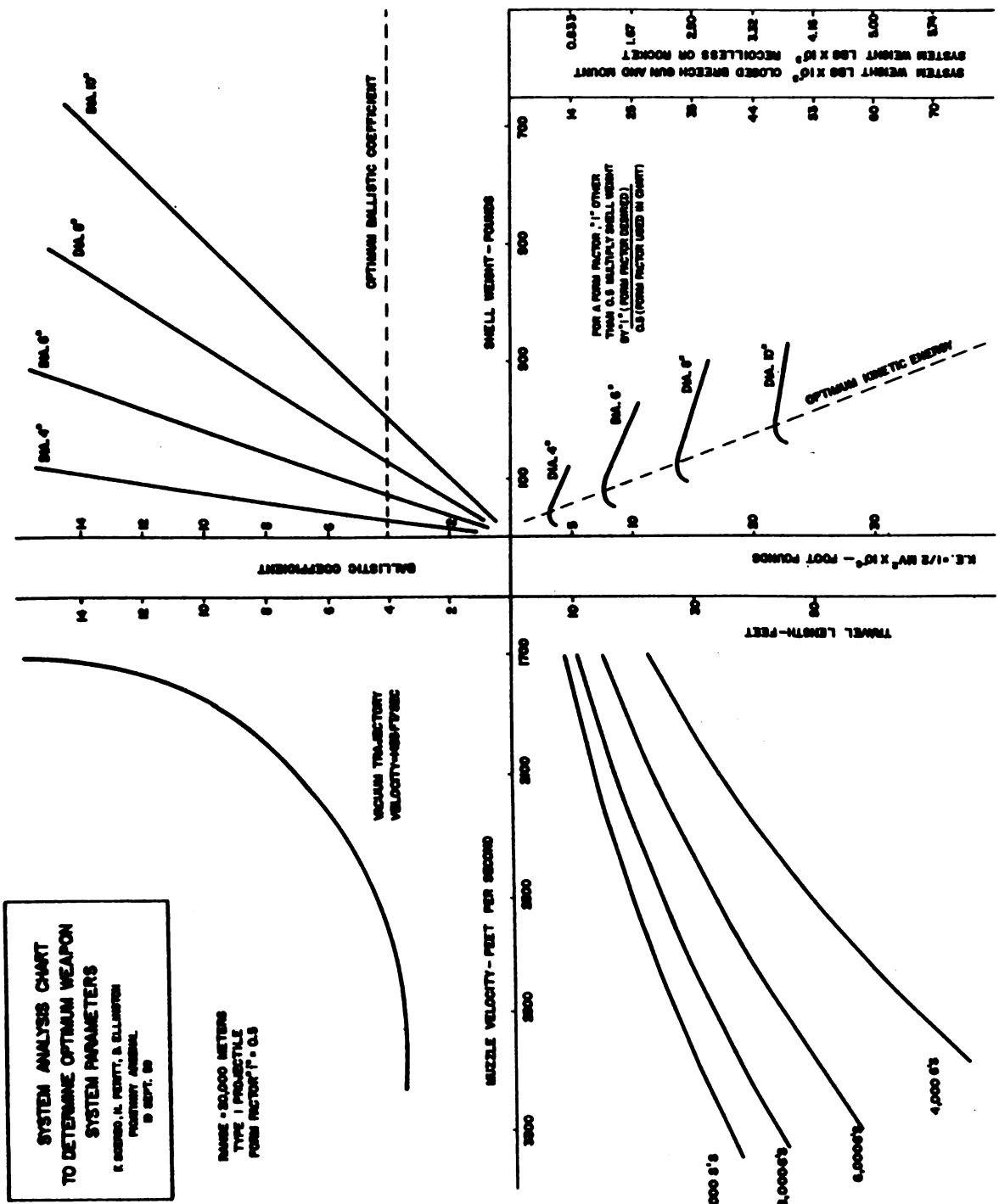


FIGURE 3.

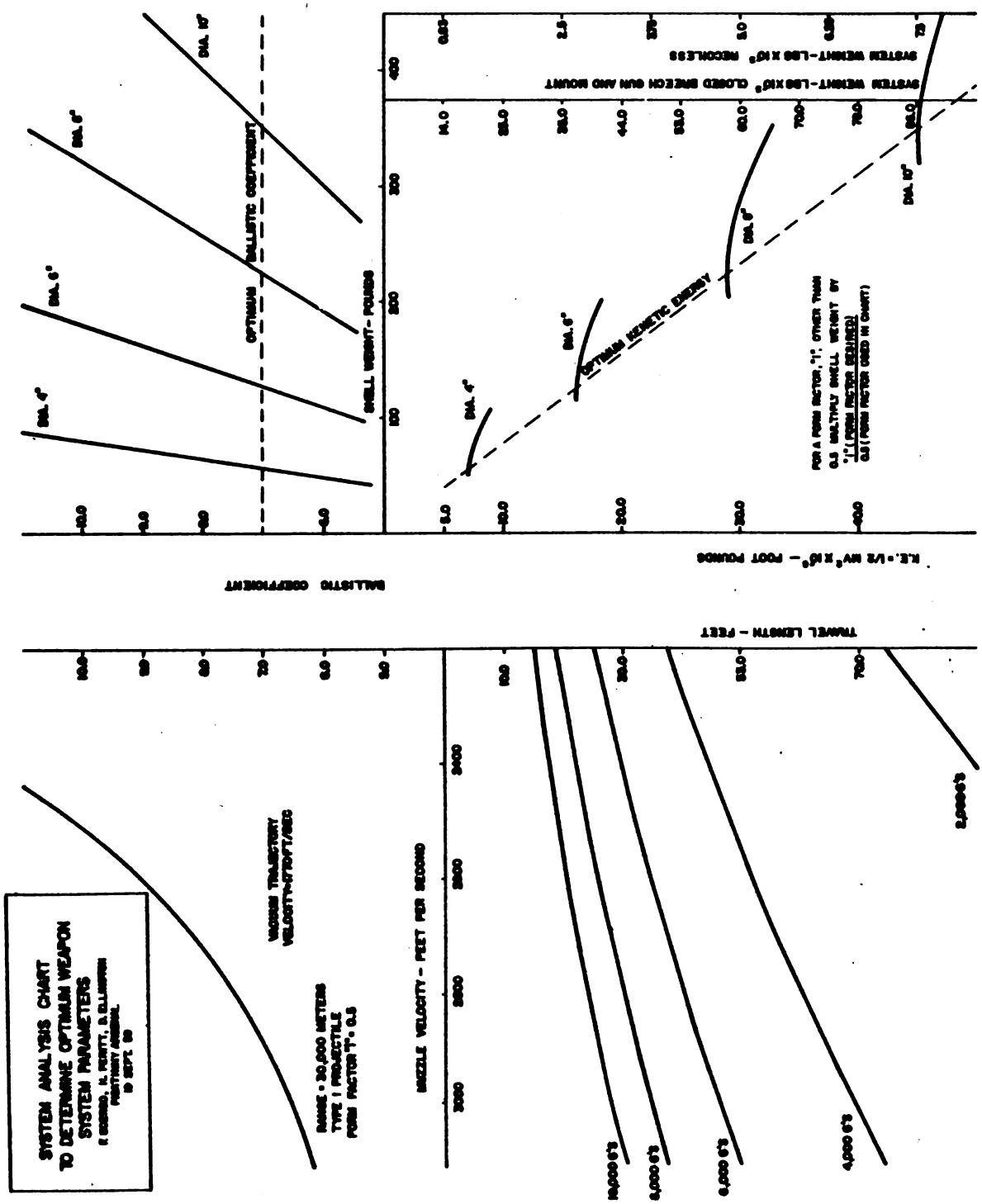


FIGURE 4.

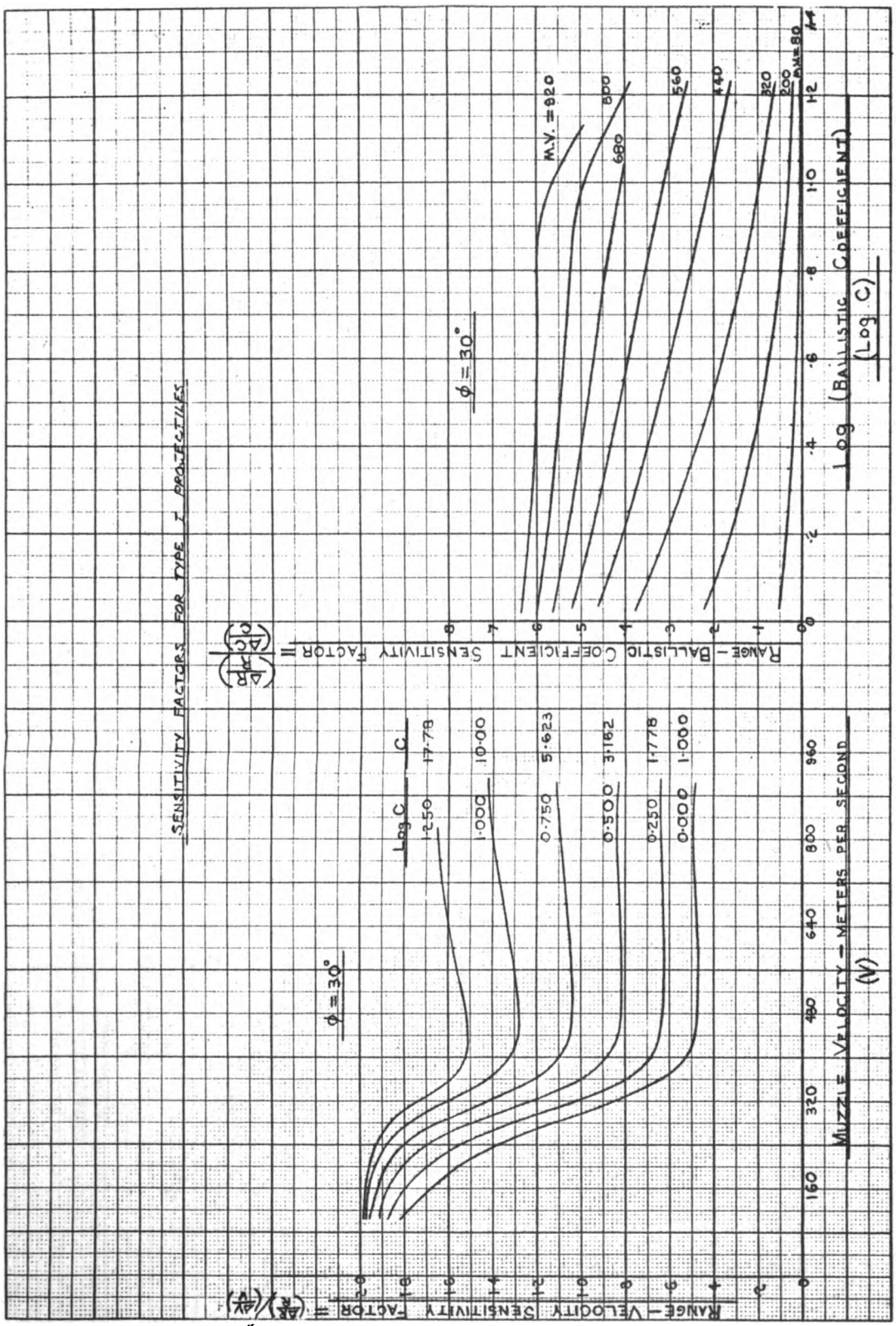


FIGURE 5.

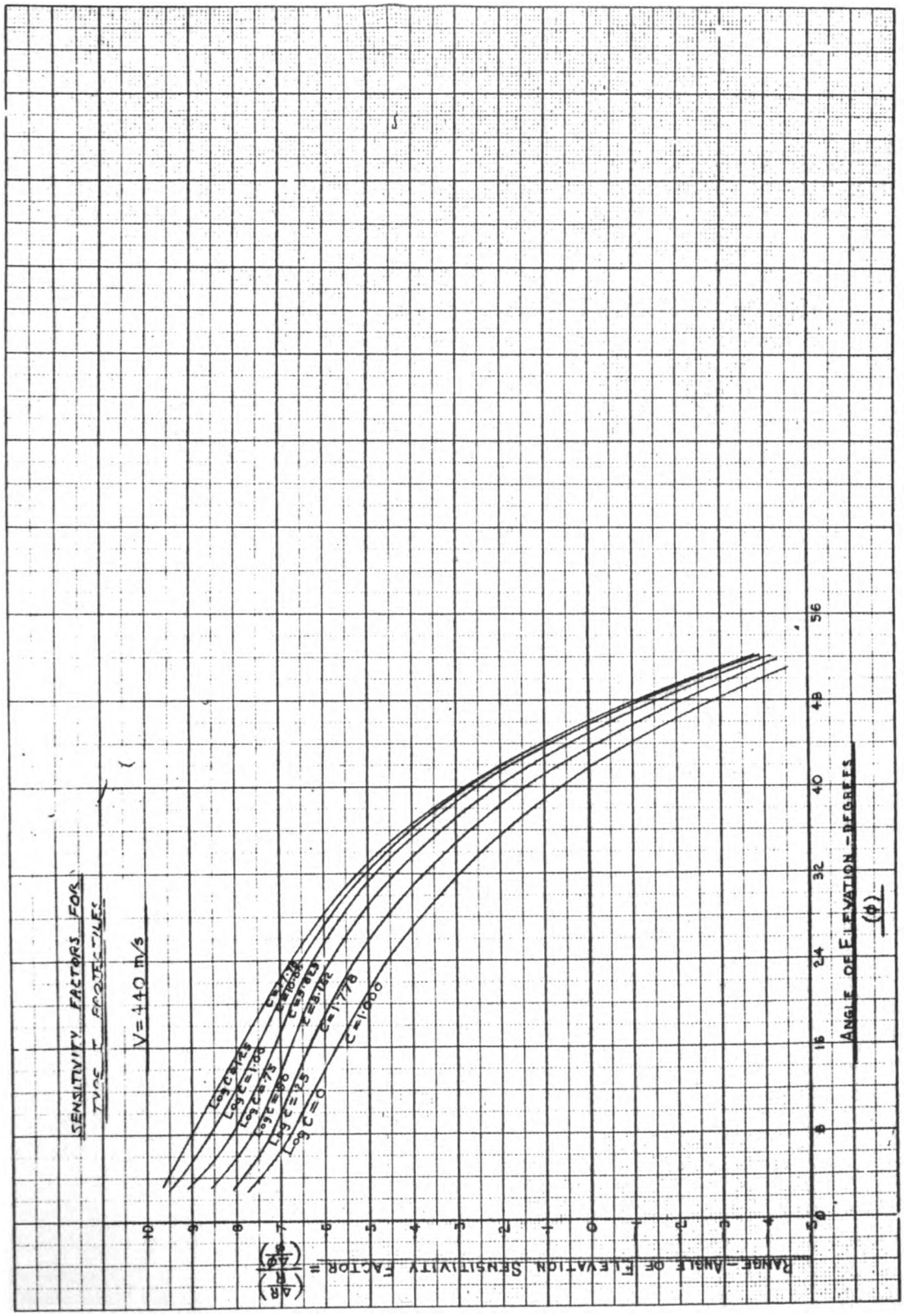


FIGURE 6.

REFERENCES

1. TRAJECTORIES, DIFFERENTIAL EFFECTS AND DATA FOR PROJECTILES, Ordnance Engineering Design Handbook, ORDP 20-140, Ordnance Corps, October, 1957.

2. A SIMPLIFIED METHOD OF WEAPON ANALYSIS, Feltman Research and Engineering Laboratories, Report No. ORDBB-TK-390, Picatinny Arsenal, February, 1960.

3. Dubin, J. A., OBSERVATIONS CONCERNING RELATIONSHIPS OF PARAMETER VARIATIONS IN GUN FIRED TRAJECTORIES, PROCEEDINGS OF THE FOURTH ANNUAL AEROBALLISTICS COMMITTEE SYMPOSIUM, Picatinny Arsenal, April, 1959.

VARIATION OF ARTILLERY AMMUNITION EXPENDITURE WITH INTELLIGENCE

Robert H. Hobbs, Mass. Institute of Technology
Simon H. Kahn , Harvard University
Thomas H. Brylawski, Woodrow Wilson High School, Wash., D. C.
Andrew H. Levy, Phillips Andover Academy, Andover, Mass.
Peter E. Lobban, Bethesda-Chevy Chase High School, Bethesda, Md.
Michael M. Weisfield, Montgomery Blair High School, Silver Spring, Md.
(All members of the Operations Research Office 1961 Summer Research Program for Young People)

SUMMARY. The purpose of this project was to determine how the amount of intelligence available to the artillery officer affects his ammunition expenditure. An experimental procedure consisting of a map exercise was designed to solve this problem. The exercise was pre-tested by 37 artillery officers to test the effectiveness of the design.

PROBLEM. The purpose of this project is to develop a method for determining the relation between target intelligence level and artillery ammunition expenditure. In this study target intelligence was defined as the percentage of the total number of enemy targets on the battlefield detected by friendly forces.

BACKGROUND. Ammunition supply is a critical problem in any theater of operations, particularly during the months immediately following the outbreak of a war. Indeed, it has been estimated that in a general war a period of twelve months would be required to attain maximum munitions production. It is therefore necessary that ammunition stockpiling be sufficient to insure adequate supply during the initial period of production adjustment. In 1958 a study was conducted at C&GSC, Ft. Leavenworth, which attempted to determine supply levels which conform with modern war requirements. However this study assumed perfect intelligence, a theoretical level never attained on the battlefield. In 1961 the Operations Research Office was asked to study this problem. In their study the battlefield was divided into segments and a weighted probability of detection was assigned over these segments. However still no attempt was made to study how a wide variation in intelligence might affect artillery ammunition expenditure and therefore it is desirable to make such a study.

METHODOLOGY. Artillery ammunition expenditure varies as a function of several factors: target detection, ammunition supply rate, terrain, phase (attack, static, retrograde), type of war (nuclear, nonnuclear), support (air,

naval), target identification (location, size, and type of target), and TOE (table of organization and equipment). The problem was evaluated by holding all of these variables constant with the exception of the first two. The interaction between these variables is evaluated by a three by three factorial experiment employing a two-way analysis of variance.

In the specific design of the experiment several other variables may indirectly affect the results of the exercise and must be taken into account. These variables include the rank, general background, training, combat experience, and status (Regular Army or Reserve Component) of the artillery officers. Because these variables can affect the results of test subjects in such a way as to mask the effects of the main variables, it is desirable in designing the details of the experiment to allow for these effects and to attempt to control them. This can be done by careful subject selection or by a distribution of subjects such that the effects of these variables can be taken into account in the analysis and true relations between the main variables may be found.

In evaluating the problem, it is desirable to utilize the experience and judgement of artillery officers. Two experimental procedures are possible: a full-scale war game, and a static map exercise.

The advantage of a war game is that is allows the problem to be evaluated over a long period of time in a dynamic situation. However, this superior yield is obtained at the expense of the sampling. Because of the large individual variances among, a large sampling is necessary to obtain significant results. Such a large sampling can best be obtained by means of a map exercise, which requires a short period of testing time, and can be administered to many subjects simultaneously.

The validity of the experimental design was tested by applying the method to a specific situation - a nonnuclear attack against hastily organized defenses. Employment of nonnuclear weapons is based on the Ft. Leavenworth study, which assumed the continuing importance of conventional war, and conducted studies based on both nuclear and nonnuclear weapon employment. In subsequent applications, the procedure could easily be modified to include nuclear armaments.

An attack situation was chosen because the scheduling of attack fires (preparation, counterbattery, harassing, and interdiction) demands more individual judgement on the part of the commanding officer than scheduling of other types of fire. The importance of the four types of fire considered

is demonstrated by statistics on artillery usage in World War II. Studies show that, in the attack phase, preparation, counterbattery, harassing, and interdiction fires constituted 43.1 percent of all 105mm howitzer missions, 70.6 percent of all 155mm howitzer missions, and 81.6 percent of all 8-in. howitzer missions.

Several situations were considered for the exercise: engagements from World War II and the Korean conflict; school problems from the AMS, Ft. Sill, and the C&GSC, Ft. Leavenworth; and an original situation. The situation selected was taken from the Battle of the Bulge - the attack on the 28th U.S. Infantry Division by German forces on 16 December 1944. This particular engagement was chosen for several reasons. It was felt that an actual tactical situation would yield the most significant data, and of the units considered, the 28th Division had the most nearly complete after action reports. In addition, the high dispersion of the 28th Division over its sector closely approximates the deployment anticipated in future wars conducted under threat of nuclear employment.

Because of the completeness of data concerning 28th Division positions, the U.S. deployment is assigned to Red (enemy) forces in the exercise. Moreover, the officers' familiarity with U.S. tactics and deployment makes a detailed description of enemy tactics and deployment unnecessary.

Blue (friendly forces) were organized according to ROAD 65 specifications. ROAD 65 was employed because it represents an estimate of the type of tactical organization which may be employed in future warfare, and because it deploys friendly artillery in homogeneous groups (i.e., no "mixed" battalions, such as rockets/howitzer or 8-in./155mm).

Division artillery is equipped with six battalions; three 105mm howitzer battalions, one 155 howitzer battalion, one 8-in. howitzer battalion, and one Honest John battalion. Each 105mm and 155mm howitzer battalion has three batteries of six guns each; the 8-in. battalion has three four-gun batteries. Three artillery battalions are attached to division artillery; two 155mm howitzer battalions, and an 8-in. howitzer battalion.

Although a nonnuclear situation is considered, the war is assumed fought under threat of nuclear weapon employment. A battalion of Honest John rockets equipped with nuclear warheads is included in division artillery to conform to ROAD 65 specifications, but it was not employed in the exercise.

Enemy target deployment down to the company level was based on 28th Division positions as given in after action reports for 16 December 1944. Positions were broken down to platoon level by the study's military advisor. Support units not listed in division records, such as supply points, were added. Positions were then plotted on a 1:50,000 map of the combat area.

In the exercise three representative levels of the target intelligence parameter were used: 25 percent, 50 percent and 75 percent. The probability that any individual target would be detected was determined from data giving probability of detection as a function of distance from the FEBA (forward edge of battle area). A mean probability of detection was then determined for each intelligence level and type of target. The number of each type of target, weighted in proportion to the mean probabilities of detection selected for use in the exercise, was determined for each intelligence level and a random number generator employed to determine the actual targets detected for each intelligence level.

Statistics were obtained on quality of visual-contact report information. These probability were then employed in conjunction with a random number generator to determine the type of identification assigned to each target. Some of the targets were identified perfectly as to size and type of target; some were identified only with respect to size; some only with respect to type of unit and some only as enemy.

Three levels of ammunition supply were considered. Figures were obtained for artillery ammunition expenditure in an attack against hastily organized defenses. Additional values of 3/2 the mean level and 1/2 the mean level were also employed.

Since artillery ammunition expenditure is dependent to a high degree on available supply, three levels of supply covering a wide range were employed to avoid having the results prejudiced by too extreme a value. The use of three supply levels also permits the evaluation of a second variable - the effect of supply on artillery ammunition expenditure for a constant intelligence level.

The exercise was pretested by giving it to 37 test subjects. A ten-minute verbal briefing was given subjects to acquaint them with the procedure and to answer any questions which might arise. Subjects were given test materials, consisting of a written set of instructions, a scenario, a map and intelligence overlay, a work sheet for preparing the schedule of

fires, and an officer questionnaire form. As much time was allowed as was necessary to complete the exercise.

The scenario was designed to acquaint test subjects with the background information which they would have acquired as commanders in the field. Three forms were appended to the scenerio. The operations order describes task organization, situation, mission, and execution. The administrative order provides the subject with his available ammunition supply rate. Division SOP (standing operating procedure) for numbering concentrations is employed by subjects in the preparation of target lists. Subjects were also given a 1:50,000 map of the combat area and an accompanying overlay showing the friendly positions and enemy targets. Each subject was given one of three different overlays, depending on the intelligence to which he was assigned. The officer questionnaire form was designed to help determine any factors in a subject's background which might prejudice his results in the exercise.

RESULTS. The experimental procedure was tested on 37 subjects, of whom 24 were able to furnish usable data. It was apparent that, because of the deviations among subjects in each cell of the experimental design, a significant analysis of variance between cells could have been obtained only is sufficient subjects had been placed in each cell to establish a significant mean, or to determine that deviations among subjects were too large for a significant mean to be established by this method.

A sample of the results obtained in the pre-test are shown in Figure 1. This is one cell containing the 50 percent intelligence level and the mean ammunition supply rate. The figures shown are for the sum of the expenditures for the three weapons each corresponding to one test subject. Shown in Figure 2 are the means for each cell of the experimental design. Again the numbers shown are for the sum of the three weapons.

Figure 1

Ammunition Expenditure for All Weapons Combined

	Ammunition Supply Level
	Mean
	2466
Target Intelligence	3766
Level	1914
50%	1002
	5670

Figure 2

Ammunition Expenditure for All Weapons Combined

		Ammunition Supply Level		
		1/2 Mean	Mean	3/2 Mean
Target	25%	2460	3366	1578
Intelligence	50%	1960	2964	4195
Level	75%	4295	3363	4770

Usable data were furnished by all test subjects of rank lieutenant colonel or above, whereas only 58% of other subjects supplied employable results. Since subjects were asked to assume the position of a brigadier general commanding division artillery, a greater familiarity with division artillery fire planning techniques was necessary than was generally encountered among test subjects below the rank of lieutenant colonel. Moreover, deviations among test subjects of rank lieutenant colonel or above were less than those among other subjects. Although it was possible for subjects unfamiliar with division artillery fire planning techniques to complete the exercise, the criterion employed in determining artillery ammunition expenditure was the experienced judgment of the test subjects. It was therefore found that the exercise could provide valid results only if artillery officers of rank lieutenant colonel or above with training and experience in fire planning had been employed as test subjects.

CONCLUSIONS.

1. The amount of data obtained was not sufficient to allow a significant analysis of the validity of the experiment.

2. Based on experience in conducting the experiment and on the comments of test subjects, it is the opinion of the authors that the procedure designed is valid and applicable to the solution of the problem.

3. Because the test subject is asked to assume the position of division artillery commander, subjects should be experienced artillery officers of rank Lt. Col. or above with recent training in fire planning techniques.

AN APPROACH TO SENSITIVITY ANALYSIS OF CARMONETTE
(A Small Unit Combat Monte Carlo Simulation)

Richard J. Matteis and William C. Suhler
Research Analysis Corporation

SUMMARY.

Problem. To develop an approach to the sensitivity analysis of the Carmonette model, a small-unit combat Monte Carlo simulation.

Facts. The Carmonette model has been designed to simulate brief intense battles between tactical units of approximately company strength. The model contains a large number of variables to simulate in detail the combat activities of such elements as individual tanks and infantry squads. A sensitivity analysis is required to isolate critical variables and to determine whether the simulated environment is consistent with existing knowledge of combat situations and the assumptions on which the model was constructed.

Discussion. The approach presented in this paper is an attempt to screen a large number of variables in the Carmonette model under two contradictory constraints. The number of input variables coupled with the time requirement per play imposes an economic constraint, which forbids the use of a full factorial statistical design. Conversely, the lack of a priori knowledge concerning the interaction effects of the input variables and the known existence of large experimental errors prohibits the use of a less rigid design. Essentially the approach is divided into two distinct phases. The first phase is devoted to generating some estimates of effects and variances; the second phase will use these generated values to develop a fractional factorial of the overall model.

The first phase is accomplished as follows: (1) A list of the input variables that may be important is developed. (2) Through study of the structure of the overall simulation, the model is divided into two logical parts and the input variables grouped with respect to these parts. (3) From study of preliminary runs some of the variables are shown to have so much individual effect that they may obscure effects of other variables in the statistical analysis, so these very important variables are studied individually. (4) The remaining variables within each group are then studied simultaneously by a complete factorial.

The second phase consists of taking the variables that have been found to play a statistically significant role and placing them in a fractional factorial of the entire simulation to study the interactions between parts.

Applications. The restraints which determined the nature of this approach are common to the analysis of most Monte Carlo simulations. Consequently, the method of analysis presented in this paper should be applicable as a guide to the sensitivity analysis of Monte Carlo simulation in general.

INTRODUCTION. To acquaint the reader with the simulation to be analyzed a brief description of the objectives and structure of the Carmonette Model is presented. The Carmonette Model was developed to accomplish the following:

> "Military planners need techniques that may be used to test new ideas for Army equipment, organization, tactics, and doctrine early in the development cycle, prior to substantial investments in prototype equipment or in reorganization or retraining. The development of the Carmonette Model was therefore undertaken to provide a method for testing these ideas operationally in an environment of simulated combat." [1]

Carmonette is a Monte Carlo model of small-unit ground combat that has been programmed for the Univac Scientific 1103A digital computer. It has been designed to simulate brief intense battles between tactical units of approximately company strength. To accomplish this task the model simulates in detail the combat activities of such elements as individual tanks and infantry squads. Each combat unit is able to maneuver on the battlefield, acquire combat intelligence, select and fire on targets, and communicate certain information to other units. The activities of the individual units are guided by a tactical scenario, which may be varied from battle to battle. Since it is difficult to state fixed relations for the complex interactions of ground combat, many of the rules of play in Carmonette are probabilistic. This results in an extensive use of the Monte Carlo method in the simulation.

The structural flow diagram of the Carmonette Model is shown in Fig. 1. The sequence of events that occur within the simulation is determined by

a clock system contained within the control routine, as shown on the flow diagram. The clock system will determine whether a combat-unit operation, new tactical missions, intelligence acquisition or neutralization will be activated. The sub-routines shown on the flow diagram, such as target selection and moving, will be referred to as "submodels" throughout the rest of the paper.

It should be emphasized that, although the statistical designs are necessarily oriented to the particular variables of the Carmonette Model, the primary interest is to present a proposed approach for the sensitivity analysis of Monte Carlo simulation in general.

GENERAL DESCRIPTION OF THE PROBLEM.

Sensitivity Study Requirement. To establish the necessity of a sensitivity analysis, the basic reason for a simulation approach to problem solving should be examined. Simulation techniques should be utilized only when the problem cannot be feasibly approached by analytical or experimental methods. Of course this does not preclude the use of analytical and experimental techniques to supplement the simulation effort. In the case of small-unit combat there exists no satisfactory analytical approach to determine the "outcome" of an engagement, and naturally an experimental approach is not feasible. The point is that the simulation approach is often used when the functional relation between some or all of the variables is unknown.

In the construction of the Carmonette Model particular input variables have been included. This inclusion has been prompted by several motivations:

1. Certain variables have been included because of their known importance, at least at the submodel level.

2. Other variables have been included because intuitively they are adjudged important within the structure of at least one of the submodels.

3. Finally there are variables included for which only a conjecture can be made as to their importance to the phenomena being simulated.

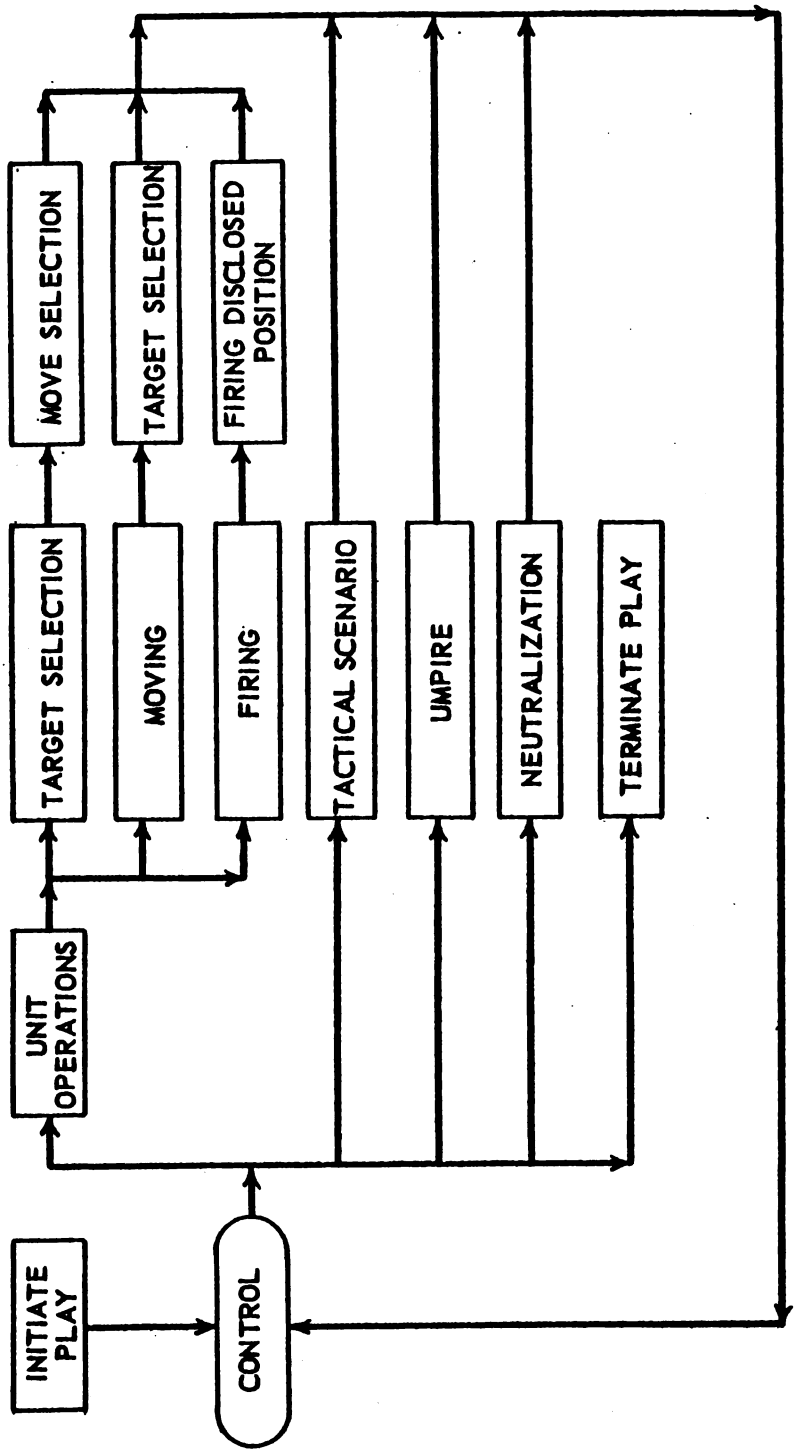


Fig. 1—Structural Flow Diagram of the CARMONETTE Model

Having incorporated these three categories of variables within the simulation, the purposes of a sensitivity analysis become clear.

1. To verify that the important known variables play their expected role.

2. To determine the relative role of the variables that were adjudged important.

3. To ascertain which members of the third class of variables have any significant effect within the structure of the completed model.

4. To determine interactions between variables.

The model was constructed based on some known facts and assumptions concerning the environment to be simulated. The sensitivity analysis should generate a list of the important variables and interactions within the simulation. It is necessary for the analysts and military advisors associated with the simulation to study this list, to decide whether the simulated environment is consistent with these known facts and assumptions. Anomalous results that deviate from the analyst's conception of the simulated environment should constitute a guide to parts of the simulation that require improvement. (Improvements can take the form of either simplifications or refinements of certain parts of the simulation.) Having determined a list of important variables and interactions that is considered appropriate and consistent with the existing knowledge of combat situations, the developers should present this list to any potential users of the simulation. Even if no knowledge exists by which the analyst can validate the importance of a particular variable, the user should be aware of the environment in which his hypotheses are being tested. Also the user may wish to examine the effects of variables that play an insignificant role in the simulation, and the model would be inadequate to measure differences in performance based on variation of these variables. The sensitivity analysis of a model is not a new concept, any discussion of model construction states the necessity of testing the relevance of the variables included. However, there is a tendency to jump from model construction to problem solving without checking model to determine what it can measure with any confidence.

Measure of Effectiveness. As discussed earlier Carmonette was developed to test new ideas for Army equipment, organizations, tactics, and

doctrine for small military units in given situations. Given this objective, a basic question arises at the outset: how does one measure the effectiveness of competing systems? In simulations based on analytical models there usually are well-defined measures of effectiveness, such as, minimization of operating cost and time.[2] In attempting to evaluate competing combat systems in Carmonette measures of effectiveness are not easily identified. The analyst might consider one or more of the following measures to base a decision as to which combat system is most effective.

1. The ratio of the casualties produced as a function of time.

2. The amount of ammunition expended per casualty.

3. Weight of ammunition per casualty.

4. Percentage attainment of final terrain objective by ratio of casualty production or by time or both.

Of course, the costs of competing systems are paramount and decisions should be based on a cost-effectiveness analysis of the competing systems.

Any analysis of combat systems based on the above-mentioned measures, or any other measure of effectiveness, would have to be related to a thorough sensitivity study of the effect of the input data on the measure of effectiveness within the simulation. The input data can be grouped into six major categories of information:

1. Information Acquisition Probabilities.

2. Hit and kill probabilities of a particular weapon system.

3. Terrain.

4. Mobility.

5. Organization.

6. Tactics.

The user must know the sensitivity of the measure of effectiveness to the input data to determine any real differences between the various combat systems studied by use of the simulation.

STATISTICAL APPROACH.

Discussion. Sensitivity analyses of computer simulations based on specific functional equations may utilize at least two different approaches, perturbation analysis and derivative analysis.[2] Since no analytic measure of effectiveness function exists in the Carmonette simulation, the avenue of attack must be through the more difficult and tedious perturbation approach.

A major restriction should be emphasized at this point. This analysis is being carried out on a specific piece of terrain and with a specific tactical meeting engagement of tanks. No attempt should be made to generalize which effects will be most critical given a different tactical situation within the Carmonette model. For example, playing dismounted infantry would require a different resolution, as well as a different tactical situation. For this new situation, the interactions within the simulation might be quite different.

The requirement of perturbation analysis, coupled with the large number of input variables, immediately suggests the use of a large statistical design. The next step is to consider which statistical design is best suited for a sensitivity analysis of the Carmonette model. The constraints on such a design are as follows:

1. Only a limited number of runs can be made due to cost and time limitations. The simulation requires, at the very minimum, one hour of computer time per run, in addition to the analyst time needed to reduce the data.

2. Limited a priori information about the interactions of the variables within the simulation is known.

3. It is known that the experimental error is large. For example, an analytic study of the growth of information within the intelligence submodel indicates a large variance, on the order of the square of the mean, in the time to attain various levels of knowledge.

Keeping these constraints in mind, a thorough literature search of the available statistical designs was made. This included a review of designs ranging from Satterthwaite's Random Balance[3] to a full factorial design. Owing to the economic constrain, a strong emphasis was placed on selecting one of the less rigid designs. It was hoped that such a design would be

sufficient to show at least the large effects with a minimum number of runs. However, as the literature search proceeded it was found that the assumptions required to use the limited designs were very restrictive. Davies states:

> ". . . .When the experimental error is large and/or when higher order interactions are expected to be appreciable, then there is no satisfactory alternative to the complete factorial design."[4]

The above is just one statement among the many such restrictive statements that may be found throughout the literature. These statements demonstrate that, in view of our particular constraints, there are no feasible alternatives to a factorial approach. Any results obtained from the limited designs would be of questionable value.

This leaves only a full factorial design for consideration. Unfortunately, the minimum number of variables that must be considered in studying sensitivity of the model is on the order of 10, requiring for a two-level factorial, 2^{10} or 1024 runs for a single replication. This clearly exceeds the economic constraint on the study. Through a careful study of the model the list of variables, shown in Table 1, were chosen for consideration in the analysis.

Table 1

BASIC VARIABLES TO BE CONSIDERED

	Levels
1. Terrain roughness	2
2. Umpire-information-gain probability tables	2
3. Umpire-information-loss probability tables	2
4. Firing-disclosed-position probability tables	2
5. Movement	2
6. Move missions	2
7. Move speeds	2
8. Organization	2
9. Hit and kill probability tables	2
10. Rates of fire	2

PROPOSED APPROACH.

General. It becomes obvious at this point that a pure statistical approach is not completely satisfactory. Fortunately this is a simulation not a physical experiment. Since it is a simulation, the analyst can turn to the basic model itself, and through study and control of the model, reduce the size of the analysis to a workable level within the constraints placed on it. Instead of considering the simulation to be a "black box", the relations of the various submodels and their corresponding input variables, can be ascertained. From these relations a grouping of the listed variables can be developed. Once these groups have been derived, they can be studied as units to find their internal effects. After these internal effects have been studied each unit grouping may be treated as a single secondary variable to study over-all effects. Jacoby also suggests

this use of secondary variables in the Project OMEGA Air Battle simulation and denotes them as "meta variables".[5]

The original list is divided into two logical groups. One group consists of those variables directly associated with information acquisition and the second group with variables directly associated with casualty production. Table 2 shows the five variables to be considered in the information acquisition analysis. Even with the division of the variables into two groups, to cover all combinations of the remaining five variables would require a 2^5 factorial, or 32 plays per replication.

Again, by appealing to the ability to control a simulation, certain variables can be approached individually, both analytically and experimentally during the rather extensive check-out period of the simulation. Much experimental data can be obtained from a series of runs that must be made to verify that the input data has been encoded correctly and that the simulation is working properly.

Table 2

INFORMATION ACQUISITION VARIABLES

1. Terrain roughness: Line-of-sight Distribution
2. Umpire-Information-Gain Probability Tables
3. Umpire-Information-Loss Probability Tables
4. Firing-Disclosed-Position Probability Tables
5. Movement-Either Stationary or Moving Targets

Finally, some variables will have to be held constant simply to force the number of remaining variables to a reasonable size. Of the list of variables in each grouping, this final requirement was imposed only after some preliminary analysis was carried out.

Information-Acquisition Analysis. To study the five variables included in the acquisition analysis a major requirement is to select a measure of effectiveness. The measure chosen is the average time to pinpoint, E(P).

(Pinpoint defines the level of knowledge about an enemy unit necessary for casualties to occur in direct-fire engagements.)

Of the five variables, perhaps the most basic is the terrain. The characteristics of the terrain determine the existence of line of sight (LOS) between units. Because of its precedence over the other variables, it has been the subject of a thorough preliminary investigation. Terrain is represented in CARMONETTE by a 36 by 36 array of grid squares. Each square is provided with a numerical measure of several characteristics: elevation, height of vegetation, and cover and concealment. In the particular simulation being carried out, a tank meeting engagement, fairly rough terrain is being approximated. The grid-square size is 100 meters.

Several related sets of LOS calculations have been performed.[6] The first of these computations is the probability of LOS by range. The values for the probability distribution for the entire battlefield are presented in Fig. 2. Tactical considerations rule out random location of the units on the terrain. An additional curve representing the specific area with a high probability of containing a unit of either side during a play is also presented in Fig. 2.

A second series of computations was carried out to determine the average duration of LOS between an observer and an enemy unit. These average durations of LOS are presented in Fig. 3. When related to the average time to pinpoint, the critical nature of the length of retention of LOS is apparent. Assuming a fixed set of acquisition probabilities, this relation can best be demonstrated by some examples. Given a moving observer looking at a partly concealed moving tank target, with constant LOS in the range 300 to 1,200 meters, the average time to pinpoint has been computed to be 40 seconds with a standard deviation of 30 seconds. Given the same conditions in the range 1,200 to 2,000 meters, the average time to pinpoint is 150 seconds with a standard deviation of 90 seconds. These values were computed using a Markov chain analysis of the intelligence acquisition submodel and are illustrated on Fig. 3. Thus an observer must have LOS to the square occupied by an enemy unit and retain LOS for sufficient time to acquire a pinpoint level of intelligence about the enemy unit.

Further analysis has been carried out on the availability of LOS during one of the preliminary check-out plays of the simulation. This is presented in histogram form in Fig. 4. Each block represents the range between

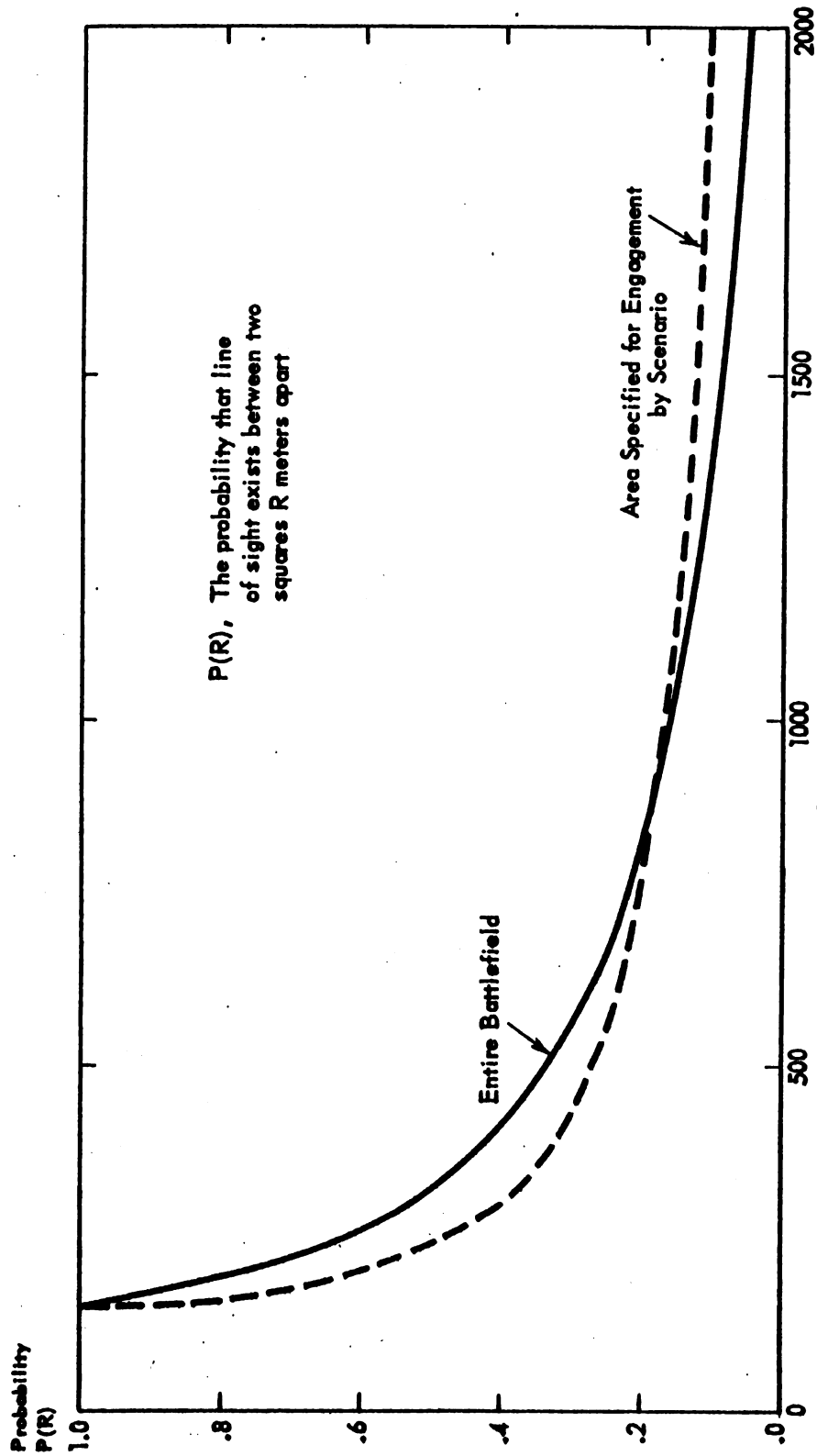


Fig. 2—Probability of Having Line of Sight

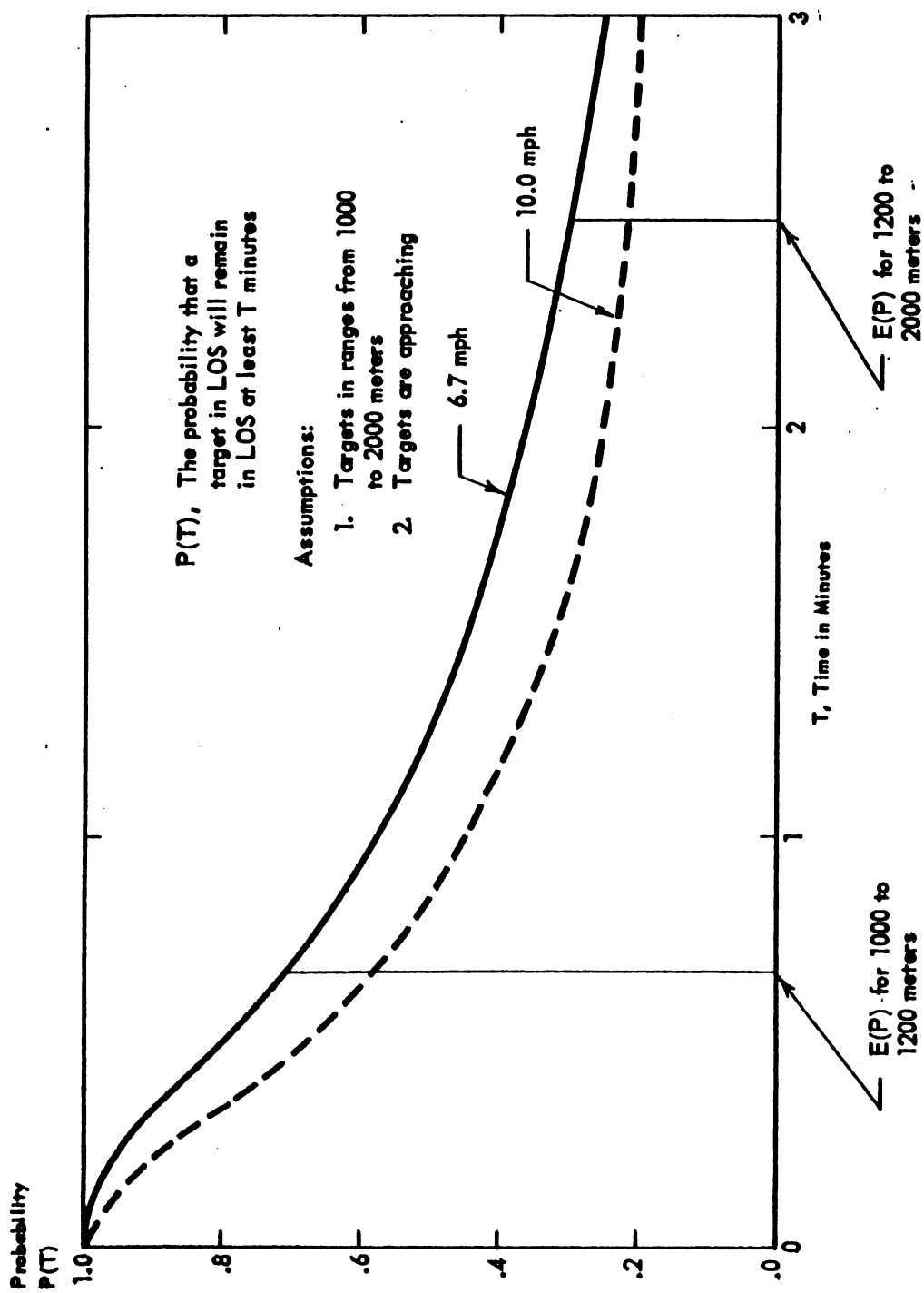


Fig. 3—Probability of Retention of Line of Sight

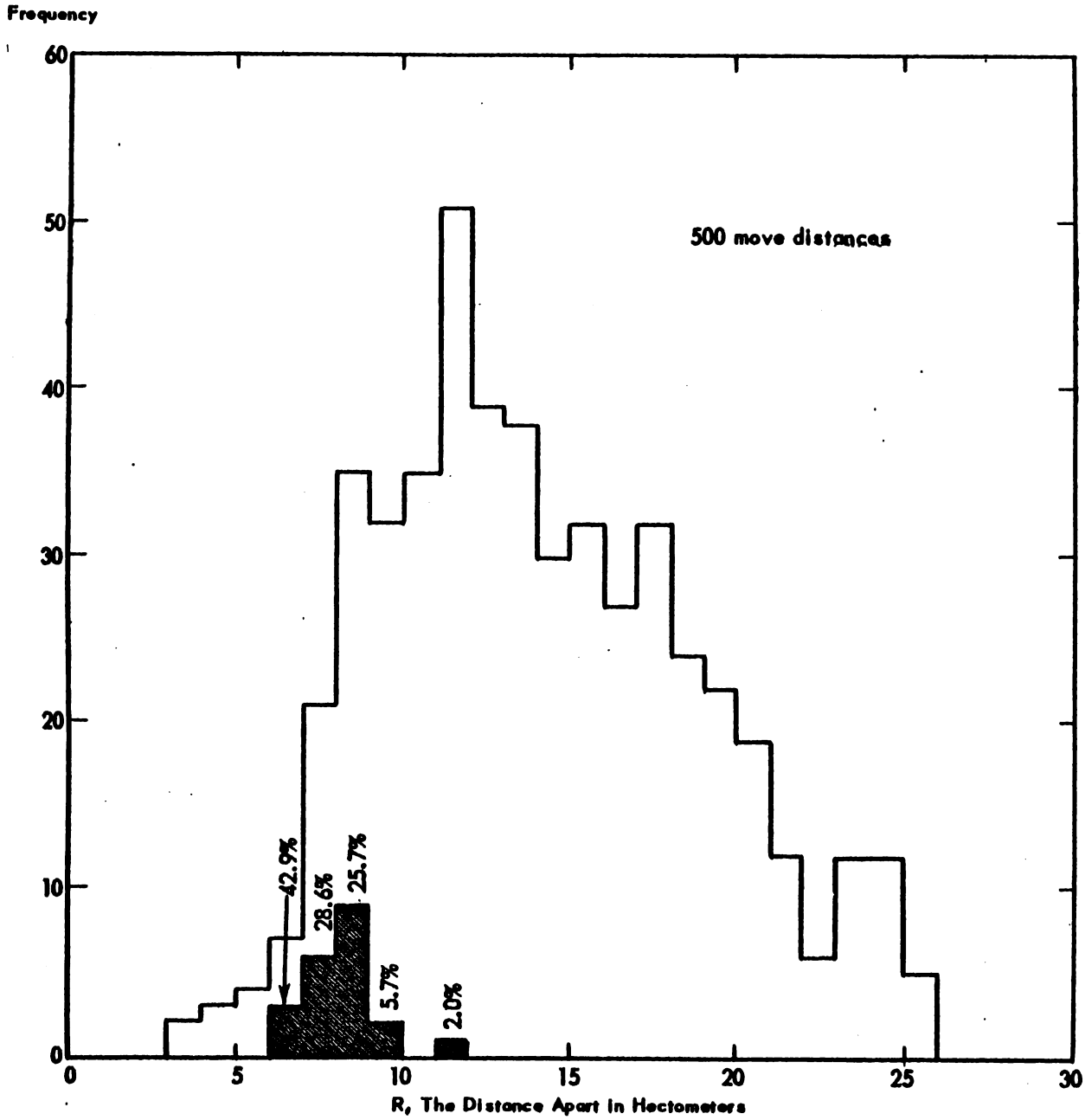


Fig. 4—Existence of Line of Sight During a Play

two opposing units at some time during the play. The crosshatched blocks represent the cases for which LOS actually existed. The percentage figures represent the percentage of total cases at a particular range for which LOS existed. The situation consisted of five tanks in the Blue force and seven tanks in the Red force. Thus, each time a blue tank changed location seven potential LOS's resulted. This figure verifies the low probability of LOS that exists on this particular piece of terrain.

From this separate study of the LOS distributions the significant effect on E(P) of the existence of LOS is apparent. In order that other variables in information acquisition be investigated a series of plays with constant LOS is required. Since LOS exists only a small percentage of the time, even in the near ranges, it is felt that the effect of the other acquisition variables would be obscured. This particular technique has been referred to by Kahn as "importance sampling".[7] Use of this technique permits a far larger sample of occurrences from the region of the sample space which requires investigation.

One other variable will be held constant during this information-acquisition study, i.e., the probability of loss of intelligence information. Little knowledge exists about the length of retention of intelligence information in a short engagement. Consequently these probabilities have been set such that little loss of information will occur during a Carmonette play (30 minutes battlefield time).

Removal of the preceding two variables reduces the list to three still to be considered, and this resolves into a three-factor factorial design requiring eight plays per replication. Table 3 is a representation of this design.

Casualty Production Analysis. On completion of the information acquisition analysis, the remaining variables, shown in Table 4, will be analyzed using casualty production as the measure of effectiveness. Given that the hit and kill probabilities of all weapons will be varied simultaneously, the total number of casualties will be the selected measure. Again the variables must be examined to reduce the analysis to a workable level. It has been previously stated that this analysis is applicable to a specific tactical situation, and no attempt should be made to generalize the results. This restriction was based on an examination of the structure of the model, which resulted in the hypothesis that changes in the tactical situation would have a significant effect on the "outcome"

of the engagement. As in the case of LOS in the information-acquisition analysis it is argued that changes in the tactical situation would obscure the effects of other variables on casualty production. This leads to the use of a fixed tactical scenario. The organization of the opposing forces is considered part of the tactical scenario, and the variable denoted as organization will be held constant in the analysis.

Table 3

INFORMATION ACQUISITION FACTORIAL DESIGN

	Umpire Probabilities			
	A_1 Low		A_2 High	
Firing Disclosed Position Probabilities	B_1 Stationary	B_2 Moving	B_1 Stationary	B_2 Moving
C_1 Low	X_{111}	X_{121}	X_{211}	X_{221}
C_2 High	X_{112}	X_{122}	X_{212}	X_{222}

Table 4

CASUALTY PRODUCTION VARIABLES

1. Move Missions
2. Move Speeds
3. Organization
4. Hit and Kill Probability Tables
5. Rates of Fire

Preliminary analysis has shown that, owing to the extremely limited choice of paths leading to intermediate terrain objectives under a fixed

scenario, little effect could be anticipated by varying the move missions. Therefore the move missions will be held constant in the analysis. This leaves three variables to be considered and results in the three-factor factorial design shown in Table 5.

Table 5

CASUALTY PRODUCTION FACTORIAL DESIGN

	Move Speed			
	Slow		Fast	
	Hit and Kill Probabilities			
Rates of Fire	Low	High	Low	High
Slow	Y_{111}	Y_{121}	Y_{211}	Y_{221}
Fast	Y_{112}	Y_{122}	Y_{212}	Y_{222}

Over-all Fractional Factorial Analysis. After completing this extensive preliminary analysis of sensitivity, sufficient information concerning measures of variation and interaction terms should be available to generate a fractional factorial of the complete simulation. It is at this point that perhaps the previously mentioned "meta" variables should be introduced to aid in the further reduction in the number of plays required in this final design. An example of a possible meta variable is the use of the expected time to pinpoint E(P) as a variable in the over-all analysis. The high and low values and the interaction effects on the expected time to pinpoint can be ascertained from the information-acquisition analysis. This fractional factorial will provide estimates of the effects of the interactions of the information-acquisition variables and the casualty-production variables when in operation simultaneously.

It should be mentioned before concluding that this approach has been established only as a quideline to a continuous process of analysis. Results at any stage in the study can alter the nature of future procedure. Any screening study of a complex structure, like Carmonette, is going to be

difficult regardless of the method of attack, but it is the authors' hope that this approach will facilitate getting the task accomplished.

REFERENCES

1. H. E. Adams, et al, "CARMONETTE: A Computer-Played Combat Simulation," Operations Research Office, ORO-T-389, Feb. 61. UNCLASSIFIED.

2. K. W. Webb, "The Mathematical Theory of Sensitivity", Presented at the Eighteenth Meeting of the Operations Research Society of America, Oct. 60.

3. F. E. Satterthwaite, "Random Ballance Experimentation", Technometrics, 1: 1960.

4. O. L. Davies, "The Construction and Use of Fractional Factorial Designs in Industrial Research", Biometrics, 6: 1950.

5. Joan E. Jacoby and Stephen Harrison, "Efficient Experimentation with Simulation Models", Technical Operations, Incorporated, Project OMEGA Tr 60-2, Jun. 60.

6. Richard E. Forrester and Dick M. Lester, Unpublished analyses of CARMONETTE input data, 1961.

7. Herman Kahn, "Use of Different Monte Carlo Sampling Techniques", Symposium on Monte Carlo Methods, John Wiley & Sons, Inc., New York, 1956.

THE COMPLEX NATURE OF RELIABILITY

A. Bulfinch
Picatinny Arsenal

INTRODUCTION. Many currently published articles on reliability still include definitions, and rightly so. People can talk for hours about reliability before discovering that they are talking about different things. The present day applications of reliability are very complex.

The basic concept of reliability is not new nor complex. Reliability has been used in a qualitative way in the form of safety factors for years. In the past engineers have been satisfied with reasonably large safety factors without calculating the reliability. They would simply conclude in a qualitative implicit way that the probability of successful functioning is directly proportional to the magnitude of the safety factor. The explicit use of safety factors was strictly limited--usually confined to mechanical or electrical loads. But nevertheless safety factors were used then as now to assure success in use, which is now called reliability. Because of the limited use of safety factors there was seldom any question concerning their interpretation. For example, if the applied stress was a tensile load, the safety factor could only be calculated in those terms. From this the only conclusion drawn would be concerning the probability of success (reliability) with respect to tensile loads.

This limited qualitative use of reliability through the use of safety factors was simple and clear cut. But the extensive use of reliability in modern weaponry is quite another story. Safety factors are built into missile components in many subtle ways. Quantitative measures of reliability with respect to an endless number of environmental conditions are now required. These requirements have created many problems although the basic concept of reliability has not changed. Safety factors are still the only way of creating reliability. Economic considerations have made the solution of these problems very difficult.

The quantitative measure of reliability contains many pit-falls. The technical difficulties and the high costs of testing make simplified short-cuts look very attractive. But all that glitters is not gold. Many measures of probability are not measures of reliability.

DEFINITION AND RESULTING CONCLUSIONS. In order to obtain quantitative measures of reliability we must be very specific and carefully

define our terms and objectives. Reliability can be defined as follows:

Reliability is the probability of the successful functioning of a measurable characteristic

1. Under specified conditions
2. For a specified length of time
3. After a specified period of storage.

From this definition the following conclusions can be drawn:

1. Every item has many reliability values simultaneously. There is one for each condition, time, and storage period, and one for every combination of these stresses and every measurable characteristic. For example a TV tube can withstand heat but not vibration. That is, it has a high reliability with respect to heat but a low reliability with respect to vibration. It has even lower reliability with respect to vibration at elevated temperatures than it has at ambient temperatures. Conversely a match cannot withstand heat but can withstand vibration.

2. An item cannot fail unless the applied stress exceeds the strength of the item. When the stress equals the strength there is no safety margin and the safety factor equals one. Under this condition the reliability is equal to 0.3678 when time is the variable (stress) and 0.50 when the level of other environmental stress is the variable.

3. There can be reliability only with respect to some stress. A stress can be applied by only an independent variable; such as, temperature, voltage or a tensile load. A stress cannot be applied by dependent variables, such as, resistance, elongation, or hardness. These are properites of systems or materials. A system or an item cannot be stressed by its own properties. Therefore:

   a. There can be reliability with respect to only independent variables.

   b. Reliability can exist only with respect to some stress.

4. There exists for each item, a set of true but unknown reliability values which remain constant for at least short periods of time.

PIT-FALLS IN RELIABILITY DETERMINATION. In reliability testing there are many pit-falls and traps for the unwary experimenter. Those

for which we have no answers are listed here as well as below under problems to be investigated. These pit-falls are:

1. Determining reliability with respect to only one environment to represent the over-all reliability.

2. Applying environments in sequence in the laboratory to simulate flight conditions.

3. In determining a system reliability, mathematically combining component reliability values that have been determined in different ways.

Pit-falls that can be avoided without further investigation are:

1. Testing without failure. The disadvantages of this procedure are that it is:
   a. Inefficient
   b. Demonstrates reliability only in proportion to the number of specimens tested.
   c. Obtains reliability at the test condition only.
   d. Does not measure the safety margin.
   e. Obtains biased estimates of the true reliability.
   f. Results cannot be mathematically manipulated when no failures are obtained.

2. Testing without stress. Testing at ambient static conditions cannot measure reliability with respect to any environmental stress.

3. Using variable data. Measuring circuit resistance is a good example of a short-cut method to measure reliability. The reason given for measuring circuit resistance is that variable data is more efficient than attribute data. Unfortunately this kind of data contains no reliability information since resistance is not a stress.

4. Using the lower limit of the 50% confidence interval. In an effort to avoid the mathematical difficulties encountered when no failures are obtained, attempts have been made to use the lower limit of the 50% confidence interval as the "best estimate" of the true value of a binomial proportion. The lower limit of any confidence interval cannot be taken as the "best estimate" of the value the interval is expected to encompass if the phrase "best estimate" is defined as follows:

a. The average of all possible values of the estimator must equal the true value. That is, the estimator must be unbiased.

b. The variance of the estimator must be less than that of any other estimator. That is the estimator is efficient.

The lower limit of a confidence interval can never be an unbiased estimate of the true value expected to be within the interval; and the variance of any lower confidence limit of a binomial proportion, which exceeds 50%, is always greater than that of any unbiased estimator.

5. Using life-tests for "one-shot" items. Life-tests cannot measure flight characteristics of missiles or missile components. The important stresses acting on a missile in-flight are the magnitudes of the induced environments--not time. However, life-testing techniques can measure storage characteristics, but only after the fact.

ONE APPROACH TO THE DESIGN OF A RELIABILITY EXPERIMENT. A great deal of planning must to into any experimental program. We must have assurance prior to data collection that the required information will be obtained. That is, our objectives must be explicitly stated in operational terms. It is easy to collect data that contains no reliability information what-so-ever. We must also have assurance before testing starts that the data will be collected in a highly efficient manner. The cost of reliability testing is high at best. Every effort must be made to keep this cost to a minimum by using the most efficient methods known.

In the development phase, the testing problem is usually one of determining the reliability of several components with respect to a large number of environmental conditions. Cost considerations immediately place a very strict limitation on the number of conditions that can be used. As a result the question of which conditions can be eliminated must first be answered. After eliminating the conditions that are obviously of little importance simply by using good engineering judgement, there usually remains several conditions that should be evaluated by more objective means.

The most efficient methods of screening variables such as environmental conditions, are the factorial experimental designs. These designs can determine the effect of several environments simultaneously with an absolute minimum of test specimens.

After having determined the effects of the several most important environments in this manner, an objective decision can b e made concerning the conditions under which to determine reliability. If cost considerations limit the number of conditions to only one, the logical condition to choose for determining reliability will be the most severe condition. This will result in determining the minimum reliability of the several reliability values with respect to separate environments. This approach then identifies the most critical environment with a minimum of test specimens.

The second phase of this approach is to determine the reliability with respect to the environment identified as the most critical. This can be done most efficiently by testing to failure using tests of increased severity. It is only in this manner that the following can be accomplished:

1. The measure of safety margins.
2. Conversion of safety margins to a measure of probability.
3. Furnish an unbiased estimate of reliability.
4. Measure ultimate reliability.
5. Measure reliability-in-use.
6. Furnish high precision.
7. Measure high reliability with small sample sizes.

LIMITATIONS. All of this is very good, especially when compared to the disadvantages of testing without failure given above. We have accomplished the objective of determining high reliabilities with sample sizes within the limitations imposed by cost considerations. But the reliability information contained in results obtained in this way is meager indeed when the over-all picture is considered.

A complicated system of any kind cannot be fully characterized or described by a single numerical value. Just as the "whole man" cannot be fully described by an intelligence quotient, a whole missile system cannot be fully described by a single reliability value. To fully characterize the expected performance of a missile, all possible reliabilities should be:

1. Determined and weighed in accordance with:

    a. Their engineering importance,
    b. Probability of the various environments occurring,
    c. Duration of the environments,

d. Presence of interaction among environments and among components, and

2. Mathematically combined:

a. In accordance with the way the environments occur (that is, simultaneously, in combinations, or in sequence,
b. In various ways to predict the probability of successful functioning of the major and minor subassemblies,
c. In accordance with the system circuitry to predict the reliabilities of the system from the components.

It is clear from this that the present state of the art leaves much to be desired.

PROBLEMS IN RELIABILITY DETERMINATION. A little experience in reliability testing raises many questions yet unaswered. A few of these encountered in the determination of reliability in the development phase of missile components are the following:

1. What is the relation between the reliability with respect to the most severe environment and the reliability with respect to the combination of environments occurring in use?

2. What is the relation between reliability values obtained by varying the level of the environment and values obtained by varying the time of application at a constant level of environment?

3. How can shelf-life be predicted from tests of a few weeks duration?

4. What is the relation between applying environments in sequence and applying them simultaneously?

5. How can system reliabilities be calculated using components whose reliabilities have been determined in different ways? For example, how can reliability values obtained by varying the level of an environment be combined in the same system with reliability values obtained by varying the time of application at a constant level of environment?

Comments leading to the solution of these problems will be appreciated.

PROBLEMS INVOLVED IN DEVELOPING AND ANALYZING DURABILITY DATA FROM FIELD TESTS OF TEXTILE AND FOOTWARE ITEMS*

Virginia W. Perry
Quartermaster Research and Engineering
Field Evaluation Agency, U. S. Army

The Quartermaster Research and Engineering Field Evaluation Agency is located at Fort Lee, Virginia, and is engaged in field testing of newly-developed Quartermaster items and concepts. A good portion of the test activity involves accelerated wear or use of the test items, while the remainder attempts to reproduce normal use as closely as possible. The problem to be presented derives from a normal wear test of women's nylon stockings, which was run by this Agency in 1958. The purpose of the test was to evaluate an experimental stretch-type nylon stocking and the standard nylon stocking in regard to several characteristics, such as comfort, fit, appearance, preference, and durability. Data collected in regard to comfort, fit, appearance, and preference is subjective in nature and generally our analysis is by non-parametric methods, whereas durability data, being objective, should lend itself more readily to analysis by parametric methods. This paper is concerned strictly with the durability data obtained, and the problems encountered in using this data. Durability data is collected in two forms. One is simply a description of various types of minor failures that occur as the stockings are worn. The other is a record of the number of days each stocking is worn before the occurrence of a failure so severe that the stocking must be withdrawn from the test. This is the basic data with which we are concerned. For about 90 percent of the sample on a nylon stocking test, the failure causing withdrawal is a run in the leg portion of the item.

Problems in connection with the analysis of this data are two-fold. The first problem arises through inability to control the test and the items completely, and concerns the proper method of adjusting data to correct for a constantly changing sample size. The other problem concerns identification of the type of distribution involved, when data from a truncated or restricted sample only is available. An analysis of the first problem has been devised and will be presented for evaluation, but no analysis of the second will be given.

---

* At the Conference this paper was presented by Harold R. Rush.

The test was run at Fort Sam Houston, Texas, utilizing 90 members of the Womens Army Corps and a sample of 180 pairs of each type of stocking. Each participant was issued four pairs of stockings, two pairs of each type. Test subjects were not allowed to launder the items themselves; stockings were to be turned in for commercial laundering after each day's wear. Following laundering and inspection by the Agency test team, they were returned to the test subject for a second day of wear. All stockings were coded on the welt with a felt marking pen, for easy identification of stocking type and test subject to whom issued. If one stocking of a pair failed so that it was no longer wearable, the test subject was issued a replacement stocking of the same type for use until failure of the other stocking of the original pair. Test subjects would be removed from the test when the last of their eight original issue stockings had failed. Since the stockings were turned in after each day of wear for laundry and examination, it was possible to maintain a fairly accurate record for each failed stocking of the number of days worn before failure. However, the test was terminated before all the test items had failed. This is a standard procedure, given a previously determined criterion for termination, and considerable work has been done in this field in regard to estimating parameters from such restricted samples. Following Cohen[1], restriction may be one of two types. A censored sample is defined as one in which sample specimens with measurements falling in certain restricted intervals of the random variable may be identified and thus counted, but not otherwise observed, while the remaining samples may be observed without restriction. Time is considered to be the random variable within the content of our problem. A truncated sample is one in which certain population values of the random variable are entirely excluded from observation. In terms of the stocking test, then, the sample would be censored if (1) the test were stopped after a pre-specified number of days of wear on each test stocking; (2) if we knew the actual number of days' wear before failure for all those removed for failure prior to the pre-specified number of days; and if (3) we knew how many stockings were still wearable at that point. The sample would be truncated if we fulfilled the first two conditions for a censored sample, but not the third; that is, the number of stockings still servicable at the time of test termination was not known.

The first problem in connection with our data is that the first condition for either a censored or truncated sample is not fulfilled, this condition being that test termination will take place after a given number of days of wear on each stocking. Decision to terminate was based

on factors not related to the amount of use of each item in the sample. Usable data at the time of test termination may be considered to fall into two categories. The first is the frequency distribution of number of days worn before failure, for those stockings that failed prior to termination of the test. The second is the frequency distribution of the number of days worn before test termination for those stockings that did not fail. This latter distribution covers any number of days of wear from 2 to 35. Actual frequencies observed for these two categories are presented in the Table for the Standard sample up to 30 days of wear. Some word of explanation is relevant at this point regarding the excessively large frequencies in the non-failed categories. Stockings appear in this category for one of the following reasons:

(1) Stockings lost during the test are entered as of the number of days of wear for which records are available.
(2) When subjects are transferred to another duty station and their stockings are turned in to the test team, it is not considered feasible to reissue the stockings to a new test subject, so they are carried as non-failed as of the number of days worn up to transfer of the test subject.
(3) Stockings are still being worn at termination of the test, but, because of test subject leave time or failure of test subject to wear stockings on all scheduled testing days, they have not received the maximum possible wear.

We would now define the expression, the t-th day of wear. This is to be considered as meaning the t-th day that each individual stocking was actually worn, and hence, timewise, could stand for several different dates. Stocking S1 could be worn for the tenth time on April 1, while stocking S10 might be worn the tenth time in May. The information obtained from the t-th day of wear of each individual stocking comprises a set.

On this test, the objective was to determine whether the standard and the stretch-type stockings are equal in durability; that is, we are testing the null hypothesis. A choice of analytical techniques is available, but prior to such analysis, it is necessary to find some method of combining into one distribution the information now contained in the two frequency distributions. At one time we contemplated analyzing the data by the following method. A cutoff point would be set arbitrarily, say 30 days of wear. The original sample was considered then to be equal to the number of stockings worn 30 days and not failing plus the number failed up

to and including 30 days of wear. All other information on the non-failed stockings was discarded, this data on approximately 30 percent of the sample being considered as non-existent. It is obvious that with this procedure, any calculated failure rates for days of wear less than 30 would be over-estimated because the sample size used to determine the failure rate is smaller than the number of stockings we know were actually subjected to wear. We then considered a procedure whereby a different sample size would be used to calculate the failure rates from day to day. This sample size for the t-th day would equal the number of stockings worn on that day plus the number that had failed prior to the t-th day. This method of adjustment, however, fails to give a cumulative failure rate of 100 percent when all items on test have failed by the end of the last test day or are withdrawn for other reasons prior to the final day of wear. Under such a condition, we would expect the cumulative percent failed to equal 100. Therefore this method of adjustment was discarded as unsatisfactory.

At present we are using the following method to determine the percent failed at each day of wear. We attempt to simulate the distribution we would have obtained had no stockings been removed from the test except for failure. In other words, we will attempt to turn our two observed distributions into a truly censored sample as defined by Cohen, with a cutoff point at 30 days. We are arbitrarily using 30 days as a cutoff point because the number of stockings worn longer than 30 days is so small as to cast doubt on the validity of estimates of the probability of failure. The adjustment is best explained by using the idea of hazard, as defined by Lehman and Anderson[2] in reports of their work on the Weibull distribution. Hazard is the instantaneous tendency to fail--that is--it is the probability of failing in a given time interval after having survived up to the beginning of that interval. If the hazard, $z(t)$, is defined as:

$$z(t) = \frac{f(t)}{1 - F(t)} \quad ; \quad t \geq 0$$

then the conditional probability of failure in the interval $(t, t + dt)$ given the individual has survived until time t, is proportional to $z(t)$.

It is obvious that z(t), or the hazard, can be obtained directly from our data, since for each t, where t = 1, 2, ...30, we know the number of stockings failing and the number worn. Knowing z(t) for each value of t, and knowing that F(t) = 0 when t = 1, we can calculate the desired probability distribution values, f(t), and the desired cumulative distribution values, F(t) successively for each value of t from 1 to 30. Some inaccuracy is involved in the above because of the necessity for having to take one day as the value of dt. This is an extremely large value in relation to the entire span of test of 30 days, but the practical difficulties involved in reducing it to minutes, or even to one or two hours, are considered to be insuperable in our test situation. Results of the above calculations for the Standard stockings are shown in the Table, and the cumulative percents failed for both samples are also shown in the graph.

In order for any analysis, either parametric or non-parametric, to be performed on this adjusted data, knowledge of the correct sample size is needed. Although we started with a sample of 360 Standard stockings, we lost almost 40 percent of these before the 30th day of wear for various reasons aside from failure. Therefore, the sample size as of the 30th day must be some value smaller than 360. An average of the number worn each day, which would give us a figure of 218 stockings as of the 30th day of wear, would seem to underestimate the true sample size value. An average of the set of 30 figures obtained by adding to the number worn each day the number failed prior to that day would give us a value of 307 for the standard sample as of the 30th day. Averaging such values for the first 15 days gives a value of 344 for the sample size as of that point. We have used this method to arrive at sample size estimates for use in testing the null hypothesis. This may not be the optimum solution of the sample size problem, because we earlier rejected this method as a preliminary step in arriving at the adjusted failure rates.

In the original analysis of this data, we limited ourselves to a determination of whether or not the 95 percent confidence limits on the difference between the cumulative proportions failed covered zero.[3] Letting:

P_s = cumulative proportion of Standard stockings failed by the t-th day,

P_e = cumulativeproportion of Experimental stockings failed by the t-th day,

N_s = adjusted Standard sample size on the t-th day,

N_e = adjusted Experimental sample size on the t-th day,

Then:

$$P_s - P_e \pm z_{\alpha/2} \cdot \sqrt{\frac{P_s \cdot (1 - P_s)}{N_s} + \frac{P_e \cdot (1 - P_e)}{N_e}}, \alpha = .05$$

gives the 95 percent conficence limits on the difference between the cumulative proportions failed by the t-th day. We found that this difference in cumulative proportions between the standard and stretch-type samples was significantly different from zero from about the eighth day of wear on to the end of the test, and we concluded only that the stretch-type stocking is more durable than the standard.

Naturally this type of conclusion leaves a good deal to be desired. Nothing is presented regarding the life expectancy of the two types of stockings. Some information regarding this appears to be mandatory before an intelligent decision regarding the importance of the difference in durability can be made. The first step in this direction involves the question of what type of distribution is involved. In order to investigate this, we first developed a simulated distribution of number failing each day using our adjusted values of f(t) and a sample size value of 300 for both types of stockings. This distribution for the Standard sample is shown in the Table. All tests yet to be described in this paper are based on this set of figures, and on the corresponding set for the stretch-type sample.

Our first effort was in accordance with Cohen's technique for determining estimators for the normal distribution.[1] We assumed our simulated distributions to be Type I Singly Censored Samples, with censoring on the right at x_o =30 days. Estimates of the mean life for the Standard and stretch-type samples were 26.5 and 30.9 days respectively. Both standard deviations were close to 15 days. Using these values, the expected frequencies were calculated for each of ten 3-day periods, and the Chi-square values for each period calculated. Results indicated very large deviations from expected in the first ten days. Actually it had not been anticipated that this data would follow a normal distribution, and our attention then turned to the exponential.

A number of tests for the validity of the assumption that the underlying distribution of life is exponential are given by Epstein.[4] Certain of these tests based on total lives were utilized and generally resulted in rejecting the null hypothesis that the underlying distribution of life was exponential. In applying total life tests for the exponential, the data in its original form was utilized and not the simulated distributions. For those tests based on total lives, items can be removed from testing at will and replaced at will, if desired. Hence, a constantly changing sample size is no problem within this framework, and the calculation of total life is perfectly straightforward from the basic data. Estimates of mean life using exponential theory were 38 and 51 days, respectively, for the Standard and Stretch samples. These estimates seem far too large. Intuitively, also, we would hardly expect this to be a straight exponential situation. It seems much more reasonable to assume that there are effects present which operate unequally at different phases of the life of the items.

Efforts were next expended in estimating parameters using a Weibull distribution. Here the work of Lehman and Anderson[2] provided guidance. Since their work postulates a censored sample, our adjusted distributions were used as the basic data. Actually, their censored samples fulfill two prespecified requirements--a given number of failures and a given amount of use of each item on test, whereas Cohen's censored samples are based on only one of these requirements.

The Weibull probability distribution may be expressed by

$$f(t) = \frac{Mt^{M-1}}{\alpha} \quad \exp\left(-\frac{t^M}{\alpha}\right),$$

and the cumulative by

$$F(t) = 1-\exp\left(-\frac{t^M}{\alpha}\right).$$

When M = 1, we have the straight exponential probability distribution. M is called the shape parameter and has been found to vary in life-testing situations from values of less than 1 up to a maximum of

approximately 2.5. We are not considering a two-parameter exponential or a three-parameter Weibull as a possibility; with our data, we certainly could postulate no guarantee time and the location parameters, A in the exponential and β in the Weibull, are considered to be zero. The parameter, θ, in the exponential is equivalent to λ 1/m in the Weibull. The shape parameter, M, is of course unknown for our data. Using values of M equal to 2, 1 1/2, and 1 1/3, we estimated the scale parameter, λ, and the mean life, μ, for our samples. Chi-square tests for goodness of fit indicated good results at M = 1 1/3 for the Standard sample, but no encouraging results were obtained for the Stretch sample at any of the values of M utilized. Mean life estimates with M = 1 1/3 were 32 and 39 days respectively for the two type stockings, with standard deviations of approximately 24 and 30. A summary of estimates of mean life for all the cases discussed is given below.

ESTIMATES OF MEAN LIFE (DAYS)

Underlying Distribution		Standard Sample	Stretch Sample
Absolute Minimum		22.1	24.5
Normal		26.5	30.9
Weibull	M = 2	28.0	33.2
	M = 1 1/2	30.9	37.2
	M = 1 1/3	31.6	39.4
Exponential		38.2	51.4

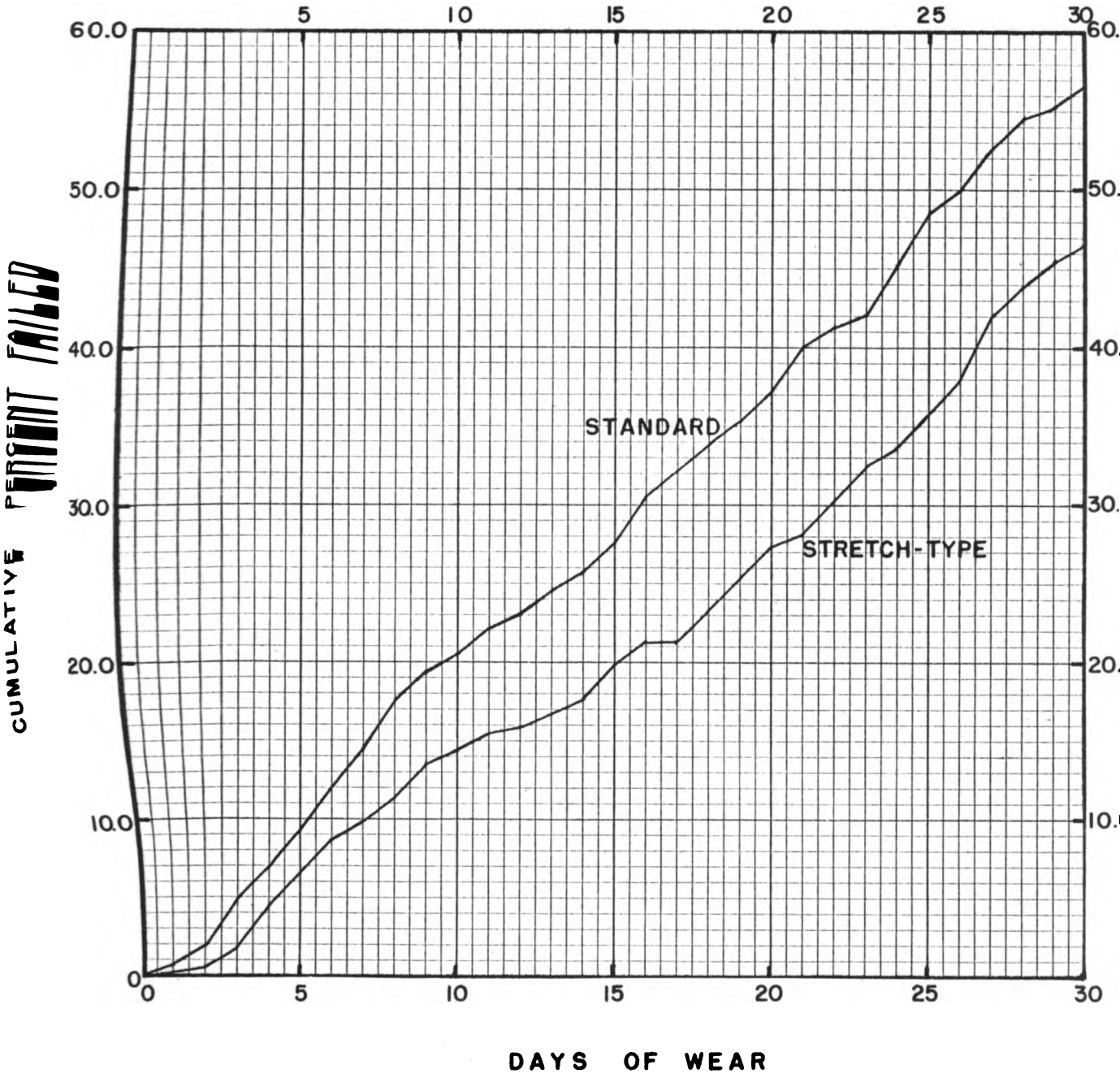
The work of Kao[5] in the field of mixed Weibull parameters may provide the solution to our particular test problem. He is working with electron tubes which can experience either of two types of failure--a catastrophic or sudden failure and a wear-out or delayed failure. He postulates a mixture of two Weibull distributions, each representing one of these types of failure. It seems reasonable that the same factors are present in our samples. However, Kao is able to assign his failures to one of the two causes and develop the distribution for each cause. It is rather doubtful that such an allocation of stocking failures to one of these two cause categories could be performed with validity, although it may not be impossible.

In conclusion, we would like to review the questions encountered in handling the above data. Firstly, can we consider that we have a valid

censored sample when the time of test termination is an arbitrary decision unrelated to the status of the test items themselves? It appears impossible for us to censor by either of the conventional methods. Subjecting all items to the same amount of use is impossible for reasons stated earlier. Termination of the test after a pre-specified number of failures is meaningless in the context of a constantly changing sample size. Within the limitations of our testing situation, it is conceivable that we could censor by a pre-specified criterion of total life. Once the test sample had accumulated this amount of total life, the test would terminate and the data could then be adjusted as described above. Next, it is necessary to determine an effective sample size different from the original size sample, and, if so, is the method presented above satisfactory? When data from a censored sample only is available, how does one attack the problem of underlying distribution? Chi-square tests of goodness of fit can indicate rejection of the hypotheses, but would acceptance ever be possible with censored sample data? Can a mixed Weibull distribution be developed when the data cannot be validly sorted into two cause-of-failure categories? And last, but far from least, come the problems attendant on testing differences between two samples. Little literature in the field of life testing seems available on this subject. How does one set up censoring criteria for a comparison test of two items? Would tests of differences between means of samples from Weibull distributions depend on whether or not the two samples have the same shape parameters, or the same scale parameters, or both?

Days Worn t	Number Worn	At End of Day: Number Failed	At End of Day: Number Withdrawn For Other Reasons	Hazard z(t)	Cumulative F(t)	Probability f(t)	Simulated Observed n = 300
1	360	2	0	.00556	0	.00556	1.668
2	358	5	3	.01397	.00556	.01389	4.167
3	350	11	0	.03143	.01945	.03082	9.246
4	339	7	5	.02065	.05027	.01961	5.883
5	327	8	1	.02446	.06988	.02275	6.825
6	318	10	3	.03145	.09263	.02854	8.562
7	305	8	4	.02623	.12117	.02305	6.915
8	293	11	4	.03754	.14422	.03213	9.639
9	278	6	1	.02158	.17635	.01777	5.331
10	271	4	2	.01476	.19412	.01189	3.567
11	265	5	4	.01887	.20601	.01498	4.494
12	256	3	3	.01172	.22099	.00913	2.739
13	250	5	3	.02000	.23012	.01540	4.620
14	242	4	9	.01653	.24552	.01247	3.741
15	229	6	7	.02620	.25799	.01944	5.832
16	216	8	5	.03704	.27743	.02676	8.028
17	203	5	2	.02463	.30419	.01714	5.142
18	196	5	2	.02551	.32133	.01731	5.193
19	189	4	5	.02116	.33864	.01399	4.197
20	180	5	10	.02778	.35263	.01798	5.394
21	165	8	5	.04848	.37061	.03051	9.153
22	152	3	4	.01974	.40112	.01182	3.546
23	145	2	12	.01379	.41294	.00810	2.430
24	131	7	10	.05344	.42104	.03094	9.282
25	114	7	8	.06140	.45198	.03365	10.095
26	99	3	4	.03030	.48503	.01559	4.677
27	92	5	7	.05435	.50122	.02711	8.133
28	80	3	7	.03750	.52833	.01709	5.307
29	70	1	11	.01429	.54602	.00649	1.947
30	58	2	-	.03448	.55251	.01543	4.629
		163	141		.56794		170.382

(56 still on test)



REFERENCES

1. Cohen, A. C., Jr., "Simplified estimators for the normal distribution when samples are singly censored or truncated," Technometrics, Vol. I (1959), pp. 217 - 237.

2. Lehman, Eugene H., Jr., and Anderson, R. L., "Estimation of the scale parameter in the Weibull distribution using samples censored by time and by number of failures," Sponsored by the Office of Naval Research, Contract NONR - 486(04) - Project NR - 042 - 202. (ASTIA 257362)

3. Dixon, W. J. and Massey, F. J., Jr., Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., New York, 1951, pp. 195 - 196.

4. Epstein, B., "Tests for the validity of the assumption that the underlying distribution of life is exponential," Parts I and II, Technometrics, Vol. 2 (1960), pp. 83 - 101 and 167 - 183.

5. Kao, John H. K., "A graphical estimation of mixed Weibull parameters in life testing electron tubes," Technometrics, Vol. I(1959), pp. 389 - 407.

A PREDICTOR MODEL FOR STABILITY ESTIMATES IN THE ROTATING DRUM

Cecil O. Eckard
U. S. Chemical Corps, Fort Detrick, Frederick, Maryland

This presentation is a preliminary and purely empirical analysis of the problem under discussion.

In biological aerosols which we at Fort Detrick are in the business of testing, two types of decay are taking place. These are biological and physical and the sum of these will be termed total decay. It is assumed that

(1) $$Y_t = Ae^{Kt}$$

where Y_t = recovery at time t
A = recovery at $t = 0$
K = decay

is a model that adequately fits the aerosol recovery data. The value of K is estimated from the experimental data by

(2) $$\frac{\sum (t_i - \bar{t})(y_i - \bar{y})}{\sum (t_i - \bar{t})^2}$$

where t_i = time
y_i = ln recovery at t_i
K = decay

If it is additionally hypothesized that biological decay is exponential with time, the additive property of decays can be stated as

(3) $$K_b = K_t - K_p$$

K_b = Biological Decay
K_t = Total Decay
K_p = Physical Decay.

If K_t and K_p are mutually independent and the factor of covariance is zero the theoretical variance of K_b is

$$(4) \qquad V(K_b) = (\partial K_b/\partial K_t)^2 V(K_t) + (\partial K_b/\partial K_p)^2 V(K_p). \quad 1/*$$

The theoretical variance for equation (3) using the relationship that exists in (4) is

$$(5) \qquad V(K_b) = V(K_t) + V(K_p).$$

If by some method or system one type of decay was negligible, it would eliminate a laboratory assay and reduce the $V(K_b)$. An approach to this is the rotating drum, where the enclosed air mass moves with the drum and the physical decay becomes negligible.

Several rotating drums have been used by the Aerobiology Division at the Biological Laboratories, Fort Detrick, Maryland, and a large amount of data collected. The question was asked of the Biomathematics Division if a predictor equation could be formulated and used in estimating physical decay (K_p) given the drum dimensions, rotational speed and particle size.

Meetings were held with the experimenters and the design engineers to formulate and discuss models and the variables which should be included in the models. The experimenter, from his observations, suggested relationships that existed between decay and single variables. These relationships, with the addition of higher order terms, were used to develop models. They are

$$(6) \qquad f(x_i) = a_o x_o + a_1 x_1 + \ldots + a_{11} x_{11}$$

$$(7) \qquad f(x_i') = a_o' x_o' + a_1' x_1' + \ldots + a_{14}' x_{14}'$$

$$(8) \qquad f(x_i'') = a_o'' x_o'' + a_1'' x_1'' + \ldots + a_{11}'' x_{11}''$$

$$(9) \qquad f(x_i''') = a_o''' x_o''' + a_1''' x_1''' + \ldots + a_{14}''' x_{14}'''$$

---

* See Literature Cited.

where

$x_o = x'_o = x''_o = x'''_o = 1$

$x_1 = x'_1 = x''_1 = x'''_1 =$ 1/drum diameter (D)

$x_2 = x'_2 = x''_2 = x'''_2 =$ rotational speed (ω)

$x_3 = x'_3 = x''_3 = x'''_3 =$ particle size (r)

$x_4 = x'_4 = x''_4 = x'''_4 = \omega^2$

$x_5 = x'_5 = x''_5 = x'''_5 = r^2$

$x_6 = x'_6 = x''_6 = x'''_6 = \omega r$

$x_7 = x'_7 = x''_7 = x'''_7 = \omega/D$

$x_8 = x'_8 = x''_8 = x'''_8 = r/D$

$x'_9 = x'''_9 = e^{kL}$ where L is drum length

$x_{10} = x''_{10} = (1/D)\ [1/(L + D)]$

$x_{11} = x''_{11} = D^2 L$

$x'_{10} = x'''_{10} = 1/D^2$

$x'_{11} = (e^{kL})^2$

$x'_{12} = e^{kL}/D$

$x'_{13} = e^{kL}/\omega$

$x'_{14} = e^{kL}/r$

$x_9 = x''_9 = L$

$x'''_{11} = L^2$

$x'''_{12} = L/D$

$x'''_{13} = L\omega$

$x'''_{14} = Lr$

The k constant in the exponential terms was assigned values which were suggested by the investigator.

The experimental data were collected from 175 trials in various sizes of rotating drums. The variables and their limits were

drum diameter (D) : $12" \leq D \leq 72"$
drum length (L) : $10" \leq L \leq 60"$
rotation speed (ω) : $0.5 \leq \omega \leq 10.5$ rpm
mass mediam diameter (r) : $1\mu, 2\mu, 4\mu$.

Aerosols of sodium fluorescein were aerosolized into a rotating drum and sufficient time was permitted for establishment of laminar flow of air within the drum before aerosol samples were taken. The sampler concentration was measured in a fluorophotometer. In each trial samples were taken at different times and the exponential model shown in equation (1) fitted to the data. The value of K was determined for each trial and it is this dependent value which we would like to predict given length diameter, revolutions per minute and particle size. Other environmental and physical variables were not considered in this analysis.

Normal equations were formed from the data for each model and the resulting set of simultaneous equations solved by the Gaussian elimination method. R^2, which is the proportion of the variation of y around $\bar{y}$ that is associated with the x variables was computed for each model. R^2 was computed by

$$R^2 = \frac{\sum(y - \bar{y})^2 - \sum(y - Y)^2}{\sum (y - \bar{y})^2} \tag{10}$$

All R^2 values were statistically significant ($P \leq .05$). The two models (6,7) with the greatest R^2 value were selected as best fitting the data. The sum of squares and the F ratio for each variable adjusted for all others were examined and the non-significant variables deleted from the models. The new models are

$$f(x_i) = a_o x_o + a_1 x_1 + \ldots + a_9 x_9 \tag{11}$$

$$f(x_i') = a_o' x_o' + a_1' x_1' + \ldots + a_{10}' x_{10}' \tag{12}$$

where

$x_o = x'_o = 1$ $\quad x_7 = \omega/D$

$x_1 = x_1' = L$ $\quad x_8 = (1/D)\ [(1/)L + D)]$

$x_2 = x'_2 = \omega$ $\quad x_9 = D^2L$

$x_3 = x'_3 = r$ $\quad x'_7 = 1/D$

$x_4 = x'_4 = \omega^2$ $\quad x'_8 = 1/D^2$

$x_5 = x'_5 = r^2$ $\quad x'_9 = L^2$

$x_6 = x'_6 = \omega r$ $\quad x'_{10} = L/D.$

A set of simultaneous equations was again solved for each model and the value of the a_i's determined. The R^2 values were computed and the loss in the amount of variance accounted for by our new models was negligible.

Table of R^2 Values

	1st Run	2nd Run
$f(x_i)$	0.7387	0.7377
$f(x'_i)$	0.7616	0.7590

A multiple F test was again run and all variables tested significant. This left us with two unwieldy models which an investigator would probably not be inclined to use. An examination of the sum of squares associated with each model indicates that the first order term of particle size and the first and second order term of rotational speed accounted for a large proportion of the total variation.

A model

$$f(x_i) = a_o x_o + a_1 x_1 + a_2 x_2 + a_3 x_3$$

where

$x_o = 1$

$x_1 =$ particle size (r)

$x_2 =$ rotational velocity (ω)

$x_3 = \omega^2$

was fitted to the experimental data and the resulting loss in R^2 was .076.

Inserting the coefficients and the variables in (13) we have

$$\ln K = 0.46565\, r + 0.035544\, \omega^2 - 0.33460\, \omega - 2.5946$$

$$= 0.46565\, r + 0.035544\, (\omega - 4.7068)^2 - 3.3820. \tag{14}$$

In canonical form (14) can be written as

$$K' = 0.46565\, r' + 0.035544\, \omega'^2 \tag{15}$$

where $K' = \ln K + 3.3820$
$r' = r$

$$\omega' = \omega - 4.7068 \tag{16}$$

Examining equation (15) we see that our surface is a parabolic cylinder whose generator is not perpendicular to any of the coordinate planes (Fig. 1).

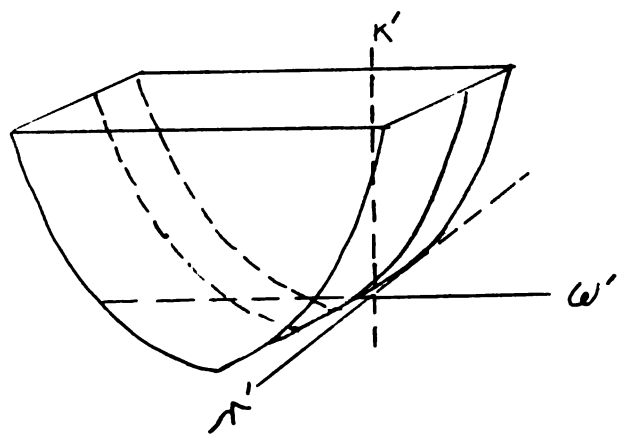


Fig. 1 Surface Generated by Equation (15)

The minimum values of K' lie on a ridge that occurs at the vertex of our parabola in the r'k' when the ω' coordinate is equal to zero. Considering the variable ω it is seen that the optimum rotational speed for minimum decay is 4.7 revolutions per minute. This value is in agreement with experimental results from other test facilities at the Biological Laboratories. 2/

Literature

1. Davies, O. L.: Statistical Methods in Research and Production, 3rd Edition, Hafner Publishing Company. New York, 1957.

2. Foster, W. D.: "Decay Rates for Four Toroid Speeds," Analysis 4804, Biological Laboratories, Fort Detrick, Maryland, 1961.

DISEASE SEVERITY QUANTITATION

Clifford J. Maloney*
U. S. Army Chemical Corps
Fort Detrick, Frederick, Maryland

I. INTRODUCTION. The economic cost of an epidemic outbreak can be divided into two portions; the first depending upon the number of deaths; the second on the number and severity of cases of sickness. The deaths are to be included among the cases of sickness prior to death.

Death is an abrupt change of state, but sickness is a continuously variable condition, so that the economic cost will vary greatly from case to case. A first approach to the problem of allowing for variation can be made by the usual process of forming discrete classes. Some such series as the following (Maloney [1959]) might be attempted:

1. Normal health.

2. Able to perform normal duties, but at a reduced output and/or restriction of off-duty activities.

3. Unable to perform normal duties, but able to perform lighter activities.

4. Unable to work but able to provide own care.

5. Requires care, but not services of special personnel or equipment.

6. Requires "conventional" medical care or facilities.

7. Requires unusual degree or type of care or facilities.

8. Dead.

Items 2, 5, 6, and 7 might easily be separated into two classes. Other classes might also be subdivided, or conversely several of the listed classes might be consolidated. The desirability of doing so would depend upon the

---

* Now at the Division of Biologics Standards, National Institutes of Health.

purpose of any study of the cost of an epidemic or of the cost of disease in general and might well be different under different circumstances. In any case, it appears plausible that the listed classes or any modification of them that might be considered can be viewed as forming an ordered or uni-dimensional series of "productivity" classes. Persons requiring care would have negative "productivity".

II. THE DIMENSIONS OF ILLNESS. Membership in one of the listed classes is a productivity characteristic. The present paper is not concerned with this subject. In turn, productivity rests entirely on the state of the individual's health, which latter is a physiological (and psychological) condition. It is composed of an indefinitely large number of symptoms and signs, including temperature, appetite, aches and pains, agglutination titer, white and red blood cell count, opsonic index, microbial recovery from blood and tissues, x-ray, fluoroscopic-and-percussive examination. One approach to the allocation to the proper class of an individual presenting a given list of symptoms might well be to calculate the multiple regression of the individual's "productive value" on a chosen sample of symptoms. Such a procedure is subject to several limitations:

1. The dependent variable would need to be measured by some sort of scoring system.

2. Generally speaking only a linear regression is manageable (though metameters could be used).

3. A different multiple regression would be needed to forecast different dependent variables.

4. Missing or doubtful values are troublesome.

5. Calculations of the "productivity score", while not difficult, would profit from simplification.

The limitations of the multiple regression approach of the last paragraph would disappear if the complex of physiological signs and symptoms could be summarized or ordered into a one-dimensional sequence of "degree of illness". Many terms from ordinary language suggest that there exists an underlying sickness gradation. We speak of "taking a turn for the worse", of "relapsing", of "being placed on" or of "being taken off" the "critical list", of being "much improved", and similar changes. Even if one master

series of symptom classes cannot profitably be devised, such series for each disease, and even for each phase such as initial attack, recovery, relapse, or secondary attack would serve. But the difficulties of differential diagnosis, the rareness of the "typical" case, the frequency of "complications" and similar factors suggest that gradation may be no more than qualitative--that attempts to render the sickness scale precise will not be successful.

Four possible mechanisms of disease severity quantitation are conceivable:

A. Signs and symptoms may exhibit a strict gradation and errors or observation may not occur;

B. Signs and symptoms may exhibit a strict gradation in every patient or experimental animal; but errors of observation may lead to the reporting of concurrences that may not exist in fact, or failure to detect some that do;

C. Signs and symptoms may not exhibit a strict gradation but some sort of stochastic model may be applicable, the expression of each symptom in any one patient being uninfluenced by any other symptom or sign.

D. A stochastic model may hold in which the expression of two or more signs or symptoms are mutually interdependent.

III. EARLIER DEVELOPMENTS IN SCALING. The possibility of applying to medical research the vast amount of development of the theory of scales in psycho-physics, intelligence, attitude, personality, aptitude, and other fields of psychology was brought home to me by a paper of Clyde Coombs at the 1958 Meeting of the AAAS at Washington, D. C.

A comprehensive summary of the work in the field of psychology has recently appeared. (Torgerson [1958]) The work of Guttman in scale analysis and of Lazarsfeld's latent structure approach are most completely covered in Guttman [1950] and Lazarafeld [1950].

Guttman's "pure scale analysis" involves a formalized, nearly mechanical procedure for developing the scale of intensity from presence or absence measurements on a number of criteria, in the strictly deterministic case A.

Scale analysis in psychological research has not escaped the necessity of choosing between theoretical concepts constructed entirely in terms of "observables" and of inferring a "reality" behind the "observables" whose properties are inferrable from the observations and which in turn ties the observations together. For further discussion, reference must be made to the excellent first chapter of Torgerson [1958].

The notion of scaling is not confined to psychology. The result of the efforts of many workers on the subject of "value" and "utility" in economics has been to reduce the structure of these topics to the same basic mathematical model as those listed under A and B of Section II. Some aspects of value theory arising from the present research on cases C and D will be treated at the end of the paper.

Cases A and B can be adequately treated by a modification and an extension of Guttman's scale analysis. Case C will be treated by essentially new methods in Section VIII. Case D is not further discussed.

IV. SCALE PROPERTIES. A set of objects whether numbers, events, human responses, or disease symptoms can only form a unidimensional scale if the members satisfy certain conditions. The basic postulates may take different forms but it is common to express the requirements in the form of three conditions on an "ordering relation" between the elements. The further characteristics of the relation will depend on the particular field of interest but the three necessary conditions are that the relation be (1) irreflexive, (2) asymmetric and (3) transitive. Further, given any two elements in the set, the relation must hold between them in one or the other order (not both). The abstract properties can be visualized in terms of an ordering relation "precedes". Then, of any two different elements in the set (or scale) one "precedes" the other. No element can precede itself, if A precedes B, then B does not precede A, and if A precedes B, and B precedes C, then A precedes C. In the application of these ideas to disease symptoms, no notion of temporal or spatial sequence is relevant. Thus if A is a rise in Temperature, and B a headache, to say that A "precedes" B merely means that (in a certain disease transmitted to a certain population in a certain way) all patients reporting a headache will exhibit a rise in temperature. The notion "precedes" may accordingly be freed entirely of geometric or temporal connotations by explaining it as an implication. That a temperature rise "precedes" a headache means merely that possession of a headache

"implies" or is always accompanied (whether earlier or indeed later) by a rise in temperature. In terms of a four-fold table then, where lower case letters indicate absence of the corresponding symptom and upper case letters presence.

	Temperature	
Headache	Absent	Present
Absent	th	Th
Present	tH	TH

Figure 1
Hypothetical frequency of co-occurrence of disease symptoms.

Here the symbol th represents the number of patients showing neither a headache nor a temperature and TH represents the number showing both. The statement "temperature precedes a headache" is equivalent to the statement that the condition $tH = 0$ in the four-fold table, i.e., no patients had a headache with absence of temperature. Any set of symptoms which leads to one zero cell in all possible four-fold tables (if consistent) forms an unidimensional scale.

This analysis provides a conceptually simple procedure for examining any universe (and in particular any symptom complex) for "scalability". The set of symptoms may be examined in pairs and does not require any over-all analysis. If, in this analysis, the proper off-diagonal cells are void then the sample forms a scale. (The other off-diagonal cell must be non-zero or the pair of symptoms would denote the same and not an ordered pair of levels of illness. The system would then constitute a semi-scale rather than a simple scale, but this contingency should be rare in practice. Alteration in detail could be made in the model described here if desired to allow for such contingencies.) While any sample whatever is necessarily scalable if it meets this test, the set of elements may not be "meaningful" in the sense of corresponding with any outside variable or variables. Moreover, agreement with theory may be due to chance alone so that while the sample is scalable the universe might not be.

V. EXPERIMENTAL RESULTS. A study of respiratory infection of B. tularense, strain Schu-5 in Osborne-Mendel rats, affords a small amount of data; more by way of an illustration than as a test of the reality of a "sickness continuum" or "morbidity scale". One hundred forty-four rats weighing between 220 and 375 grams were exposed to varying doses of aerosolized B. tularense in a Henderson apparatus. (Maloney [1959]). Following exposure, the rats were held 15-17 days in ventilated cages. Measurements taken on the rats were recovery of the causative agent by culture from the (A) spleen, (B) lung, (C) heart blood; gross pathology of the (D) spleen, (E) lung, and (F) liver; (G) agglutination titer; maximum body temperature during illness exceeding (H) 100°F and exceeding (I) 101°cF and (L) death. Where "gross pathology" is indicated for the spleen or for the liver, the observation was restricted to "enlarged" or "normal"; for the lung, observations were made as "enlarged" or "hemorrhagic" or "consolidated" and in a few cases combinations of these three. It was usually not feasible to obtain an agglutination titer for those rats which died spontaneously, since they usually died before they could have developed antibodies. A number of other factors prevented observation in many cases, so that all of the above observations were available on few rats, though various combinations of smaller sets of observations were available on larger numbers.

The ten qualitative disease symptoms listed above may be used to form disease symptom pairs. A selection of 4 such pairs is set out below in figure 2.

		Agglutination Titer		
		g	G	
Temperature 100°F or Higher	h	10	1	11
	H	51	36	87
		61	37	98

(a)

		Agglutination Titer		
		g	G	
Liver Gross Pathology	f	42	27	69
	F	0	0	0
		42	27	69

(b)

		Lung Blood Culture		
		b	B	
Spleen Blood Culture	a	71	6	77
	A	11	54	65
		82	60	142

(c)

		Agglutination Titer		
		g	G	
Lung Blood Culture	b	59	19	78
	B	1	17	18
		60	36	96

(d)

Figure 2.
Four selected four-fold tables from an experiment on rats.

Most of the tables took the form of (a) and (d) figure 2. The two other tables of figure 2 indicate some departure from the simple theory. Fairly small numbers in one off-diagonal cell, such as the unit for hG in figure 2(a) and for Bg in figure 2(d) can easily be accepted as confirmation that the simple "Guttman Criterion" for scalability has been met. Figures 2(b) and 2(c) will be discussed presently in detail. The observed

results are given in Table 1 and all possible 2 x 2 tables in Table 2. It was not difficult to determine by visual examination of the latter that the symptoms scaled as:

(1) $H < I \lesssim G \lesssim A \lesssim B \lesssim C < L < D < E < F$

The order for H and I is of course beyond doubt, since they are physical measurements. The order of all other adjacent pairs depends on a single 2 x 2 table and hence could have been reversed due to experimental error. The order of more widely separated symptoms is subject to little doubt.

VI. MISCLASSIFICATION. If either of the two symptoms relating to any one table is in error then the rat will be assigned to the wrong cell. This is case B of Section II. Wrong classification can be due to (1) anomalous behavior of the individual rat, (2) an experimental error, (3) extraneous influence, (4) laboratory or diagnostic error, or (5) clerical error. One, or at most a few, entries in the "forbidden" cell, as in figures 2(a) and 2(d) can be dismissed on these grounds. The other two tables require fuller discussion.

Figure 2(b) relates one of the most sensitive symptoms (agglutination titer) and one of the least (liver gross pathology). The zero frequencies in the two cells which indicate presence of gross pathology of the liver result because all these animals died--hence it was not possible to obtain agglutination titers for them. It is thus possible to infer from this table that, in the morbidity scale applicable to this experiment, death is a "milder" symptom than liver gross pathology--for all animals which showed liver gross pathology are in the table, and all died!

Figure 2(c) indicates a still different kind of departure from type. Presence of the causative organism in an organ of the body is based on a sampling procedure. Hence, it is quite possible for the organism to be in fact present in an organ, but in sufficiently small numbers that it is missed by chance in a sample. This is called a "false negative". If the 6 animals which show positive lung culture but negative spleen culture are in fact false negative for spleen culture, then presence in the spleen is a more sensitive indicator of disease than presence in the lung, since

the entry in the aB cell would then be zero, and that in the AB cell 60. The six observed aB entries accordingly suggest a 10 o/o incidence of false negatives in spleen cultures as the experiment was actually performed. On the other hand, the lung was the portal of entry in this experiment. It would appear quite possible for a few organisms to survive on the lung tissue, though no disease was really ever established in the animal. This would constitute a "false positive" and could account for the six entries in the aB cell of figure 2(c). This would still indicate that the spleen culture is the more sensitive "disease" indicator. The frequency of "false positive", if all six entries are such, would be about 8 o/o. Of course, false positive lung cultures and false negative spleen cultures could both be present at some lower numerical frequency. Mathematically, but not biologically, a corresponding analysis would apply to false positives in the Ab cell.

Figure 2(d) suggests however that false positive lung cultures are not frequent, since if the disease does not occur, so that no agglutination titer is produced, a false positive would fall in the Bg cell of figure 2(d). Since only a single entry occurs, such an outcome seems to have a probability of occurring at or below 2 o/o. The one entry that is observed, however, showed no titer and a peak temperature of 100.5. As near as can be judged, it is a bona fide false positive.

Examination of the records of individual rats in the aB cell of figure 2(c) confirms the inference that they represent false negatives for spleen culture, except for the one rat of the previous paragraph, which seems to belong in the cell ab. Of the remaining 5 rats, 3 died and were diagnosed as having died from tularemia. It is felt that the spleen bacteria may have been destroyed by decomposition products, or that they may have received inadequate sampling. The other two rats both showed temperatures of 104^{o}. One showed a titer of 1:40 and one a titer of 1:80. The latter showed recovery of the organism from heart blood.

In summary, there is an indication that 2 survivors showed false negative spleen cultures and one a false positive lung culture. The percentage occurrence is hence 3.6 o/o and 1.4 o/o. The three dead rats would raise the incidence of false negative spleen culture to 8.5 o/o, had none been negative due to destruction of the bacteria by decomposition products.

This example represents the greatest incidence of misclassification in the entire experiment, and involved the culturing of a very delicate organism, effective in low concentrations. Accordingly, it is concluded

that case B Section II is applicable to the observed data. The frequency of misclassification indeed is so low that any formal estimate of misclassification error seems superfluous.

VII. EXPERIMENTAL CONCLUSIONS. The scaling of degrees of morbidity by the foregoing analytical procedure is in general agreement with the opinions of pathologists. Thus, temperature rise (H) and (I) and agglutination titer (G) are more sensitive indicators of onset of disease than recovery of organisms from tissues of sacrificed animals. If these conclusions regarding the appraisal of the relative states of severity of disease are not new, at least, the analysis carried out in this example serves to strengthen those based on professional opinion.

That tissue samples being taken for culturing of the organism were not completely adequate is a further lesson learned from the analysis.

In general, the experimentalists felt that the results of the experiment (which is of a type extremely difficult to carry out) were dissappointing. The above analysis revealed that, on the contrary, the results were very clear-cut, and the experiment well executed.

VIII. STOCHASTIC CASE. While the data of this particular experiment seem to follow the deterministic model (with minor allowance for errors of observation) in an entirely satisfactory way, the few non-conforming readings suggest the desirability of extending the analysis to include a chance component in symptom expression as an alternative. This situation is listed under cases C and D of Section II.

Consider the "trace" for frequency of appearance of a specified symptom over the full range of illness from perfect health (0) to maximum illness (say - 1) in the deterministic situation.

If an animal presents the symptom, it yields a positive response. Now, if s is the sickness level that always evokes the given symptom, then any level of illness less than s will never evoke it, and all greater levels always will. The profile of the functional relation between disease severity and symptom occurrence is thus a step function.

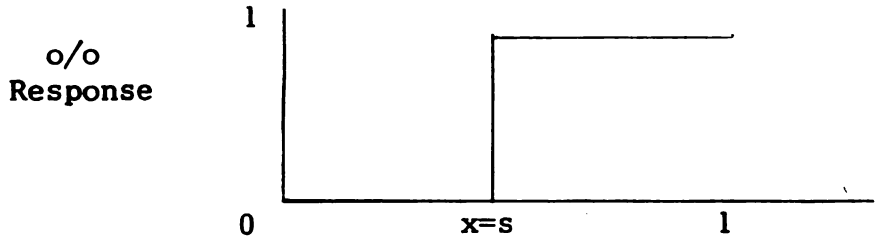


Figure 3.
Graph of deterministic symptom dependence on level of disease severity.

To generalize this to a chance relation in the psychological context, Lazarsfeld [1950] considers a series of polynomials of successively higher degree. He finds the first and second degree curves particularly interpretable and helpful.

In the biological context, it is more natural to consider a relation

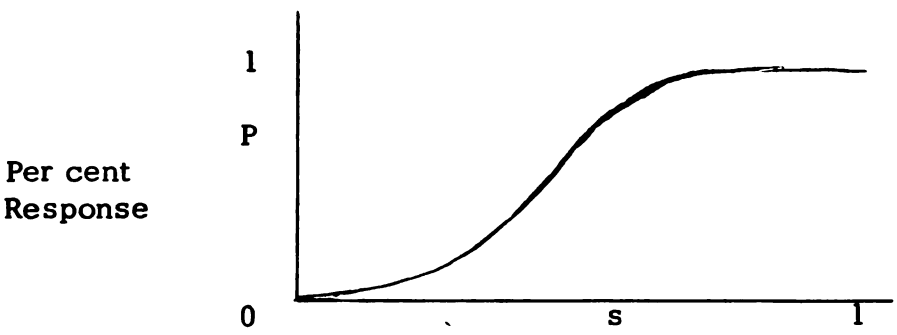


Figure 4.
Curve relating frequency of occurrence of a specified symptom on level of disease severity

between illness severity and mean frequency of occurrence of symptom that takes the form of an intergrated normal, a logistic, or some similar curve. In a large population of animals all at an equal disease severity level, s, the proportion that would show a particular symptom would be P. At other sickness levels, the proportion showing the chosen symptom would be greater or less. A separate curve would relate disease severity and frequency of occurrence for each different symptom studied. In the example of this paper nine symptoms were observed (not in every rat), hence the figure, to be complete, would contain nine such graphs.

In a given animal any combination of symptoms might appear, depending on those features that contribute the stochastic variation around the curve relating mean frequency of occurrence and disease severity. A definition of scalability is needed in these new circumstances.

Scalability at a fixed level of disease severity will be defined first as holding, not between the actual symptoms presented by an individual animal, but between the mean frequency of symptoms in the population of animals at that fixed level of disease severity. As, on the basis of this mathematical model, the proportions of animals showing the different symptoms (being numbers between zero and one) necessarily form a scale, scalability is necessarily exhibited at any fixed level of disease severity when mean proportionate symptom occurrence in an animal population, not the occurrence of symptoms in individual animals, is used.

Consider the next set of mean symptom frequencies at two different levels of disease severity. The order of mean symptom frequencies at the two different disease severity levels may or may not match. If they do, the symptom complex will be said to scale at these two disease severity levels. If this condition holds at all possible pairs of disease severity levels, then the symptom complex for that particular disease in that particular host will be called scalable without qualification as to disease severity level.

Scalability is defined in terms of mean symptom frequency. Deviation from expected frequencies will occur in small samples because of a number of considerations. In this paper it will be assumed that such deviations are independently distributed in the binomial distribution, i.e., case C holds.

IX. MATHEMATICAL FORMULATION. The set of expected probabilities in case C can be expressed as a matrix.

$$
(2)\qquad
\begin{array}{ccccccc}
P_{1,1} & & P_{1,2} & \cdots & P_{1,i} & \cdots & P_{1,K} \\
P_{2,1} & & P_{2,2} & \cdots & P_{2,j} & \cdots & P_{2,K} \\
\cdot & & \cdot & \cdots & \cdot & \cdots & \cdot \\
P_{i,1} & & P_{2,2} & \cdots & P_{i,j} & \cdots & P_{i,K} \\
\cdot & & \cdot & \cdots & \cdot & \cdots & \cdot \\
P_{L,1} & & P_{L,2} & \cdots & P_{L,j} & \cdots & P_{L,K}
\end{array}
$$

where each column (indexed by the second subscript) contains the probabilities of each of the L symptoms (indexed by the first subscript) at a fixed level of disease severity. The K columns represent a division of the sickness scale into K levels. The successive entries across each row yield the probabilities that the i'th symptom (or sign) will be shown by an animal or patient at each of the K sickness levels. Choose an arbitrary row of the matrix. Subject the matrix to column interchanges until the entries in this one row are non-decreasing to the right. If the entries in all other rows are also non-decreasing to the right, the matrix (and the physical phenomenon underlying it, disease severity or something else) is scalable in the sense of Section VIII. Next choose an arbitrary column and subject the matrix to row interchanges until that one column is non-increasing from top to bottom. It is a matter of common observation that then all columns are also ordered, for otherwise symptoms would more frequently appear for less ill animals or patients. This operation will order the signs and symptoms from top to bottom in order of increasing severity. Since the sickness levels at the left of the matrix are low, the

probabilities here are small. At each of the successive columns to the right, the sickness level increases, and hence these probabilities are larger.

In practice, the population values in (2) would have to be estimated from observation, and very large amounts of data and an adequate estimation method would be required for reliable estimates. Two simplifying assumptions will greatly reduce the volume of such data required and yet yield estimates differing little if at all from those derived from known values of matrix (2).

Consider a single row of (2), say the i'th and assume that the sickness level, z, is directly observable. Then, it would be standard bioassay practice to assume that a metameter of z, say s, and another of P, say Y, could be found, so that

$$Y_i = a_i s + b_i \qquad 1 \leq i \leq L \tag{3}$$

would yield the probability of exhibiting the i'th symptom at a sickness level corresponding to a value of s. The K values of P_i have here been replaced by the standard two values treated in bioassay, a and b. Since (3) is a linear equation, its graph is, of course, a straight line. In Section VIII, we have defined scalability to mean that the greatest population probability of symptom occurrence at one sickness level is greater at all other sickness levels. Together with the fact that each symptom occurrence is related to disease severity by a linear relation of the form (3), this scalability condition is enough to show that the several straight lines relating the metameter of mean frequency of occurrence of each sign or symptom to the metameter of disease severity are parallel, (9), the system of lines being represented by the system of equations:

$$Y_i = a\ s + b_i \qquad 1 \leq i \leq L \tag{4}$$

where, now, a is common to the L equations.

Equation (4) relates a transformation of symptom frequency to the metameter of sickness level s. However, no "operational method" exists by

which illness can be measured directly in terms of physical processes. The present procedure of scaling by means of symptom complexes can at most reduce the measurement of illness level to isomorphism with the real number system. Hence, no metameter of disease severity is appropriate in the adopted relation between disease level and frequency of occurrence of a given symptom; the s in equations (4) will serve directly as the measure of sickness level, as well as any other.

The choice of origin and scale in (4) is still arbitrary. The b_i serve to relate the several parallel lines relative to a chosen one. While greater experience might lead to a different choice, convenience would seem to suggest choice of one of the central lines, denoting one of the common, but not the mildest, symptom. For this one symptom, the equation becomes

$$Y_o = a \cdot s \tag{5}$$

Now, in nearly all realistic applications Y ranges between 0 and 10. If we choose the scale so that $a = 1$, the sickness level for the base symptom likewise will range between 0 and 10, and system (4) will become

$$Y_i = s + b_i \qquad 1 \leq i \leq L \tag{6}$$

In the application of these equations to a particular complex of symptoms, the b_i are considered known. How they may be estimated from accumulated data will be considered in Section XII. The relation between population per cent occurrence of the i'th symptom and its probit is

$$P_i = \frac{1}{\sqrt{2\sigma}} \int_{-\infty}^{Y_i - 5} e^{-\frac{t^2}{2}} \, dt \tag{7}$$

From which

$$\frac{dP_i}{dY_i} = \frac{1}{\sqrt{2\sigma}} \left(e^{-\frac{(Y_i - 5)^2}{2}} \right) \tag{8}$$

X. AN ILLNESS INDEX. In practice, we do not start by being given the sickness level and inferring the symptoms, but conversely, observe the symptoms and seek to infer the sickness level. A specific patient might or might not exhibit a "typical case".

A given individual, at any point in the course of an illness, will however exhibit certain signs and symptoms. It is on the basis of this set of symptoms that an index of illness must be constructed. The role of dynamic factors can be accomodated, to some extent at least, in the usual fashion by regarding shifts in readings over fixed periods of time as being themselves readings. If a patient who showed any one of these symptoms showed all milder symptoms, then case A would apply and illness would scale in a deterministic fashion. The essence of the stochastic case is that one or more milder symptoms may be absent when a severer symptom is present. The discussion in this section is limited to case C of Section II.

To get a first feel for the situation, consider the mildest five symptoms for a particular disease. Suppose a patient exhibits the first three, fails to show the fourth, but does show the fifth. A little reflection will show that this patient must be ranked as sicker than a second patient who shows the four mildest symptoms and no others. Failure of the fourth symptom in the presence of the fifth in the first patient is considered due to chance, so that that level of illness is chosen which maximizes the chance of their joint occurrence. This will be at some level of illness greater than a 50-50 probability for the fourth symptom and less than a 50-50 for the fifth. But in the case of the second patient, the non-occurrence of the fifth symptom and occurrence of the fourth implies that the level of illness is estimated to be at (or even below) the 50-50 point of the fourth symptom.

An illness index can accordingly be set up by forming the maximum likelihood criterion for the particular combination of symptoms exhibited by each patient. Further discussion will be facilitated by reference to figure 5.

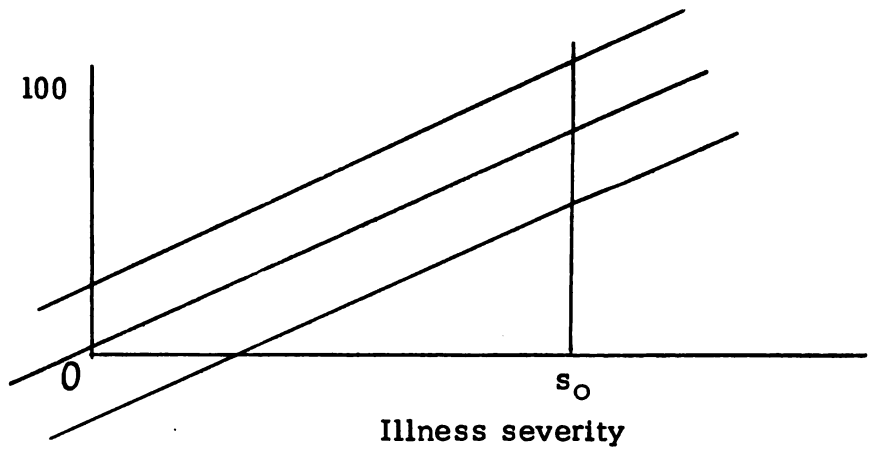


Figure 5.
Hypothetical relation between disease severity and frequency of several symptoms.

For simplicity, curves of frequency of symptom occurrence versus disease severity are shown in units of probit of response as straight lines, (see Section IX) and only three lines are shown. In the experiment used to illustrate this paper, nine symptoms were used. The position of one patient is shown by a vertical line on the figure at s_o. The probability of each symptom for this patient is given by the point of intersection of its curve (line) with the vertical line at s_o.

If the various probabilities of the matrix (2) were known from prior experience or otherwise than the probability of the whole set of symptoms, for a sickness level j would be:

$$P_j = \prod_i P_{ij}^{x} \; Q_{ij}^{(1-x)} \qquad 1 \leq j \leq k \tag{9}$$

where

$$Q_{ij} = 1 - P_{ij} \tag{10}$$

and

$x = 1$, if the i'th symptom occurred, and

(11) $x = 0$, if the i'th symptom failed.

If the values from each of the j columns of (2) are substituted in (9), then that column whose values led to the largest P would be from the column at the most likely sickness level.

Taking the logarithm of equation (9)

$$(12) \qquad L_j = \log P_j = \sum x \log P_{ij} + (1 - x) \sum \log Q_{ij}$$

whence,

$$(13) \qquad \frac{dL_j}{ds} = 0 = \sum \frac{x}{P_{ij}} \cdot \frac{dP_{ij}}{dY_{ij}} - \sum \frac{1-x}{Q_{ij}} \cdot \frac{dP_{ij}}{dY_{ij}} .$$

since

$$(14) \qquad \frac{dY_{ij}}{ds} = 1 \qquad \text{all } i, j \qquad \text{(from 6)}$$

An exact solution of equation (13) employing the usual relation between each P and its corresponding Y from equation (7) can be carried out by successive approximations. However, since for L symptoms there are only 2^L possible outcomes, it would seem that a simple table would always be used. In practice then a glance at the appropriate table entry would immediately yield an estimate of illness severity as soon as the appropriate symptoms occurrences were noted.

XI. UNSELECTED PATIENT FREQUENCIES. The previous discussion assumed either (a) that the illness level was known, or (b) that the b_i were known. Either situation relates to a later time in the sequence of steps involved in developing an illness index. The raw data consist of a mass of individuals exhibiting assorted symptom complexes at a variety of unknown degrees of illness. The specific illness of each individual at a specific time is not known directly, but further, not even the nature of the distribution of illness in the population is known, though it appears very plausable that it is non-normal, and even constitutes an extreme form of a J-shaped distribution, most of the frequency residing at zero illness. The first point to be established is that the order of symptoms in a large group of patients unselected for disease severity is identical with the order at any single level.

Let the distribution of the K disease severity levels in the population be given by a set of K weights $w_1 \ldots w_K$ assigned to the successive columns of (2), where $\sum w_j = 1$, and each $w_j \geq 0$. For typographic simplicity, let P'_j and P''_j be any two probabilities selected from the j'th row of (2).

Set

$$P' = \sum_j w_j P'_j \tag{15}$$

and

$$P'' = \sum w_j P''_j \tag{16}$$

P' and P'' are row marginal probabilities of symptom combinations in the population. If P' is to represent the more frequent combination, the more common, hence the milder, symptom then

$$P' - P'' = \sum w_j P'_j - \sum w_j P''_j$$

$$= \sum w_j (P'_j - P''_j) \tag{17}$$

Since the w_j are arbitrary, equation (17) yields the requirement

$$P'_j > P''_j \tag{18}$$

for each j. Conversely the condition is clearly sufficient for any one set of w_j, representing a distribution of illness in a population. Scalability in stochastic model C then depends only on the reasonable assumption that the more frequently appearing symptom at one level of illness will also be more prevalent at each other level, and is independent of the distribution of disease severity among the cases.

In a sample of data, symptom frequencies might not exhibit an order thought to be appropriate, or two sets of data might exhibit conflicting orders. Tests of order are discussed by Chassan [1960].

XII. DEVELOPMENT OF THE EQUATIONS. The previous sections provide a rationale by which the severity of illness can be inferred from the complement of symptoms exhibited by a particular patient. So that this may be done, the set of displacement constants d_i must be provided. For this purpose a set of symptom occurrences on a fairly large number of patients will be required. In general, it should be easy to secure many times the 140 records used in this small study--and hence the material should be adequate.

Section XI indicates that the symptoms can be ordered on the overall proportionate frequency of their occurrence in the data disregarding considerations of disease severity. This being done, those records

showing few symptoms and those records showing nearly all of the more severe symptoms, may be temporarily set aside, in confidence that, of the remaining records, (nearly) all pertain to intermediate levels of disease severity. Now, it is well known (Finney [1947]) that, over the range from 20 to 80 percent response, P is nearly proportional to Y. Hence, we can put

$$Y_i = 10\ P_i \tag{19}$$

since $0 \leq Y_i \leq 10$, and $0 \leq P \leq 1$,

nearly always.

The d_i are the differences between successive Y_i, when $s = 5$; for which, for this first approximation., we substitute the P_i from those records showing intermediate levels of illness in equations (9) . The resulting equations can then be used as first approximations to obtain a measure of disease severity for each case record, now using them all. These in turn can now be used to obtain final values for the parameters in equations (13), by a process of successive approximation.

XIII. REMARKS. Four possible mathematical models for the scaling of disease severity were enumerated at the end of Section II. Data from a controlled experiment on 144 rats were shown to conform closely to the second model which postulates a strict scaling of the symptoms but provides for instrumental and observational error. The theory for model C was developed in Sections VIII to XII.

As indicated earlier (Section VII) the analysis showed that the results of this experiment were clear-cut and illuminating, in constrast to the informed but subjective impressions of the experimenters. The latter were, of course, cognizant of a variety of deviations from the strict experimental protocol, and were inclined to see in these an explanation for the seeming anomalies in the results. The role of statistics vis a vis clinical judgment in medical practice is currently under discussion (Sarbin [1942] and Meehl [1954]). Perhaps the function of the development of a disease severity index which transcends in importance even its service in day-to-day application, lies in its contribution to the banishment of the mysterious

and the occult from one more corner of medical practice by drawing attention to a new class of statistical regularities and by providing a tool by which, in any given case, the obscuring effect of chance can be minimized.

XIV. LARGER IMPLICATIONS. The study of the problem of scaling disease severity had not progressed far before the analogy with the problem of providing measures of value and utility in economics became obvious. The problems of non-metric factor analysis (3) should be analogous, but have not been fully examined.

In each of these cases, as in the cases of force and electricity in physics, unobservable concepts are developed on the basis of (apparently) more immediate observables. In the last generation (with roots going back at least to Henry James) stress has been laid on the importance of the role of a measurement technique in establishing the existence of a concept (Bridgman [1927]). In this sense, disease severity is necessarily quantitative, since we have provided a means of measuring it. But one does this for the fruits expected to follow from the concept. The anticipated benefits in the case of disease severity are discussed in section II. An elaboration of all implications in the larger context will have to await another occasion.

XV. ACKNOWLEDGMENT. I am indebted to Dr. Eiglesbach, Dr. Guss, Dr. Rooney and their associates for the experimental phases of this work, and to Mr. Kennedy for help in the analysis of the data and in the preparation of the paper.

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

H SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED -=NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				First Experiment									
1	-	-	3	N	N	N	-	-	-	N	N	N	S
2	-	-	3	N	N	N	-	-	-	N	P	N	S
3	-	-	3	N	N	N	-	-	-	N	P	N	S
4	-	-	3	N	N	N	-	-	-	N	P	N	S
5	-	-	3	N	N	N	-	-	-	N	N	N	S
6	-	-	3	N	N	N	-	-	-	N	P	N	S
7	-	-	3	N	N	N	-	-	-	N	P	P	S
8	-	-	3	N	N	N	-	-	-	N	P	P	S
9	-	-	3	N	N	N	-	-	-	N	N	N	S
10	-	-	3	N	N	N	-	-	-	N	P	P	S
11	-	-	3	N	N	N	-	-	-	N	P	N	S
12	-	-	3	N	N	N	-	-	-	N	P	N	S

KEY*

1. (A) Spleen (B) Lung, (C) Heart Blood. Readings of Cultures.
2. (D) Spleen, (E) Lung, (F) Liver. Observations of Gross Pathology.
3. (G) Titer of Serum Agglutinins. Titer of 1:40 or more - P.
4. (H) Maximum Body Temperature of 100°F or more -P.
5. (I) Maximum Body Temperature of 101°F or more -P.
6. (L) S = Survival to and of experimental period of 16 days. Figure(s) in Column (L) correspond to day of death after treatment.

TABLE 1 (PAGE 2)

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N=SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED -=NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				First Experiment									
13	-	-	30	N	N	N	-	-	-	N	N	N	S
14	-	-	30	N	N	N	-	-	-	P	P	P	S
15	-	-	30	N	N	N	-	-	-	N	P	P	S
16	-	-	30	N	N	N	-	-	-	N	P	P	S
17	-	-	30	N	N	N	-	-	-	N	N	N	S
18	-	-	30	P	P	P	-	-	-	-	P	P	4
19	-	-	30	N	N	N	-	-	-	N	P	P	S
21	-	-	30	N	N	N	-	-	-	-	P	N	S
22	-	-	30	N	N	N	-	-	-	N	P	P	S
24	-	-	30	N	N	N	-	-	-	P	P	P	S
25	-	-	3000	-	-	-	-	-	-	P	N	N	S
26	-	-	3000	N	N	N	-	-	-	P	P	P	S
27	-	-	3000	P	P	N	-	-	-	P	P	P	9
28	-	-	3000	N	N	N	-	-	-	-	P	P	S
29	-	-	3000	P	N	P	-	-	-	P	P	P	S
30	-	-	3000	P	P	P	-	-	-	-	P	P	S
31	-	-	3000	N	N	N	-	-	-	P	P	P	S
32	-	-	3000	N	N	N	-	-	-	P	P	P	S
34	-	-	3000	P	N	N	-	-	-	P	P	P	S
35	-	-	3000	-	-	-	-	-	-	N	P	P	S
36	-	-	3000	P	P	P	-	-	-	-	P	P	15

*For key see page 1 of this table.

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N=SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED -=NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms*									Survival* or Death (L)
				A	B	C	D	E	F	G	H	I	
				Second Experiment									
1	M	315	15-90	P	P	P	N	N	N	-	P	P	8
2	M	300	15-90	P	P	P	N	N	N	-	P	P	10
3	M	270	15-90	P	P	P	N	N	N	P	P	P	S
4	M	240	15-90	P	P	P	P	N	N	-	P	P	7
5	M	330	15-90	N	N	N	N	N	N	P	P	P	S
6	M	280	15-90	N	N	N	N	N	N	P	P	P	S
7	M	300	15-90	N	N	N	N	N	N	P	P	P	S
8	M	250	15-90	N	N	N	N	N	N	P	P	P	S
9	M	335	16300	P	P	P	P	P	N	-	P	P	7
10	M	305	16300	P	P	N	P	N	N	-	P	P	12
11	M	280	16300	N	P	N	N	N	N	-	P	P	9
12	M	275	16300	P	P	P	P	N	N	-	P	P	6
13	M	305	16300	P	P	P	P	P	N	-	P	P	8
14	M	290	16300	P	P	P	P	P	N	-	P	P	8
15	M	285	16300	P	P	P	P	P	N	-	P	P	8
16	M	292	16300	N	P	P	P	P	N	-	P	P	6
17	F	255	16300	P	P	P	N	N	N	-	P	P	7
18	F	215	16300	P	P	P	P	P	N	-	P	P	8
19	F	215	16300	P	P	P	P	P	N	-	P	P	6

*For key see page 1 of this table.

TABLE 1 (PAGE 4)

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N=SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED -=NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				Second Experiment									
20	F	235	16300	P	P	P	P	P	N	-	P	N	9
21	M	306	2.37×10^6	P	N	N	P	N	N	-	P	P	4
22	M	315	2.37×10^6	P	N	N	P	N	N	-	P	P	4
23	M	285	2.37×10^6	P	P	P	P	N	N	-	P	N	4
24	M	280	2.37×10^6	N	P	P	N	N	N	-	P	P	4
25	M	305	2.37×10^6	P	P	P	P	N	N	-	P	P	4
26	M	290	2.37×10^6	P	P	P	N	P	N	-	P	P	5
27	M	290	2.37×10^6	P	P	P	N	N	N	-	P	P	6
28	M	295	2.37×10^6	P	P	P	N	N	N	-	P	P	5
29	F	225	2.37×10^6	P	P	P	P	N	N	-	P	N	4
30	F	230	2.37×10^6	P	P	P	P	N	N	-	P	P	2
31	F	210	2.37×10^6	P	N	P	P	N	N	-	P	P	4
32	F	240	2.37×10^6	P	P	P	P	N	N	-	N	N	3
				Third Experiment									
1	M	265	11	N	N	N	P	N	N	N	P	N	S
2	M	330	11	N	N	N	N	N	N	N	P	N	S
3	M	250	11	N	N	N	N	N	N	N	P	P	S
4	M	260	11	P	P	P	N	P	P	-	P	P	7
5	M	270	11	P	P	P	P	N	N	-	P	P	6
6	M	350	11	N	N	N	N	N	N	N	P	N	S

*For key see page 1 of this table.

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N=SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED - =NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				Third Experiment									
7	M	360	11	N	N	N	N	N	N	N	N	N	S
8	M	235	11	N	N	N	N	N	N	N	N	N	S
9	M	310	11	N	N	N	N	N	N	N	N	N	S
10	M	360	11	N	N	N	N	N	N	F	P	P	S
11	M	250	11	P	P	N	N	N	N	P	P	P	S
12	M	260	11	N	N	N	N	N	N	N	N	N	S
13	F	220	11	P	N	N	N	N	N	P	P	P	S
14	F	220	11	N	N	N	N	N	N	N	P	P	S
15	F	235	11	N	N	N	N	N	N	N	P	N	S
16	F	240	11	P	N	N	N	N	N	N	P	P	S
17	M	335	20	N	N	N	N	N	N	N	P	N	S
18	M	280	20	N	N	N	N	N	N	N	P	P	S
19	M	360	20	N	N	N	N	N	N	N	P	P	S
20	M	240	20	N	N	N	N	N	N	N	P	P	S
21	M	350	20	N	N	N	N	N	N	N	P	N	S
22	M	360	20	N	N	N	N	N	N	N	P	P	S
23	M	240	20	N	N	N	N	N	N	N	P	P	S
24	M	255	20	N	N	N	N	N	N	N	P	P	S
25	M	340	20	N	N	N	N	N	N	N	N	N	S
26	M	300	20	N	N	N	N	N	N	N	P	P	S

*For key see page 1 of this table.

TABLE 1 (PAGE 6)

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N=SYMPTOM NOT EXHIBITED P=SYMPTOM EXHIBITED -=NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms*									Survival* or Death (L)
				A	B	C	D	E	F	G	H	I	
				Third Experiment									
27	M	265	20	N	N	N	N	N	N	N	P	P	S
28	M	230	20	N	N	N	N	N	N	N	P	P	S
29	F	255	20	N	N	N	N	N	N	N	P	P	S
30	F	248	20	N	N	N	N	N	N	N	P	P	S
31	F	275	20	N	P	N	P	-	-	P	P	P	S
32	F	235	20	N	N	N	N	N	N	N	P	P	S
33	M	285	36	N	N	N	N	N	N	N	P	P	S
34	M	260	36	P	P	P	P	P	P	-	P	P	S
35	M	335	36	P	N	N	N	N	N	P	P	P	S
36	M	335	36	N	P	N	N	N	N	N	P	N	S
37	M	275	36	N	N	N	N	N	N	N	P	P	S
38	M	255	36	N	N	N	N	N	N	N	P	P	S
40	M	350	36	N	N	N	N	N	N	N	P	N	S
41	M	290	36	P	N	P	N	N	N	P	P	P	S
42	M	310	36	N	N	N	N	N	N	N	P	P	S
43	M	290	36	N	P	P	N	N	N	P	P	P	S
44	M	280	36	N	N	N	N	N	N	P	P	P	S
45	F	255	36	P	N	N	N	N	N	P	P	P	S
46	F	230	36	P	P	N	N	N	N	P	P	P	S

*For key see page 1 of this table.

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N = SYMPTOM NOT EXHIBITED P = SYMPTOM EXHIBITED - = NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				Third Experiment									
47	F	230	36	P	P	N	N	N	N	P	P	P	S
48	F	260	36	N	N	N	N	N	N	N	P	P	S
49	M	350	111	P	P	P	P	P	N	-	P	P	5
50	M	375	111	N	N	N	N	N	N	N	P	P	S
51	M	240	111	P	P	N	N	N	N	P	P	N	S
52	M	255	111	N	N	N	N	N	N	N	P	N	S
53	M	285	111	P	P	N	N	N	N	P	P	P	S
54	M	285	111	N	N	N	N	N	N	N	P	N	S
55	M	325	111	N	N	N	N	N	N	N	P	N	S
56	M	380	111	N	N	N	N	N	N	N	P	P	S
57	M	285	111	P	P	N	N	N	N	P	P	P	S
58	M	270	111	N	N	N	N	N	N	N	P	P	S
59	M	300	111	P	P	N	N	N	N	P	P	P	S
60	M	355	111	P	P	P	N	N	N	P	P	P	S
61	M	260	111	P	P	P	N	N	N	P	P	P	S
62	M	235	111	N	N	N	N	N	N	N	P	N	S
63	M	270	111	P	P	P	N	N	N	P	P	P	S
64	M	285	1967	N	N	P	N	N	N	P	P	P	S

*For key see page 1 of this table.

RECORDS OF INDIVIDUAL RATS EXPOSED TO AEROSOLIZED PASTEURELIA TULARENSIS IN THREE SEPARATE EXPERIMENTS

N = SYMPTOM NOT EXHIBITED P = SYMPTOM EXHIBITED - = NO OBSERVATION MADE

Rat No.	Sex	Weight (gms)	No. Organisms Inhaled	Symptoms.* A	B	C	D	E	F	G	H	I	Survival* or Death (L)
				Third Experiment									
66	M	380	1967	P	P	N	N	N	N	P	P	P	S
67	M	280	1967	P	P	P	N	N	N	P	P	P	S
68	M	240	1967	P	P	P	P	P	-	-	P	P	11
69	M	285	1967	P	P	N	N	P	P	-	P	P	7
70	M	305	1967	P	P	P	P	P	-	-	P	P	11
71	M	253	1967	P	P	P	P	P	P	-	P	P	8
72	M	285	1967	P	P	P	N	P	N	-	P	P	7
73	M	305	1967	P	P	P	P	P	-	-	P	P	11
74	M	265	1967	P	P	P	N	N	N	-	P	P	S
75	M	290	1967	P	P	P	N	N	N	P	P	P	S
76	M	275	1967	P	P	P	P	P	P	-	P	P	9
77	M	240	1967	P	P	P	P	P	N	-	P	P	7
78	M	265	1967	P	P	P	P	P	P	-	P	P	12
79	M	295	1967	P	N	N	N	N	N	P	P	P	S
80	M	280	1967	P	P	N	N	N	N	P	P	P	S

*For key see page 1 of this table.

TABLE 2 (PAGE 1)

COMBINATIONS OF PAIRS OF DISEASE SYMPTOMS

The Numbers in Each Cell of the Following 2 x 2 Tables are the Totals Which Correspond to the Indicated Combinations.*

		Culture of Lung	
		b	B
Culture of Spleen	a	71	6
	A	11	54

		Culture of Heart Blood	
		c	C
Culture of Spleen	a	73	4
	A	20	45

		Spleen (Gross Pathology)	
		d	D
Culture of Spleen	a	49	3
	A	31	29

		Lung (Gross Pathology)	
		e	E
Culture of Spleen	a	50	1
	A	39	20

		Liver (Gross Pathology)	
		f	F
Culture of Spleen	a	51	0
	A	50	6

		Agglutination Titer	
		g	G
Culture of Spleen	a	59	14
	A	1	22

*Lower case letters denote symptom not exhibited.
Upper case letters denote symptom exhibited.

TABLE 2 (PAGE 2)

		Temperature 100°F or higher	
		h	H
Culture of Spleen	a	10	67
	A	1	64

		Temperature 101°F or higher	
		i	I
Culture of Spleen	a	29	48
	A	5	60

		Culture of Heart Blood	
		c	C
Culture of Lung	b	78	4
	B	15	45

		Spleen (Gross Pathology)	
		d	D
Culture of Lung	b	51	4
	B	29	28

		Lung (Gross Pathology)	
		e	E
Culture of Lung	b	55	0
	B	34	21

		Liver (Gross Pathology)	
		f	F
Culture of Lung	b	55	0
	B	46	6

		Agglutination Titer	
		g	G
Culture of Lung	b	59	19
	B	1	17

		Temperature 100°F or higher	
		h	H
Culture of Lung	b	10	72
	B	1	59

		Temperature 101°F or higher	
		i	I
Culture of Lung	b	28	54
	B	6	54

		Lung (Gross Pathology)	
		e	E
Culture of Heart Blood	c	64	1
	C	25	20

		Agglutination Titer	
		g	G
Culture of Heart Blood	c	60	26
	C	0	10

		Temperature 101°F or higher	
		i	I
Culture of Heart Blood	c	30	63
	C	4	45

		Spleen (Gross Pathology)	
		d	D
Culture of Heart Blood	c	61	5
	C	19	27

		Liver (Gross Pathology)	
		f	F
Culture of Heart Blood	c	64	1
	C	37	5

		Temperature 100°F or higher	
		h	H
Culture of Heart Blood	c	10	83
	C	1	48

		Lung (Gross Pathology)	
		e	E
Spleen(Gross Pathology)	d	77	4
	D	13	17

		Liver (Gross Pathology) f	Liver (Gross Pathology) F			Agglutination Titer g	Agglutination Titer G
Spleen (Gross Pathology)	d	79	2	Spleen (Gross Pathology	d	41	27
	D	23	4		D	1	1

		Temperature 100°F or higher h	Temperature 100°F or higher H			Temperature 101°F or higher i	Temperature 101°F or higher I
Spleen (Gross Pathology)	d	5	75	Spleen (Gross Pathology	d	17	63
	D	1	31		D	5	27

		Liver (Gross Pathology) f	Liver (Gross Pathology) F			Agglutination Titer g	Agglutination Titer G
Lung (Gross Pathology)	e	90	0	Lung (Gross Pathology)	e	42	27
	E	12	6		E	0	0

		Temperature 100°F or higher h	Temperature 100°F or higher H			Temperature 101°F or higher i	Temperature 101°F or higher I
Lung (Gross Pathology)	e	6	83	Lung (Gross Pathology)	e	21	68
	E	0	21		E	1	20

		Agglutination Titer	
		g	G
Liver (Gross Pathology)	f	42	27
	F	0	0

		Temperature 100°F or higher	
		h	H
Liver (Gross Pathology)	f	6	95
	F	0	6

		Temperature 101°F or higher	
		i	I
Liver (Gross Pathology)	f	22	79
	F	0	6

		Temperature 100°F or higher	
		h	H
Agglutination Titer	g	10	51
	G	1	36

		Temperature 101°F or higher	
		i	I
Agglutination Titer	g	28	33
	G	2	35

TABLE 2 (PAGE 6)

		Death	
		1	L
Culture of Spleen	a	74	3
	A	24	41

		Death	
		1	L
Culture of Lung	b	79	3
	B	18	41

		Death	
		1	L
Culture of Heart Blood	c	87	6
	C	11	38

		Death	
		1	L
Spleen (Gross Pathology)	d	69	12
	D	2	29

		Death	
		1	L
Lung (Gross Pathology)	e	70	19
	E	0	21

		Death	
		1	L
Liver (Gross Pathology)	f	70	31
	F	0	6

		Death	
		1	L
Agglutination Titer	g	61	0
	G	36	1

		Death	
		1	L
Temperature 100°F or higher	h	11	1
	H	89	43

		Death	
		1	L
Temperature 101°F or higher	i	31	4
	I	69	40

REFERENCES

Bridgman, Percy W. [1927]. The logic of modern physics. Mac Millian.

Chassan, J. B. [1960]. On a test for order. Biometrics, Vol 16, No 1, March, 119-121.

Coombs, Clyde H. [1958]. An application of a nonmetric model for multi-dimensional analysis of similarities. Psychological Reports, 4, 511-518. Southern Universities Press.

Coombs, Clyde H. [1958]. Paper before the Industrial Mathematics Session, AAAS Annual Meeting, Washington, D. C.

Finney, David [1947]. Probit analysis. Cambridge University Press.

Guttman, L. in Stauffer, et al [1950]. Measurement and prediction. Princeton University Press.

Lazarsfeld, P. F. in Stauffer, et al [1950]. Measurement and prediction. Princeton University Press.

Maloney, Clifford J. [1959]. Disease severity quantitation. Paper read at annual meeting of American Statistical Association, Washington, D. C., December.

Maloney, Clifford J. [1957]. Mathematical developments in the theory of human lethal dose. Annual meeting of the Institute of Mathematical Statistics, Atlantic City, N. J.

Meehl, Paul E. [1954]. Clinical versus statistical prediction. University of Minnesota Press.

Sarbin, R. R. [1942]. A contribution to the study of actuarial and individual methods of prediction. American Journal Sociology 48: 593-602.

Torgerson, Warren S. [1958]. Theory and methods of scaling. Wiley.

STATISTICAL STUDIES OF PLAQUE RESULTS IN VIRUS ASSAY

Francis M. Wadley and Walter D. Foster
U. S. Chemical Corps, Fort Detrick, Frederick, Maryland

Animal disease viruses have not been successfully grown on artificial media, and their evaluation presents many difficulties. Virus preparations are commonly assayed by injection of each of several suitable dilutions into a group of experimental animals. The proportions succumbing to the dilutions are analyzed by dosage mortality methods, and the dilution to which 50 percent (or the average animal) will succumb is estimated. This indicates the concentration of the original undiluted material, measured in animal units. The method is laborious and not very precise and involves waiting some days for effects to show. Wadley (Proc. 3rd Army Exp. Design Conf., 1957) finds that with 40 to 60 animals precision is not better than 1 1/2 or 2 fold; whereas with a moderate number of bacterial plates, estimates within 10 o/o can be made. With the best technique an occasional failure occurs with 40 to 60 animals per assay.

The plaque method consists in preparing plates with monolayer tissue culture preparations, and after 48 hours incubation, counting the dead spots or "plaques" which appear in the tissue. Lennett (Jour. Publ. Health 31 (4), 1961) reviews the technique. It is hoped that each living virus particle will produce a plaque, and evaluation can be made on the same basis as with bacterial plate counts. If this method can be made to yield stable results, a considerable gain will be made.

The plates are made from several dilutions as with bacteria, and the most successful dilution is counted. Plaques per plate are fewer than bacterial colonies; under good conditions 25 to 50 per plate are secured. It is obvious that great care is needed with this relatively new technique. Even with every care taken there is some risk of failure. Tissues die, plaques unaccountably fail to appear, or contamination occurs; plaques may even be "too numerous to count." However, some successful assays are being made. The technique should improve with practice, whereas the animal injection technique has probably been pushed to as precise a point as is feasible, since it has long been practiced.

A detailed experiment, by George Scott of Fort Detrick, in assay of slurries of an encephalomyelitis virus preparation, has been studied. The original slurry was made up in several dilutions, and the dilutions assayed several times by both animal injection (using mice), and plaque

assay using chick fibroblasts. Logs of dilution were 0, 2, 4, 6, 7. The first result noted was surprising and somewhat disappointing; plaque-forming units ("PFU") were consistently fewer than mouse units. This obviously indicates that not every particle formed a plaque, as had been hoped. It seems that a "PFU" requires more particles than a mouse unit, or that not every particle is capable of forming a plaque. Mr. Scott calculated the concentrations in log no. PFU's per ml. They are shown with the comparable figures derived by dosage-mortality methods from mouse assays. Sixty-four mice per trial were used.

TABLE I

Mouse and Plaque Results, Averages

Log Dilution	0	2	4	6	7
Log Mouse units/ml	9.57	7.20	5.33	3.52	2.47
Log PFU/ml	8.84	6.70	4.37	2.46	1.28

Even if the exact measure is not the same, plaque results if consistent can be used in various virus studies; hence further studies are made of these results.

In making his determinations Mr. Scott used every available plate for averaging, even some with only 2 or 3 plaques. In subsequent analysis, the authors have used only the most successful dilution. In each unit 2 or 3 plates with favorable numbers were used. One of two technicians was more successful than the other, and most of the results are his.

However, for one preliminary study, it seemed desirable to use all possible plates. The variation between plates within a group is of much interest. The counts were arbitrarily grouped as shown below. Variances in each group of 2 or 3 plates were averaged and compared with the mean of the plates. These individual variances and means were averaged and compared with the mean of the plates. These individual variances and means were averaged for each size group.

TABLE II

Average Mean and Variance of Plates Within a Unit

Size Group, Plaques Per Plate	Degrees of Freedom	Average Variance	Average Mean
0 - 5	64	4.09	2.86
5 - 20	71	16.96	10.37
20 - 50	22	46.43	26.98

Thus the variance of plates within group, for plaques, was but little higher than Poisson expectation, and compares favorably in this respect with bacterial plates. Of course, the greater numbers obtainable on bacterial plates give some advantage in precision over the plaques.

For further analysis only the best set of plates in each trial was used. There were 5 treatments (see Table I) in 6 trials, and 30 units should be available. The mouse assay was unusually poor; only 18 units out of 30 gave standard dosage-mortality curves, while 4 more could be estimated by interpolation. Thus, 22 out of 30 trials are available; the plaques gave readings in 25 out of 30 trials.

These gaps in the data made a symmetrical analysis (5 dilutions by 6 replications) impracticable. Hence analysis was carried out first for between and within replications, then with the same data for between and within dilutions. Logarithms of estimated number of units for each were used in analysis. Dilution showed large differences as expected. When the log dilution factor was added to the figures to bring them all to the same basis, no significant differences were found. With this procedure the only difference to be expected among dilutions would be some secondary effect caused by manipulation.

TABLE III

Analyses of Variance of Log No. Units per ml.

Source of Variation	Mice		Plaques	
	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square
Between Replication	5	0.2596	5	0.0887
Within Replications	16	0.0882	19	0.1114
Total	21	0.1290	24	0.1067

There is no indication of pronounced differences among replications, dilutions, or between mice and plaques in variance, although with mice replications reached the margin of significance. The latter fact may be due to day to day variation in mouse susceptibility. Over-all mean squares are: mice 0.1290, Plaques 0.1067.

This approximate equality of precision is not what would be predicted from the low value of variance for plates within groups. Plaque plates show variance similar to bacterial plates, and should thus give much more precise results than the mouse injection assay. For 20 to 50 plates, the variance is about 1.7 times the mean, and for a favorable condition with 30 plaques per plate, a variance of about 51 would be expected. This may be transformed into expected variance of log count by the approximation of the "propagation of error" (Deming, "Statistical Adjustment of Data" N. Y., 1943). If $y = f(X)$, the variance of y may be estimated as $\left(\frac{dy}{dx}\right)^2 \cdot$ variance X, or 0.0107. Subsequent operations, in going to log no. units per ml, will involve only addition of constants and will not alter the variance. For the mouse assay, a simplification of the formula for internal variance of log LD-50 is $1/b^2$ SNW where b is probit regression, about 1.5 in this material; n is number of mice per point; and w is probit weight, which usually averages about 0.4. This may be further simplified into 1/0.9N, where N is total number of mice. For 64 mice this comes out 0.0174, as a variance under optimum conditions.

These approximations would lead us to think a single plate a little better than 100 mice on a within-trial basis. However, in the analysis of

variance between trials, 64 mice give about the same precision as 2 or 3 plates. This suggests a larger between-trial component of variance for plaques than for mice. Since this is the preferred basis, an effort is made to arrange the experimental material to study this question in more detail.

Each individual plate count was converted to an estimate of log no. units per ml., and analysis was made of variance of groups and of plates within groups. For the mice no such division is possible, but an estimate of internal variance in each trial is given by the computed variance of log LD50. These variances were averaged for the 18 trials that permitted calculation. They were higher than the 0.0174 estimated as a theoretical minimum above, though some approach it; a few large variances bring the average up.

TABLE IV

Internal and External Variances

	Mice		Plaques	
Source of Variation	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square
Between Trials	21	0.1290	24	0.1067
Within Trials	36	0.0996	43	0.0057*

* This is stepped down to mean value for comparison with variance of group means; the value on an individual basis is 0.0156.

Thus it appears that plaques varied much less within trials than mice, but showed a larger between-trial component of variance.

Work with the plaques on this agent is being continued, and has been carried out as a supplement to several recent experiments. Recent results appear somewhat like those already discussed. Plates as a rule are very close together within a trial, and assay values are lower than those from the mouse assays. In one recent experiment, however, plaques came near mice in assay value. Plaques have given fairly stable results in assaying stock preparations of "Controls", but have not so far been successful in assaying aerosol samples. This is disappointing, but work is continuing in efforts to solve the difficulty.

To sum up; plaque technique has been carried out for this encephalomyelitis virus with fair success. Plates within a group vary about as much as bacterial plates, promising a gain in precision over animal assay, but a large between-trial component of variance is found. Plaques give lower assay values than animals, and have not yet been successful in assaying aerosol samples. It is hoped that these weaknesses can be improved with further work.

A CONFIDENCE INTERVAL FOR THE RELIABILITY OF MULTI-COMPONENT SYSTEMS

John K. Abraham
Weapons System Laboratory, Ballistic Research Laboratories,
Aberdeen Proving Ground

INTRODUCTION. The reliability, P, of a device may be defined simply as the probability that it functions correctly under specified conditions. For many devices an interval estimate of this number P is desired. It is sometimes impractical, if not impossible, to test many assembled devices, and the suggestion has been made that parts be tested separately, and the results combined to form an interval estimate. The present study indicates a means of constructing an interval for P with any desired confidence coefficient, preferably, however, greater than .90. As is the case with all non-Neyman-shortest confidence intervals based on discrete variables, the actual probability of coverage will generally be somewhat higher than the selected coefficient.

For any given device with k distinct parts, P may be written in terms of $q_1, \ldots, q_k$, where q_i denotes the probability of the i^{th} part failing. The parts are assumed to function independently of one another within the device in the same way and under the same conditions as when tested separately. The problem considered here is that of finding a confidence interval for P, given that n_i Bernoulli trials have been conducted on the i^{th} part (i=1, . . ., k) and that X_i failures have been observed in the n_i trials.

The approach of this study is to consider the problem first for the simple series case, and then extend the results to the general case.

For a simple series device consisting of k parts, P (expressed in terms of the q_i) can be minimized and maximized under the restriction $\sum n_i q_i = \xi$ (Throughout this and the following sections, the range of all sums and products is assumed to extend from i equals 1 to k unless otherwise indicated). Thus P will be bounded on either side by functions of ξ. Since $X = \sum X_i$ has expectation ξ and is approximately Poisson (or binomially) distributed, one can easily find a confidence interval for ξ based on X, and by appropriate computations, for the two bounding functions of ξ, and hence for P.

The extension from the simple series to the more complicated devices follows from factoring P into its series and non-series parts. Both parts can be parametrically bounded, the series part as outlined above and the non-series part using several approaches. From these bounds one can write upper and lower bounds for P, and by using several independent confidence intervals, a confidence interval for P may be obtained.

Numerical examples are given which illustrate the procedure and provide comparisons with the results of some other currently used methods.

The reader who wishes to apply the results must be willing to perform some numerical computations. Situations undoubtedly exist for which the present approach is of little use, and in many other cases the experimenter's supply of ingenuity may be heavily taxed to provide shortcuts peculiar to the problem at hand. The present discussion is designed to suggest several solutions to the problem, and to urge the reader to choose the best one available. Most likely this choice cannot be made without first performing trial and error calculations.

I. PARAMETRIC BOUNDS FOR P IN THE SIMPLE SERIES CASE. A very simple case is a device consisting of k different parts connected in series. Assuming that the parts function independently, the reliability P equals $\prod (1-q_i)$ where q_i equals the probability of the i^{th} part failing. We find (for reasons which may not yet be obvious) the range of possible values that $\prod (1-q_i)$ can assume given that $\sum n_i q_i$ equals ξ, where the n_i are positive integers, and ξ is a positive fixed number.

The problem of finding the range of P given ξ can be solved several ways, perhaps most easily by simple trial and error calculations. For known n_i and fixed ξ, the minimum and maximum values of P, expressed in terms of functions of ξ and the n_i, may be found by numerical trial and error of various values of the q_i. When P is near one (say greater than one half), it is clear from the algebraic expansion of $\prod (1-q_i)$ that $1 - \sum q_i$ will dominate, and it follows that the minimum and maximum of P (at worst, approximately) will be attained when $\sum q_i$, subject to $\sum n_i q_i = \xi$, is respectively maximized and minimized.

For example, if $k=3$, $n_1=500$, $n_2=250$, $n_3=300$, $\xi=1.5$, the condition is $500q_1 + 250q_2 + 300q_3 = 1.5$, and $q_1 + q_2 + q_3$ is minimized when $q_1 = 1.5/500$, $q_2 = q_3 = 0$, and maximized when $q_1 = q_3 = 0$, $q_2=1.5/250$. A little arithmetic shows that P is indeed maximized and minimized, respectively, at these points. Similar results easily may be seen to hold for all values of ξ less than 250. Hence for any fixed value of ξ in this range,

$$1 - \xi/250 \leq P \leq 1 - \xi/500$$

and P may assume any value in this interval, which has width $\xi/500$.

The general form of the bounds for P can be somewhat complicated, depending on the values of the n_i given. However, the following two general cases are of greatest interest:

A. If $n_i = \bar{n} = \sum n_i/k$ for $i=1, \ldots, k$,

$$1 - \xi/\bar{n} \leq P \leq (1 - \xi/\bar{n}k)^k$$

B. If $0 < \sum n_i - k\min_i n_i$ and $0 < \xi < \min(\min_i n_i, \sum n_i - k\min_i n_i)$,

$$1 - \xi/\min_i n_i \leq P \leq 1 - \xi/\max_i n_i, \text{ where } i=1, \ldots, k.$$

II. SAMPLING AND USE OF THE BOUNDS. Suppose that the results of n_i Bernoulli trials are known for the i^{th} part of a simple series device. If X_i denotes the number of failures observed, X_i is binomially distributed with parameters n_i and q_i. The variable $X = \sum X_i$ does not follow a distribution having a simple form, but it is easy to verify that the expectation and variance of X are ξ, and

$\xi - \sum n_i q_i^2$ respectively. Using the method of Lagrange multipliers, it is also easily shown that given $\xi = \sum n_i q_i$, $\sum n_i q_i^2$ attains its minimum value when $q_i = \xi / \sum n_i$ for all i, and hence the variance is largest when all the q_i's are equal. But under the latter circumstances, X is binomially distributed with parameters $\sum n_i$ and $\xi / \sum n_i$. If X were Poisson distributed with parameter ξ, then the variance would also be ξ. Writing the variance of X under the above binomial assumptions as $V_{max}(X)$, and the variance of X under Poisson assumptions as $V_P(X)$, the following inequality holds for all values of n_i and q_i, $i=1,\ldots k$:

$$V(X) \leq V_{max}(X) \leq V_P(X) \quad \text{or}$$

$$\xi - \sum n_i q_i^2 \leq \xi - \xi^2 / \sum n_i \leq \xi .$$

Thus referring X to either the Poisson or the appropriate binomial distribution, one should not be surprised to find greater probabilities for the extreme tail values than under the true distribution. If so, when using binomial or Poisson confidence intervals with X, one may expect them to be conservative in the sense that the probability of containing the unknown parameter ξ will be at least as high as the confidence coefficient of the binomial or Poisson intervals. In the next section, comparisons will be presented concerning these probabilities.

For the present, assuming that in referring X to the Poisson or binomial confidence intervals one will not be led astray, it is a simple matter to look up one or two-sided confidence limits for P, once the appropriate parametric bounds have been chosen. For example, if the α-level two-sided Poisson or binomial confidence limits for ξ based on X turn out to be t_1, t_2, then whenever $t_1 \leq \xi \leq t_2$ it is also true that (using the case of $n_i = \bar{n}$ here)

$$1 - t_2/\bar{n} \leq 1 - \xi/\bar{n} \leq P \leq (1 - \xi/\bar{n}k)^k \leq (1 - t_1/\bar{n}k)^k$$

and the resulting confidence interval covers P with probability at least $1 - \alpha$. There will be a growth in actual confidence coefficient in addition to that due to "too-conservative" intervals because P is a fixed number between two functions of ξ, and the confidence interval covers both of these functions.

In many cases, the q_i will be known to be close to zero, and the values of n_i will be such that ξ may be safely assumed to be within the range of case B, Section I, or preferably, case A. There is a way of converting all cases into the equal sample size case, which unfortunately throws away part of the available information and subjects the confidence interval for P to additional fluctuation. This will be discussed further in Section IV.

III. THE BINOMIAL AND POISSON APPROXIMATIONS. In this section the distribution of $X = \sum X_i$ is in specific cases compared to the distribution of X when referred to the Poisson or appropriate binomial distribution discussed in Section II. A comparison of the confidence coefficients for ξ under these circumstances is also made.

For the following cases, the density of X has been computed to five decimals:

(1) $n_1 = n_2 = 15$, $n_1q_1 + n_2q_2 = 3$ for $q_1 = 0, .01, .05, .1$ and $.2$ and the Poisson density with parameter 3.

(2) $n_1 = 5$, $n_2 = 15$, $n_1q_1 + n_2q_2 = 3$, for $q_1 = 0, .03, .15, .30$, and $.60$.

TABLE 1

Density of $X_1 + X_2$ when X_i: $B(n_i, q_i)$; $n_1 = n_2 = 15$,

$$n_1 q_1 + n_2 q_2 = 3$$

X	q_1=0,.2	q_1=.01,.19	q_1=.05,.15	q_1=.1	Poisson density
0	.03518	.03646	.04047	.04239	.04979
1	.13194	.13380	.13908	.14130	.14936
2	.23090	.23046	.22868	.22766	.22404
3	.25014	.24741	.23952	.23609	.22404
4	.18760	.18542	.17949	.17707	.16803
5	.10318	.10297	.10248	.10230	.10082
6	.04299	.04389	.04635	.04736	.05041
7	.01382	.01468	.01705	.01804	.02160
8	.00346	.00390	.00520	.00576	.00810
9	.00067	.00083	.00133	.00156	.00270
10	.00010	.00014	.00029	.00036	.00081
11	.00001	.00002	.00005	.00007	.00022
12			.00001	.00001	.00006
13					.00001

TABLE II

Density of $X_1 + X_2$ when X_i: $B(n_i, q_i)$; $n_1 = 5$, $n_2 = 15$,

$$n_1 q_1 + n_2 q_2 = 3$$

X	q_1=0, q_2=.2	q_1=.03, q_2=.19	q_1=q_2=.15	q_1=.3, q_2=.1	q_1=.6, q_2=0
0	.03518	.03640	.03876	.03460	.01024
1	.13194	.13371	.13680	.13183	.07680
2	.23090	.23047	.22934	.23200	.23040
3	.25014	.24753	.24283	.25089	.34560
4	.18760	.18554	.18212	.18711	.25920
5	.10318	.10300	.10285	.10240	.07776
6	.04299	.04386	.04537	.04272	
7	.01382	.01463	.01601	.01394	
8	.00346	.00388	.00459	.00362	
9	.00067	.00083	.00108	.00075	
10	.00010	.00014	.00021	.00013	
11	.00001	.00002	.00004	.00002	

In the two tables, $n_1q_1 + n_2q_2$ is always three, and the frequency functions may be compared directly with the approximating binomial and Poisson frequency functions. In Table I, the maximal variance of **X** occurs when $q_1 = .1$, and in Table II, when $q_1 = .15$. It will be noted in both tables that for the extreme tail values of **X**, the approximating bionomial and Poisson probabilities are too large. It is also apparent that since Prob (X=k) is a continuous function of q_1, that for all values of q_1 (and hence q_2) in both tables, the binomial and Poisson probabilities will be too large in both tails, and for more values of **X** in the upper than in the lower tail. Also, in the tables, the binomial and Poisson tail probabilities are greater than the corresponding true probabilities for nearly the same values of **X** (differing at most by one point), regardless of the value of q_1. The actual probabilities in both tails increase as q_1 and q_2 approach equality, although they are always less than the corresponding Poisson probabilities, suggesting that for arbitrary k, the distribution of **X** when all the q_1's are equal is "closer" to the corresponding Poisson distribution than when two q_i's differ. This is not surprising, for in case the q_i's are all equal, **X** is binomially distributed with parameters $\sum n_i$, $\xi / \sum n_i$, and if ξ is small relative to $\sum n_i$, the Poisson approximation to the binomial is known to be very good.

In this optimum case, wherein **X** is binomially distributed, one may easily compare the tabled binomial and Poisson densities. For a single binomial variable **X** having parameters n, $\bar{\xi} = \xi/n$, the tail probabilities have been compared with the corresponding tail probabilities of the Poisson distribution (parameter $n\bar{\xi} = \xi$) in the following cases:

(1) $n\bar{\xi}=1$; n=5, $\bar{\xi}=.2$; n=10, $\bar{\xi}=.1$; n=100, $\bar{\xi}=.01$

(2) $n\bar{\xi}=10$; n=50, $\bar{\xi}=.2$; n=100, $\bar{\xi}=.1$; n=1000, $\bar{\xi}=.01$

(3) $n\bar{\xi}=50$; n=100, $\bar{\xi}=.5$; n=500, $\bar{\xi}=.1$; n=1000, $\bar{\xi}=.05$

In all cases, exactly the same remarks as made for the sum of two binomial variables apply, with the obvious additional observation that an n increases, the tail probabilities of the binomial distribution

approach the corresponding Poisson probabilities. Both the binomial and Poisson distributions are asymptotically normally distributed (i.e., as n and ξ, respectively, increase) and hence if X is large, one may expect the difference between the two sets of confidence limits to be relatively small.

In any particular problem, the preceding indicates that one will achieve narrower confidence bounds for ξ and hence for P, by using the binomial rather than Poisson limits. When ξ is small and $\sum n_i$ is relatively large, the difference between the two approximating distributions becomes quite small. When $\sum n_i$ is large and X turns out to be small, the binomial confidence limits are easily and quite accurately approximated by the Poisson limits. In the following tables and remarks, comparisons are made of the Poisson and binomial intervals for certain small values of $n = \sum n_i$, α, and $X \leq 30$. For large values of n, the Poisson limits will be much closer to the binomial limits than for those values tabled.

These tables have been drawn from three widely available sources:

Tables III and VI are from the binomial graphs and Poisson confidence limit tables in "Biometrika Tables for Statisticians", Volume I, edited by E. S. Pearson and H. O. Hartley. They are based on the equal-tails approach, which chooses, for a given value of X (say c) the values of a Poisson parameter λ, such that, in the two-sided case,

$$\sum_{i=c}^{\infty} \frac{e^{-\underline{\lambda}} \underline{\lambda}^i}{i!} = \frac{\alpha}{2} = \sum_{i=0}^{c} \frac{e^{-\overline{\lambda}} \overline{\lambda}^i}{i!}$$

where $(\underline{\lambda}, \overline{\lambda})$ is to be the α-level confidence interval. The two-sided intervals are tabled for Poisson $X = 0\ (1)\ 30\ (5)\ 50$, $\alpha = .10, .05, .02, .01, .002$, and binomial $X \leq n$, for $n \leq 1000$, $\alpha = .05, .01$.

Tables IV and VII are based on tabled published by Crow and Gardner in Biometrika, as follows:

(1) Crow, E. L. "Confidence Intervals for a Proportion", Biometrika, Vol. 43 (1956), pp 423-435. (For binomial $X \leq n$, n=1 (1) 30, α = .10, .05, .01)

(2) Crow, E. L. and Gardner, R. S. "Confidence Intervals for the Expectation of a Poisson Variable", Biometrika, Vol. 46 (1959), pp 441-453. (For X=0 (1) 300, α = .20, .10, .05, .01, .001) The system used, described in detail in (1), is optimum in a geometrical sense and generally yields bounds of less width than does the equal-tails method.

Tables V and VIII are based on tables published by Colin R. Blyth and David W. Hutchinson in Biometrika, as follows:

(1) "Tables of Neyman-shortest Unbiased Confidence Intervals for the Binomial Parameter", Biometrika, Vol 47 (1960), pp 381-391. (For $X \leq n$, n=2 (1) 24 (2) 50, α = .05, .01)

(2) "Tables of Neyman-shortest Unbiased Confidence Intervals for the Poisson Parameter", Biometrika, Vol. 48 (1961), pp 191-194. (For X=0 (1) 250, α = .05, .01)

The optimum property of these intervals may be described as follows: Among all unbiased α-level confidence intervals, the tabled intervals uniformly minimize the probability of covering false values. An unbiased interval A is defined such that if P_θ denotes probability when the distribution parameter is θ, that

$$P_\theta(\theta \in A) \geq 1 - \alpha \text{ for all } \theta \text{ and } P_\theta(\theta' \in A) \leq P_\theta(\theta \in A) \text{ for all } \theta, \theta'$$

In the binomial portions of each table, the entries are nc_1, nc_2, where c_1, c_2 are the tabled α-level, two-sided confidence limits for p.

When using the Neyman-shortest tables, it is necessary to choose a random number between zero and one and add it to X and then read the appropriate entry.

TABLE III

Two-sided Poisson and Binomial Confidence Limits: Equal Tails, $\alpha = .05$

X	n=10		n=20		n=30		Poisson	
0	.00	3.1	.00	3.4	.00	3.5	.000	3.69
1	.02	4.4	.02	5.0	.06	5.2	.0253	5.57
2	.25	5.6	.25	6.3	.22	6.6	.242	7.22
3	.68	6.5	.65	7.5	.63	8.0	.619	8.77
4	1.2	7.4	1.2	8.7	1.1	9.3	1.09	10.24
5	1.9	8.1	1.7	9.8	1.7	10.4	1.62	11.67
6	2.6	8.8	2.4	10.8	2.3	11.6	2.20	13.06
7	3.5	9.3	3.0	11.8	3.0	12.7	2.81	14.42
8	4.4	9.8	3.8	12.8	3.6	13.7	3.45	15.76
9	5.6	10.0	4.6	13.7	4.4	14.8	4.12	17.08
10	6.9	10.0	5.4	14.6	5.2	15.8	4.80	18.39
11			6.3	15.4	5.9	16.8	5.49	19.68
12			7.2	16.2	6.8	17.8	6.20	20.96
13			8.2	17.0	7.6	18.8	6.92	22.23
14			9.2	17.6	8.4	19.6	7.65	23.49
15			10.2	18.3	9.4	20.6	8.40	24.74
16			11.3	18.8	10.4	21.6	9.15	25.98
17			12.5	19.4	11.2	22.4	9.90	27.22
18			13.7	19.8	12.2	23.2	10.67	28.45
19			15.0	20.0	13.2	24.1	11.44	29.67
20			16.6	20.0	14.2	24.8	12.22	30.89
21					15.2	25.6	13.00	32.10
22					16.3	26.4	13.79	33.31
23					17.3	27.0	14.58	34.51
24					18.4	27.7	15.38	35.71
25					19.6	28.3	16.18	36.90
26					20.7	28.9	16.98	38.10
27					22.0	29.4	17.79	39.28
28					23.4	29.8	18.61	40.47
29					24.8	29.9	19.42	41.65
30					26.5	30.0	20.24	42.83

TABLE IV

Two-sided Limits: Crow and Gardner, $\alpha = .05$

X	n=5		n=10		n=20		n=30		Poisson	
0	.000	2.5	.000	2.67	.000	2.86	.000	3.00	.0	3.285
1	.050	3.28	.05	3.97	.06	4.44	.06	4.89	.051	5.323
2	.380	4.06	.37	6.03	.36	5.86	.36	6.15	.355	6.686
3	.945	4.62	.87	6.19	.84	7.02	.84	7.32	.818	8.102
4	1.72	4.95	1.50	7.33	1.42	8.22	1.41	8.76	1.366	9.598
5	2.50	5.00	2.22	7.78	2.08	9.34	2.04	9.72	1.970	11.177
6			2.67	8.50	2.80	10.7	2.73	10.9	2.613	12.817
7			3.81	9.13	2.86	11.8	3.00	12.1	3.285	13.765
8			3.97	9.63	4.18	13.0	3.93	13.2	3.285	14.921
9			6.03	9.95	4.44	14.1	4.89	14.3	4.460	16.768
10			7.33	10.0	5.86	14.1	5.25	15.7	5.323	17.633
11					5.86	15.6	6.15	16.8	5.323	19.050
12					7.02	15.8	7.08	17.9	6.686	20.335
13					8.22	17.1	7.32	19.1	6.686	21.364
14					9.34	17.2	8.76	20.3	8.102	22.945
15					10.7	17.9	9.72	20.3	8.102	23.762
16					11.8	18.6	9.72	21.2	9.598	25.400
17					13.0	19.2	10.9	22.7	9.598	26.306
18					14.1	19.6	12.1	22.9	11.177	27.735
19					15.6	19.9	13.2	23.8	11.177	28.966
20					17.1	20.0	14.3	24.8	12.817	30.017
21							15.7	25.1	12.817	31.675
22							16.8	26.1	13.765	32.277
23							17.9	27.0	14.921	34.048
24							19.1	27.3	14.921	34.665
25							20.3	28.0	16.768	36.030
26							21.2	28.6	16.77	37.67
27							22.7	29.2	17.63	38.16
28							23.8	29.6	19.05	39.76
29							25.1	29.9	19.05	40.94
30							27.0	30.0	20.33	41.75

TABLE V

Two-sided Neyman-shortest Limits: $\alpha = .05$

X+Y	n=5		n=10		n=20		n=30		Poisson	
.0	.0	.0	.0	.0	.0	.0	.0	.0	.00	.00
.5	.0	2.2	.0	2.5	.0	2.6	.0	2.7	.0	2.8
1.0	.0	2.6	.0	3.0	.0	3.2	.0	3.3	.0	3.5
1.5	.0	3.4	.0	4.0	.0	4.3	.0	4.4	.0	4.7
2.0	.0	3.8	.1	4.4	.0	4.8	.0	5.1	.0	5.4
2.5	.2	4.3	.1	5.2	.1	5.7	.0	5.8	.1	6.4
3.0	.5	4.5	.4	5.6	.4	6.2	.3	6.6	.4	7.1
3.5	.7	4.8	.6	6.3	.5	7.1	.6	7.4	.5	8.0
4.0	1.2	5.0	1.0	6.6	.8	7.6	.9	7.8	.8	8.6
4.5	1.6	5.0	1.2	7.2	1.1	8.3	1.0	8.6	1.0	9.5
5.0	2.4	5.0	1.7	7.5	1.6	8.8	1.5	9.3	1.4	10.1
5.5			2.0	8.0	1.7	9.3	1.6	9.9	1.6	10.9
6.0			2.5	8.3	2.2	9.8	2.1	10.5	2.0	11.5
6.5			2.8	8.8	2.4	10.4	2.4	11.1	2.2	12.3
7.0			3.4	9.0	3.0	10.8	2.7	11.4	2.6	12.9
7.5			3.7	9.4	3.2	11.4	3.0	12.3	2.9	13.7
8.0			4.4	9.6	3.6	11.8	3.6	12.6	3.3	14.3
8.5			4.8	9.9	4.0	12.4	3.9	13.2	3.6	15.0
9.0			5.6	9.9	4.6	12.8	4.2	13.8	3.9	15.6
9.5			6.0	10.0	4.8	13.4	4.5	14.4	4.2	16.3
10.0			7.0	10.0	5.4	13.8	5.1	14.7	4.6	16.9

TABLE V

Two-sided Neyman-shortest Limits: $\alpha = .05$

(Continued)

X+Y	n=5	n=10	n=20		n=30		Poisson	
11			6.2	14.6	6.0	15.9	5.3	18.2
12			7.2	15.4	6.6	16.8	6.0	19.5
13			8.2	16.4	7.5	18.0	6.8	20.8
14			9.2	17.0	8.4	18.9	7.5	22.1
15			10.2	17.8	9.3	19.8	8.2	23.3
16			11.2	18.4	10.2	20.7	9.0	24.6
17			12.4	19.2	11.1	21.6	9.7	25.8
18			13.8	19.6	12.0	22.5	10.5	27.1
19			15.2	20.0	13.2	23.4	11.3	28.3
20			16.8	20.0	14.1	24.0	12.0	29.5
21					15.3	24.9	12.8	30.7
22					16.2	25.8	13.6	31.9
23					17.4	26.4	14.4	33.1
24					18.6	27.3	15.2	34.3
25					19.5	27.9	16.0	35.5
26					20.7	28.5	16.8	36.7
27					22.2	29.1	17.6	37.9
28					23.4	29.7	18.4	39.1
29					24.9	30.0	19.3	40.3
30					26.7	30.0	20.1	41.5

TABLE VI

Two-sided Equal-tails Limits: $\alpha = .01$

X	n=10		n=20		n=30		Poisson	
0	.00	4.1	.00	4.7	.00	4.9	.000	5.30
1	.00	5.4	.00	6.2	.00	6.6	.00501	7.43
2	.12	6.5	.12	7.8	.08	8.2	.103	9.27
3	.37	7.4	.40	8.9	.38	9.6	.338	10.98
4	.77	8.1	.74	10.1	.75	10.9	.672	12.59
5	1.3	8.7	1.1	11.2	1.1	12.2	1.08	14.15
6	1.9	9.2	1.7	12.2	1.6	13.3	1.54	15.66
7	2.6	9.6	2.3	13.1	2.6	14.4	2.04	17.13
8	3.5	9.9	2.9	14.0	2.8	15.4	2.57	18.58
9	4.6	10.0	3.6	14.8	3.4	16.5	3.13	20.00
10	5.9	10.0	4.4	15.7	4.1	17.5	3.72	21.40
11			5.2	16.4	4.7	18.4	4.32	22.78
12			6.0	17.1	5.6	19.4	4.94	24.14
13			6.9	17.7	6.3	20.2	5.58	25.50
14			7.8	18.3	7.1	21.2	6.23	26.84
15			8.8	18.9	8.0	22.1	6.89	28.16
16			9.9	19.3	8.8	22.9	7.57	29.48
17			11.1	19.6	9.8	23.7	8.25	30.79
18			12.2	19.9	10.6	24.4	8.94	32.09
19			13.8	20.0	11.6	25.3	9.64	33.38
20			15.3	20.0	12.5	25.9	10.35	34.67
21					13.5	26.6	11.07	35.95
22					14.6	27.2	11.79	37.22
23					15.6	27.4	12.52	38.48
24					16.7	28.4	13.25	39.74
25					17.8	28.9	14.00	41.00
26					19.1	29.2	14.74	42.25
27					20.4	29.6	15.49	43.50
28					21.8	29.9	16.24	44.74
29					23.4	30.0	17.00	45.98
30					25.1	30.0	17.77	47.21

TABLE VII

Two-sided Limits: Crow and Gardner, $\alpha = .01$

X	n=5		n=10		n=20		n=30		Poisson	
0	.000	3.01	.00	3.12	.00	4.18	.00	4.53	.000	4.771
1	.010	3.89	.01	5.12	.02	5.86	.00	6.18	.010	6.914
2	.165	4.47	.16	6.24	.16	7.50	.15	7.68	.149	8.727
3	.530	4.84	.48	7.03	.46	8.48	.45	9.30	.436	10.473
4	1.11	4.99	.93	7.82	.88	10.0	.84	10.4	.823	12.347
5	1.99	5.00	1.50	8.50	1.38	11.5	1.35	11.6	1.279	13.793
6			2.18	9.07	1.96	12.0	1.89	12.9	1.785	15.277
7			2.97	9.52	2.58	12.7	2.49	14.1	2.330	16.801
8			3.76	9.84	3.26	14.1	3.12	15.2	2.906	18.362
9			4.88	9.99	4.00	14.5	3.81	16.1	3.507	19.462
10			6.24	10.0	4.18	15.8	4.53	17.1	4.130	20.676
11					5.48	16.0	4.53	18.4	4.771	22.042
12					5.86	16.7	5.94	19.6	4.771	23.765
13					7.26	17.4	6.18	20.1	5.829	24.925
14					7.98	18.0	7.47	20.8	6.668	25.992
15					8.48	18.6	7.68	22.3	6.914	27.718
16					10.0	19.1	9.24	22.5	7.756	28.852
17					11.5	19.5	9.87	23.8	8.727	29.900
18					12.5	19.8	10.4	24.1	8.727	31.839
19					14.1	20.0	11.6	25.5	10.009	32.547
20					15.8	20.0	12.9	25.5	10.473	34.183
21							13.9	26.2	11.242	35.204
22							14.8	26.9	12.347	36.544
23							15.9	27.5	12.347	37.819
24							17.1	28.1	13.793	38.939
25							18.4	28.6	13.793	40.373
26							19.6	29.2	15.28	41.39
27							20.7	29.6	15.28	42.85
28							22.3	29.8	16.80	43.91
29							23.8	30.0	16.80	45.26
30							25.5	30.0	18.36	46.50

TABLE VIII

Two-sided Neyman-shortest Limits: $\alpha = .01$

X+Y	n=5		n=10		n=20		n=30		Poisson	
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.5	.0	3.1	.0	3.6	.0	4.0	.0	4.2	.0	4.4
1.0	.0	3.4	.0	4.1	.0	4.6	.0	4.8	.0	5.1
1.5	.0	4.1	.0	5.1	.0	5.7	.0	6.2	.0	6.6
2.0	.0	4.3	.0	5.5	.0	6.2	.0	6.6	.0	7.2
2.5	.0	4.7	.0	6.2	.0	7.3	.0	7.6	.0	8.4
3.0	.2	4.8	.2	6.6	.2	7.8	.3	8.1	.1	9.1
3.5	.3	5.0	.2	7.2	.2	8.5	.3	9.2	.2	10.2
4.0	.7	5.0	.5	7.5	.4	9.0	.6	9.6	.4	10.8
4.5	.9	5.0	.7	8.1	.6	9.7	.6	10.4	.5	11.8
5.0	1.6	5.0	1.1	8.3	1.0	10.2	.9	10.8	.8	12.4
5.5			1.2	8.8	1.1	10.9	1.0	11.6	1.0	13.4
6.0			1.7	8.9	1.4	11.2	1.5	12.0	1.3	14.0
6.5			1.9	9.3	1.6	11.8	1.5	12.9	1.4	14.9
7.0			2.5	9.5	2.0	12.2	2.1	13.2	1.8	15.5
7.5			2.8	9.8	2.2	12.8	2.1	14.1	2.0	16.4
8.0			3.4	9.8	2.8	13.2	2.7	14.4	2.3	17.0
8.5			3.8	10.0	3.0	13.8	2.7	15.0	2.5	17.8
9.0			4.5	10.0	3.4	14.0	3.3	15.6	2.9	18.4
9.5			4.9	10.0	3.8	14.6	3.6	16.2	3.1	19.2
10.0			5.9	10.0	4.2	15.0	3.9	16.5	3.5	19.8

TABLE VIII

Two-sided Neyman-shortest Limits: $\alpha = .01$

X+Y	n=5	n=10	n=20		n=30		Poisson	
11			5.0	15.8	4.8	17.4	4.1	21.2
12			6.0	16.6	5.4	18.6	4.8	22.6
13			6.8	17.2	6.3	19.5	5.4	24.0
14			7.8	18.0	6.9	20.4	6.0	25.3
15			8.8	18.6	7.8	21.3	6.7	26.7
16			9.8	19.0	8.7	22.2	7.4	28.0
17			11.0	19.6	9.6	23.1	8.1	29.3
18			12.2	19.8	10.5	23.7	8.8	30.6
19			13.8	20.0	11.4	24.6	9.5	31.9
20			15.4	20.0	12.6	25.2	10.2	33.2
21					13.5	26.1	10.9	34.5
22					14.4	26.7	11.6	35.8
23					15.6	27.3	12.4	37.0
24					16.8	27.9	13.1	38.3
25					18.0	28.5	13.8	39.6
26					19.2	29.1	14.6	40.8
27					20.4	29.4	15.3	42.1
28					21.9	29.7	16.1	43.3
29					23.4	30.0	16.8	44.6
30					25.2	30.0	17.6	45.8

It is apparent from visually comparing the Poisson limits with the corresponding binomial limits, that except for a few cases (a total of 16), when $\alpha = .05$ or $.01$, the former completely cover the binomial limits for $n = 10, 20, 30$ and in four of the tables, for $n = 5$. In Tables III and VI the only discrepancies occur for $X \leq 2$, where the graphed values, multiplied by n, may be necessarily too imprecise to compare meaningfully with the Poisson readings. In Tables IV and VII, for the discrepancies occuring when $X \leq 2$, again impreciseness of the binomial tables may be reponsible. For the larger values of X the behavior of both intervals is somewhat erratic, and this plus the fact that both the tabled Poisson and binomial intervals generally cover the unknown parameters with probability greater than .95 or .99, indicates that an occasional slight non-inclusion for a particular value of X is not serious. The procedure of referring to the binomial or Poisson limits will still be expected to yeild confidence coefficient very near to, if not greater than $1 - \alpha$.

In all the tables, as n increases, the end-points uniformly expand towards the corresponding values in the Poisson case. For $\alpha = .05$ or $.01$, the fit generally seems to be closer for the Neyman-shortest intervals, which is not surprising since it is the only one of the three methods that exactly attains its confidence coefficient for all parameter values.

For values of X and n other than those tabled, several approximations to the binomial and Poisson distributions exist, a few of which appear in "Biometrika Tables for Statisticians".

As a check on the inclusion properties of the binomial and Poisson approximations one may for specific examples, involving known parameters, compute the true probabilities of covering P using the appropriate binomial or Poisson confidence limits (α-level). If the Poisson limits include the binomial limits, which in turn are "too wide" for P, then one would expect in most cases that the Poisson intervals would have the highest confidence coefficient, followed by the "binomial" confidence coefficient, both of which would be expected to be greater than $1 - \alpha$. Calculations, based on the distribution of X under the various values of q_i appearing in Tables I and II, were made to obtain the probabilities that the Poisson limits and the appropriate binomial limits (α-level) cover $n_1 q_1 + n_2 q_2$. These probabilities are given in Tables IX and X.

TABLE IX

Probability of the Poisson Limits Covering $n_1q_1 + n_2q_2 = 3$

$n_1 = n_2 = 15 \quad \alpha = .05$

Interval	$q_1 = 0$	$q_1 = .01$	$q_1 = .05$	$q_1 = .10$	Poisson
Equal-tails	.99575	.99509	.99312	.99221	.98809
Crow and Gardner	.98193	.98041	.97607	.97417	.96649
Neyman-shortest	.970	.969	.964	.961	.952 (.95)

$n_1 = n_2 = 15 \quad \alpha = .01$

Interval	$q_1 = 0$	$q_1 = .01$	$q_1 = .05$	$q_1 = .10$	Poisson
Equal-tails	.99921	.99899	.99832	.99797	.99619
Crow and Gardner	.99921	.99899	.99832	.99797	.99619
Neyman-shortest	.994	.994	.993	.992	.990 (.99)

$n_1 = 5, n_2 = 15 \quad \alpha = .05$

Interval	$q_1 = 0$	$q_1 = .03$	$q_1 = .15$	$q_1 = .30$	$q_1 = .60$
Equal-tails	.99575	.99514	.99408	.99549	1.00000
Crow and Gardner	.98193	.98051	.97807	.98155	1.00000
Neyman-shortest	.970	.969	.966	.971	.994

$n_1 = 5, n_2 = 15 \quad \alpha = .01$

Interval	$q_1 = 0$	$q_1 = .03$	$q_1 = .15$	$q_1 = .30$	$q_1 = .60$
Equal-tails	.99921	.99902	.99867	.99911	1.00000
Crow and Gardner	.99921	.99902	.99867	.99911	1.00000
Neyman-shortest	.994	.994	.994	.994	.999

TABLE X

Probability of the Binomial Limits Covering $n_1q_1 + n_2q_2 = 3$

$n_1 = n_2 = 15 \quad \alpha = .05$

Interval	$q_1 = 0$	$q_1 = .01$	$q_1 = .05$	$q_1 = .1$
Equal-tails	.99575	.99509	.99312	.99221
Crow and Gardner	.99575	.99509	.99312	.99221
Neyman-shortest	.962	.960	.954	.951 (.95)

$n_1 = n_2 = 15 \quad \alpha = .01$

Interval	$q_1 = 0$	$q_1 = .01$	$q_1 = .05$	$q_1 = .1$
Equal-tails	.99921	.99899	.99832	.99797
Crow and Gardner	.99575	.99509	.99312	.99221
Neyman-shortest	.993	.993	.991	.990 (.99)

$n_1 = 5, n_2 = 15 \quad \alpha = .05$

Interval	$q_1 = 0$	$q_1 = .03$	$q_1 = .15$	$q_1 = .30$	$q_1 = .60$
Equal-tails	.99575	.99514	.99408	.99549	1.00000
Crow and Gardner	.96057	.95874	.95532	.96089	.98976
Neyman-shortest	.956	.953	.949(.95)	.956	.992

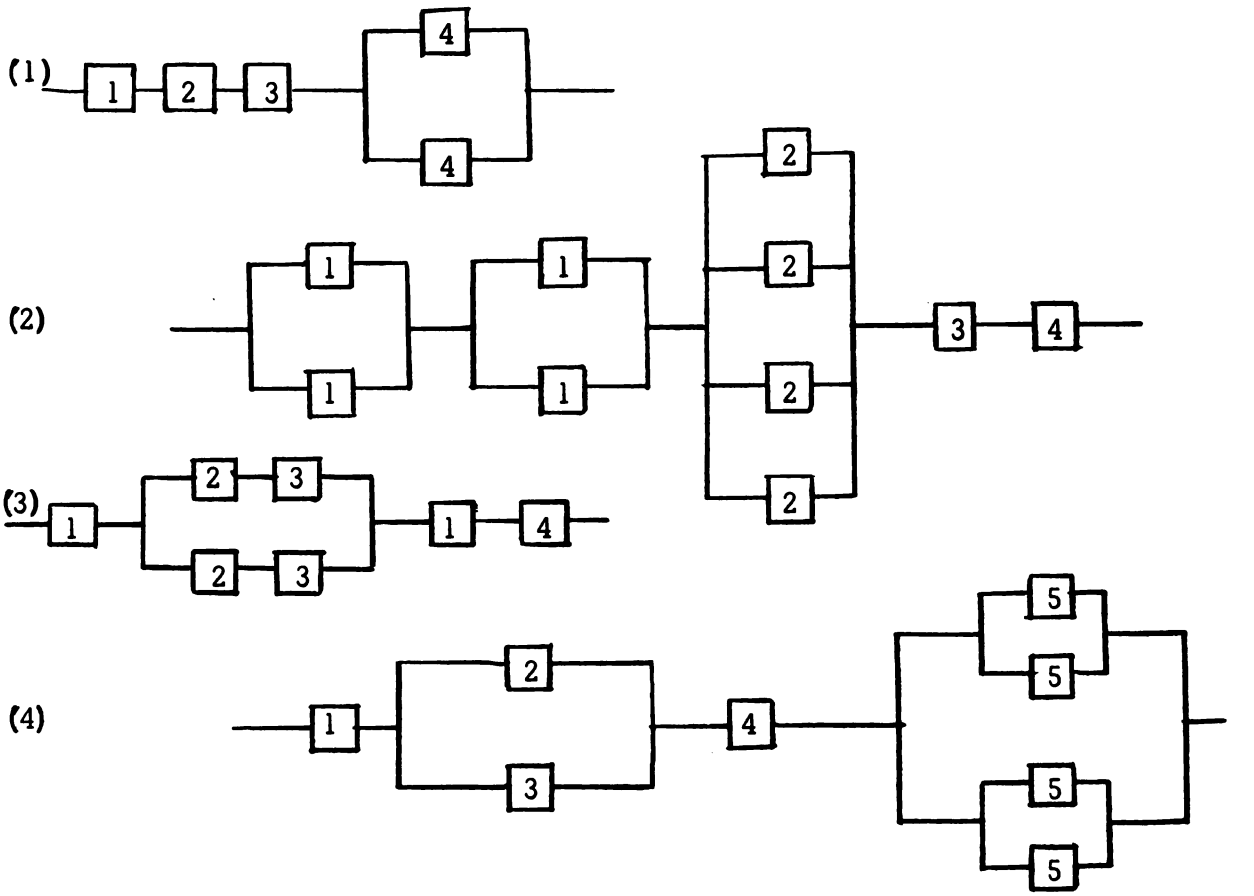
$n_1 = 5, n_2 = 15 \quad \alpha = .01$

Interval	$q_1 = 0$	$q_1 = .03$	$q_1 = .15$	$q_1 = .30$	$q_1 = .60$
Equal-tails	.99921	.99902	.99867	.99911	1.00000
Crow and Gardner	.99575	.99514	.99408	.99549	1.00000
Neyman-shortest	.991	.991	.990(.99)	.991	.998

Again, one may notice that the smallest probabilities occur when $q_1 = q_2 = .1$ and $q_1 = q_2 = .15$, where the variance of X is largest. Also, the Neyman-shortest intervals have the smallest probabilities of all three systems, in most cases, which is due mainly to the exactness of these intervals. Owing to the number of significant decimals published in tables of the latter intervals, the probabilities of coverage may be in error by a few places in the last decimal. For example, in the Poisson case for $\alpha = .05$, $n_1 = n_2 = 15$, the computed probability using linear interpolation turns out to be .952 whereas the true probability is exactly .95. The corresponding probability when $\alpha = .01$ is correst to three decimals.

When confronted with a value of X resulting from lengthy experimental testing, the procedure adopted might be to select from the available tables the limits which are narrowest for the particular value of X observed, and not to choose the bounds with respect to any optimal mathematical property. Such a procedure defines a kind of composite interval about which little may confidently be said concerning the probability of coverage. Preferably, a single type of interval will be decided upon in advance of the experiment, or else the results using different limits will be presented.

IV EXTENSION TO COMPLEX SYSTEMS. Unfortunately, the case of a system consisting solely of non-identical parts in series is relatively rare. More often one is interested in systems having a mixture of series and parallel-connected parts. The following diagrams illustrate some of the possibilities:



The squares containing numbers denote parts of the i^{th} type, and natural subdivisions within systems may be called components. For example, the first figure represents a system of three parts in series and one component of two identical parts connected in parallel.

In the following examples the methods used are those suggested for use in more general models.

In the first illustration, suppose that for the i^{th} part, n trials have been conducted. Then from Section I, $P = \prod_{i=1}^{3} (1-q_i)\,(1-q_4^{\,2})$,

$$\xi = \sum_{i=1}^{3} n_i q_i, \quad \text{and} \quad (1-q_4^2)\,(1-\xi/n) \le P \le (1-\xi/3n)^3(1-q_4^2).$$

Assuming that the numbers X_i are available, it is easy to find two-sided α-level bounds for ξ, and β-level bounds for q_4 based on $X_1 + X_2 + X_3$ and X_4 respectively. These bounds are such that the events, say $c_1 \le \xi \le c_2$, $d_1 \le q_4 \le d_2$ independently occur with probabilities $1-\alpha$, $1-\beta$ respectively. But if both occur at the same time, then

$$1-c_2/n \le \prod_{i=1}^{3} (1-q_i) \le (1-c_1/3n)^3, \quad \text{and} \quad 1-d_2^2 \le 1-q_4^2 \le 1-d_1^2$$

and hence

$$(1-d_2^2)\,(1-c_2/n) \le P \le (1-c_1/3n)^3(1-d_1^2)$$

with probability $(1-\alpha) \times (1-\beta)$. If β is chosen small enough then

$$(1-d_2^2)\,(1-c_2/n), \qquad (1-c_1/3n)^3(1-d_1^2)$$

will form a confidence interval for P with coefficient very nearly $1-\alpha$. As β approaches zero, the interval d_1, d_2 will, of course, widen. But $1-d_1^2$, $1-d_2^2$ will change by a relatively small amount. Hence one may choose β very near to zero, and the resulting confidence interval for P will not be appreciably widened.

It may be noted that one may also lower α a bit in order not to make the bounds so wide, at the expense of a slight widening of the entire interval, to ensure the desired confidence coefficient. One is really free to choose α and β just so long as $(1-\alpha) \times (1-\beta)$ attains the desired number. In this example, the width will be found to suffer least if α is lowered as little as possible, and β is made nearly zero. If $(1-\alpha) \times (1-\beta)$ is slightly below the desired coefficient, the probability of coverage will quite likely be higher, due to the raising of the actual confidence coefficient noted previously.

For the second illustration, $P = \prod_{i=3}^{4} (1-q_i)\,(1-q_2^{\,4})\,(1-q_1^{\,2})^2$. Since $q_2^{\,4}$ will generally be very small, one may be willing to ignore it altogether and consider $P = \prod_{i=3}^{4} (1-q_i)\,(1-q_1^{\,2})^2$, in which case the parametric bounds are (assuming n_i observations for the i^{th} part have been taken, and say

$$\xi < \sum_{i=3}^{4} n_i - 2\min_{i=3,4} n_i \quad \text{and} \quad \min_{i=3,4} n_i,$$

where

$$\xi = n_3 q_3 + n_4 q_4)$$

$$(1-q_1^{\,2})^2 (1-\xi/\min_{i=3,4} n_i) \leq P \leq (1-\xi/\max_{i=3,4} n_i)\,(1-q_1^{\,2})^2.$$

If c_1, c_2 and d_1, d_2 are α and β-level two-sided confidence limits for ξ and q_1 respectively, then

$$(1-d_2^{\,2})^2 (1-c_2/\min n_i), \quad (1-c_1/\max n_i)\,(1-d_1^{\,2})^2$$

are two-sided confidence limits for P with coefficient greater than or equal to $(1-\alpha) \times (1-\beta)$. If one wishes to take account of q_2, then γ-level limits b_1, b_2 for q_2 lead to

$$(1-b_2^4)(1-d_2^2)^2(1-c_2/\min n_i), \quad (1-c_1/\min n_i)(1-d_1^2)^2(1-b_1^4)$$

as the limits for P, with confidence coefficient at least $(1-\alpha) \times (1-\beta) \times (1-\gamma)$, where γ is chosen to be very nearly zero.

In the third illustration, $P = (1-q_1)^2(1-q_4)\left(1-\left[1-(1-q_2)(1-q_3)\right]^2\right)$. In this case there are several directions in which one can proceed, some of which are listed below:

A. One can use the rather primitive inequality

$$(1-q_1)^2(1-q_2^2)(1-q_3)(1-q_4) \leq P \leq (1-q_1)^2(1-q_2^2)(1-q_3^2)(1-q_4)$$

and find four independent confidence limits at low enough levels so that the product of the four confidence coefficients equals $1-\alpha$. The limits will vary considerably in width according to the size of the X_i, n_i, and the levels selected. If the q_i happen to be very close to zero, this may provide narrow limits.

B. If for example $1-\xi/\min_{i=2,3} n_i \leq (1-q_2)(1-q_3) \leq 1-\xi/\max_{i=2,3} n_i$, that is $0 \leq \xi < \min\left(\min_{i=2,3} n_i, \sum_{i=2,3} n_i - 2\min_{i=2,3} n_i\right)$, where $\xi = n_2q_2 + n_3q_3$, then it follows that

$$(1-q_1)^2(1-q_4)\left(1-\left[\xi/\min_{i=2,3} n_i\right]^2\right) \leq P \leq \left(1-\left[\xi/\max_{i=2,3} n_i\right]^2\right)(1-q_1)^2(1-q_4).$$

Using three independent confidence limits, one will in general be able to arrive at narrower confidence bounds for P than by using procedure A.

C. If one is willing to ignore terms in the expansion of P which involve products of three or more q_i, then one may write

$$P = 1 - (2q_1 + q_4) + q_1^2 + 2q_1q_4 - q_2^2 - 2q_2q_3 - q_3^2$$

and hence

$$1 - (2q_1 + q_4) - (q_2 + q_3)^2 \leq P \leq 1 - (2q_1 + q_4) + q_1^2 + 2q_1q_4 .$$

With arbitrary n_i one will have difficulty finding confidence limits for $(2q_1 + q_4)$ and $(q_2 + q_3)$. However in the special case that $n_1 = 2n_4$ and $n_2 = n_3$, the problem can be given a simple solution. For in this case one may refer $X_1 + X_4$ and $X_2 + X_3$ to the binomial or Poisson confidence intervals, using the following argument: X_1 is distributed binomially with parameters n_1 and q_1 and variance $n_1q_1 - n_1q_1^2$. If one refers X_1 to the binomial distribution with parameters $n_1/2$ and $2q_1$, which has variance $n_1q_1 - 2n_1q_1^2 \leq n_1q_1 - n_1q_1^2$, one will be making a conservative approximation in the sense that (as in the case of the Poisson and binomial approximations to X) the tail probabilities will really be smaller than under the approximation. Thus a confidence interval for $2q_1$ based on the approximation, will be expected to be too wide. $X_1 + X_4$ may be referred to as binomial (parameters $n_1/2 + n_2$, $\xi / [n_1/2 + n_2]$) or Poisson (parameter $\xi = (n_1/2)(2q_1) + n_2q_2$) for a confidence interval for ξ. But $n_1 = 2n_4$ means that from any confidence

interval for ξ one also derives one for $2q_1 + q_4$ since $\xi = n_4 (2q_1 + q_4)$ and one may divide the end points through by n_4. Similarly for $q_2 + q_3$, one may derive any level confidence interval desired. If c_1, c_2 are two-sided α-level bounds for $2q_1 + q_4$ and d_1, d_2 are two-sided β-level bounds for $q_2 + q_3$, then the event $c_1 \leq 2q_1 + q_4 \leq c_2$ and $d_1 \leq q_2 + q_3 \leq d_2$ implies $-(q_2 + q_3)^2 \geq -d_2^2$ and $q_1^2 + 2q_1q_4 \leq (2q_1q_4)^2/2 \leq c_2^2/2$

and $1-c_2 \leq 1 - (2q_1 + q_4) \leq 1-c_1$ and thus $1-c_2-d_2^2 \leq P \leq 1-c_1+c_2^2/2$ occurs with probability greater than or equal to $(1-\alpha) \times (1-\beta)$. This procedure also will generally yield shorter intervals than procedure A.

As an illustration of how $B(n_1, q_1)$ compares with $B(n_1/2, 2q_1)$ the following frequencies have been tabled:

X	B(100,.02)	B(50,.04)
0	.13262	.12989
1	.27065	.27059
2	.27342	.27623
3	.18227	.18416
4	.09021	.09016
5	.03535	.03456
6	.01142	.01080
7	.00313	.00283
8	.00074	.00063
9	.00016	.00013
10	.00002	.00002
11	.00001	.00000

D. Using the condition $\sum_{i=1}^{4} n_i q_i = \xi$, one can find the minimum and maximum values of P. These will be functions of ξ, and using X one can find confidence limits for these functions, and hence for P.

Unfortunately, there are several reasons why this procedure is unsatisfactory: First, the bounds are not simple to derive in many cases; and second, their form depends upon the particular range of values within which ξ happens to lie and in instances this may be quite uncertain; and third, the resultant parametric interval is wider than under alternative procedures.

In general, one can often find bounds for P which are almost as narrow, and in many cases narrower, by considering groups of functions of the q_i such as illustrated in procedure B above, which are easier to derive. Terms of the form $\prod (1-q_i)$ are often the main contributors to the value of P, while the remaining terms may be roughly bounded with little cost in terms of width of the confidence interval. As to which of two or more groups of parameters deserve the smallest confidence level, trial and error calculations may best provide the answer.

One can see that with computation, values of n_i may be chosen to minimize the length of the interval. If one also knows the approximate size of the q_i, a better choice of testing procedure may be made.

In the fourth illustration, $P = \prod_{i=1,4} (1-q_i)\,(1-q_2q_3)\,(1-q_5^4)$. Usually q_5^4 will be so small as to be entirely negligible, and P may be taken to be $\prod_{i=1,4} (1-q_i)\,(1-q_2q_3)$.

If one tries to minimize and maximize the entire expression, subject to the restraint, the result, even when all the n_i are equal to $\bar{n}$, is in many situations unsatisfactory. The bounds for one range of values of $\xi = n_1q_1 + n_4q_4$ are $\xi/\bar{n}$, 1, which, if $\bar{n}$ is large relative to ξ, is much too wide to be of any practical value. In the general case, however, one can often use the bounds for $(1-q_1)\,(1-q_4)$ and q_2q_3 to provide bounds for P. For example, if $\bar{\xi}_1 = n_1q_1 + n_4q_4$, $\bar{\xi}_2 = n_2q_2 + n_3q_3$,

$(1-\bar{\xi}_1/\min_{i=1,4} n_i)\,(1-\bar{\xi}_2^2/4n_2n_3) \le P \le (1-\bar{\xi}_1/\max_{i=1,4} n_i)$ is valid when $\bar{\xi}_1 < \min\left(\sum_{i=1,4} n_i - 2\min_{i=1,4} n_i,\ \min_{i=1,4} n_i\right)$, and $\bar{\xi}_2 < \min(n_2, n_3)$.

$X_1 + X_4$ gives bounds for $\bar{\xi}_1$ and $X_2 + X_3$ gives bounds for $\bar{\xi}_2$, where the product of the two confidence coefficients is $1-\alpha$. In this case a narrower interval will result if the confidence coefficient for $\bar{\xi}_1$ is taken much closer to $1-\alpha$ than the coefficient for $\bar{\xi}_2$.

In summary, to find a confidence interval for the reliability of a given system, the following approaches are suggested. Some numerical trial and error may be necessary to select the most promising parametric interval.

(1) The upper and lower bounds for P are found under the restriction $\sum n_i q_i = \xi$. The upper bound for one range of values of ξ will be the solutions of the equations $\partial P/\partial q_i = \lambda n_i$ $(i=1,\ldots k)$, $\lambda \neq 0$, and $\xi = \sum n_i q_i$. The bounds for other ranges of ξ may be found from trial and error of various numerical quantities in P. The lower and upper bounds will both be functions of ξ and the known constants. A confidence interval for ξ based on $\sum X_i$ can be appropriate algebraic manipulation determine the confidence interval for P.

(2) Alternatively, P is separated into products of simpler functions, such as $\left[\prod_{i=a}^{b} (1-q_i)\right]\left[\prod_{i=c}^{d} (1-q_i)^2\right]$, and the parametric bounds for each product found, as in illustrations 3 and 4. Then by trial and error the confidence coefficients of limits for the bounds of each product are determined so that the probability of covering P will be at least $1-\alpha$, and the interval will have small, if not minimum width. If there exists prior information that certain q_i's are small then one may be willing to neglect powers of q_i greater than two, simplifying the computations. With a small amount of trial and error and even a vague idea about the size of the q_i, one may determine the various confidence coefficients in the product.

(3) If both (1) and (2) are unsatisfactory approaches, one may expand P in terms of the q_i, which may then be divided into homogeneous polynomials of ascending dimension. The first group will be a linear combination of the q_i, the second will be a quadratic form, etc. Often the q_i

may be felt to be so small that forms of higher order can safely be neglected. At any rate, the linear group will contribute most to the value of P, and by the device illustrated in method C of the third illustration, one may obtain bounds for this group, and rougher bounds for the other groups, and hence for P. In this case one must have at least some of the n_i in certain known ratios to each other. If the experiment has already been conducted and this is not the case, one may pick a new set of n'_i such that the n'_i are properly related to each other and $n'_i \leq n_i$ for all i. Then if one randomly selects n'_i observations from the n_i Bernoulli observations previously obtained using a table of random numbers, the new total number of failures in n'_i, say X'_i, will be distributed binomially with parameters n'_i and q_i. One may then proceed as before, using the approach of method C, illustration 3, having thrown away some of the available information. However, this method is a variation of an approach to the confidence interval problem, due to Dr. D. H. Evans of Bell Telephone Laboratories, appearing in Reference 2.

(4) If none of the previous attempts succeeds, one may search for weaker bounds for P and use independent confidence intervals, such as described in method A of the third illustration.

Prior investigation into the desirable sizes of the n_i in general will often simplify computations, in addition to giving a narrower confidence interval. All the remarks and examples given above also apply to the case in which one-sided intervals are desired. The lower parametric bound of P is usually easy to determine from a small amount of numerical trial and error.

V. NUMERICAL EXAMPLES. In the following computations the Poisson confidence limits were used in preference to the narrower binomial intervals for the following reasons: (1) For sample sizes as large as in the examples, there is little difference between the two; (2) The binomial limits are not easy to compute, it being necessary to use approximating formulas for the incomplete beta function in some cases; and (3) The Poisson limits are immediately available.

In problems where the sample size is, say less than 50, it is advisable to use binomial limits.

Consider the first illustration in the preceding section, and suppose that $n_1 = n_2 = n_3 = n_4 = 500$, and the numbers of failures observed are $X_1 = 3$, $X_2 = 1$, $X_3 = 10$, and $X_4 = 2$ and hence $\sum_{i=1}^{4} X_i = 16$. If one fixes $\xi = \sum_{i=1}^{4} n_i q_i$ for various values and investigates the behavior of P, it may be seen that for $\xi < 500$, $1 - \xi/500 \leq P \leq 1 - (\xi/500)^2$, where $P = \prod_{i=1}^{3} (1-q_i)(1-q_4^2)$. In this case, the .05 level two-sided Poisson confidence bounds for 500ξ are 9.598, 25.400 (Crow and Gardner). Accordingly, the confidence bounds for P are .9492, .9996.

Alternatively, one may make use of the inequality

$$(1-q_4^2)(1-\xi/500) \leq P \leq (1-\xi/1500)^3 (1-q_4^2)$$

where $\xi = \sum_{i=1}^{3} n_i q_i$. The .05 level two-sided confidence bounds for ξ are 8.102, 22.945 (Crow and Gardner) and the $.0^4 1$ level two-sided interval for q_4 (equal-tails Poisson using the χ^2 distribution as tabled in "Biometrika Tables for Statisticians") is $(.0^5 6, .0345)$. The resulting .05 level interval for P is (.9530, .9839), appreciably narrower than before.

A third procedure in this illustration involves expanding P in terms of the q_i and ignoring the non-quadratic and linear parts. The result is that $P = 1 - \xi + \sum_{i<j} q_i q_j - q_4^2$, and maximizing $\sum_{i<j} q_i q_j$ subject to $500 \sum_{i=1}^{3} q_i = \xi$, one arrives at the inequality

$$1-\xi/500-q_4^{\,2} \leq P \leq 1-\xi/500+\xi^2/3(500)^2.$$

In this procedure the one-sided $.0^41$ level upper bound for q_4 may be used with the two-sided .05 level bounds for ξ. The bounds for ξ are again 8.102, 22.945 and the upper bound for q_4 is .0335 (using the χ^2 tables in Biometrika "Tables"). Performing the calculations, the .05-level bounds for P under the above assumptions are .9529, .9845.

From this example, it is again shown that the method of minimizing the entire expression $P(q_1, \ldots q_k)$ subject to $\sum n_i q_i = \xi$ does not necessarily result in the shortest confidence intervals. For a particular set of n_i values it will be advantageous to gain an idea of how wide the resulting parametric interval will be under more than one of the approaches suggested. There is nothing "illegal" about picking the method which yields the narrowest parametric interval for a particular problem.

There have been various procedures suggested for finding confidence intervals for P, two of which have appeared in Reference 2. The first method described in the report depends on the Tchebycheff Inequality, and for the previous values of X_i, n_i, yields two-sided .05 level bounds of .9393, 1.0000. Assuming that the mode equals the mean, the interval becomes .9503, .9940. The second method, due to D. H. Evans, depends on a randomization procedure and in the above illustration necessitates using only 250 observations for each of the first three components. The number of failures Y out of 250 complete systems constructed ranges between 0 and 16. For a few of the values of Y (the expected value of Y is approximately 7) the intervals are as follows:

	Y	Interval Bounds (Poisson, Crow and Gardner)	
α= .05	3	.968	.997
	7	.945	.987
	10	.930	.979
	15	.905	.968

In comparing these results with the previous Poisson limits, it is to be noted that the randomization approach suffers from the disadvantage of being unable to utilize all the information available in the samples. On the other hand, if $n_1 = n_2 = n_3 = 500$ and $n_4 = 1000$ one may obtain a more meaningful comparison of the results of the two procedures. In this case, Y may vary from 10 to 15, and gives the following table of intervals:

TABLE XI

	Y	P(Y)	Confidence Interval (Poisson, Crow and Gardner)		Length
	15	.0009	.9525	.9838	.0313
	14	.9158	.9541	.9838	.0297
$\alpha = .05$	13	.0808	.9573	.9866	.0293
	12	.0024	.9593	.9866	.0273
	11	.0000			
		1.0000			Expected Length = .0297

The previous Poisson limits, using the bounds

$$(1-q_4^2)\,(1-\xi/500) \leq P \leq (1-\xi/1500)^3(1-q_4^2)$$

give the .05-level interval (.9538, .9839), with length .0301. While the randomization procedure almost always yields a slightly shorter interval (with probability .9991), the Poisson interval is for any given data a fixed interval. Only the interval corresponding to the most probable value of Y is completely included in the Poisson interval in this example.

For more complicated models it becomes difficult to calculate the probability density of Y even when the most favorable sample size relations are attained, and hence to compare the resultant limits with the Poisson limits. However the mean and variance of Y may be computed and confidence intervals for values of Y reasonably near EY may be compared to the Poisson limits with respect to width and centering. It appears that when

information must be thrown away to use the randomization approach, the Poisson interval based on parametric bounds may be appreciably shorter. In the optimum case for the randomization method, the interval seems to be slightly shorter on the average than the corresponding Poisson interval, at the price, however, of an increased variability.

If in the same illustration, $n_1 = 500$, $n_2 = 250$, $n_3 = 300$, $n_4 = 500$ and $X_1 = X_2 = X_3 = X_4 = 0$, then for $\bar{\xi}_2 < \sum_{i=1}^{3} n_i - 3\min_{i=1,2,3} n_i$, that is, $\bar{\xi}_2 < 300$, one can use $1 - \bar{\xi}_1/250 \leq P \leq 1 - (\bar{\xi}_1/500)^2$ or

$$(1 - \bar{\xi}_2/250)(1-q_4^2) \leq P \leq (1 - \bar{\xi}_2/500)(1-q_4^2)$$

where $\bar{\xi}_1 = \sum_{i=1}^{4} n_i q_i$, $\bar{\xi}_2 = \sum_{i=1}^{3} n_i q_i$. If a .025-level lower bound for P is desired, one can use the upper bound in the Poisson equal-tails .05 confidence limits for $\bar{\xi}_1$, $\bar{\xi}_2$. For all $X_i = 0$, the lower bounds for P turn out to be (using the $.0^4 1$ upper Poisson bound for q_4) .9852 and .9847 respectively. In this system, the first inequality will provide lower limits when $X_4 = 0$ and the second, when X_4 is positive. As remarked before, it is undesirable to select the confidence bounds on the basis of the particular sample point observed, and hence the greater lower parametric bound attained by the second interval (for $q_4 < .7$) seems to make it preferable.

REFERENCES

1. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, 1946.

2. De Cicco, H., The Reliability of Weapon Systems Estimated from Component Test Data Alone, Ordnance Special Weapons-Ammunition Command, Technical Note No. 1, December, 1959.

RELIABILITY OF COMPLIANCE WITH ONE-SIDED SPECIFICATIONS LIMITS WHEN DATA IS NORMALLY DISTRIBUTED

E. L. Bombara
Army Rocket and Guided Missile Agency
U. S. Army Ordnance Missile Command
Redstone Arsenal

ABSTRACT

This report describes the theory and application of demonstrating reliability of compliance with single specification limits. Two statistical methods are used: the variables sampling plan method and the tolerance limit method. These methods, respectively, determine sample size and demonstrate reliability of compliance with specification limits at a prescribed confidence level. They may be applied to internal and external ballistics data of weapons systems.

INTRODUCTION

This reliability report explains the use of two sets of tables for the purpose of demonstrating reliability of compliance with one-sided specification limits when data is normally distributed. The variables sampling plan factors presented in table I may be used when sample sizes are to be selected to demonstrate such reliability. Factors for one-sided tolerance limits in table II may be used when the purpose is not to select sample size, but only to demonstrate reliability at a fixed confidence level. Both tables I and II are prepared using two different methods:

1. The approximate method of one-sided tolerance limits (modern-type values), and

2. The exact method using the noncentral t-statistic (bold-face-type values).

Comparison of these data gives an indication of accuracy of the much-used one-sided tolerance factors.

The best demonstration of reliability of a complete missile system is based on the number of successes and failures of the system to accomplish its intended mission. The same also applies to component reliabilities; however, in many instances it is economically impossible to obtain a sufficient sample size for realistic demonstration. The problem does not appear to be resolvable for a complete system,[1] but practical methods are available for partial reliability demonstration; for example, demonstration of compliance with specification limits. Such demonstration is based on a measurement that characterizes each sample unit with respect to a numerical scale. Points on the scale on one side of the specification limit should describe success, and the other points should describe failure of the component characterized. Sometimes extreme values (high and low) of a variable describe failure and intermediate values describe success. At other times, failure is described by just one extreme. This report is concerned only with the latter category.

The primary assumption for the variable measured is that it be normally distributed. (Methods outlined herein give invalid conclusions if the distribution is other than normal.)[2] A specification limit should be set so that values below the limit describe success and values above the limit describe failure or vice-versa. The procedure, then, is to test the component to determine whether its performance is or is not on the successful side of the specification limit. Suppose that in hydrostatic testing 12 pressure vessels to failure, all fail above the minimum stress level (2200 psi) specified in accordance with its expected environment. The success-to-trials ratio of 12/12 demonstrates a reliability of .825 at the 90% confidence level, but we have not considered the distribution of the observed burst pressures as compared to the specification limit. The reliabilities demonstrated in figures 1 and 2 are both .825 at 90% confidence on the basis of success-to-trials, but is this a fair or complete evaluation? Obviously, the probability that a burst will occur below 2200 psi is much greater for the results shown in figure 1 than for the results shown in figure 2. Thus, the reliability of motor cases described by figure 2 is superior to those described by figure 1.

---

[1]Where large sample sizes are available, system reliability is best demonstrated using attributes, the ratio of success to total trials. Small -sample methods outlined in this paper cannot substitute for this method

[2]Investigation of distribution should precede the use of methods outlined herein to insure normality, particularly in the tail of the distribution.

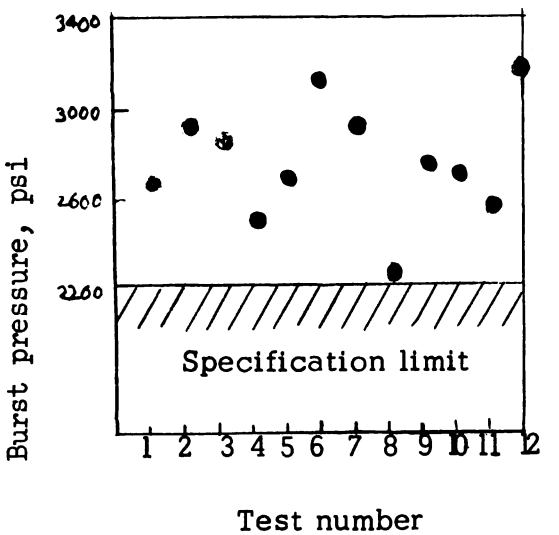


Figure 1. - High variability

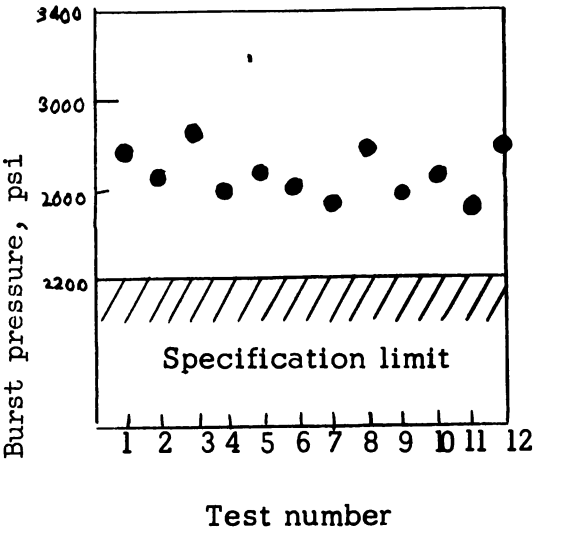


Figure 2. - Low variability

(Using the variables approach and table II, .90 reliability is demonstrated at 90% conficence in figure 1, but more than .999 reliability is demonstrated at this confidence in figure 2.) It is assumed, of course, that sampling is random, that no shift will occur, and that manufacturing quality control is rigid. This improvement will show up on a success-to-trials basis only if sample sizes are very large. As a general rule, reliability demonstration on a variables basis does not require as large a sample size as does demonstration on the basis of success to trials.

Two procedures based on variables are outlined below, tables for which are described in Discussion. Equations used in constructing these tables are given in Appendix A, and derivations of these are presented in Appendices B and C, for the two respective procedures. The first procedure, using table I, is established for selecting sample size and testing reliability hypotheses. It should be used when systems are tested to determine whether reliability is or is not as high as a prescribed value, with the power of discrimination from some lower minimum acceptable value. The second procedure, using table II, is established to demonstrate reliability at a fixed confidence level. The latter should be used for

reliability demonstration when the number of tests has been previously fixed. This procedure, using table II, permits demonstration of reliability and confidence without power of discrimination; only the former procedure, using table I, should be employed to select sample size.

DISCUSSION

VARIABLES SAMPLING AND TOLERANCE LIMIT METHODS. Tables I and II were prepared to simplify computations using the equations of Appendix A. These tables are applicable only when the data to be analyzed is normally distributed. Modern type values in table I give the solutions to equations (3) and (4) and in table II give the solution to equation (7). Values appearing in bold-face type, computed from the noncentral t-statistic, are more accurate than the modern type values. Illustrations of the use of the tables are given below.

Variables Sampling Plan Method.

1. Suppose we wish to fire a number of rocket motors in static tests for the primary purpose of deciding whether reliability of compliance with a specification limit is higher or lower than specified values; that is, we want to discriminate between two reliability levels. For example, suppose that because of g-load limitations, thrust must not exceed a single upper specification limit, U = 12,000 lbs. Suppose further that the absolute minimum g-load reliability requirement (reliability here is the probability of not exceeding U) is $R_L = .95$, and $R_H = .9995$ is set as a goal. We will make one of these two decisions:

a. Reliability is at least .9995, with risk β when it is actually .95.

b. Reliability is less than .95, with risk α when it is actually .9995.

Suppose we let $\alpha = .05$ and $\beta = .05$. From table I, $\alpha = .05$, $\beta = .05$, $R_H = .9995$, and $R_L = .95$ gives K = 2.47. If we do not know the standard deviation, σ, from previous experience with the tested motor or a similar motor, the number of motors to be tested is n = 17. If we know σ based on a large amount of experience, the number of motors to be tested is n' = 5. As a general rule, we recommend that μ be assumed unknown. After the motors are tested, compute the mean value of maximum thrust, $\bar{X}$. If the standard deviation is unknown, the quantity of $\bar{X} + Ks$ is computed, where s is the sample standard deviation. If the standard deviation, σ, is known, $\bar{X} + K\sigma$ is computed instead, and

utilized in the same manner as $\bar{X} + Ks$ to arrive at a decision (K is the same in either case). Suppose the standard deviation is unknown, $\bar{X}$ obtained from 17 static firings is 11,150 lbs., and s is 320 lbs. Then $\bar{X} + Ks = 11,150 + (2.47)(320) = 11,940$ lbs. Because this value is not greater than the upper specification limit, U = 12,000 lbs, we decide that reliability is at least .9995, but with 5% risk ($100\beta\%$) of making this decision when it is .95. (The demonstrated reliability here is not .9995, because there is a $100\beta\% = 5\%$ chance of deciding .9995 reliability when actually it is .95). If $\bar{X} + Ks$ should be greater than U, we would decide that reliability is less than .95, but with 5% risk ($100\alpha\%$) of this decision when it is .9995. The two risks, α and β were both chosen equal in this example. If the risk, β, of accepting unreliable components must be less than the risk, α, of rejecting reliable components we would choose $\beta < \alpha$, and conversely.

3. Table I may also be used when the specification calls out a single lower limit, L. Values of K, n, and n' are the same; only the decision procedure is changed. If $\bar{X} - Ks \geq L$ we make a decision such as that in paragraph 1a above. Otherwise, we make a decision such as that in paragraph 1b. Again, σ is substituted for s when the standard deviation is known. It is seen that if a normal distribution of variables data can be assumed, it is possible to demonstrate high reliability using a relatively small sample size.

4. It may be noted that n or n' = 5 is the smallest sample size shown in table I. This is because it is considered poor practice to use a sample size smaller than this. Also, the approximation for K using equation (3) in Appendix A is very poor when $n < 5$.

Tolerance Limit Method.

1. Table II is provided for use when the number of test firings has been fixed in advance, not on the basis of reliability hypotheses. It should not be used for selecting sample size. It permits demonstration of reliability at a given confidence level, but does not permit discrimination between two reliability levels as is the case for table I.

2. Suppose that, in firing a rocket motor, reliability of .998 of not exceeding the maximum allowable g-load must be demonstrated at the 90% confidence level. Suppose the g-load limit corresponds to an upper thrust limit of U = 28,800lbs. Thrust data is available from 15 motors of the same type. From table II, 90% confidence, sample size 15, reliability .998

gives the factor $K_C = 3.9$.[1] Suppose the mean and standard deviation of the 15 thrust values are 26,000 lbs and 620 lbs, respectively. Then $\bar{X} + K_C s = 26,000 + (620)(3.9) = 28,400$ lbs. Because this value does not exceed $U = 28,800$ lbs, a reliability of at least .998 is demonstrated at the 90% confidence level, assuming normally distributed average thrusts. if $\bar{X} + K_C s$ had exceeded 28,800 lbs, it would have been concluded that the .998 reliability was not demonstrated at 90% confidence. If the specification were a lower limit, L, table II would be used in the same manner to obtain $K_C = 3.9$, but the decision procedure would be modified: only if $\bar{X} - K_C s$ were not less than L would reliability .998 be demonstrated at the 90% confidence level.

3. It is not good practice to use table II when n is less than 5 because approximations used in derivation of statistical tolerance methods are invalid for such small sample sizes. Also from a practical standpoint, it is dangerous to use such small sample sizes in making decisions.

ACCURACY OF FACTORS USED IN VARIABLES SAMPLING PLAN. The round-off in table I was based on a comparison of K-values in table I to those in reference 3, table 1.2. The latter, available for $\alpha = .05$, $\beta = .10$, were computed from tables of the noncentral t-distribution by Johnson and Welch (reference 10). They are shown in bold-face type. Comparison of these tables also indicated that n (not n') in table I might be increased by one as a conservative measure. (Table 1.2 in reference 3 was constructed using an iterative procedure and the noncentral t-statistic. A simplified iterative method is given in reference 9, pp. 10-16). Modern type K-values in table I are more accurate than modern type K_C values in table II, because division of two quantities biased in the same direction (see equations (1.3) to (1.5), Appendix B) reduces the bias of the quotient. The bias in n is not reduced because n is obtained by subtracting equation (1.4) from equation (1.3), not dividing.

---

[1] If sample size (n) is 15 or less, it is recommended that n - 1 be used to find K_C from modern-type values in table II (except at the 95% confidence level.)

ACCURACY OF FACTORS USED IN ONE-SIDED TOLERANCE LIMITS. Data in table II was rounded off, such that although the error in the last digit of values in modern type is at times as much as 3 or 4, it is usually not more than 2. The round-off was based on a comparison of K_C-values in table II to those obtained from noncentral t-values in reference 9. The latter values are shown in bold-face type in table II. Round-off procedures were extrapolated for reliabilities in the ranges .01 to .70, and .9991 to .9999. Bold-face factors in table II are more accurate than factors in modern type because the former were computed from the exact noncentral t-distribution. They were obtained by using equations(10) and (11) below in conjunction with reference 9. Reverence 9 indicates that bold-face values are accurate to all three digits except for an occasional inaccuracy of 1 or 2 in the third digit. Some of the factors in bold-face type were checked by comparison with reference 7 and also reference 6, in which a recursion formula is utilized for computation of tolerance factors with confidence coefficients of .90, .95, and .99. On occasions, differences were as much as 2 in the third significant digit, but usually not more than 1. For all levels of confidence in table II (except 95%), n = 5 to 15, it appears that the K-factor for n-1 is a closer approximation to the true value than the factor for n itself. Therefore, it is good practice in this region to subtract 1 from the sample size before finding the factor. This bias in the tables is based on comparison with K-factors shown in bold-face type.

From equation (2.1), Appendix C, we see that for the case C = .50, $Z_C = 0$ when $\bar{X} + K_C s$ is assumed normally distributed with mean $\mu + K_C \sigma$. The value of K_C becomes $K_C = \sigma Z_R / E(s)$, where E(s) is σ times a function of sample size, n, and the ratio of gamma functions of n. Thus, a large portion of the bias of the sample standard deviation is taken into account although some bias remains due to departure from assumed normality. The larger portion of the bias is accounted for by the bias of the sample standard deviation when one-sided tolerance factors are used. Note that using equation (7), which is biased because of the bias in the standard deviation, when C = .50, $Z_C = 0$ and $K_C = Z_R$. From a normal distribution table such as reference 4, it is seen that the bias is greater between modern type data and Z_R than between modern type and bold-face type data. (Modern type data for 50% confidence in table II was adjusted for bias of the standard deviation as explained above in this paragraph.)

Appendix A

EQUATIONS FOR OBTAINING FACTORS GIVEN IN TABLES I AND II

EQUATIONS FOR FACTORS USED IN VARIABLES SAMPLING PLAN.[1] Suppose that, in accordance with specifications, a measured variable must not exceed a single upper specification limit, U. Suppose further that the minimum acceptable reliability associated with this variable (for this purpose "reliability" is defined as the probability of not exceeding U) is $R_L = .99$, and $R_H = .999$ is set as a goal. Let us test the hypothesis, H_0 that the true reliability is R_H against the hypothesis, H_1, that the true reliability is R_L. If we make the decision (based on our sample) that reliability is at least R_H, we would like to be sure that only $100\beta\%$ of the time will we make this decision when the true reliability is R_L. Conversely, if we make the decision that reliability is less than R_L, we would like to be sure that only $100\alpha\%$ of the time will we make this decision when the true reliability is R_H. Decision procedure is as follows:

if

(1) $\bar{X} + Ks \leq U$, accept H_0 (decide $R \geq R_H$)

if

(2) $\bar{X} + Ks > U$, reject H_0 (decide $R < R_L$)

$\bar{X}$ and s are the sample mean and sample standard deviation, respectively, taken from a sample of size n, U is the upper specification limit and

$$(3) \qquad K \cong \frac{Z_{RL}Z_{1-\alpha} + Z_{RH}Z_{1-\beta}}{Z_{1-\alpha} + Z_{1-\beta}}$$

When $\alpha = \beta$ this becomes

$$(3a) \qquad K \cong (Z_{RL} + Z_{RH}) \ /2;$$

---

[1] These equations were utilized to construct modern type data in table I, therefore, the results obtained by using them are no more accurate than data in table I.

(4)
$$n \cong \left(\frac{Z_{1-\alpha} + Z_{1-\beta}}{Z_{RH} - Z_{RL}} \right)^2 (1 + K^2/2)$$

All Z's are one-sided normal deviates. In the event that the true standard deviation (σ) is known from past experience, s in (1) and (2) should be replaced by σ, and the quantity $(1 + K^2/2)$ in (4) is deleted. K remains unchanged.

If the specification should call out a one-sided lower limit, L, only the decision procedure is changed. Equations (1) and (2) become

(5) If
$$\bar{X} - Ks \geq L, \text{ accept } H_0 \text{ (decide } R \geq R_H).$$

(6) If
$$\bar{X} - Ks < L, \text{ reject } H_0 \text{ (decide } R < R_L).$$

Derivation of this plan is presented in Appendix B. Theory and applications are also presented in references 1, 3, and 9.

As an illustration of the equations above, let $R_L = .99$, $R_H = .999$, $\alpha = .10$, $\beta = .05$. From a normal distribution table, $Z_{RL} = 2.326$, $Z_{RH} = 3.090$, $Z_{1-\alpha} = 1.282$ and $Z_{1-\beta} = 1.645$. Then

$$K \cong \frac{(2.326)(1.282) + (3.090)(1.645)}{1.282 + 1.645} = 2.76$$

For sigma unknown,

$$n \cong \left(\frac{1.282 + 1.645}{3.090 - 2.326} \right)^2 \cdot \left(1 + \frac{2.76^2}{2} \right) = 70.4 \cong 71$$

For sigma known (deleting right-hand parenthesis),

$$n' = 14.68 \cong 15.$$

These are the same results as given by table I for $\alpha = .10$. $\beta = .05$, $R_H = .999$, $R_L = .99$.

EQUATIONS FOR FACTORS USED IN TOLERANCE LIMITS.[1] In review, the equations for factors given in table I were applicable only when a sample size was to be selected in advance for the primary purpose of discrimination between a minimum acceptable reliability, R_L, and goal reliability, R_H. One either decided that reliability was as good as R_H with risk β when it was actually R_L or decided that reliability was worse than R_L with risk α when it was actually R_H. Specifically, an inference was made that the sampled population was "positioned" such that either the proportion R_H of the population lie on the acceptable side of the specification limit, or the proportion R_L lie on the acceptable side of such limit.

Regarding equations for factors given in table II, suppose that a test program has been designed for purposes other than reliability evaluation. The number of firings in this case has not been established to test reliability hypotheses. In this situation we desire only to demonstrate reliability at a given confidence level. (Confidence is the probability that we are correct when we say that reliability of at least R is demonstrated). The information obtained does not permit discrimination between two reliabilities as in the former case. Also, the population is not positioned with respect to a specification limit according to a reliability hypothesis.

Suppose there is a single upper specification limit, U, above which material is defined as defective. In order to demonstrate a given reliability at a selected confidence, the quantity $\bar{X} + K_c s$ is computed such that there is confidence C that this quantity contains at least the proportion R of the true population below it. If $\bar{X} + K_c s \leq U$, it may be stated that reliability of at least R is demonstrated at confidence level C. $\bar{X}$ and s are the sample mean and sample standard deviation, respectively, taken from a sample size n. K_c is defined by

---

[1] These equations were utilized to construct modern type data in table II; therefore, the results obtained by using them are no more accurate than the data in table II.

(7)
$$K_C \cong \left\{ Z_R + Z_C \sqrt{\frac{1}{n}\left[1 - \frac{Z_C^2}{2(n-1)}\right] + \frac{Z_R^2}{2(n-1)}} \right\} \Big/ \left[1 - \frac{Z_C^2}{2(n-1)}\right]$$

Where n is sample size and Z_R, Z_C are the respective one-sided standard normal deviates corresponding to R and C, available in a normal distribution table.

As an illustration, it is desired that reliability of at least .998 of not exceeding the maximum allowable g-load in firing a rocket motor be demonstrated at the 90% confidence level. Suppose the g-load limit corresponds to a thrust limit of U = 28,800 lbs. When R = .998, Z_R = 2.878 (look up 2(.998) - 1 = .996 in reference 4). Similarly, when C = .90, Z_C = 1.282. Suppose funds permit that a sample of n = 15 motors of the same type are to be fired under homogeneous conditions. From (7) above, K_C = 3.9. This result is the same as that given in table II, for 90% confidence, .998 reliability, n = 15.

If the specification limit were a lower limit, L, below which material is classified as defective, equation (7) still applies, but decision procedure is modified. If $\bar{X} - K_C s \geq L$, reliability R is demonstrated at confidence level C. Otherwise reliability R is not demonstrated.

Derivation of equations for reliability estimates based on a one-sided tolerance limit is presented in Appendix C. Theory and applications in other areas are presented in references 2, 3, and 6 through 10. Other applications in ordnance engineering are presented in reference 5.

EQUATIONS FOR FACTORS IN BOLD-FACE TYPE IN TABLE II. Equation 2.1, Appendix C, may be written as:

(8)
$$P\left\{\left[(\mu - \bar{X})\sqrt{n} + \sigma Z_R \sqrt{n}\right] / s \leq K_C \sqrt{n}\right\} = C.$$

The left-hand side of the inequality is the noncentral t statistic with n - 1 degrees of freedom and parameter of noncentrality $Z_R\sqrt{n}$. This is seen more clearly by writing the left-hand side of the inequality in the form

(9)
$$t_{n-1,\, Z_R\sqrt{n}} = \frac{\frac{\mu - \bar{X}}{\sigma/\sqrt{n}} + Z_R\sqrt{n}}{s/\sigma}$$

Reference 9, tables of the noncentral t-distribution, gives the probability integral

(10)
$$P\left\{ t_{n-1,\, K_p\sqrt{f+1}} \leq x\sqrt{f} \right\} = C.$$

Since $Z_R = K_p$, and $f = n-1$ in this case, the left-hand member of (9) is identical to the left-hand sides of inequalities in both (8) and (10). The relationship between a one-sided tolerance factor and the noncentral t-statistic tabulated in reference 9 is therefore given by

(11)
$$K_C = x\sqrt{\frac{n-1}{n}}.$$

Appendix B

DERIVATION OF A VARIABLES SAMPLING PLAN FOR TESTING RELIABILITY HYPOTHESES

STANDARD DEVIATION UNKNOWN. Suppose we have material which is defective if a measure describing the material exceeds a single upper specification limit, U. Suppose also that the hypothesized population proportion of reliable material is R_H, so that $\mu_{RH} + Z_{RH}\,\sigma = U$. The population standard deviation, σ, and the population mean, μ, are unknown. Based on a sample of n observations, we wish to test the null hypothesis $R = R_H$ against $R = R_L$, where $R_L < R_H$.

These sketches are intended for assistance in deriving the sampling plan:

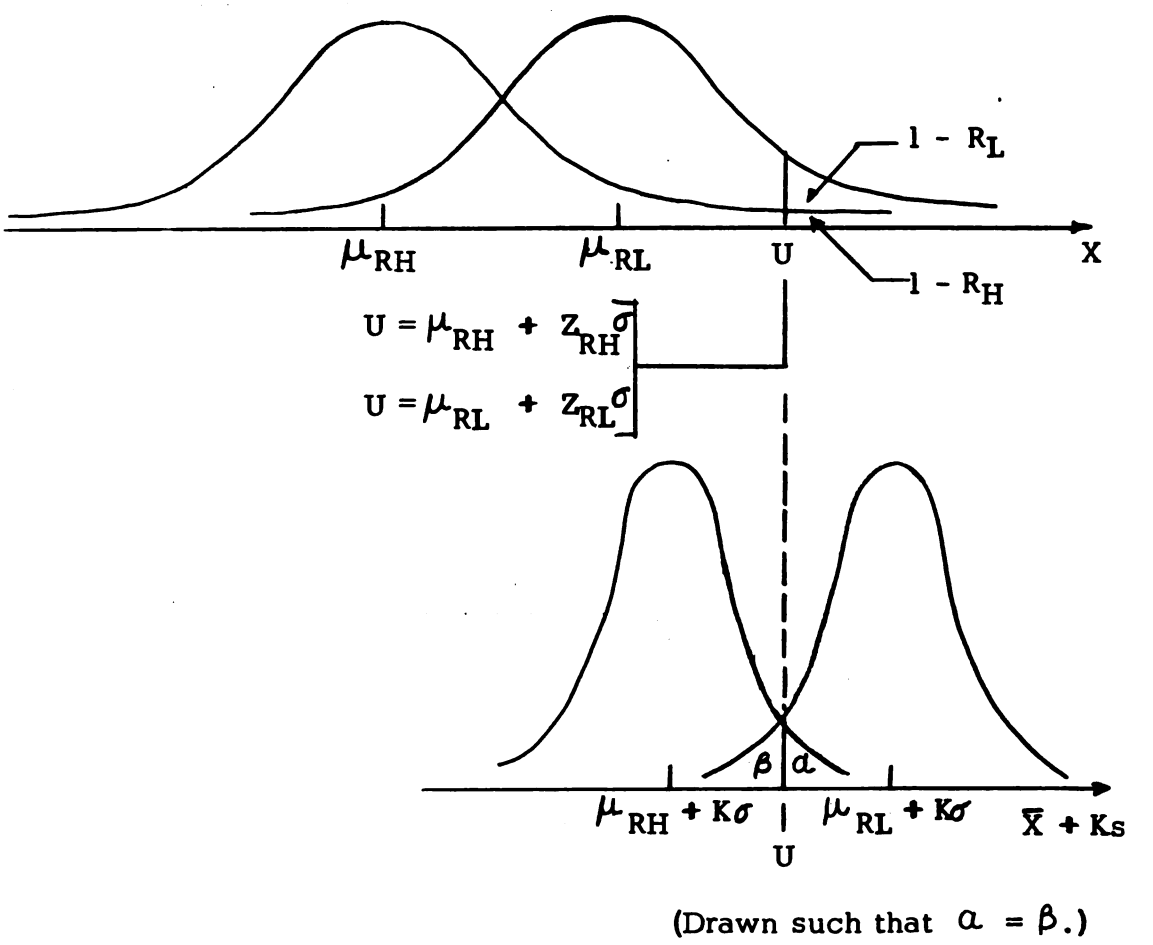


Figure 3. - Distributions of X and $\bar{X}$ + Ks

This plan assumes that X is normally distributed with mean μ and standard deviation σ and that $\bar{X}$ + Ks is approximately normally distributed with mean μ + K σ and standard deviation

$$\sigma \sqrt{\frac{1}{n} + \frac{K^2}{2(n-1)}}$$

See reference 1.

The null and alternate hypotheses are

$$H_o: R = R_H \quad \text{or} \quad \mu + K\sigma = \mu_{RH} + K\sigma$$
$$H_1: R = R_L < R_H \quad \text{or} \quad \mu + K\sigma = \mu_{RL} + K\sigma$$

Then

(1.1) $$P\left[\bar{X} + Ks \leq U \mid R = R_H\right] = 1 - \alpha$$

(1.2) $$P\left[\bar{X} + Ks \leq U \mid R = R_L\right] = \beta$$

Where $\bar{X}$ and s are the sample mean and sample standard deviation, respectively. Substituting $U = \mu_{RH} + Z_{RH}\sigma$, and $U = \mu_{RL} + Z_{RL}\sigma$, the solutions of (1.1) and (1.2) are

$$\frac{(\mu_{RH} + Z_{RH}\sigma) - (\mu_{RH} + K\sigma)}{\sigma\sqrt{\frac{1}{n} + \frac{K^2}{2(n-1)}}} \cong Z_{1-\alpha}$$

or

(1.3) $$\frac{Z_{RH} - K}{\sqrt{\frac{1}{n} + \frac{K^2}{2(n-1)}}} \cong Z_{1-\alpha};$$

$$\frac{(\mu_{RL} + Z_{RL}\sigma) - (\mu_{RL} + K\sigma)}{\sigma\sqrt{\frac{1}{n} + \frac{K^2}{2(n-1)}}} \cong Z_{\beta},$$

or

(1.4) $$\frac{Z_{RL} - K}{\sqrt{\frac{1}{n} + \frac{K^2}{2(n-1)}}} \cong Z_{\beta}$$

Dividing (1.3) by (1.4) and substituting $Z_{1-\beta} = -Z_{\beta}$

$$(1.5) \qquad K \cong \frac{Z_{RL}Z_{1-\alpha} + Z_{RH}Z_{1-\beta}}{Z_{1-\alpha} + Z_{1-\beta}}$$

When $\alpha = \beta$, this becomes

$$(1.5a) \qquad K \cong (1/2)\ (Z_{RL} + Z_{RH})$$

Subtracting (1.4) from (1.3), squaring and substituting $Z_{1-\beta} = -Z_{\beta}$,

$$(1.6) \qquad n \cong \left(\frac{Z_{1-\alpha} + Z_{1-\beta}}{Z_{RH} - Z_{RL}}\right)^2 \cdot \left(1 + \frac{K^2}{2}\right).$$

The procedure for testing the null hypothesis above is:

If

(1.7) $\bar{X} + Ks > U$, reject H_0 with probability α when H_0 is true.

If

(1.8) $\bar{X} + Ks \leq U$, accept H_0 with probability β when H_1 is true.

KNOWN STANDARD DEVIATION. By following the same type of derivation as above, but assuming the standard deviation, σ, known, one obtains K identical to that of (1.5) above (the denominators of (1.3) and (1.4) become $1/\sqrt{n}$) and effect on sample size is to eliminate the latter parenthesis from (1.6). Thus, the sample size assuming a known standard deviation becomes

$$(1.9) \qquad n' \cong \left(\frac{Z_{1-\alpha} + Z_{1-\beta}}{Z_{RH} - Z_{RL}}\right)^2$$

Procedure for testing the null hypothesis is the same except that σ is substituted for s in (1.7) and (1.8). Computation of s, the sample standard deviation, becomes unnecessary.

It is recommended that these procedures for testing reliability hypotheses not be used for sample sizes (n) less than 5.

Appendix C

DERIVATION OF EQUATIONS FOR RELIABILITY DEMONSTRATION BASED ON A ONE-SIDED TOLERANCE LIMIT

Let μ and σ be the true population mean and standard deviation, respectively, of some ballistic performance parameter, and let Z_R be chosen (from a normal distribution table) such that the proportion R of the population will lie below $\mu + Z_R\sigma$. Let $\bar{X}$ and s represent the estimated mean and standard deviation, respectively, obtained from a sample of size n taken from this population.

If a constant, K_c (greater than Z_R when $C > .50$), can be found such that 100 C% of the time $\bar{X} + K_c s \geq \mu + Z_R\sigma$, we have confidence, C, that at least the proportion R of the population will lie below $\bar{X} + K_c s$. Therefore, provided that $\bar{X} + K_c s$ is not greater than a single upper specification limited, U, we have demonstrated reliability of at least R at the 100 C% confidence level. The above reliability and confidence may be expressed by

$$(2.1) \qquad P(\bar{X} + K_c s \geq \mu + Z_R\sigma) = C.$$

In order to find K_c to satisfy this equation, we must specify a distribution for $\bar{X} + K_c s$. As stated in reference 1, we may assume that, if $\bar{X}$ is normally distributed, $\bar{X} + K_c s$ is approximately normally distributed with mean $\mu + K_c\sigma$ and standard deviation

$$\sigma\sqrt{\frac{1}{n} + \frac{K_c^2}{2(n-1)}}$$

when $n \geq 5$, say. Let Z_c be the number of

such standard deviations between the mean of this population, $\mu + K_c\sigma$, and the value $\mu + Z_R\sigma$. It is apparent that $\mu + K_c\sigma$ must be greater than $\mu + Z_R\sigma$ to satisfy equation (1) in the case when C is greater than .50. This sketch on the next page shows the distribution of X and the corresponding distribution of $\bar{X} + K_c s$, as figure 4.

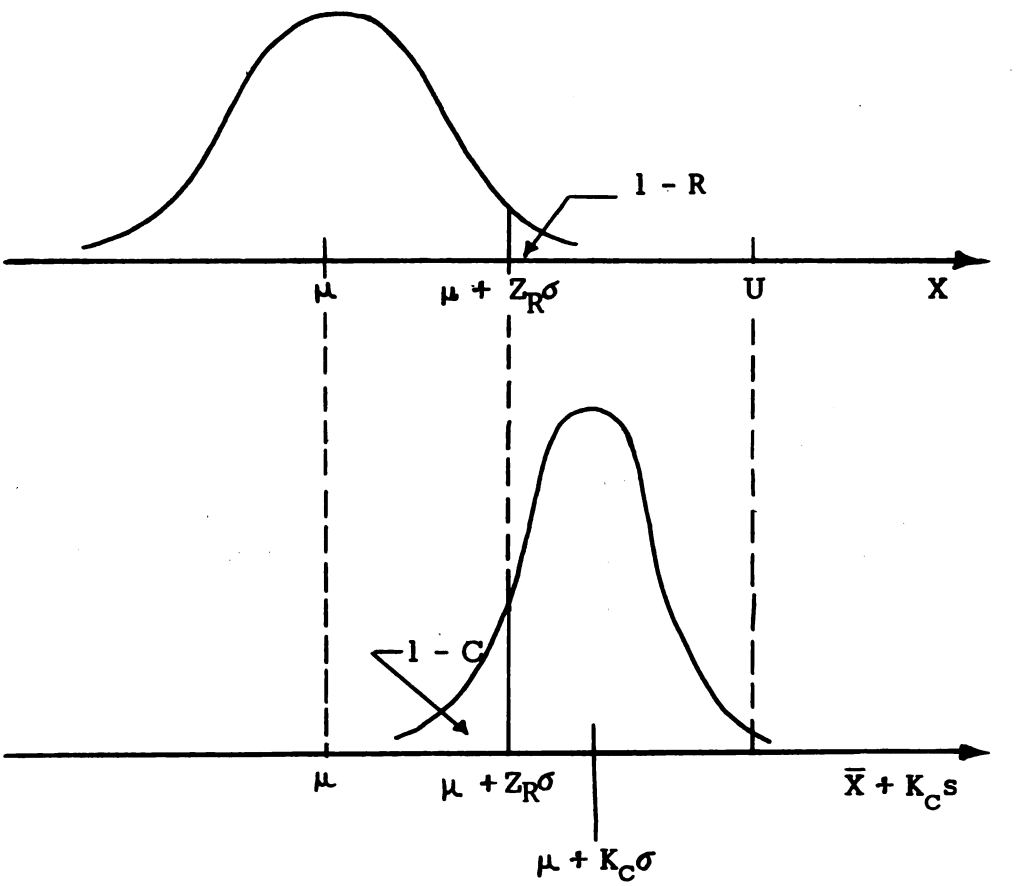


Figure 4. - Distributions of X and $\bar{X} + K_c s$

K_c must be found such that the proportion of area under the curve above $\mu + Z_R\sigma$ is C, the probability that $\bar{X} + K_c s \geq \mu + Z_R\sigma$. The solution to equation (2.1) is obtained from the standard normal deviate

$$(2.2) \qquad Z_C = \frac{\mu + K_C\sigma - (\mu + Z_R\sigma)}{\sigma\sqrt{1/n + K_C^2/2(n-1)}}$$

which gives,

$$(2.3) \qquad K_C \cong \left\{ Z_R + Z_C\sqrt{\frac{1}{n}\left[1 - \frac{Z_C^2}{2(n-1)}\right] + \frac{Z_R^2}{2(n-1)}} \right\} \Big/ \left[1 - \frac{Z_C^2}{2(n-1)}\right]$$

(See reference 2.) Due to symmetry, equation (2.3) also applies to a lower specification limit, L. If $\bar{X} - K_C s \geq L$, reliability R is demonstrated at the 100 C% confidence level.

It is recommended that this procedure for obtaining reliability estimates not be used for sample sizes n less than 5, due to inaccuracy of approximations assumed in writing equation (2.2).

ACKNOWLEDGMENTS. The author wishes to acknowledge his indebtedness to the following persons: Mr. George Nagy, Goodyear Aircraft Corporation, Mr. Waldemar Heinzelmann, Virginia Polytechnic Institute, for technical suggestions, Miss Mary Mattey and Mrs. Mary Pase, Thiokol Chemical Corporation, for programming, and Mr. Everett Robertson, ARGMA, for editing.

TABLE I

VARIABLES SAMPLING PLAN

Factors for Testing Reliability Hypotheses

$\alpha = .01$
$\beta = .01$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.26	81	79	.999	.80	1.97	13	5
.70	.60	0.38	317	295	.999	.85	2.06	17	6
.80	.50	0.42	33	31	.999	.90	2.19	23	7
.80	.60	0.54	72	63	.999	.95	2.37	40	11
.80	.70	0.68	265	216	.999	.99	2.71	174	38
.85	.50	0.51	23	21	.999	.995	2.83	411	82
.85	.60	0.64	43	36	.9995	.50	1.6	5	5
.85	.70	0.78	108	83	.9995	.60	1.8	7	5
.85	.80	0.93	825	573	.9995	.70	1.9	8	5
.90	.50	0.64	16	14	.9995	.80	2.07	12	5
.90	.60	0.76	27	21	.9995	.85	2.16	15	5
.90	.70	0.90	54	38	.9995	.90	2.29	20	6
.90	.80	1.06	175	112	.9995	.95	2.47	33	8
.90	.85	1.16	598	358	.9995	.99	2.81	115	24
.95	.50	0.82	11	8	.9995	.995	2.93	225	43
.95	.60	0.94	17	12	.9995	.999	3.19	3262	536
.95	.70	1.08	28	18	.9999	.50	1.9	5	5
.95	.80	1.24	60	34	.9999	.60	2.0	6	5
.95	.85	1.34	111	59	.9999	.70	2.1	7	5
.95	.90	1.46	341	165	.9999	.80	2.28	10	5
.99	.50	1.2	7	4	.9999	.85	2.38	12	5
.99	.60	1.29	10	6	.9999	.90	2.50	16	5
.99	.70	1.43	14	7	.9999	.95	2.68	24	6
.99	.80	1.58	23	10	.9999	.99	3.02	63	12
.99	.85	1.68	32	14	.9999	.995	3.15	99	17
.99	.90	1.80	53	20	.9999	.999	3.41	372	55
.99	.95	1.99	139	47	.9999	.9995	3.51	844	119
.995	.50	1.3	6	5	.99999	.50	2.1	5	5
.995	.60	1.4	9	5	.99999	.60	2.3	5	5
.995	.70	1.55	12	6	.99999	.70	2.4	6	5
.995	.80	1.71	18	8	.99999	.80	2.6	8	5
.995	.85	1.81	25	10	.99999	.85	2.65	10	5
.995	.90	1.93	37	13	.99999	.90	2.77	12	5
.995	.95	2.11	81	25	.99999	.95	2.96	17	5
.995	.99	2.45	137	347	.99999	.99	3.30	38	6
.999	.50	1.5	5	5	.99999	.995	3.42	52	8
.999	.60	1.7	7	5	.99999	.999	3.68	122	16
.999	.70	1.8	9	5	.99999	.9995	3.78	186	23
					.99999	.9999	3.99	652	73

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .01$
$\beta = .05$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.22	59	58	.999	.85	1.89	11	5
.70	.60	0.37	229	215	.999	.90	2.03	15	5
.80	.50	0.35	24	23	.999	.95	2.24	27	8
.80	.60	0.50	52	46	.999	.99	2.64	122	28
.80	.70	0.66	191	157	.999	.995	2.79	292	60
.85	.50	0.43	17	15	.9995	.50	1.4	5	5
.85	.60	0.58	31	26	.9995	.60	1.5	5	5
.85	.70	0.74	77	61	.9995	.70	1.7	5	5
.85	.80	0.92	595	418	.9995	.80	1.9	8	5
.90	.50	0.53	11	10	.9995	.85	1.97	10	5
.90	.60	0.68	19	15	.9995	.90	2.11	13	5
.90	.70	0.84	38	28	.9995	.95	2.33	22	6
.90	.80	1.02	124	82	.9995	.99	2.73	80	17
.90	.85	1.14	430	261	.9995	.995	2.87	159	31
.95	.50	0.7	8	6	.9995	.999	3.17	2356	391
.95	.60	0.83	11	9	.9999	.50	1.5	5	5
.95	.70	0.99	19	13	.9999	.60	1.7	5	5
.95	.80	1.17	42	25	.9999	.70	1.8	5	5
.95	.85	1.29	78	43	.9999	.80	2.0	6	5
.95	.90	1.43	243	120	.9999	.85	2.1	8	5
.99	.50	1.0	5	5	.9999	.90	2.29	10	5
.99	.60	1.1	6	5	.9999	.95	2.50	16	5
.99	.70	1.3	9	5	.9999	.99	2.90	43	9
.99	.80	1.46	15	8	.9999	.995	3.05	69	13
.99	.85	1.57	22	10	.9999	.999	3.35	264	40
.99	.90	1.71	36	15	.9999	.9995	3.47	604	87
.99	.95	1.93	98	34	.99999	.50	1.8	5	5
.995	.50	1.1	5	5	.99999	.60	1.9	5	5
.995	.60	1.2	6	5	.99999	.70	2.1	5	5
.995	.70	1.4	8	5	.99999	.80	2.3	5	5
.995	.80	1.56	12	6	.99999	.85	2.4	6	5
.995	.85	1.67	16	7	.99999	.90	2.5	8	5
.995	.90	1.82	25	10	.99999	.95	2.73	11	5
.995	.95	2.03	56	19	.99999	.99	3.13	25	5
.995	.99	2.43	997	253	.99999	.995	3.28	36	6
.999	.50	1.3	5	5	.99999	.999	3.58	85	12
.999	.60	1.4	5	5	.99999	.9995	3.69	131	17
.999	.70	1.6	5	5	.99999	.9999	3.95	465	53
.999	.80	1.8	9	5					

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .05$
$\beta = .01$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.31	61	58	.999	.85	2.24	14	5
.70	.60	0.41	233	215	.999	.90	2.34	19	5
.80	.50	0.49	25	23	.999	.95	2.49	31	8
.80	.60	0.60	54	46	.999	.99	2.77	131	28
.80	.70	0.71	197	157	.999	.995	2.88	307	60
.85	.50	0.61	18	15	.9995	.50	1.9	5	5
.85	.60	0.71	33	26	.9995	.60	2.0	6	5
.85	.70	0.82	81	61	.9995	.70	2.1	7	5
.85	.80	0.96	608	418	.9995	.80	2.28	10	5
.90	.50	0.75	13	10	.9995	.85	2.36	12	5
.90	.60	0.86	21	15	.9995	.90	2.46	16	5
.90	.70	0.97	41	28	.9995	.95	2.61	26	6
.90	.80	1.10	131	82	.9995	.99	2.89	88	17
.90	.85	1.18	443	261	.9995	.995	3.00	170	31
.95	.50	1.0	9	6	.9995	.999	3.21	2399	391
.95	.60	1.07	13	9	.9999	.50	2.2	5	5
.95	.70	1.18	22	13	.9999	.60	2.3	5	5
.95	.80	1.31	46	25	.9999	.70	2.4	6	5
.95	.85	1.39	84	43	.9999	.80	2.5	8	5
.95	.90	1.50	254	120	.9999	.85	2.61	10	5
.99	.50	1.4	6	5	.9999	.90	2.71	13	5
.99	.60	1.5	8	5	.9999	.95	2.86	19	5
.99	.70	1.58	11	5	.9999	.99	3.14	49	9
.99	.80	1.71	18	8	.9999	.995	3.25	76	13
.99	.85	1.79	25	10	.9999	.999	3.46	279	40
.99	.90	1.89	41	15	.9999	.9995	3.54	626	87
.99	.95	2.04	106	34	.99999	.50	2.5	5	5
.995	.50	1.5	6	5	.99999	.60	2.6	5	5
.995	.60	1.6	7	5	.99999	.70	2.7	6	5
.995	.70	1.73	10	5	.99999	.80	2.8	7	5
.995	.80	1.86	15	6	.99999	.85	2.9	8	5
.995	.85	1.94	20	7	.99999	.90	3.03	10	5
.995	.90	2.04	30	10	.99999	.95	3.18	14	5
.995	.95	2.19	62	19	.99999	.99	3.46	30	5
.995	.99	2.47	1024	253	.99999	.995	3.57	41	6
.999	.50	1.8	5	5	.99999	.999	3.78	93	12
.999	.60	1.9	6	5	.99999	.9995	3.86	141	17
.999	.70	2.0	8	5	.99999	.9999	4.04	485	53
.999	.80	2.16	11	5					

TABLE I.-(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .05$
$\beta = .05$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.26	41	40	.999	.85	2.1	9	5
.70	.60	0.39	159	148	.999	.90	2.19	12	5
.80	.50	0.42	17	16	.999	.95	2.37	20	6
.80	.60	0.55	36	32	.999	.99	2.71	87	19
.80	.70	0.68	133	108	.999	.995	2.83	206	41
.85	.50	0.52	12	11	.9995	.50	1.6	5	5
.85	.60	0.65	22	18	.9995	.60	1.8	5	5
.85	.70	0.78	54	42	.9995	.70	1.9	5	5
.85	.80	0.94	413	287	.9995	.80	2.1	6	5
.90	.50	0.6	8	7	.9995	.85	2.2	8	5
.90	.60	0.77	14	11	.9995	.90	2.29	10	5
.90	.70	0.90	27	19	.9995	.95	2.47	17	5
.90	.80	1.06	88	56	.9995	.99	2.81	58	12
.90	.85	1.16	299	179	.9995	.995	2.93	113	22
.95	.50	0.8	6	5	.9995	.999	3.19	1632	268
.95	.60	0.9	9	6	.9999	.50	1.9	5	5
.95	.70	1.09	14	9	.9999	.60	2.0	5	5
.95	.80	1.24	30	17	.9999	.70	2.1	5	5
.95	.85	1.34	56	30	.9999	.80	2.3	5	5
.95	.90	1.46	171	83	.9999	.85	2.4	6	5
.99	.50	1.2	5	5	.9999	.90	2.5	8	5
.99	.60	1.3	5	5	.9999	.95	2.68	12	5
.99	.70	1.4	7	5	.9999	.99	3.02	32	6
.99	.80	1.58	12	5	.9999	.995	3.15	50	9
.99	.85	1.68	16	7	.9999	.999	3.41	186	28
.99	.90	1.80	27	10	.9999	.9995	3.51	423	60
.99	.95	1.99	70	24	.99999	.50	2.1	5	5
.995	.50	1.3	5	5	.99999	.60	2.3	5	5
.995	.60	1.4	5	5	.99999	.70	2.4	5	5
.995	.70	1.6	6	5	.99999	.80	2.6	5	5
.995	.80	1.7	9	5	.99999	.85	2.7	5	5
.995	.85	1.81	13	5	.99999	.90	2.8	6	5
.995	.90	1.93	19	7	.99999	.95	3.0	9	5
.995	.95	2.11	41	13	.99999	.99	3.30	19	5
.995	.99	2.45	694	174	.99999	.995	3.42	26	5
.999	.50	1.5	5	5	.99999	.999	3.68	61	8
.999	.60	1.7	5	5	.99999	.9995	3.78	93	12
.999	.70	1.8	5	5	.99999	.9999	3.99	326	37
.999	.80	2.0	7	5					

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .05$
$\beta = .10$

(Numbers in bold-face type computed from non-central t)

R_H	R_L	K	n	n'
.70	.50	0.23	32	32
.70	.60	0.37	125	117
.80	.50	0.37	13	13
.80	.60	0.51	28	25
.80	.70	0.66	104	86
.85	.50	0.5	9	8
.85	.60	0.60	17	14
.85	.70	0.75	42	33
.85	.80	0.93	325	227
.90	.50	0.6	7	6
.90	.60	0.70	11	9
.90	.70	0.86	21	15
.90	.80	1.03	68	45
.90	.85	1.14	235	142
.95	.50	0.7	5	5
.95	.60	**0.9162**	**6**	5
.95	.60	0.9	7	
.95	.70	**1.0386**	**11**	7
.95	.70	1.02	11	
.95	.80	**1.2047**	**24**	14
.95	.80	1.19	23	
.95	.85	**1.3129**	**44**	24
.95	.85	1.30	43	
.95	.90	**1.4408**	**133**	66
.95	.90	1.44	133	
.99	.50	1.0	5	5
.99	.60	1.2	5	5
.99	.70	1.3	5	5
.99	.80	1.5	9	5
.99	.85	1.60	12	6
.99	.90	**1.7643**	**21**	8
.99	.90	1.74	20	
.99	.95	**1.9527**	**54**	19
.99	.95	1.94	54	
.995	.50	1.1	5	5
.995	.60	1.3	5	5
.995	.70	1.4	5	5
.995	.80	1.6	7	5
.995	.85	1.7	9	5
.995	.90	1.85	14	6
.995	.95	**2.0723**	**32**	10
.995	.95	2.05	31	
.995	.99	**2.4359**	**547**	138
.995	.99	2.44	544	
.999	.50	1.4	5	5
.999	.60	1.5	5	5
.999	.70	1.6	5	5
.999	.80	1.8	5	5
.999	.85	1.9	6	5
.999	.90	2.1	9	5
.999	.95	**2.3059**	**16**	5
.999	.95	2.28	15	
.999	.99	**2.6725**	**68**	15
.999	.99	2.66	67	
.999	.995	2.80	160	33
.9995	.50	1.4	5	5
.9995	.60	1.6	5	5
.9995	.70	1.7	5	5
.9995	.80	1.9	5	5
.9995	.85	2.0	6	5
.9995	.90	2.2	8	5
.9995	.95	2.37	13	5
.9995	.99	2.75	44	10
.9995	.995	2.89	87	17
.9995	.999	3.18	1283	213
.9999	.50	1.6	5	5
.9999	.60	1.8	5	5
.9999	.70	1.9	5	5
.9999	.80	2.1	5	5
.9999	.85	2.2	5	5
.9999	.90	2.3	6	5
.9999	.95	2.6	9	5
.9999	.99	2.94	24	5
.9999	.995	3.08	38	7
.9999	.999	3.37	145	22
.9999	.9995	3.48	330	47
.99999	.50	1.9	5	5
.99999	.60	2.0	5	5
.99999	.70	2.2	5	5
.99999	.80	2.3	5	5
.99999	.85	2.5	5	5
.99999	.90	2.6	5	5
.99999	.95	2.8	7	5
.99999	.99	3.18	14	5
.99999	.995	3.32	20	5
.99999	.999	3.60	47	7
.99999	.9995	3.72	72	10
.99999	.9999	3.96	254	29

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .05$
$\beta = .20$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.18	23	23	.999	.85	1.7	5	5
.70	.60	0.35	90	85	.999	.90	1.9	6	5
.80	.50	0.29	10	9	.999	.95	2.13	10	5
.80	.60	0.45	20	18	.999	.99	2.59	46	11
.80	.70	0.63	74	62	.999	.995	2.75	112	24
.85	.50	0.4	7	6	.9995	.50	1.1	5	5
.85	.60	0.52	12	11	.9995	.60	1.3	5	5
.85	.70	0.70	30	24	.9995	.70	1.5	5	5
.85	.80	0.91	231	164	.9995	.80	1.7	5	5
.90	.50	0.4	5	5	.9995	.85	1.8	5	5
.90	.60	0.6	7	6	.9995	.90	2.0	5	5
.90	.70	0.78	15	11	.9995	.95	2.2	8	5
.90	.80	0.99	48	32	.9995	.99	2.65	30	7
.90	.85	1.12	167	103	.9995	.995	2.82	61	13
.95	.50	0.6	5	5	.9995	.999	3.16	917	154
.95	.70	0.9	7	5	.9999	.50	1.3	5	5
.95	.80	1.11	16	10	.9999	.60	1.4	5	5
.95	.85	1.24	30	17	.9999	.70	1.6	5	5
.95	.90	1.41	94	47	.9999	.80	1.8	5	5
.99	.50	0.8	5	5	.9999	.85	1.9	5	5
.99	.60	1.0	5	5	.9999	.90	2.1	5	5
.99	.70	1.1	5	5	.9999	.95	2.3	6	5
.99	.80	1.3	6	5	.9999	.99	2.80	16	5
.99	.85	1.5	8	5	.9999	.995	2.96	26	5
.99	.90	1.64	14	6	.9999	.999	3.30	101	16
.99	.95	1.88	37	14	.9999	.9995	3.44	233	34
.995	.50	0.9	5	5	.99999	.50	1.4	5	5
.995	.60	1.0	5	5	.99999	.60	1.6	5	5
.995	.70	1.2	5	5	.99999	.70	1.8	5	5
.995	.80	1.4	5	5	.99999	.80	2.0	5	5
.995	.85	1.6	6	5	.99999	.85	2.1	5	5
.995	.90	1.72	10	5	.99999	.90	2.3	5	5
.995	.95	1.96	21	8	.99999	.95	2.5	5	5
.995	.99	2.41	387	99	.99999	.99	3.0	9	5
.999	.50	1.0	5	5	.99999	.995	3.15	13	5
.999	.60	1.2	5	5	.99999	.999	3.49	32	5
.999	.70	1.4	5	5	.99999	.9995	3.62	50	7
.999	.80	1.6	5	5	.99999	.9999	3.90	179	21

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .10$
$\beta = .01$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.34	51	48	.999	.85	2.36	12	5
.70	.60	0 43	194	178	.999	.90	2.45	16	5
.80	.50	0.54	22	19	.999	.95	2.58	27	7
.80	.60	0.63	46	38	.999	.99	2.82	111	23
.80	.70	0.73	164	130	.999	.995	2.91	258	50
.85	.50	0.67	15	13	.9995	.50	2.1	5	5
.85	.60	0.76	28	22	.9995	.60	2.2	5	5
.85	.70	0.85	68	50	.9995	.70	2.3	7	5
.85	.80	0.97	506	345	.9995	.80	2.4	9	5
.90	.50	0.83	11	8	.9995	.85	2.49	11	5
.90	.60	0.92	18	13	.9995	.90	2.58	14	5
.90	.70	1.01	35	23	.9995	.95	2.71	23	5
.90	.80	1.13	110	68	.9995	.99	2.95	75	14
.90	.85	1.20	369	216	.9995	.995	3.04	143	26
.95	.50	1.1	8	5	.9995	.999	3.22	1993	323
.95	.60	1.15	12	7	.9999	.50	2.4	5	5
.95	.70	1.23	19	11	.9999	.60	2.5	5	5
.95	.80	1.36	39	21	.9999	.70	2.6	6	5
.95	.85	1.43	71	36	.9999	.80	2.7	8	5
.95	.90	1.52	213	99	.9999	.85	2.8	9	5
.99	.50	1.5	6	5	.9999	.90	2.85	12	5
.99	.60	1.6	7	5	.9999	.95	2.98	17	5
.99	.70	1.69	10	5	.9999	.99	3.22	42	7
.99	.80	1.80	16	6	.9999	.995	3.31	65	10
.99	.85	1.87	22	8	.9999	.999	3.50	234	33
.99	.90	1.96	35	12	.9999	.9995	3.57	524	72
99	.95	2.08	90	29	.99999	.50	2.8	5	5
.995	.50	1.7	5	5	.99999	.60	2.8	5	5
.995	.60	1.8	7	5	.99999	.70	2.9	5	5
.995	.70	1.8	9	5	.99999	.80	3.0	7	5
.995	.80	1.96	13	5	.99999	.85	3.1	8	5
.995	.85	2.03	17	6	.99999	.90	3.2	9	5
.995	.90	2.12	26	8	.99999	.95	3.33	13	5
.995	.95	2.25	53	16	.99999	.99	3.58	26	5
.995	.99	2.49	853	209	.99999	.995	3.67	36	5
.999	.50	2.0	5	5	.99999	.999	3.85	80	10
.999	.60	2.1	6	5	.99999	.9995	3.92	120	14
.999	.70	2.2	7	5	.99999	.9999	4.07	406	44
.999	.80	2.29	10	5					

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .10$
$\beta = .05$

R_H	R_L	K	n	n'
.70	.50	0.29	33	32
.70	.60	0.41	127	117
.80	.50	0.47	14	13
.80	.60	0.58	29	25
.80	.70	0.70	107	86
.85	.50	0.58	10	8
.85	.60	0.69	18	14
.85	.70	0.81	44	33
.85	.80	0.95	330	227
.90	.50	0.7	7	6
.90	.60	0.83	11	9
.90	.70	0.95	22	15
.90	.80	1.09	71	45
.90	.85	1.17	240	142
.95	.50	0.9	5	5
.95	.60	1.0	7	5
.95	.70	1.15	12	7
.95	.80	1.29	25	14
.95	.85	1.38	46	24
.95	.90	1.49	137	66
.99	.50	1.3	5	5
.99	.60	1.4	5	5
.99	.70	1.5	6	5
.99	.80	1.68	10	5
.99	.85	1.76	14	6
.99	.90	1.87	22	8
.99	.95	2.03	57	19
.995	.50	1.4	5	5
.995	.60	1.6	5	5
.995	.70	1.7	5	5
.995	.80	1.8	8	5
.995	.85	1.90	11	5
.995	.90	2.01	16	6
.995	.95	2.17	34	10
.995	.99	2.47	555	138
.999	.50	1.7	5	5
.999	.60	1.8	5	5
.999	.70	2.0	5	5
.999	.80	2.1	6	5
.999	.85	2.2	7	5
.999	.90	2.30	10	5
.999	.95	2.46	17	5
.999	.99	2.76	71	15
.999	.995	2.86	166	33
.9995	.50	1.8	5	5
.9995	.60	1.9	5	5
.9995	.70	2.1	5	5
.9995	.80	2.2	5	5
.9995	.85	2.3	7	5
.9995	.90	2.4	9	5
.9995	.95	2.57	14	5
.9995	.99	2.87	48	10
.9995	.995	2.98	92	17
.9995	.999	3.20	1300	213
.9999	.50	2.1	5	5
.9999	.60	2.2	5	5
.9999	.70	2.3	5	5
.9999	.80	2.5	5	5
.9999	.85	2.5	6	5
.9999	.90	2.7	7	5
.9999	.95	2.81	10	5
.9999	.99	3.11	26	5
.9999	.995	3.22	41	7
.9999	.999	3.44	151	22
.9999	.9995	3.53	339	47
.99999	.50	2.4	5	5
.99999	.60	2.5	5	5
.99999	.70	2.6	5	5
.99999	.80	2.8	5	5
.99999	.85	2.9	5	5
.99999	.90	3.0	6	5
.99999	.95	3.1	8	5
.99999	.99	3.42	16	5
.99999	.995	3.53	22	5
.99999	.999	3.75	50	7
.99999	.9995	3.84	76	10
.99999	.9999	4.03	262	29

TABLE I. -(Continued)

Factors for Testing Reliability Hypotheses

$\alpha = .20$
$\beta = .10$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.32	18	17	.999	.85	2.3	5	5
.70	.60	0.42	67	62	.999	.90	2.4	5	5
.80	.50	0.5	8	7	.999	.95	2.52	10	5
.80	.60	0.61	16	14	.999	.99	2.79	38	8
.80	.70	0.72	57	45	.999	.995	2.89	89	18
.85	.50	0.6	6	5	.9995	.50	2.0	5	5
.85	.60	0.73	10	8	.9995	.60	2.1	5	5
.85	.70	0.83	24	18	.9995	.70	2.2	5	5
.85	.80	0.96	175	120	.9995	.80	2.3	5	5
.90	.50	0.8	5	5	.9995	.85	2.4	5	5
.90	.60	0.9	6	5	.9995	.90	2.5	5	5
.90	.70	0.10	12	8	.9995	.95	2.6	8	5
.90	.80	1.11	38	24	.9995	.99	2.91	26	5
.90	.85	1.18	127	75	.9995	.995	3.01	49	5
.95	.50	0.10	5	5	.9995	.999	3.21	688	112
.95	.60	1.1	5	5	.9999	.50	2.2	5	5
.95	.70	1.2	7	5	.9999	.60	2.3	5	5
.95	.80	1.33	14	7	.9999	.70	2.5	5	5
.95	.85	1.40	25	13	.9999	.80	2.6	5	5
.95	.90	1.50	73	35	.9999	.85	2.7	5	5
.99	.50	1.4	5	5	.9999	.90	2.8	5	5
.99	.60	1.5	5	5	.9999	.95	2.9	6	5
.99	.70	1.6	5	5	.9999	.99	3.17	14	5
.99	.80	1.7	6	5	.9999	.995	3.27	22	5
.99	.85	1.8	8	5	.9999	.999	3.47	81	12
.99	.90	1.91	12	5	.9999	.9995	3.55	180	25
.99	.95	2.06	31	10	.99999	.50	2.6	5	5
.995	.50	1.6	5	5	.99999	.60	2.7	5	5
.995	.60	1.7	5	5	.99999	.70	2.8	5	5
.995	.70	1.8	5	5	.99999	.80	2.9	5	5
.995	.80	1.9	5	5	.99999	.85	3.0	5	5
.995	.85	2.0	6	5	.99999	.90	3.1	5	5
.995	.90	2.1	9	5	.99999	.95	3.2	5	5
.995	.95	2.21	18	6	.99999	.99	3.5	9	5
.995	.99	2.48	294	73	.99999	.995	3.60	12	5
.999	.50	1.9	5	5	.99999	.999	3.80	27	5
.999	.60	2.0	5	5	.99999	.9995	3.88	41	5
.999	.70	2.1	5	5	.99999	.9999	4.05	140	16
.999	.80	2.2	5	5					

TABLE I. -(Concluded)

Factors for Testing Reliability Hypotheses

$\alpha = .20$
$\beta = .20$

R_H	R_L	K	n	n'	R_H	R_L	K	n	n'
.70	.50	0.26	11	11	.999	.85	2.1	5	5
.70	.60	0.39	43	39	.999	.90	2.2	5	5
.80	.50	0.4	5	5	.999	.95	2.4	6	5
.80	.60	0.55	10	9	.999	.99	2.70	23	5
.80	.70	0.68	35	29	.999	.995	2.83	54	11
.85	.50	0.5	5	5	.9995	.50	1.6	5	5
.85	.60	0.6	6	5	.9995	.60	1.8	5	5
.85	.70	0.78	15	11	.9995	.70	1.9	5	5
.85	.80	0.93	109	75	.9995	.80	2.1	5	5
.90	.50	0.6	5	5	.9995	.85	2.2	5	5
.90	.60	0.8	5	5	.9995	.90	2.3	5	5
.90	.70	0.9	7	5	.9995	.95	2.5	5	5
.90	.80	1.06	23	15	.9995	.99	2.81	16	5
.90	.85	1.16	79	47	.9995	.995	2.93	30	6
.95	.50	0.8	5	5	.9995	.999	3.19	428	71
.95	.60	0.9	5	5	.9999	.50	1.8	5	5
.95	.70	1.1	5	5	.9999	.60	1.9	5	5
.95	.80	1.2	8	5	.9999	.70	2.1	5	5
.95	.85	1.34	15	8	.9999	.80	2.3	5	5
.95	.90	1.46	45	22	.9999	.85	2.4	5	5
.99	.50	1.2	5	5	.9999	.90	2.5	2	5
.99	.60	1.3	5	5	.9999	.95	2.7	5	5
.99	.70	1.4	5	5	.9999	.99	3.0	9	5
.99	.80	1.6	5	5	.9999	.995	3.15	13	5
.99	.85	1.7	5	5	.9999	.999	3.40	49	8
.99	.90	1.8	7	5	.9999	.9995	3.51	111	16
.99	.95	1.99	19	5	.99999	.50	2.1	5	5
.995	.50	1.3	5	5	.99999	.60	2.3	5	5
.995	.60	1.4	5	5	.99999	.70	2.4	5	5
.995	.70	1.6	5	5	.99999	.80	2.6	5	5
.995	.80	1.7	5	5	.99999	.85	2.7	5	5
.995	.85	1.8	5	5	.99999	.90	2.8	5	5
.995	.90	1.9	5	5	.99999	.95	3.0	5	5
.995	.95	2.11	11	5	.99999	.99	3.3	5	5
.995	.99	2.45	182	46	.99999	.995	3.4	5	5
.999	.50	1.5	5	5	.99999	.999	3.68	16	5
.999	.60	1.7	5	5	.99999	.9995	3.78	25	5
.999	.70	1.8	5	5	.99999	.9999	4.00	86	10
.999	.80	2.0	5	5					

TABLE II

K_c - FACTORS FOR ONE-SIDED TOLERANCE LIMITS*

95% confidence level

Sample size	Reliability																							
	.9999	.9998	.9996	.9995	.9994	.9992	.9991	.999	.998	.996	.995	.994	.992	.99	.98	.97	.96	.95	.94	.93	.92	.91	.90	.88	.85
5	9.0	8.6	8.1	8.0	7.9	7.7	7.6	**7.51** 7.5	7.0	**6.49** 6.5	6.3	6.2	5.9	**5.75** 5.8	5.1	4.7	**4.44** 4.4	4.2	4.0	3.8	3.7	3.5	**3.41** 3.4	3.1	**2.89** 2.8
6	7.9	7.5	7.1	7.0	6.9	6.7	6.6	**6.61** 6.6	6.1	**5.72** 5.7	5.5	5.4	5.2	**5.06** 5.0	4.5	4.1	**3.92** 3.9	3.7	3.5	3.3	3.2	3.1	**3.00** 3.0	2.8	**2.55** 2.5
7	7.2	6.9	6.5	6.4	6.3	6.1	6.1	**6.06** 6.0	5.6	**5.24** 5.2	5.1	4.9	4.7	**4.64** 4.6	4.1	3.8	**3.59** 3.5	3.4	3.2	3.1	2.9	2.8	**2.76** 2.7	2.5	**2.33** 2.3
8	6.7	6.4	6.1	6.0	5.9	5.7	5.7	**5.69** 5.6	5.3	**4.92** 4.9	4.7	4.6	4.5	**4.35** 4.3	3.8	3.5	**3.37** 3.3	3.1	3.0	2.9	2.7	2.6	**2.88** 2.5	2.4	**2.19** 2.1
9	6.4	6.1	5.8	5.7	5.6	5.5	5.4	**5.41** 5.4	5.0	**4.69** 4.6	4.5	4.4	4.2	**4.14** 4.1	3.7	3.4	**3.21** 3.2	3.0	2.8	2.7	2.6	2.5	**2.45** 2.4	2.3	**2.07** 2.0
10	6.2	5.9	5.6	5.5	5.4	5.3	5.2	**5.20** 5.2	4.8	**4.50** 4.5	4.3	4.2	4.1	**3.98** 3.9	3.5	3.2	**3.07** 3.0	2.9	2.7	2.6	2.5	2.4	**2.35** 2.3	2.2	**1.99** 2.0
11	6.0	5.7	5.4	5.3	5.2	5.1	5.0	**5.03** 5.0	4.7	**4.36** 4.3	4.2	4.1	3.9	**3.85** 3.8	3.4	3.1	**2.97** 2.9	2.8	2.6	2.5	2.4	2.3	**2.28** 2.2	2.1	**1.92** 1.9
12	5.8	5.5	5.3	5.2	5.1	5.0	4.9	**4.90** 4.9	4.5	**4.24** 4.2	4.1	4.0	3.8	**3.74** 3.7	3.3	3.1	**2.89** 2.9	2.7	2.6	2.5	2.4	2.3	**2.21** 2.2	2.0	**1.87** 1.8
13	5.7	5.4	5.1	5.0	5.0	4.8	4.8	**4.78** 4.7	4.4	**4.14** 4.1	4.0	3.9	3.7	**3.66** 3.6	3.2	3.0	**2.82** 2.8	2.6	2.5	2.4	2.3	2.2	**2.15** 2.1	2.0	**1.82** 1.8
14	5.6	5.3	5.0	4.9	4.9	4.8	4.7	**4.69** 4.7	4.3	**4.06** 4.0	3.9	3.8	3.7	**3.58** 3.6	3.2	2.9	**2.77** 2.7	2.6	2.5	2.4	2.3	2.2	**2.11** 2.1	1.9	**1.77** 1.8
15	5.5	5.2	4.9	4.9	4.8	4.7	4.6	**4.61** 4.6	4.3	**3.98** 4.0	3.8	3.8	3.6	**3.52** 3.5	3.1	2.9	**2.71** 2.7	2.5	2.4	2.3	2.2	2.1	**2.07** 2.0	1.90	**1.74** 1.72
20	5.1	4.9	4.6	4.6	4.5	4.4	4.3	**4.32** 4.3	4.0	**3.73** 3.7	3.6	3.5	3.4	**3.29** 3.3	2.9	2.7	**2.53** 2.5	2.4	2.3	2.2	2.1	1.99	**1.93** 1.91	1.77	**1.62** 1.60
25	4.9	4.7	4.5	4.4	4.3	4.2	4.2	**4.14** 4.1	3.8	**3.58** 3.6	3.5	3.4	3.2	**3.15** 3.1	2.8	2.6	**2.43** 2.4	2.3	2.2	2.07	1.98	1.90	**1.84** 1.83	1.68	**1.54** 1.53
30	4.80	4.57	4.34	4.26	4.19	4.09	4.04	**4.02** 4.01	3.74	**3.47** 3.46	3.36	3.28	3.15	**3.07** 3.05	2.71	2.50	**2.35** 2.34	2.21	2.10	2.00	1.92	1.84	**1.78** 1.77	1.64	**1.48** 1.47
40	4.62	4.40	4.17	4.10	4.03	3.93	3.89	**3.87** 3.85	3.60	**3.34** 3.32	3.23	3.15	3.03	**2.94** 2.93	2.60	2.40	**2.25** 2.24	2.12	2.01	1.92	1.83	1.76	**1.70** 1.69	1.57	**1.41** 1.40
50	4.50	4.29	4.06	3.99	3.93	3.83	3.79	**3.76** 3.76	3.51	**3.25** 3.24	3.15	3.07	2.95	**2.86** 2.85	2.54	2.33	**2.19** 2.18	2.06	1.95	1.86	1.78	1.71	**1.64** 1.64	1.52	**1.37** 1.36
100	4.24	4.04	3.83	3.76	3.70	3.64	3.57	3.53	3.30	3.04	2.96	2.89	2.77	2.68	2.38	2.18	2.04	1.92	1.82	1.74	1.66	1.59	1.52	1.41	1.26
1000	3.87	3.69	3.49	3.43	3.37	3.30	3.25	3.20	3.00	2.77	2.69	2.62	2.52	2.43	2.15	1.97	1.84	1.73	1.63	1.55	1.48	1.41	1.35	1.24	1.10

*Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values.

TABLE II. - Continued

95% confidence level

Sample size	Reliability																							
	.83	.80	.75	.70	.68	.65	.63	.60	.58	.55	.50	.48	.45	.40	.38	.35	.30	.28	.25	.20	.15	.10	.05	.21
5	2.7	2.4	2.15 2.1	1.8	1.7	1.5	1.5	1.3	1.1	1.1	0.90	0.83	0.72	0.55	0.48	0.38	0.22	0.16	0.062	-0.10	-0.28	-0.49	-0.78	-1.3
6	2.3	2.1	1.90 1.8	1.6	1.5	1.4	1.3	1.2	1.1	0.96	0.79	0.72	0.62	0.46	0.40	0.30	0.15	0.090	-0.003	-0.16	-0.34	-0.55	-0.84	-1.4
7	2.1	2.0	1.73 1.7	1.5	1.4	1.2	1.2	1.1	0.98	0.87	0.71	0.64	0.55	0.40	0.34	0.25	0.099	0.039	-0.052	-0.21	-0.39	-0.60	-0.89	-1.4
8	2.0	1.8	1.62 1.6	1.4	1.3	1.2	1.1	0.98	0.91	0.81	0.65	0.59	0.50	0.35	0.29	0.20	0.058	-0.001	-0.092	-0.25	-0.43	-0.64	-0.93	-1.5
9	1.9	1.7	1.54 1.5	1.3	1.2	1.1	1.0	0.92	0.85	0.76	0.60	0.54	0.45	0.31	0.25	0.17	0.024	-0.034	-0.12	-0.28	-0.46	-0.67	-0.97	-1.5
10	1.8	1.7	1.46 1.4	1.2	1.2	1.0	0.97	0.87	0.81	0.72	0.56	0.51	0.42	0.28	0.22	0.14	-0.004	-0.062	-0.15	-0.31	-0.48	-0.70	-1.00	-1.54
11	1.8	1.6	1.41 1.4	1.2	1.1	1.0	0.94	0.84	0.77	0.68	0.53	0.48	0.39	0.25	0.20	0.11	-0.028	-0.086	-0.17	-0.33	-0.51	-0.72	-1.02	-1.57
12	1.7	1.6	1.37 1.3	1.1	1.07	0.97	0.90	0.81	0.74	0.65	0.51	0.45	0.37	0.23	0.17	0.091	-0.049	-0.11	-0.19	-0.35	-0.53	-0.74	-1.05	-1.59
13	1.7	1.5	1.33 1.31	1.11	1.04	0.94	0.87	0.78	0.72	0.63	0.48	0.43	0.35	0.21	0.15	0.072	-0.068	-0.12	-0.21	-0.37	-0.54	-0.76	-1.07	-1.62
14	1.64	1.49	1.30 1.28	1.09	1.02	0.91	0.85	0.76	0.70	0.61	0.46	0.41	0.33	0.19	0.14	0.055	-0.084	-0.14	-0.23	-0.38	-0.56	-0.78	-1.08	-1.64
15	1.61	1.46	1.27 1.25	1.06	0.99	0.89	0.83	0.74	0.68	0.59	0.45	0.39	0.31	0.17	0.12	0.040	-0.10	-0.16	-0.24	-0.40	-0.58	-0.79	-1.10	-1.66
20	1.50	1.36	1.17 1.15	0.97	0.91	0.81	0.75	0.66	0.60	0.52	0.38	0.33	0.25	0.12	0.063	-0.017	-0.15	-0.21	-0.30	-0.45	-0.63	-0.85	-1.16	-1.73
25	1.43	1.29	1.11 1.09	0.92	0.85	0.76	0.70	0.61	0.56	0.47	0.34	0.29	0.21	0.076	0.024	-0.056	-0.19	-0.25	-0.34	-0.49	-0.67	-0.89	-1.21	-1.79
30	1.38	1.24	1.06 1.05	0.88	0.82	0.72	0.66	0.58	0.52	0.44	0.31	0.26	0.18	0.047	-0.005	-0.084	-0.22	-0.28	-0.36	-0.52	-0.70	-0.92	-1.24	-1.83
40	1.31	1.18	1.00 0.99	0.83	0.76	0.67	0.62	0.53	0.48	0.40	0.26	0.21	0.14	0.007	-0.045	-0.12	-0.26	-0.32	-0.40	-0.56	-0.74	-0.97	-1.29	-1.89
50	1.27	1.14	0.96 0.96	0.79	0.73	0.64	0.58	0.50	0.45	0.37	0.24	0.18	0.11	-0.021	-0.073	-0.15	-0.29	-0.34	-0.43	-0.59	-0.77	-1.00	-1.32	-1.93
100	1.17	1.05	0.87	0.71	0.65	0.56	0.51	0.43	0.37	0.29	0.17	0.11	0.039	-0.088	-0.140	-0.22	-0.35	-0.41	-0.50	-0.66	-0.84	-1.11	-1.41	-2.04
1000	1.02	0.903	0.733	0.580	0.523	0.440	0.386	0.307	0.255	0.178	0.052	0.002	-0.074	-0.201	-0.253	-0.332	-0.470	-0.527	-0.618	-0.782	-0.973	-1.21	-1.57	-2.23

* Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values.

TABLE II. -(Continued)*

90% confidence level

Sample size	Reliability																							
	.9999	.9998	.9996	.9995	.9994	.9992	.9991	.999	.998	.996	.995	.994	.992	.99	.98	.97	.96	.95	.94	.93	.92	.91	.90	.88	.85
5	6.9	6.6	6.2	6.1	6.0	5.9	5.8	6.11 5.8	5.4	5.28 5.0	4.8	4.7	4.5	4.67 4.4	3.9	3.6	3.60 3.4	3.2	3.1	2.9	2.8	2.7	2.75 2.6	2.4	2.31 2.2
6	6.3	6.0	5.7	5.6	5.5	5.4	5.4	5.56 5.3	5.0	4.80 4.6	4.5	4.4	4.2	4.24 4.0	3.6	3.3	3.27 3.1	3.0	2.8	2.7	2.6	2.5	2.49 2.4	2.2	2.10 2.0
7	6.0	5.7	5.4	5.3	5.2	5.1	5.1	5.20 5.0	4.7	4.49 4.3	4.2	4.1	3.9	3.97 3.8	3.4	3.1	3.06 2.9	2.8	2.6	2.5	2.4	2.3	2.33 2.2	2.1	1.96 1.9
8	5.7	5.5	5.2	5.1	5.0	4.9	4.8	4.96 4.8	4.5	4.28 4.1	4.0	3.9	3.8	3.78 3.7	3.3	3.0	2.91 2.8	2.7	2.5	2.4	2.3	2.2	2.22 2.1	2.0	1.86 1.8
9	5.5	5.3	5.0	4.9	4.9	4.7	4.7	4.77 4.6	4.3	4.12 4.0	3.9	3.8	3.7	3.64 3.5	3.2	2.9	2.80 2.7	2.6	2.4	2.3	2.2	2.2	2.13 2.1	1.9	1.79 1.7
10	5.4	5.1	4.9	4.8	4.7	4.6	4.6	4.63 4.5	4.2	3.99 3.9	3.8	3.7	3.6	3.53 3.4	3.1	2.8	2.71 2.6	2.5	2.4	2.3	2.2	2.1	2.07 2.0	1.9	1.74 1.7
11	5.3	5.0	4.8	4.7	4.6	4.5	4.5	4.51 4.4	4.1	3.90 3.8	3.7	3.6	3.5	3.44 3.4	3.0	2.8	2.65 2.6	2.4	2.3	2.2	2.1	2.0	2.01 2.0	1.8	1.69 1.6
12	5.2	4.9	4.7	4.6	4.5	4.4	4.4	4.42 4.3	4.0	3.82 3.7	3.6	3.6	3.4	3.37 3.3	2.9	2.7	2.59 2.5	2.4	2.3	2.2	2.1	2.0	1.96 1.9	1.8	1.65 1.6
13	5.1	4.9	4.6	4.5	4.5	4.3	4.3	4.34 4.3	4.0	3.75 3.7	3.6	3.5	3.4	3.30 3.2	2.9	2.7	2.55 2.5	2.4	2.2	2.1	2.0	2.0	1.93 1.9	1.8	1.61 1.6
14	5.0	4.8	4.5	4.5	4.4	4.3	4.2	4.27 4.2	3.9	3.69 3.6	3.5	3.4	3.3	3.26 3.2	2.8	2.6	2.51 2.5	2.3	2.2	2.1	2.0	1.9	1.90 1.9	1.7	1.59 1.6
15	5.0	4.7	4.5	4.4	4.3	4.2	4.2	4.21 4.1	3.9	3.64 3.6	3.5	3.4	3.3	3.22 3.2	2.8	2.6	2.46 2.4	2.3	2.2	2.1	2.0	1.9	1.86 1.8	1.7	1.57 1.5
20	4.7	4.5	4.3	4.2	4.1	4.0	4.0	4.01 4.0	3.7	3.46 3.4	3.3	3.2	3.1	3.05 3.0	2.7	2.5	2.34 2.3	2.2	2.1	2.0	1.9	1.81	1.76 1.7	1.62	1.47 1.45
25	4.6	4.4	4.2	4.1	4.0	3.9	3.9	3.88 3.8	3.6	3.35 3.3	3.2	3.1	3.0	2.95 2.9	2.6	2.4	2.26 2.2	2.1	2.01	1.91	1.83	1.75	1.70 1.69	1.56	1.41 1.40
30	4.51	4.30	4.08	4.00	3.94	3.84	3.80	3.80 3.77	3.51	3.27 3.25	3.16	3.08	2.96	2.88 2.86	2.54	2.34	2.20 2.19	2.06	1 96	1.87	1.79	1.71	1.66 1.64	1.52	1.38 1.36
40	4.39	4.18	3.96	3.89	3.83	3.73	3.69	3.68 3.66	3.41	3.17 3.15	3.06	2.99	2.87	2.79 2.78	2.47	2.27	2.13 2.12	2.00	1.90	1.81	1.73	1.66	1.60 1.59	1.47	1.32 1.31
50	4.31	4.10	3.89	3.82	3.76	3.66	3.62	3.60 3.59	3.36	3.11 3.09	3.01	2.93	2.82	2.73 2.72	2.42	2.22	2.09 2.08	1.96	1.86	1.77	1.69	1.62	1.56 1.55	1.43	1.29 1.28
100	4.12	3.92	3.71	3.65	3.59	3.50	3.46	3.43	3.20	2.95	2.87	2.80	2.69	2.59	2.30	2.11	1.97	1.86	1.76	1.68	1.60	1.53	1.47	1.35	1.20
1000	3.84	3.65	3.46	3.34	3.34	3.26	3.22	3.19	2.97	2.74	2.66	2.60	2.49	2.41	2.13	1.95	1.82	1.71	1.62	1.54	1.46	1.40	1.34	1.23	1.09

* Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values. Round - off for sample size n of 15 or less is based on comparison of bold-face value for n to modern type value for n-1. (See discussion on accuracy, p. 7)

TABLE II. -(Continued)*

90% confidence level

Sample size	Reliability																							
	.83	.80	.75	.70	.68	.65	.63	.60	.58	.55	.50	.48	.45	.40	.38	.35	.30	.28	.25	.20	.15	.10	.05	.01
5	2.0	1.9	1.70 1.6	1.4	1.3	1.2	1.1	0.98	0.91	0.80	.64	.58	.49	.34	.28	.19	.049	-.0094	-.10	-.26	-.43	-.64	-.93	-1.5
6	1.9	1.7	1.54 1.5	1.3	1.2	1.1	0.99	0.89	0.82	0.73	.57	.51	.43	.28	.23	.14	-.0010	-.059	-.15	-.30	-.48	-.69	-.94	-1.5
7	1.8	1.6	1.43 1.4	1.2	1.1	0.99	0.93	0.83	0.76	0.67	.52	.46	.38	.24	.18	.10	-.040	-.097	-.19	-.34	-.52	-.73	-1.03	-1.6
8	1.7	1.5	1.36 1.3	1.1	1.0	0.94	0.88	0.78	0.72	0.63	.48	.43	.34	.21	.15	.069	-.071	-.13	-.22	-.37	-.55	-.76	-1.1	-1.6
9	1.6	1.5	1.30 1.3	1.1	1.0	0.90	0.84	0.74	0.68	0.59	.45	.40	.31	.18	.12	.042	-.096	-.15	-.24	-.40	-.57	-.79	-1.1	-1.6
10	1.6	1.4	1.26 1.2	1.0	0.97	0.87	0.80	0.71	0.65	0.57	.43	.37	.29	.15	.10	.020	-.12	-.17	-.26	-.42	-.59	-.81	-1.1	-1.7
11	1.5	1.4	1.22 1.2	1.0	0.94	0.84	0.78	0.69	0.63	0.54	.40	.35	.27	.14	.082	.0012	-.14	-.19	-.28	-.44	-.61	-.83	-1.1	-1.7
12	1.5	1.4	1.19 1.2	0.98	0.91	0.82	0.76	0.67	0.61	0.52	.38	.33	.25	.12	.065	-.015	-.15	-.21	-.30	-.45	-.63	-.85	-1.2	-1.7
13	1.5	1.3	1.16 1.14	0.96	0.89	0.80	0.74	0.65	0.59	0.51	.37	.31	.24	.10	.050	-.030	-.17	-.22	-.31	-.47	-.64	-.86	-1.2	-1.7
14	1.5	1.32	1.14 1.12	0.94	0.87	0.78	0.72	0.63	0.57	0.49	.35	.30	.22	.090	.037	-.043	-.18	-.24	-.32	-.48	-.66	-.87	-1.2	-1.8
15	1.43	1.30	1.12 1.10	0.92	0.86	0.76	0.70	0.62	0.56	0.48	.34	.29	.21	.078	.026	-.054	-.19	-.25	-.33	-.49	-.67	-.89	-1.2	-1.8
20	1.36	1.23	1.04 1.03	0.86	0.80	0.71	0.65	0.56	0.51	0.43	.29	.24	.16	.033	-.019	-.098	-.23	-.29	-.38	-.53	-.71	-.94	-1.26	-1.85
25	1.31	1.18	1.00 0.99	0.82	0.76	0.67	0.61	0.53	0.47	0.39	.26	.21	.13	.031	-.049	-.13	-.26	-.32	-.41	-.56	-.74	-.97	-1.29	-1.89
30	1.27	1.14	0.96 0.96	0.79	0.73	0.64	0.59	0.50	0.45	0.37	.24	.19	.11	-.019	-.071	-.15	-.29	-.34	-.43	-.59	-.77	-.99	-1.32	-1.93
40	1.22	1.10	0.93 0.92	0.75	0.69	0.61	0.55	0.47	0.41	0.33	.20	.15	.077	-.050	-.10	-.18	-.32	-.37	-.46	-.62	-.80	-1.03	-1.36	-1.97
50	1.19	1.07	0.89 0.89	0.73	0.67	0.58	0.53	0.44	0.39	0.31	.18	.13	.056	-.072	-.12	-.20	-.34	-.39	-.48	-.64	-.83	-1.05	-1.39	-2.01
100	1.12	1.00	0.82	0.67	0.61	0.52	0.47	0.39	0.33	0.26	.13	.078	.0025	-.512	-.18	-.26	-.39	-.45	-.54	-.70	-.88	-1.12	-1.46	-2.10
1000	1.00	0.889	0.720	0.568	0.511	0.428	0.374	0.295	0.243	0.167	.041	-.010	-.085	-.212	-.264	-.344	-.482	-.539	-.630	-.795	-.987	-1.23	-1.58	-2.25

* Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values. Round - off for sample size n of 15 or less is based on comparison of bold-face value for n to modern type value for n-1. (See discussion on accuracy, p. 7)

TABLE II. - (Continued)*

50% confidence level

Sample size	Reliability																								
	.9999	.9998	.9996	.9995	.9994	.9992	.9991	.999	.998	.996	.995	.994	.992	.99	.98	.97	.96	.95	.94	.93	.92	.91	.90	.88	.85
5	3.96	3.77	3.57	3.50	3.45	3.36	3.32	3.36 3.29	3.06	2.88 2.82	2.74	2.67	2.56	2.52 2.47	2.19	2.00	1.90 1.86	1.75	1.65	1.57	1.50	1.43	1.39 1.36	1.25	1.12 1.10
6	3.91	3.72	3.52	3.46	3.40	3.32	3.28	3.30 3.25	3.02	2.83 2.79	2.71	2.64	2.53	2.48 2.44	2.16	1.98	1.86 1.84	1.73	1.63	1.55	1.48	1.41	1.36 1.35	1.24	1.10 1.09
7	3.88	3.69	3.50	3.43	3.38	3.29	3.25	3.27 3.22	3.00	2.81 2.76	2.69	2.62	2.51	2.45 2.43	2.14	1.96	1.84 1.83	1.72	1.62	1.54	1.47	1.40	1.34 1.34	1.23	1.08 1.08
8	3.85	3.67	3.47	3.41	3.36	3.27	3.23	3.24 3.20	2.98	2.78 2.75	2.67	2.60	2.50	2.43 2.41	2.13	1.95	1.83 1.81	1.71	1.61	1.53	1.46	1.39	1.34 1.33	1.22	1.08 1.07
9	3.84	3.65	3.46	3.40	3.34	3.26	3.22	3.22 3.19	2.97	2.76 2.74	2.66	2.59	2.49	2.42 2.40	2.12	1.94	1.82 1.81	1.70	1.60	1.52	1.45	1.38	1.33 1.32	1.21	1.07 1.07
10	3.82	3.64	3.45	3.38	3.33	3.25	3.21	3.21 3.18	2.96	2.75 2.73	2.65	2.58	2.48	2.41 2.39	2.11	1.93	1.81 1.80	1.69	1.60	1.52	1.44	1.38	1.33 1.32	1.21	1.07 1.07
11	3.81	3.63	3.44	3.37	3.32	3.24	3.20	3.19 3.17	2.95	2.74 2.72	2.64	2.58	2.47	2.40 2.39	2.11	1.93	1.80 1.80	1.69	1.59	1.51	1.44	1.38	1.32 1.31	1.21	1.07 1.06
12	3.81	3.62	3.43	3.37	3.31	3.23	3.19	3.18 3.16	2.94	2.73 2.71	2.64	2.57	2.46	2.39 2.38	2.10	1.92	1.80 1.79	1.68	1.59	1.51	1.44	1.37	1.32 1.31	1.20	1.06 1.06
13	3.80	3.61	3.42	3.36	3.31	3.22	3.19	3.17 3.16	2.94	2.73 2.71	2.63	2.57	2.46	2.39 2.38	2.10	1.92	1.80 1.79	1.68	1.59	1.51	1.44	1.37	1.32 1.31	1.20	1.06 1.06
14	3.80	3.61	3.42	3.36	3.30	3.22	3.18	3.17 3.15	2.93	2.72 2.70	2.63	2.56	2.46	2.38 2.37	2.09	1.92	1.79 1.79	1.68	1.59	1.51	1.43	1.37	1.31 1.31	1.20	1.06 1.06
15	3.79	3.60	3.41	3.35	3.30	3.21	3.18	3.16 3.15	2.93	2.71 2.70	2.62	2.56	2.45	2.38 2.37	2.09	1.92	1.79 1.78	1.68	1.58	1.50	1.43	1.37	1.31 1.31	1.20	1.05 1.06
20	3.77	3.59	3.40	3.33	3.28	3.20	3.16	3.14 3.13	2.92	2.70 2.69	2.61	2.55	2.44	2.37 2.36	2.08	1.91	1.78 1.77	1.67	1.58	1.50	1.42	1.36	1.31 1.30	1.19	1.05 1.05
25	3.76	3.58	3.39	3.33	3.27	3.19	3.15	3.14 3.12	2.91	2.70 2.68	2.60	2.538	2.43	2.36 2.35	2.08	1.90	1.77 1.77	1.66	1.57	1.49	1.42	1.36	1.30 1.30	1.19	1.05 1.05
30	3.75	3.57	3.38	3.32	3.27	3.18	3.15	3.13 3.12	2.90	2.68 2.68	2.60	2.53	3.43	2.35 2.35	2.07	1. 90	1.77 1.77	1.66	1.57	1.49	1.42	1.35	1.30 1.29	1.18	1.05 1.04
40	3.74	3.56	3.37	3.31	3.26	3.18	3.14	3.12 3.11	2.90	2.68 2.67	2.59	2.53	2.42	2.35 2.34	2.07	1.89	1.77 1.76	1.66	1.56	1.48	1.41	1.35	1.29 1.29	1.18	1.04 1.04
50	3.74	3.56	3.37	3.31	3.26	3.17	3.14	3.11 3.11	2.89	2.67 2.67	2.59	2.52	2.42	2.35 2.34	2.06	1.89	1.76 1.76	1.65	1.56	1.48	1.41	1.35	1.29 1.29	1.18	1.04 1.04
100	3.73	3.55	3.36	3.0	3.25	3.16	3.13	3.10	2.88	2.66	2.58	2.52	2.42	2.33	2.06	1.89	1.76	1.65	1.56	1.48	1.41	1.34	1.28	1.18	1.04
1000	3.720	3.541	3.354	3.292	3.240	3.157	3.122	3.091	2.879	2.653	2.577	2.513	2.410	2.327	3.055	1.881	1.751	1.645	1.555	1.476	1.405	1.341	1.282	1.175	1.036

* This table accounts for the bias of the estimated standard deviation as a function of sample size.

Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values. Round-off for sample size n of 15 or less is based on comparison of bold-face value for n to modern type value for n-1. (See discussion on accuracy, p. 7)

Sample size	Reliability																						
	.83	.80	.75	.70	.68	.65	.63	.60	.58	.55	.50	.48	.45	.40	.38	.35	.30	.28	.25	.20	15	.10	.05	.01
5	1.02	.895	**0.72** .718	.558	.498	.410	.353	.269	.215	.134	0	-.0534	-.134	-.269	-.325	-.410	-.558	-.620	-.718	-.895	-1.10	-1.86	-1.75	-2.47
6	1.00	.884	**0.71** .709	.551	.492	.405	.349	.266	.212	.132	0	-.0528	-.132	-.266	-.321	-.405	-.551	-.612	-.709	-.884	-1.09	-1.35	-1.73	-2.44
7	.995	.877	**0.70** .703	.547	.488	.402	.346	.264	.210	.131	0	-.0523	-.131	-.264	-.318	-.402	-.547	-.608	-.703	-.877	-1.08	-1.34	-1.72	-2.43
8	.989	.872	**0.70** .699	.543	.485	.399	.344	.262	.209	.130	0	-.0520	-.130	-.262	-.317	-.399	-.543	-.604	-.699	-.872	-1.07	-1.33	-1.71	-2.41
9	.984	.868	**0.70** .696	.541	.483	.398	.342	.261	.208	.130	0	-.0518	-.130	-.261	-.315	-.398	-.541	-.601	-.696	-.868	-1.07	-1.32	-1.70	-2.40
10	.981	.865	**0.69** .693	.539	.481	.396	.341	.260	.208	.129	0	-.0516	-.129	-.260	-.314	-.396	-.539	-.599	-.693	-.865	-1.06	-1.32	-1.69	-2.39
11	.978	.863	**0.69** .692	.538	.480	.395	.340	.260	.207	.129	0	-.0515	-.129	-.260	-.313	-.395	-.538	-.598	-.692	-.863	-1.062	-1.314	-1.69	-2.39
12	.976	.861	**0.69** .690	.536	.478	.394	.340	.259	.207	.129	0	-.0514	-.129	-.259	-.313	-.394	-.536	-.596	-.690	-.861	-1.060	-1.311	-1.68	-2.38
13	.974	.859	**0.69** .689	.535	.478	.393	.339	.259	.206	.128	0	-.0513	-.128	-.259	-.312	-.393	-.535	-.595	-.689	-.859	-1.058	-1.309	-1.68	-2.38
14	.972	.858	**0.69** .688	.535	.477	.393	.338	.258	.206	.128	0	-.0512	-.128	-.258	-.311	-.393	-.535	-.594	-.688	-.858	-1.056	-1.307	-1.677	-2.37
15	.971	.857	**0.69** .687	.534	.476	.392	.338	.258	.206	.128	0	-.0511	-.128	-.258	-.311	-.392	-.534	-.593	-.687	-.857	-1.055	-1.305	-1.675	-2.37
20	.967	.853	**0.68** .683	.531	.474	.390	.336	.257	.205	.127	0	-.0509	-.127	-.257	-.310	-.390	-.531	-.590	-.683	-.853	-1.050	-1.299	-1.667	-2.36
25	.964	.850	**0.68** .682	.530	.473	.389	.335	.256	.204	.127	0	-.0507	-.127	-.256	-.309	-.389	-.530	-.589	-.682	-.850	-1.047	-1.295	-1.662	-2.35
30	.963	.849	**0.68** .680	.529	.472	.389	.335	.256	.204	.127	0	-.0506	-.127	-.256	-.308	-.389	-.529	-.588	-.680	-.849	-1.045	-1.293	-1.659	-2.346
40	.960	.847	**0.68** .679	.528	.471	.388	.334	.255	.203	.127	0	-.0505	-.127	-.255	-.307	-.388	-.528	-.587	-.679	-.847	-1.043	-1.290	-1.656	-2.341
50	.959	.846	**0.68** .678	.527	.470	.387	.334	.255	.203	.126	0	-.0505	-.126	-.255	-.307	-.387	-.527	-.586	-.678	-.846	-1.041	-1.289	-1.653	-2.338
100	.957	.844	.676	.526	.469	.386	.333	.254	.202	.126	0	-.0503	-.126	-.254	-.306	-.386	-.526	-.584	-.676	-.844	-1.039	-1.285	-1.649	-2.332
1,000	.954	.842	.675	.525	.468	.385	.332	.253	.202	.126	0	-.0502	-.126	-.253	-.306	-.385	-.525	-.583	-.675	-.842	-1.036	-1.282	-1.645	-2.327

* Values in bold-face type, computed from the noncentral t-statistic, are more accurate than modern type values. Round-off for sample size n of 15 or less is based on comparison of bold-face value for n to modern type value for n-1. (See discussion on accuracy, p. 7)

REFERENCES

1. A. Hald, Statistical Theory with Engineering Application, John Wiley & Sons, N. Y. 1960, p.303.

2. Ibid. p. 311.

3. Eisenhart, Hastay and Wallace, Techniques of Statistical Analysis, McGraw-Hill, 1947, Chapter 1.

4. Tables of Normal Probability Functions, U. S. Department of Commerce, National Bureau of Standards.

5. National Bureau of Standards Report 4817-A, Manual on Experimental Statistics for Ordnance Engineers, p. 55.

6. One-Sided Tolerance Factors for Normal Distributions, A Thesis submitted to the Graduate Council of the University of Tennessee in Partial Fulfillment of the Requirements for the Degree of Master of Science, Parkhurst B. Wood, August, 1954.

7. Tables for One-Sided Statistical Tolerance Limits, by Gerald J.Lieberman, Technical Report No. 34, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California, 1 November 1957.

8. Use of Tolerance Limits in Missile Evaluation, Albert Madansky, the RAND Corporation, 14 July 1958.

9. Resnikoff, George J., and Lieberman, Gerald J., Tables of the Non-Central t-Distribution, Stanford University Press, Stanford, California (1957).

10. Johnson, N. L., and Welch, B. L., Applications of the Non-Central t-Distribution, Biometrika, Vol. 31 (1940).

A GENERAL APPROACH TO ENGINEERING TOLERANCE SPECIFICATIONS

Sheldon G. Levin
Diamond Ordnance Fuze Laboratories

INTRODUCTION. When a complex device or a system composed of a set of components is to be mass produced, the limits within which the components are to be made must be specified in order to have assurance that the system will function as intended. These limits are called either "tolerance" limits or "specification" limits and a lack of proper distinction between the two terms causes confusion. Both express uncertainty, but statistical tolerance limits are a precise statement associated with the uncertainty of individual observations, usually based on the variance of the underlying population, while specification limits are merely instructions to the component maker and convey the range of values within which all components must lie to be acceptable.

This discussion will provide techniques for estimating the limits within which a given percent of the systems responses will fall for a specified set of Engineering Tolerance Specifications (ETS). While none of the techniques is new, the assembly of methods which lead to the selection of the set of ETS on the individual components, either to achieve minimum ETS in the system for a given cost or to minimize cost for a given set of ETS, should be a useful contribution.

Given a set of nominal requirements R_i, $i = 1, \ldots, n$, on the individual components, there is usually associated with each R_i a symmetric interval $R_i \pm \delta_i$ within which the values of the component must lie to be acceptable. The interval length for the R_i component* is $2\delta_i$ and the limits will be written as $L_{1i} = R_i - \delta_i$ and $L_{2i} = R_i + \delta_i$. The limits on the system are not necessarily symmetrical and will simply be written as L_{1y} and L_{2y}; they will be specified in advance with the requirement that a fixed percent of the systems produced fall in the interval, or they will be selected to cut off equal percentages from the tails of the distribution. An illustration for $n = 3$ is given below.

---

*Component refers throughout the discussion to measurements of a characteristic of the component.

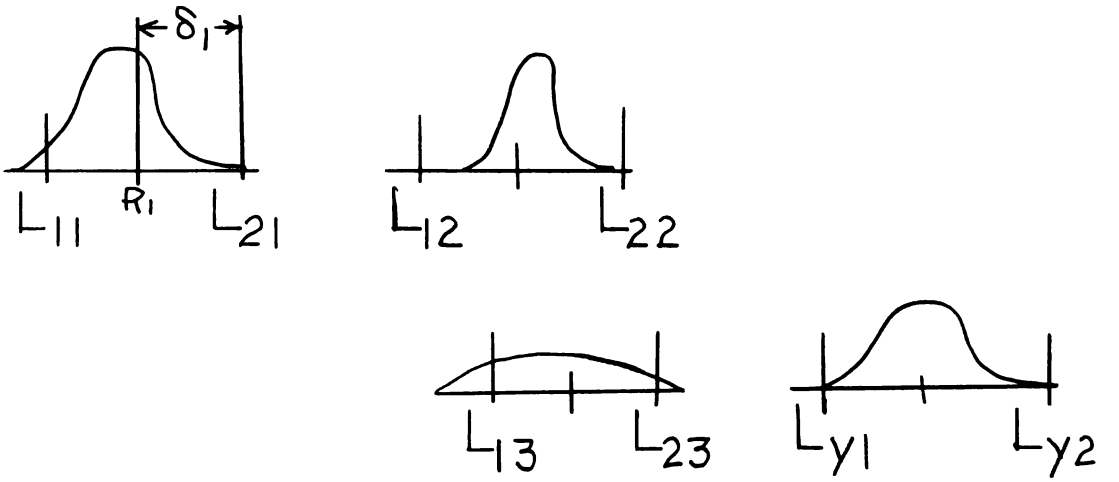


Figure 1. Component and system distributions and ETS limits.

Each of the components X_i is assumed to be a random variable with distribution overlapping part of the ETS limits. The response Y is a function of these random variables and is, therefore, itself a random variable. The first part of this discussion will be an attempt to find the distribution of Y, given information about the distribution of the X_i and about the response function.

When the distribution of the X_i and the response function

$$Y = f(X_1, X_2, \ldots, X_n) \tag{1}$$

are known, an exact expression for the distribution of Y can often be obtained by the method of characteristic functions described in Kullback [1].

In most practical situations, the exact form of the distributions of the X_i is not known nor is there available an explicit expression for the response function; hence, a true statistical tolerance interval cannot be obtained. Instead, the moments of the Y must be obtained from estimates of the moments of the X_i and estimates of the derivatives of the response function. The resulting Pr. $(L_{1Y} \leq Y \leq L_{2Y})$ is an estimate based on

mathematical approximations described herein.

The present discussion considers the cases in which the distributions and the response function are not known.

PROPAGATION OF MOMENTS --LINEAR FUNCTION. Assume a response function of the form

$$(2) \qquad Y = K_1X_1 + K_2X_2 + \ldots + K_nX_n$$

and that the X_i are normally distributed and mutually independent random variables. In this case we can write

$$(3) \qquad E(Y) = K_1\mu_1 + K_2\mu_2 + \ldots + K_n\mu_n$$

and

$$(4) \qquad VAR(Y) = K_1^2\,\sigma_1^2 + K_2^2\,\sigma_2^2 + \ldots + K_n^2\,\sigma_n^2 .$$

Here the K_i's are really $(\partial Y/\partial X_i)$, i.e. the change in Y produced by a change in component I. Equation (4) is sometimes called the classical propagation of errors formula and can be used as a first approximation to a more complex function, as will be shown later.

Since the first two moments determine the normal distribution, and a linear function of normally distributed variables is also normally distributed, the requirements for the complete specification of the ETS interval in this case are simply:

1. Knowledge that the functional form is linear as in (2)
2. The set of values μ_i and σ_i
3. The set of values $\partial Y/\partial X_i$
4. Tables of the normal distribution

The requirements for the more general case will be developed subsequently.

PROPAGATION OF MOMENTS -- GENERAL METHOD. The response function will now be permitted to be any function in n variables that can be represented by an n variable Taylor series. For the two-variable case this can be written as

$$Y = f(X_1, X_2) = f(x_1, x_2) + f_1 x_1 + f_2 x_2 + \frac{1}{2!}(f_{11}x_1^2 + f_{22}x_2^2 + 2f_{12}x_1x_2) + \frac{1}{3!}(f_{111}x_1^3 + 3f_{112}x_1^2x_2 + 3f_{122}x_1x_2^2 + f_{222}x_2^3) + \frac{1}{4}(f_{1111}x_1^4 + 4f_{1112}x_1^3x_2 + 6f_{1122}x_1^2x_2^2 + 4f_{1222}x_1x_2^3 + f_{2222}x_2^4) + \ldots \quad (5)$$

using the derivative notation where for example

$$f_{12} = \left.\frac{\partial^2 f(X_1, X_2, \ldots, X_n)}{\partial X_1 \partial X_2}\right|_{\mu_i} \quad \text{and} \quad f_{1123} = \left.\frac{\partial^4 f(X_1, X_2, \ldots, X_n)}{\partial^2 X_1 \partial X_2 \partial X_3}\right|_{\mu_i}$$

The derivatives are to be evaluated at the appropriate $X_i = \mu_i$. This series represents the response function in terms of its components and can be operated on as a power series. Taking the expected value gives

$$K_1(y) = E(Y) = E\ f(X_1, X_2) = E\left[f(x_1, x_2)\right] + E\left[f_1 x_1 + f_2 x_2\right] + \frac{1}{2!}E\left[f_{11}x_1^2 + f_{22}x_2^2 + 2f_{12}x_1x_2\right] + \ldots \quad (6)$$

The following notation puts the expression in more familiar terms

$$\mu_i = E(X_i)$$

$$\sigma_i^2 = E(X_i^2) - \mu_i^2$$

$$\rho_{ij} = E\left[(X_i - \mu_i)(X_j - \mu_j)\right] / \sigma_i \sigma_j$$

$$\gamma_{ijk} = E\left[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)\right] / \sigma_i \sigma_j \sigma_k ; \gamma_i = E\left[(X_i - \mu_i)^3\right] / \sigma_i^3$$

$$\Gamma_{ijke} = E\left[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_e - \mu_e)\right] / \sigma_i \sigma_j \sigma_k \sigma_e$$

$$\Gamma_i = E\left[(X_i - \mu_i)^4\right] / \sigma_i^4$$

$$x_{ij} = X_{ij} - \mu_i, \quad E(x_{ij}) = 0$$

Using this notation equation (6) becomes

$$K_1(Y) = f(0,0) + \frac{1}{2!}\left[f_{11}\sigma_1^2 + f_{22}\sigma_2^2 + 2f_{12}\rho_{12}\sigma_1\sigma_2\right]$$

$$+ \frac{1}{3!}\left[f_{111}\gamma_1\sigma_1^3 + 3f_{112}\gamma_{112}\sigma_1^2\sigma_2 + 3f_{122}\gamma_{122}\sigma_1\sigma_2^2 + f_{222}\gamma_1\sigma_1^3\right]$$

$$+ \frac{1}{4!}\left[f_{1111}\Gamma_1\sigma_1^4 + 4f_{1112}\Gamma_{1112}\sigma_1^3\sigma_2 + 6f_{1122}\Gamma_{1122}\sigma_1^2\sigma_2^2\right.$$

$$(7) \quad \left. + 4f_{1222}\Gamma_{1222}\sigma_1\sigma_2^3 + f_{2222}\Gamma_2\sigma_2^4\right] + \text{terms of order} \geq \sigma^5$$

The extension of this formula to the case of n variables is now apparent. A detailed discussion of the propagation formulas will be found in Tukey [2].

$$K_1(Y) = E(Y) = E\left[f(X_1, X_2, \ldots, X_n)\right] = f(0_i) + \frac{1}{2!}\left[\sum f_{ii}\sigma_i^2 + \sum{}^* f_{ij}\rho_{ij}\sigma_i\sigma_j\right]$$

$$+ \frac{1}{3!}\left[\sum F_{iii}\gamma_i\sigma_i^3 + 3\sum{}^* f_{iij}\gamma_{iij}\sigma_i^2\sigma_j + 6\sum{}^* f_{ijk}\gamma_{ijk}\sigma_i\sigma_j\sigma_k\right] \quad (8)$$

$$+ \frac{1}{4!}\left[\sum f_{iiii}\Gamma_{iiii}\sigma_i^4 + 6\sum{}^* f_{iijj}\Gamma_{iijj}\sigma_i^2\sigma_j^2 + 4\sum{}^* f_{iiij}\Gamma_{iiij}\sigma_i^3\sigma_j\right.$$

$$\left. + 12\sum{}^* f_{iijk}\Gamma_{iijk}\sigma_i^2\sigma_j\sigma_k + 24\sum{}^* f_{ijke}\Gamma_{ijke}\sigma_i\sigma_j\sigma_k\sigma_e\right] + \text{terms of order}$$

$$\geq \sigma^5$$

By a similar argument, expressions for all the moments or cumulants can be obtained.

The second moment can be written

$$K_2(Y) = E\left[(Y - E(Y))^2\right] = \text{VAR}(Y) = \sum f_i^2\sigma_i^2 + 2\sum{}^* f_i f_j \rho_{ij}\sigma_i\sigma_j$$

$$+ \sum f_i f_{ii}\gamma_i\sigma_i^3 + \sum f_i f_{jj}\gamma_{ijj}\sigma_i\sigma_j^2 + 2\sum{}^* f_i f_{ij}\gamma_{iij}\sigma_i^2\sigma_j$$

$$+ 2\sum{}^* f_i f_{jk}\gamma_{ijk}\sigma_i\sigma_j\sigma_k + \frac{1}{4}\sum f_{ii}^2(\Gamma_i - 1)\sigma_i^4 + \frac{1}{2}\sum{}^* f_{ii}f_{jj}(\Gamma_{iijj} - 1) \quad (9)$$

$$\times\sigma_i^2\sigma_j^2$$

$$+ \sum{}^* f_{ii}f_{ij}(\Gamma_{iiij} - \rho_{ij})\sigma_i^3\sigma_j + \sum{}^* f_{ii}f_{jk}(\Gamma_{iijk} - \rho_{ij})\sigma_i^2\sigma_j\sigma_k +$$

10 additional terms of order $< \sigma^5$ + terms $\geq \sigma^5$

In equations (8) and (9), the $\sum^*$ notation means that each distinct symbol appears once and only once in the sum, and that subscripts with different symbols are always different. Thus

$\sum^* \sigma_i^2 \sigma_j^2$ excludes σ_4^4 and includes $\sigma_4^2 \sigma_5^2$ once, i.e. not as $\sigma_4^2 \sigma_5^2 + \sigma_5^2 \sigma_4^2$. On the other hand

$\sum^* \sigma_i^3 \sigma_j^2$ excludes σ_4^5 but includes both $\sigma_4^3 \sigma_5^2 + \sigma_5^3 \sigma_4^2$.

The formulas for propagating the first four cumulants will be found in Tukey [3].

It can be seen that these propagation formulas are very cumbersome and even with large-scale computing equipment, the complete evaluation of all the terms required would be difficult. There are several ways of reducing the number of terms in these formulas, and probably the most useful is the assumption of independence of the components. In any manufacturing operation it is difficult to conceive of a situation where two separate components could be manufactured so that their distributions would not be independent, except possibly for the case where several different characteristics of a single component are considered to be the random variables X_i; thus it is entirely possible that the several characteristics of an electron tube would not be independent. If the response function includes more than one of the nonindependent random variables, then the more elaborate formulas as given in Tukey [3] are required. When the components are assumed to be independently distributed, all the covariance terms drop out and the expressions for the first four cumulants become

$$K_1(Y) = E(Y) = f(0_1, 0_2) + \frac{1}{2} \sum f_{ii} \sigma_i^2 + \frac{1}{6} \sum f_{iii} \gamma_i \sigma_i^3$$

$$+ \frac{1}{24} \sum f_{iiii} \Gamma_i \sigma_i^4 + \frac{1}{4} \sum^* f_{iijj} \sigma_i^2 \sigma_i^2 + \text{terms of order} > \sigma^5 \tag{10}$$

$$K_2(Y) = E\left[(Y - E(Y)^2\right] = \sum f_i^2 \sigma_i^2 + \sum f_i f_{ii} \gamma_i \sigma_i^3 + \frac{1}{3}\sum f_i f_{iii} \Gamma_i \sigma_i^4$$

$$(11) \qquad + \frac{1}{4}\sum f_{ii}^2 (\Gamma_i - 1)\sigma_i^4 + \sum{}^*(f_i f_{ijj} + f_{ij}^2 + f_{iij} f_j)\sigma_i^2 \sigma_j^2 + \cdots$$

terms of order $\geq \sigma^5$

Several of the terms in formula (11) are among those not included in (9).

$$K_3(Y) = E\left[(Y - E(Y))^3\right] = \sum f_i^3 \gamma_i \sigma_i^3 + \frac{3}{2}\sum f_i^2 f_{ii}(\Gamma_i - 1)\sigma_i^4$$

$$(12) \qquad + 6\sum{}^* f_i f_j f_{ij} \sigma_i^2 \sigma_j^2 + \text{terms of order} > \sigma^5$$

$$(13) \qquad K_4(Y) = E\left[(Y - E(Y))^4\right] - 3K_2(Y) = \sum f_i^4 (\Gamma_i - 3)\sigma_i^4 +$$

terms of order $> \sigma^5$.

These are the formulas that permit evaluation of the first four cumulants of the distribution of Y in terms of the moments of the distribution of the X_i and the derivatives of the response function. Examination of the formulas shows the assumptions required to obtain the sample propagation of errors formulas that have been used for many years, see Shewhart [4]. These assumptions are:

(1) Independence of the X_i components permits deletion of the cross product terms.

(2) Linearity of the response function in each of its variables deletes all but the first term in each of the formulas.

(3) Distribution functions of the X_i determined by the first two moments, as in the case of the normal, rectangular and exponential, deletes formulas (12) and (13); and the resulting propagation formulas are then identical to equations (3) and (4).

(4) Normal distribution function of Y allows the evaluation of $\Pr\left[L_{1y} < Y < L_{2y}\right]$ from existing tables. Burr [5] shows that this assumption is valid if $|\mu_3^2/\mu_2^3| < .03$ and if $2.6 < (\mu_4/\mu_2^2) < 3.4$.

The remainder of this discussion will be predicated on the more general propagation formulas 10 to 13 and will show how to obtain the quantities required for their application.

MOMENTS OF COMPONENT DISTRIBUTIONS. If the exact distribution of the X_i are known*, as may be the case when the components have been produced for a long time by a controlled process whose history is known, then the moments can be estimated from the parameters of the distributions. Usually, however, sample data from production must be used to estimate the moments of the component distributions. If the distributions of the X_i are of a known form, or forms, then it is necessary to compute only as many sample moments as would be required to evaluate the parameters, e.g. one for the exponential, two for the normal distributions. Most standard textbooks develop the formulas for unbiased estimates of the moments from sample observations (See Cramer 6).

THE RESPONSE FUNCTION AND ITS DERIVATIVES. If the response of the system can be expressed as an explicit function of its components, as in the case of simple electrical circuits or mechanical linkages, then the partial derivatives of the function are required for use in the propagation formulas. When the response function is not known, it is necessary to determine the derivatives experimentally. A factorial-type experiment is most appropriate, since the main effects and interactions are directly the required estimates of the partial derivatives. The number of levels will be one more than the order of the highest derivative required, and the experiment should center about the nominal value of each of the components. Fractional factorial designs can be used effectively here to reduce the number of observations, since only a few cross derivatives, i.e. interactions, are required.

---

*It may be possible to use the parameters directly if a mathematical expression for the response function is available.

ESTIMATES OF $P(L_1 \leq Y \leq L_2)$. Of the several methods for determining the required probabilities, given estimates of the first four moments, the two involving a minimum of computation will be described. The first is an outgrowth of the Pearson method of moments.

Pearson and Hartley (7) have tabulated the .5, 1.0, 5.0, 95, 99, 99.5% points for values of $0 \leq \beta_1 \leq 1$ and $1.8 \leq \beta_2 \leq 5.0$ where

$$\beta_1 = \mu_3^2/\mu_2^3 = [K_3(Y)]^2/[K_2(Y)]^3$$

$$\beta_2 = \mu_4/\mu_2^2 = K_4(Y)/[K_2(Y)]^2 + 3$$

From these tables, 1, 2, and 10% confidence intervals can be obtained directly. The wide spacing of the β entries, the limited number of percentage points and the fact that the percentage points, not the integrals, are listed, limit the usefulness of the tables. However, the values that are given make these tables very easy to use, since the required computations are minimized.

The alternative to the Pearson system is the Edgeworth Series. The distribution function of Y is obtained as a power series in terms of its cumulants and the derivatives of the normal distribution function Cramer (6).

$$F(Y) = \Phi - \frac{\gamma_1}{3!}\Psi^{(2)} + \frac{\gamma_2}{4!}\Psi^{(3)} + \frac{10\gamma_1^2}{6!}\Psi^{(5)} + \ldots \tag{14}$$

Terms involving cumulants above the fourth have been dropped because of the magnitude of the sampling errors of their estimates.

Here $\gamma_1 = \mu_3 / \mu_2^{3/2}$ and $\gamma_2 = (\mu_4/\mu_2^2) - 3$

$$\Phi(z) = \int_{-\infty}^{z} \frac{\exp(-y^2/2)\, dy}{2\pi}$$

$$\Psi^{(n)} = \frac{d^n}{dy^n} \left(\frac{\exp(-y^2/2)}{2\pi} \right)$$

This series is easy to apply since only a few terms need by computed and it can be used in many situations where the Pearson tables of β are not applicable.

INTERMEDIATE RESULTS. Up to this point, the techniques for making probability statements about the values of Y have been presented. This is as far as one need go in a wide variety of applications. With these techniques, use can be made of the production records of the distribution of the X_i, and the experimental results showing the response Y to the factors X_i, and the probability that Y will be in the interval $(L_{1y} \leqq Y \leqq L_{2y})$ can be computed; or alternately, the limits can be found within which Y falls for a given percentage of the systems manufactured. This can, for example, demonstrate the manufacturer's ability to produce a particular system within required tolerance, i.e. to determine whether the specifications on the end item are realistic. It will also permit the engineer, on paper, to "juggle" the types of components in an effort to reduce the system tolerance.

ECONOMICS OF ETS. For most products, the cost of manufacture increases when the engineering tolerances decrease; hence, unless a system is to be made having "smallest possible tolerance regardless of cost", as in vital subsystems of missiles, it is desirable to minimize costs and still meet specifications on the system or alternately minimize the system tolerance for a given cost. To find the combination of component limits that will minimize the total cost of the system it is first necessary to express the cost C_i of the X_ith component as a function of the interval size

δ_i of the ith component.

(15) $$C_i = C_i(\delta_i)$$

The cost of the whole system can then be written

(16) $$C_y = \sum_1^n C_i + C_a$$

where C_a is the cost of assembly. This information can often be obtained by fitting a polynominal to cost data, supplied by the manufacturer or by the production department of an "in-house" item. A simple relation should suffice since the function must be monotonic decreasing; if not, the cheaper item with smaller δ would always be made.

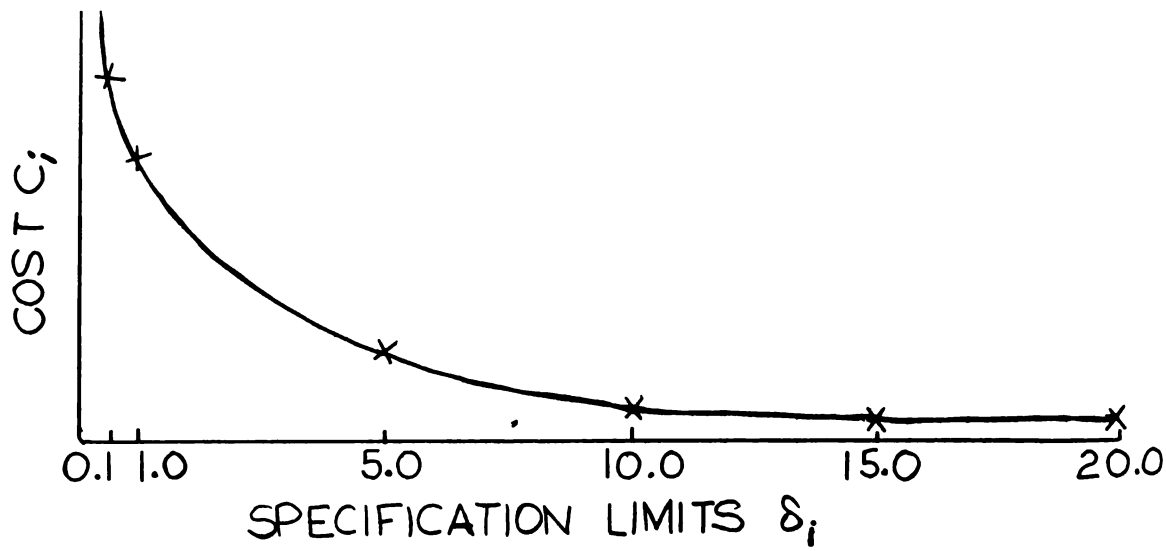


Figure 2. Component cost as a function of ETS interval length.

Figure 2 illustrates such a situation, which could be expressed by a function of the form

$$C_i = K_{1i} + K_{2i}/\delta_i$$

If it is possible to express

(17) $$\delta_Y = \mathcal{M}(\delta_i)$$

i.e. the Y interval length as a monotonic increasing function of the individual δ_i interval lengths, it will be possible to minimize C_y or δ_Y. In general it is very difficult to obtain an expression for δ_Y. However, the first term of the Taylor series may sometimes be used, if the response function can be approximated by a linear function in the vicinity of the required δ_i limits. See Scarborough (8).

When it is assumed that the X_i are n independent, normally distributed, random variables and that the response function is linear in each variable, the optimum set of σ_i can be obtained. An application of this method to machine design can be found in Pike and Silverberg (9). The response function is

(18) $$Y = \sum_i^n X_i$$

and the variance of Y is

(19) $$\sigma_y^2 = \sum_i^n \left(\frac{\partial f}{\partial X_i}\right)^2 \sigma_i^2$$

In this case the cost c_i can be related to σ_i^2 rather than δ_i since minimizing σ_i^2 will minimize C_i, and the total cost is

(20) $$C_y = \sum_i^n C_i(\sigma_i^2) + Ca$$

The method of Lagrange Multipliers, found in Shewhart (4), yields the minimum total cost subject to the constraint $\sigma_y^2 = K_\sigma$ where K is the specified value of σ_y^2. The function

$$\psi_a = C_y + \lambda_a (\sigma_y^2 - K_\sigma), \tag{21}$$

where λ_a is the Lagrange multiplier, will be differentiated with respect to σ_i^2 to obtain the necessary conditions for a minimum C_y:

$$\frac{\partial \psi_a}{\partial \sigma_i^2} = \frac{\partial C_y}{\partial \sigma_i^2} + \lambda_a \frac{\partial (\sigma_y^2 - K_\sigma)}{\partial \sigma_i^2} = 0 \tag{22}$$

$$i = 1, \ldots, n$$

Now using equations (19) and (20) for C_y and σ_y^2, equation (22) becomes

$$\frac{\partial(\sum C_i (\sigma_i^2) + C_a)}{\partial \sigma_i^2} = -\lambda_a \frac{\partial \sum (\frac{\partial f}{\partial X_i})^2 \sigma_i^2 - K_\sigma)}{\partial \sigma_i^2}$$

The differentiation results in the Lagrange equations

$$\frac{\partial C_i (\sigma_i)^2}{\partial \sigma_i^2} = -\lambda_a \left(\frac{\partial f}{\partial X_i}\right)^2 \qquad i = 1, \ldots, n \tag{23}$$

and the constraint is

$$C_y - K_\sigma = 0, \tag{24}$$

This set of n + 1 equations is then solved simultaneously for the set of σ_i^2 which will yield a minimum total cost C_y.

The alternate method is to find a set of σ_i^2, $i = 1, \ldots, n$, which will minimize the tolerance on Y or equivalently minimize σ_y^2 for a specified cost C_y. In this case σ_y^2 is treated as the variable which is to be minimized and the Lagrange equation can be written as

$$\psi_b = \sigma_y^2 + \lambda_b (C_y - K_c) \tag{25}$$

where K_c is the specified cost. Substitution of the values of C_y and σ_y^2 from equations (19) and (20) and performance of the differentiation yields the following set of equations

$$\left(\frac{\partial f}{\partial X_i}\right)^2 = -\lambda_b \frac{\partial C_i(\sigma_i^2)}{\partial \sigma_i^2} \qquad i = 1, \ldots, n \tag{26}$$

and

$$C_y - K_c = 0 \tag{27}$$

which must be solved simultaneously for the σ_i^{*2}. These equations differ from (23) and (24) only in the Lagrange multiplier λ and the constraints. The set of values σ_i^{*2}, $i = 1, \ldots, n$ will differ from the σ_i^2 set, and will be that set of values which minimizes the variance of the system for a fixed total cost C_y.

SUMMARY. If the performance, i.e. response, of a system can be expressed as a function of its n components, it can be approximated by a Taylor series. When a mathematical expression for the function is not available, experiments must be performed to obtain estimates of the partial derivatives of the function. The Taylor series in n variables is used to obtain expressions for the moments of the system's distribution in terms of the moments of the component distributions. The Edgeworth series is then employed to obtain the probability limits for the system distribution using estimates of the moments previously obtained. Finally, using Lagrange

multipliers, it is possible in some cases to find the set of ETS limits for the components such that the cost is minimized for given system limits, or alternately, such that the system tolerance interval is minimized for a given cost.

(1) Kullback, S., An Application of Characteristics Functions to the Distribution Problem of Statistics, Annals of Mathematics Statistics, Vol V, No. 4, 1934, pp. 263-307.

(2) Tukey, J. W., The Propagation of Errors, Fluctuations and Tolerances, Basic Generalized Formulas, Technical Report No. 10, Statistical Techniques Research Group, Princeton University 1958.

(3) Tukey, J. W., The Propagation of Errors, Fluctuations and Tolerances--Supplementary Formulas, Technical Report No. 11, Statistical Techniques Research Group, Princeton University 1958.

(4) Shewhart, W. A., Economic Control of Quality of Manufactured Product, D. Van Nostrand Co., Inc., New York, 1931, p. 259.

(5) Burr, I. W., Tolerances for Mating Parts, Industrial Quality Control, September 1958, pp. 18-22.

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

(7) Pearson, E. S. and Hartley, H. O., Biometrica Tables for Statisticians, Volume I, Cambridge University Press, 1956.

(8) Scarborough, J. B., Numerical Mathematical Analysis, Johns Hopkins Press, Baltimore, 1930.

(9) Pike, E. W., and Silverberg, T. R., Assigning Tolerances for Maximum Economy, Machine Design, Vol. 25, September 1953, pp. 139-146.

Acton, F. S. and Olds, E. G., Tolerance-Additive or Pythagorean?, Industrial Quality Control, November 1948, pp. 6-12.

Epstein, B., Tolerances in Assemblies, American Machinist, Vol. 80, No. 1, January 1946, pp. 119-121.

Evans, D. H., Optimum Tolerance Assignment to Yield Minimum Manufacturing Cost, Bell Telephone Labs. Journal, Vol. 37, March 1958, pp. 401-484.

Gilson, J., A New Approach to Engineering Tolerances, Machinery Publishing Co., Ltd., London, 1951.

Hailes, G. M., Some Statistical Principles of Tolerances, Industrial Quality Control, May 1951, pp. 77-81.

Hanna, R. W. and Varnam, G. C., Interference Risk when Normal Distributions Overlap, Industrial Quality Control, Vol. 7, September 1950, pp. 26-28.

Kendall, M. G., The Advanced Theory of Statistics, Charles Griffin and Co., Ltd., London, 1946

Rice, W. B., Setting Tolerances Scientifically, Mechanical Engineering Vol. 66, December 1944, p. 801-803.

Zelen, M., Bounds on a Distribution Function that are Functions of Moments to Order Four, Journal of Research of the National Bureau of Standards, Vol. 53, No. 6, December 1954, Research Paper 2556, pp. 377-381.

A FURTHER ANALYSIS OF MISSILE RANGE TRACKING SYSTEMS

Oliver Lee Kingsley
Integrated Range Mission
Range Instrumentation Development Division
White Sands Missile Range, New Mexico

I. INTRODUCTION. One objective of the Range Instrumentation Development Division has been the calibration of range instrumentation tracking systems. At last year's conference I enjoyed the opportunity to speak to you of our first attempts at evaluation of range tracking systems. We have now completed our analysis of the second flight test and are prepared to embark on a third flight test to study improvements in the tracking systems and reduction procedures. The third test should include one of the new systems being installed at the Range. It is called the Integrated Trajectory System (ITS) and obtains space position estimates by computations based on angular (phase) and range measurements (phase also) at two instrument sites spaced about 20 miles apart.

Today, I intend to briefly cover some of the high points of our analysis work during the past year. The discussion will cover results obtained from four tracking systems, namely: Askania cinetheodolite, Ballistic Camera, Doppler, Velocity and Position (DOVAP), and the FPS-16 radars.

The first results presented will be a comparison between the data of flight tests one and two. Then the results of flight test number two when handled by a new technique will be presented. The final results presented will be concerning precision estimates for part of the Askania cinetheodolite system and part of the DOVAP system.

The coverage for the first flight test consisted of four tracking systems obtaining simultaneous data for 28 seconds of the missile trajectory. The second flight test coverage consisted of four tracking systems obtaining simultaneous data for three segments of the missile trajectory. The segments I, II, and III included 18, 17, and 19 seconds of trajectory data respectively.

II. DISCUSSION OF RESULTS.

COMPARISON OF THE TWO "Operation Precise" TRACKING TESTS

The first set of contrasts I would like to present are for each of the tracking systems individually. The analysis involved a study of the shift

in the bias error as the tracking proceeds. This was performed by using a one-way classification of trajectory sub-segments. The error data used in the analysis were derived by comparison with data from the Ballistic Camera for each respective system. In all but the DOVAP analysis, this technique should give an error term based essentially on the particular system.

For the Askania system, the first test results indicated a significant bias shift along the trajectory for each coordinate studied, whereas, for the second test, only the X-component of segment II indicated a significant shift. The reasons for more variability in first case can be attributed to many possible sources none of which has been isolated as the source.

The Askania bias error indications for each component studied in the first test were only dominant in the Z-component for the second test.

The DOVAP system indicated significant sub-segment variability in X and Z during the first test and significant sub-segment variability in X, Y, and Z during the second test. These tests indicate that the DOVAP systems needs improvement. There have been changes in the data reduction procedures that have not been completely evaluated and the needed improvement may well be accomplished. Our next flight test should affirm or deny this conjecture.

The over-all bias error inherent in the reduced data from the first test was as inherent in the reduced data from the second test. Again the changes in data reduction may have removed the significant bias error inherent in the reduced data.

The Radar system consisted of data from four radars, three of which were the same for the two flight tests. Significant sub-segment variability was indicated in the Z component during the first test and some significant X and Z component variability during the second test.

The Radar system data for X and Y components exhibited bias errors that were significant during the first flight tests. There was a significant bias exhibited for all components during the second flight tests.

The table below summarizes some of the precision estimates obtained from the two tests. In general, the results are quite similar.

Table 1

Standard Deviation Estimates by Simon-Grubbs Method

Tracking System and Test \ Coordinate		X (ft)	Y (ft)	Z (ft)
Askania				
	First Test	11	11	8
	Second Test	10	15	14
Ballistic Camera				
	First Test	2	6	10
	Second Test	4	6	9
DOVAP				
	First Test	0	4	8
	Second Test	0	0.1	7
Radars				
(N=3)	First Test	15	21	12
(N=4)	Second Test	17	19	26

NOTE: Pooled estimates over the trajectory are presented for the second test.

COMPARISON OF TWO TECHNIQUES FOR BIAS ESTIMATION.

We have explored some new data reduction techniques. One technique as suggested by an RCA report by Dr. Duncan was for the study of short term bias error estimates and another by H. Dibble and C. Carroll, Jr. suggest a way of combining trajectory data.

Dr. Duncan's suggestions have not been fully explored on the second test because the data to fit the complete model was not obtained. For the third test, this data will be obtained and the method applied.

The Dibble and Carroll report was used to combine data from three of the systems, viz. Askania, DOVAP, and Ballistic Camera. The value of the technique depends on the adjustment of the tracking systems for bias. Adjustment of the data by a standard for bias was not accomplished. Also, the radar data was not included in the computations because of computer limitations.

Tables 2, 3, 4, and 5 exhibit data for the two methods of bias estimation. The Ballistic Camera bias estimates seem reasonable except for the Z components for Segments I and III using the B.E.T. There is more to learn about the application of this technique for obtaining a best estimate of trajectory.

TRACKING CORRECTIONS, DIAL READINGS, AND CYCLE COUNTS

Where possible, each of the tracking systems were explored in detail. Three detailed studies are summarized, (1) the tracking corrections for the Askania cinetheodolites, (2) dial readings for the Askania cinetheodolites, and (3) the DOVAP cycle counting methods.

The dial readings were expected to have a standard deviation of about 0.001 degrees. This was verified by having three operators read the same set of films and applying the Simon-Grubbs technique (Table 6). The effect of film reader (machines) was not expected to contribute to the variability and was confounded in the test.

Table 2

Ballistic Camera System Mean Bias Error Estimate

Coordinate / Bias Estimate	X (ft)	Y (ft)	Z (ft)
SEGMENT I			
Direct Method	0.0	0.0	0.0
B.E.T. Method	-3.8	0.7	26.7
SEGMENT II			
Direct Method	0.0	0.0	0.0
B.E.T. Method	-2.3	-6.3	-1.4
SEGMENT III			
Direct Method	0.0	0.0	0.0
B.E.T. Method	0.0	-0.2	76.8

Table 3

Askania (Mode 7 Instruments) Mean Bias Error Estimates

Coordinate / Bias Estimate	X (ft)	Y (ft)	Z (ft)
SEGMENT I			
Direct Method	7.9	-2.0	-41.6
B.E.T. Method	4.1	-1.5	-14.8
SEGMENT II			
Direct Method	0.6	1.9	-17.9
B.E.T. Method	-1.5	-4.6	-19.3
SEGMENT III			
Direct Method	-1.0	-9.1	-29.6
B.E.T. Method	-1.2	-9.2	47.3

Table 4

DOVAP (Seven Receivers)
Mean Bias Error Estimates

Bias Estimate \ Coordinate	X (ft)	Y (ft)	Z (ft)
SEGMENT I			
Direct Method	23.5	-1.8	-51.2
B.E.T. Method	19.7	-1.2	-24.4
SEGMENT II			
Direct Method	20.0	15.3	-51.1
B.E.T. Method	17.9	8.6	-49.6
SEGMENT III			
Direct Method	27.6	8.6	-101.2
B.E.T. Method	27.4	8.5	-24.5

Table 5

Radars (N=3)
Mean Bias Error Estimates

Coordinate / Bias Estimate	X (ft)	Y (ft)	Z (ft)
SEGMENT I			
Direct Method	-0.7	4.3	-17.7
B.E.T. Method	-4.6	4.5	0.4
SEGMENT II			
Direct Method	17.0	16.3	48.0
B.E.T. Method	14.8	9.6	26.8
SEGMENT III			
Direct Method	44.0	21.3	12.0
B.E.T. Method	43.9	21.5	76.4

The bore-sight readings from the first flight test exhibited a standard deviation of about 0.003 degrees which was approximated during the second flight test (Table 7). These were obtained using the same technique as for the dial readings above.

As part of the editing of the above data for analysis, estimates of gross reading error was made. An example of a gross error would be that a dial reading differed from the other two by one-half degree--the two close together are assumed correct and the other is then the anomaly. The reader is mechanized so that a one-half degree recognition must take place by the operator and so recorded. Errors of this kind ranged from a low of zero percent to a high of one-half percent for the sets of dial data. Similar percentages were derived for the bore-sight data analyzed.

Table 6

Standard Deviation Estimates for Dial Readings (in degrees)

	Azimuth	Elevation
G-39	.0016	.0015
G-75	.0013	.0011

Table 7

Standard Deviation Estimates for Bore-sight Readings (in degrees)

	Azimuth	Elevation
G-39	.0027	.0032
G-75	.0019	.0023

The DOVAP cycle count has been read by three methods. The most recent method is by a DOVAP digitizer which replaces the Putman reader. The initial method for reading was observing the cycles recorded on film.

Test data indicated that the digitizer data is, indeed, the best; Putman next; and the hand reader last (Table 8).

Table 8

Standard Deviation of DOVAP Cycle Count

Station / Method	D-13	D-16	D-350
Hand	0.06	0.05	0.05
Putman	0.04	0.04	0.04
Digitizer	0.00	0.03	0.02

Converted into loop range a standard deviation of 0.02 cycles means about 0.26 feet.

III. SUMMARY. The results are quite similar for both flight tests though there were definite changes in flight trajectory, instrumentation system instruments, and many other day-to-day changes. The data reduction procedures were essentially the same. Changes are to be made in the DOVAP data reduction procedures for the data to be obtained in test three.

The most dominant feature of the test is the large bias present in the data for the Z component. In general, the standard deviation is larger for the Z component for a given tracking system.

REFERENCES

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

2. Churchman, C.W., Statistical Manual: Methods of Making Experimental Inferences, 2nd Ed. Frankford Arsenal, Phil., Pa., June 1951.

3. Deming, W. E., Statistical Adjustment of Data, John Wiley & Sons, Inc., New York, 1938.

4. Dibble, H. L., Carroll, C. L., Jr., "A Best Estimate of Trajectory using Reduced Data from Various Instruments at a Single Point in Time," AFMTC-TR-60-12, May 1960, Patrick Air Force Base, Florida.

5. Dixon, W. J., and Massey, F. J., Jr., Introduction to Statistical Analysis, McGraw-Hill Book Co., Inc., New York, 1951.

6. Duncan, David B., "On the Optimum Estimation of a Missile Trajectory from Several Tracking Systems," AFMTC-TR-60-16, 24 Aug 1960, Patrick Air Force Missile Test Center, Florida.

7. "Final Data Report No. 10178, Radar Trajectory Data for Nike Hercules Flight 78 HE Missile 13560 Launched 12 Sep 1960" (U) IRM-DRD, WSMR, N. Mex. (25 Oct 1960) Classified Confidential.

8. "Final Data Report No. 10368. Ballistic Camera Data for Nike Hercules Flight 78 HE Missile 13560 Launched 12 Sep 1960"(U) IRM-DRD, WSMR, N. Mex. (10 Jan 1961) Classified Confidential.

9. "Final Data Report No. 10495. Askania Trajectory Data for Nike Hercules Flight 78 HE Missile 13560 Launched 12 Sep 1960" (U) IRM-DRD, WSMR, N. Mex. (29 Dec 1960) Classified Confidential.

10. "Final Data Report No. 10525, Refraction Error Data for Nike Hercules Flight 78 HE Missile 13560 Launched 12 Sep 1960" (U) IRM-DRD, WSMR, N. Mex. (18 Jan 1961) Unclassified.

11. "Final Data Report No. 10526. DOVAP Trajectory Data for Nike Hercules Flight 78 HE Missile 13560 Launched 12 Sep 1960" (U) IRM-DRD, WSMR, N. Mex. (20 Jan 1961) Classified Confidential.

12. Grubbs, F. E., "On Estimating Precision of Measuring Instruments and Product Variability." J.A.S.A. Vol 43, (1948), PP. 243-264.

13. Hoel, P. G., Introduction to Mathematical Statistics, 2nd Ed., John Wiley & Sons, Inc., New York, 1954.

14. Kendall, M. G., The Advanced Theory of Statistics, Vol II, 3rd Ed., C. Griffin & Co., Ltd., London, 1951.

15. Kingsley, O. L., "Analysis of Some Trajectory Measuring Instrumentation Systems," Paper presented at Sixth Conference on the Design of Experiments in Army Research, Development and Testing, Oct. 1960, Aberdeen Proving Ground, Aberdeen, Maryland.

16. Mood, A. F., Introduction to the Theory of Statistics, McGraw-Hill Book Co., Inc., New York, 1950.

17. Simon, L. E., "On the Relation of Instrumentation to Quality Control," Instruments, Vol. 19, Nov 1946.

18. Snedecor, G. W., Statistical Methods, 4th Ed. Iowa State College Press, Ames, Iowa, 1946.

19. Weiss, J. E., Kingsley, O. L., "Study of Accuracy and Precision of Trajectory Measuring Systems," 30 June 1961 (U) ORDBS-IRM Task 5-4-2. WSMR, N. Mex. Unclassified.

PANEL DISCUSSION ON SIMULATION

Chairman: Dr. Robert M. Thrall, The University of Michigan
Panel Members: Colonel Alfred W. DeQuoy, Chief,
Strategy and Tactics Analysis Group, Department of Army
Dr. John Hammersley, Oxford University and Princeton University
Mr. John Moss, Research Analysis Corporation
Dr. Gustave Rabson, The University of Michigan

Colonel Alfred DeQuoy, John Moss and Gustave Rabson were each asked by Professor Robert Thrall to talk on various aspects and applications of simulation. Dr. John Hammersley commented on each of these addresses, and then added several pertinent remarks on some applications of Monte Carlo methods. Following these comments, the Chairman opened the meeting to questions from the floor. The papers by DeQuoy, Moss and Rabson are made part of this publication

U. S. ARMY STRATEGY AND TACTICS ANALYSIS GROUP
THE U. S. ARMY'S WAR GAMING ORGANIZATION

Colonel Alfred W. DeQuoy

MISSION, TASKS, AND FUNCTIONS. In July 1960, General George Decker, then Vice Chief of Staff and now Chief of Staff, directed the establishment of the U. S. Army Strategy and Tactics Analysis Group (STAG) with the mission of "supporting Department of the Army operational planning and evaluation activities by war gaming and allied techniques."

The reasons for creating STAG were twofold:

1. The ever-increasing complexity of land warfare and cost of new weapons and equipment required that operations research techniques be used in attacking Department of the Army problems and that a digital computer be utilized to the maximum extent possible.

2. It was considered desirable that the Army have an "in-house" war gaming capability in the interests of greater efficiency and lower costs.

STAG is a Class II field activity operating under the control and supervision of the Deputy Chief of Staff for Military Operations. It is located

in Bethesda, Maryland, adjacent to Washington, D. C. Its tasks and functions are stated in AR 15-14, dated 11 September 1961:

1. Develop a land combat war gaming model for testing Army plans. To the extent feasible, the model will be developed for application to a large scale computer.

2. Conduct studies of Department of the Army problems using war gaming and allied techniques as feasible.

3. Advise and provide technical assistance to other Army and Army supported agencies in war gaming matters on request.

4. Provide Army participation, to include war gaming models, in joint war games as required.

5. Maintain liaison with other agencies engaged in war gaming activities as required in furtherance of STAG's mission.

6. Perform other duties as directed.

Three terms mentioned in STAG's mission, tasks, and functions may need more detailed explanation:

1. War gaming. An operations research technique whereby the various courses of action involved in a problem are subjected to analysis under prescribed rules of play representing actual conditions and employing planning factors which are as realistic as possible. (FM 101-51, p. 44)

2. Operations research. The analytical study of military problems, undertaken to provide responsible commanders and staff agencies with scientific basis for decision on action to improve military operations. Also known as operational research, operations analysis. (AR 320-5)

3. Model. A mathematical representation of the most important elements of a problem, their inter-relationship, and their logical sequences.

ORGANIZATION. STAG's authorized strength for Fiscal Year 1962 is 92 personnel--41 military and 51 civilians. Plans call for a phased increase in both military and civilian strength during the next few years. The military consist of 26 Field Grade Officers, 1 Warrant Officer, and 14 Enlisted Men.

The Officers possess a wide variety of backgrounds and experience, and represent both the Combat Arms and the Technical Services.

Most of the civilian personnel in STAG are in the professional category. They are either Operations Research Analysts or Mathematicians. As the demand for such personnel is extremely competitive both within and without the Government, the acquisition of personnel has been slow and an under-strength exists. Many applications are received--few can meet the exacting qualifications.

Although highly trained in their individual fields of military science, mathematics, physics, engineering, etc., it is necessary that all personnel receive additional training in other fields associated with computerized war gaming after joining STAG. Initial training consists of 85 hours of classroom instruction covering computer familiarization, formula translation, operations research techniques, and probability theory. More advanced training for selected individuals in some of these areas is given as necessary.

To accomplish its mission and objectives, STAG is organized functionally to provide the military, scientific, and computer integration necessary for solving operational problems.

Over-all direction is furnished by the Office of the Chief, which includes the Chief and Deputy Chief, both colonels, and a civilian Technical Director. A small administrative office handles the housekeeping functions and performs the necessary administration. In addition, there are three divisions, each consisting of two or more branches.

The Plans Division makes feasibility and applicability studies of DA problems to determine the advisability of using war gaming techniques for their solution. This Division advises and provides technical assistance to war gaming activities throughout the Army, and also furnishes Army participation in joint war games. It develops and recommends the policies and over-all procedures governing the general methods and techniques to be used in war gaming by STAG itself, and develops the analyses, evaluations and supporting information to be furnished the Army Staff. In addition, it evaluates, interprets, and appraises reports on the rules of play developed, the approach and method being used in constructing war game models as well as the conduct of the play itself.

The Operations Division has as its primary function the development of a land warfare model capable of encompassing all types of land combat operations. Here, as elsewhere in STAG, the "team" system is used in which a military officer is paired up with a civilian operations analyst. The

military officer provides the military logic and works with the civilian in developing the mathematical equation translation for communicating with the computer. This land warfare model will be accomplished through the preparation and integration of submodels developed by various teams such as Infantry, Air Defense, Nuclear, Logistics, etc. The Operations Division also is charged with the development of models for miscellaneous DA programs designated by the Plans Division through its feasibility and applicability study procedure as suitable for computerized war gaming.

The Computer Division is responsible for the operation of an IBM 7090 and 1401 computer system and a newly designed visual data display generating system which will be used in conjunction with the 7090. It has the additional responsibility of analyzing various programming systems, computer operating systems, and mathematical and statistics techniques, and of implementing those which could be most effective. The Division must program and code data and provide programming support for the various teams engaged in developing war gaming models.

PHYSICAL SITE. Specialized facilities are required for STAG. It occupies five floors of a new eight-story building. The seventh floor houses the 1401 and 7090 computer systems. The computer room was specially designed for the 34 components comprising the systems, and includes a raised floor which serves both as a cable-way for the interconnecting circuitry and as a floor plenum to distribute conditioned air throughout the room. Since the system generates considerable heat, an air conditioning system separate from the building cooling system is required for the computer room in order to maintain the permissible ranges of temperature and humidity, which are extremely limited.

The entire eighth floor is devoted to war gaming. The Control Room is a two-story facility containing two 15x15 foot screens. These screens reflect the situation maps and statistical data as processed and updated by the computer. The displays are electronically generated by a display generator unit upon impulse from the computer, printed by cathode ray tube on selenium discs, and projected as color symbology and alphanumeric characters over background reference maps being projected simultaneously. Each of the other two war rooms (Red and Blue) will be similarly equipped except with smaller 7x7 foot screens. The Controller can, at any time, project the true game situation on one screen and either the Red or Blue operations map on the other. Or, by selective switching control, he can show the Red or Blue commanders on their own screen the true situation or their opponent's operations map. He can also recall on the screen within a few seconds any

information previously shown. Direct inputs to the computer are possible from each of the war rooms and the Control room making possible the play of of either free or controlled games. An audio system provides the capability of briefing and discussion without the necessity of assembling the players.

LAND WAREFARE GAME - CENTAUR. STAG's major task is the development of a model which will permit computerized war gaming of land warfare sufficiently well to allow various analyses to be made and eventually to be used to test the Army's operational plans, including the Army portion of joint plans. This is a tremendously complicated task and will require many years before it is complete. While the game will be computerized to the maximum extend practicable, it must allow for human decisions to be interjected whenever necessary or desirable. This has given rise to the name "CENTAUR"--half man, half beast. The use of the computer is not, of course, to replace the commander or his staff. It merely provides the planners with speed in calculation (the 7090 can add up to 240,000 ten digit figures a second), a tremendous "memory" (millions of items) and reliability in performance infinitely superior to man. Instead of having the results of only one long hand-played game, the game can be played over and over on the computer in relatively brief periods of time until statistically reliable results or at least a better understanding of the problem is obtained. Changes can easily be made in the many variables to determine their criticality and to perform other worthwhile analyses.

The complete land warfare game will be the largest and most complicated computerized game ever constructed. While it will take years to complete it, a less detailed CENTAUR game will be completed within a comparatively short time so that initial work may be performed on some of the problems confronting the DA staff, CONARC and the field commanders.

FIELD ARMY BALLISTIC MISSILE DEFENSE SYSTEM - (FABMDS) SIMULATION. The Army recently awarded contracts to various civilian organizations to determine whether it was possible to defend a field army against ballistic missiles, and if so, the nature of such a system. An operational evaluation of of these contractors" proposals was necessary in addition to R&D's technical evaluation. STAG undertook this assignment since the model could, when completed, be capable of expansion into a simulation of other air defense systems of the field army and could thus be incorporated into the land warfare game. The FABMDS simulation is now being run on the computer and an operational evaluation of the proposed FABMDS is being made.

RELATIONSHIP WITH OTHER WAR GAMING AGENCIES. Since STAG has been established as the DA's operational war gaming agency, the question might arise as to its relationship with other Army agencies presently engaged in war gaming and with civilian Agencies under contract to the Army. Studies are now going on to determine which agencies should engage in war gaming concerning National Defense. Insofar as the Army is concerned, the Chief of Staff has assigned this responsibility to STAG. Whether some civilian agencies will be given authority to war game operational plans under certain circumstances has yet to be determined. CONARC, the Army Electronic Proving Ground, and other agencies, as well as some of our service schools, are also interested in war gaming. Some for analytical purposes, some for training, others for both. However, they have different spheres of interest than STAG. CONARC, for example, is concerned with the development, testing, and evaluation of doctrine (AR 10-7). STAG is concerned with the Army's plans in the event of a war occurring anywhere in the world. No other war gaming agency has been assigned this task.

One of STAG's functions is to assist other agencies in war gaming. In addition to providing technical advice, STAG will acquire information regarding all Army war gaming and will serve as a focal point for the dissemination of such information, thus helping to prevent unnecessary duplication. No such facility has previously existed. STAG will also recommend certain courses of action so that installations may prepare models in a form usable by the others.

It is not desirable that all war gaming be concentrated at STAG. It would be ideal if, in addition to other research, the Service Schools and other agencies could devote some effort to maintaining the portion of STAG's land warfare model pertaining to their specialties (Artillery, Communications, Logistics, etc.). These agencies have a vital interest in these portions of the problem and are best qualified to analyze them. Some of the outputs of these agencies would be inputs to STAG and vice versa. To some extent, this will serve to integrate the war gaming being done by many different organizations and will assure a steady forward progress rather than a series of disorganized efforts. What is of even greater importance, a composite picture rather than glimpses of isolated areas will thus be made available.

STAG is also studying means of establishing direct contact with various major overseas commands, CONARC, the Army War College, etc., through the Signal Corps' Data Link Transmission System for the purpose of exchanging data during the play of war games. STAG could process the data on its computer and furnish required information directly to the co-participating agency on which the latter could base its decisions.

Thus, STAG does not replace any existing war gaming activity. It complements them. There are today throughout the Army so many problems suitable for war gaming that the difficulty is not one of discovery, but rather of selection. There will always be more requests for war gaming than it will be possible for all existing organizations to perform.

STAG'S POLICIES AND FUTURE OPERATIONS. It is essential that STAG be highly respected in both military and scientific circles. This objective can be attained only by turning out a superior product.

In order to attain and maintain the highest standards, STAG has adopted the following policies:

1. It will attempt to be completely objective in its work whether the results are favorable to prevailing view points or not. It will then be up to the DA staff to determine how the finished product should be used.

2. A close working relationship will be established with the DA staff and other agencies, so that they may observe STAG's work as it develops and assist it with their comments.

3. STAG will prepare parametric models, i.e., models which can be expanded, contracted, or changed with comparative ease and simplicity. As a general rule, models are developed for only one particular purpose, e.g., to test NIKE HERCULES. To change such a model to one capable of handling other weapons systems or other factors is such a time-consuming task that war gamers usually prefer to construct an entirely new model. Preparation of a parametric model is difficult. One has to think of all possible eventualities and provide for them in the program. The results, however, are well worth the extra time and effort.

4. STAG plans to have the computer analyze the results of games in certain cases rather than have the analysts do so. This will take much thought and much time to formulate and program but will end up in a net gain in time and effort.

5. All models will be documented, i.e., STAG will prepare a detailed explanation of the thinking which went into each model, what the various courses of action were and why a particular one was adopted, why some elements were incorporated and why some were not. Although this will be time-consuming, it will enable others to provide constructive criticism and

will permit STAG replacement personnel to understand, modify and improve the models. Bases for war gaming models prepared by various other agencies have not usually been recorded. When the original designers of such models departed, their successors, lacking the time to analyze the models in detail perpetuated errors in, or failed to improve the models which they inherited. Rather than spend considerable time analyzing existing models, agencies have considered it preferable to develop new ones. This has led to unnecessary duplication of models and much unnecessary expense. Undocumented models are not subject to criticism because only the builders know their failings. No one will undertake the task of examining them in detail because of the terrific amount of time needed to do so. In its documentation, STAG will present its thoughts and approach and will point out the limitations of the model as well as its capabilities. This policy appears to be a novel one.

6. Most reports given the DA staff by civilian agencies are written in highly technical form. The great majority of officers who receive these reports for review have been away from mathematics too long to understand readily the calculations leading to the conclusions and recommendations. STAG proposes to issue reports in two parts. One for the military stressing the logic involved and another in technical form for the scientific community.

In conclusion, STAG is a field activity whose primary interest is in operational planning as distinquished from research and development. In the accomplishment of its mission, both hand-played and computerized games will be used as tools, but the productive force will stem from the military and scientific knowledge and experience of its personnel.

NOTES ON THE SELECTION OF A COMPUTER FOR SIMULATION PURPOSES

John Hunton Moss

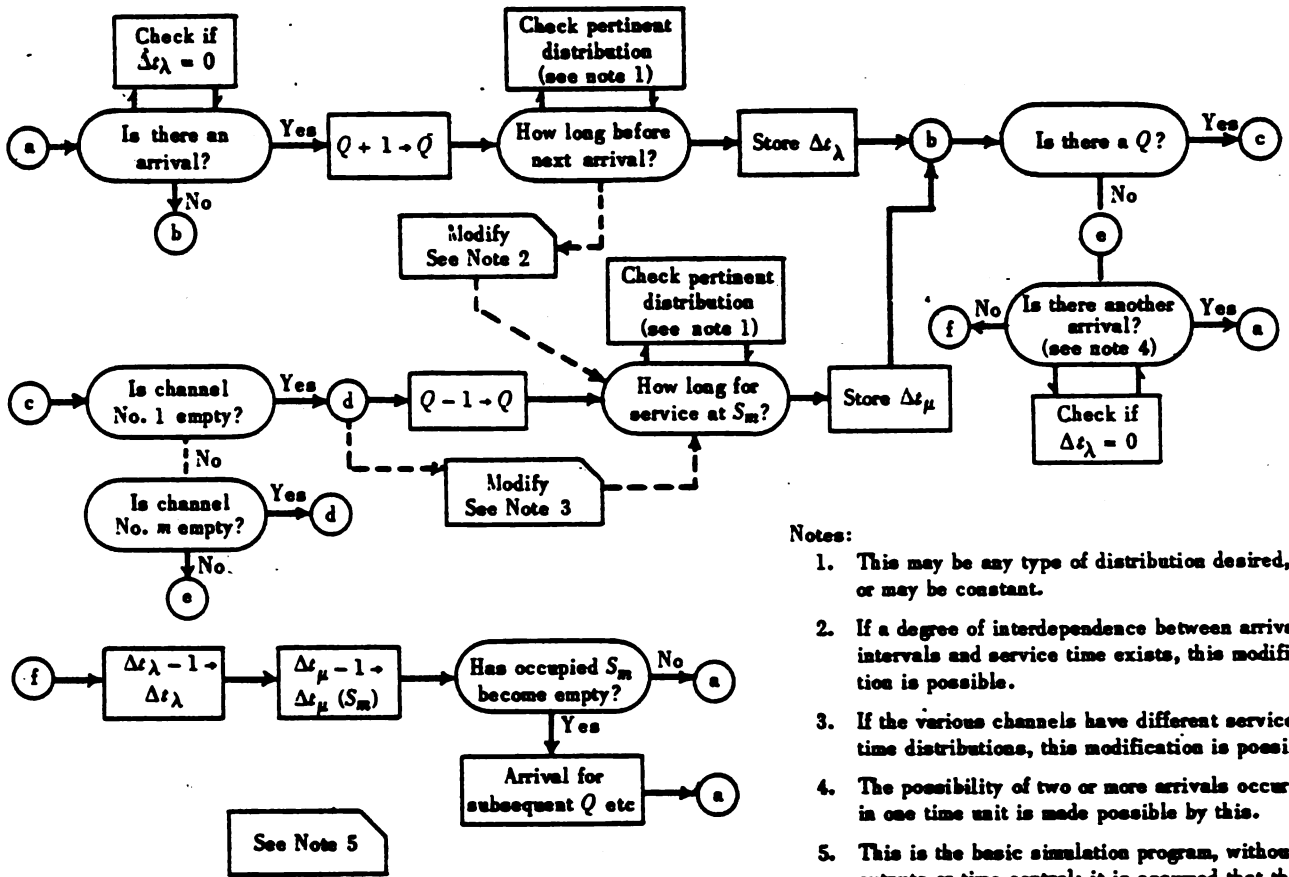
Recently, in a small group undergoing instruction in simulation techniques, a portion of the course was devoted to simulating the queuing characteristics of the Baltimore Tunnel. Various approaches to simulation were considered, and it was decided to exploit an "interval examination" method, examining the system at regular periods according to the resolution required in the problem. The conventional approach toward defining the problem was employed, empirical data being gathered to provide insight into the nature of the arrival and service distributions (both were found to be negative expotential in respect to interval between events). A time resolution of one second was employed, again consistent with the assumptions requisite to exploiting the Poisson curve, and the perturbations that might be expected in the system as a function of time of day, day of week and other such variables, were duly taken into consideration.

As is meet and proper, a flow chart [see Fig. 1 for general form] was drawn up to the satisfaction of the class; the problem then arose concerning the choice of the computer to be employed in the simulation. As is customary, the requirements imposed by the simulation were listed and the consequent problem was simply to match these requirements to available computers in the area. The characteristics of simulations which define the variables involved in selecting the computer are as follows:

1. Arithmetic requirements
2. Logic requirements
3. Timing requirements
4. Storage requirements
5. Special requirements (random number generators, special read-out devices, etc.).

Also pertinent but not directly characteristic of the simulation were:

1. Programming characteristics (learning time, etc.)
2. Time available (deadline of problem)
3. Economics.



Δt_λ = Interval of time till next arrival
Δt_μ = Interval of time for service
Q = Number in queue
S_m = Service channel

Notes:

1. This may be any type of distribution desired, or may be constant.
2. If a degree of interdependence between arrival intervals and service time exists, this modification is possible.
3. If the various channels have different service time distributions, this modification is possible.
4. The possibility of two or more arrivals occuring in one time unit is made possible by this.
5. This is the basic simulation program, without outputs or time control; it is assumed that these factors be introduced consistent with requirements.

Figure 1

Computer Flow Chart of Simulation Approach

Available in the area were computers running the gamut from desk models up through the largest available digital computer facilities. Obtaining short periods of time on these computers does not usually pose serious problems for educational institutions, and accordingly the economic aspects were not critical as they might have been for a commercial activity. Accordingly, it might be worthwhile to list what turned out to be the real variables in selecting the computer. These were as follows:

1. Availability of bootleg time
2. Geographic placement of the facility
3. Nationalities of the participants in the program (red tape associated with classified facilities and foreign nationals)
4. Ease of programming.

Consistent with the limitations imposed by these restraints, a small, nearby limited instruction, relatively slow, limited storage, drum-type computer was selected, and the above problem was programmed for it. The results were interesting in several respects. First, the real time mean interval between arrivals at certain hours was on the order of three seconds; the computer time interval for simulating the same period was on the order of three minutes. The ratio of real time to computer time was approximately 1:60, which is to say that it would have taken a week and a half[1] to simulate one hour of traffic flow through the system, or well over a half a year to simulate a day in the life of the Baltimore Tunnel. Secondly, if such a day had been simulated, it would have cost about $6,000 for one run of one tunnel-day. Thirdly, the resolution of the output left a great deal to be desired. Fourthly, changes in the parameters in the simulation would have required radical reprogramming.

The same program had previously been put on one of the high-speed, large-capacity digital computers; running time for a comparable tunnel day was on the order of 30 minutes. Based on a $320-an-hour rental, the cost per run was on the order of $160; the output had high resolution and the program was easily subjected to changes in parameters. A third and still larger facility actually required a longer time (42 minutes) at a cost per run of $295.

The moral here lies not in the inadequacy of small computers per se, but rather in the choice of computer to meet the requirements imposed by the type of problem. Other types of problems might have been more efficiently and more effectively handled by this small computer than by

[1] Based on an eight-hour day, five-day week.

larger. This particular simulation, as can be seen in the flow chart, is top-heavy with random number generators, access to random variates, and logical manipulations. The computer selected by the class group was designed primarily for arithmetic exercises and was woefully inefficient when exploited for this particular problem. It was selected, as might have been expected, according to the variables that affected the decision; the computer was not selected to match the problem but rather to match some artifactual but pertinent restraints unassociated with the problem itself.

In general, the problem characteristics initially defined should be taken into consideration in evaluating the proper choice of computer for a problem and further discussion and clarification of these characteristics might profitably be undertaken. Generally speaking, it is possible to combine timing and storage. Computers of the large-scale digital type may be described as having four types of storage capacity, rapid access (usually a core), medium access (usually a drum), slow access (usually tapes), and very slow access (punch cards). If a computer has relatively high-speed operation times but relatively limited rapid-access storage capacity, a large percentage of the operating time will be devoted (in a normal healthy program) to moving segments of the program in and out of rapid-access storage for manipulation; the apparent gain in efficiency because of the speed of the operations is cancelled by the limitation on rapid-access storage space. The relationship between these two variables--speed of operation and storage capacities--will, on consideration, be seen to hold for most types of programs, and accordingly these two variables may not be held as being independent when considered in the light of operational efficiency.

A second factor that must be taken into consideration in evaluating computer selection is the weighting of the arithmetic operations to bring about a reasonable estimate of the effective arithmetic timing of the computer. In general, the operations of a computer may be broken down into five categories:

1. Simple arithmetic (addition, subtraction)
2. Complex arithmetic (multiplication, division)
3. Logical (greater than/ less than, positive or negative, etc.)
4. Housekeeping operations (starts, stops, jumps, etc.)
5. Special (shifting operations, masking operations, etc.).

Many computer operations are, of course, combinations of the above, depending generally on the sophistication of the computer logic. In general, however, it is possible to examine a program and obtain percentages of the various types of operations employed in the problem and on this basis to derive weighting factors which result in a more realistic evaluation of the capability of the computer to fulfill the operational requirements imposed by the problem or simulation.

In respect to logical requirements, generally it is possible to simulate the most complex logical requirements with the most elementary capability. For example, a computer having only one logical operation--a sign test--can be programmed to use this test as a device for testing for equality, zero, relative magnitude and the like. The larger computers have these manipulations as separate commands, thus in some respects simplifying the programming, albeit not adding significantly to the potential versatility of the equipment. The mere number of logical operations does not necessarily reflect increased capability, but rather programming ease. Another example of programmable versatility may be applied to those computers not having direct Boolean[1] capability. Almost without exception it is possible to program such computers so that they may operate in Boolean mode, despite the fact that the initial list of commands does not include this capability.

It has been suggested that timing and capacity can be related, arithmetic operations can and should be weighted to fit the program requirements and that the logic in a computer can be expanded at some cost in programming time and ease; the effect of these factors is that differences between computers are not so significant as might appear at first glance. (Interesting in this respect is the capability given to computers by manufacturers to simulate much larger, more versatile and technologically improved, newer models.) Computer manufacturers tend to play leap frog such that one after another comes out a little bit ahead of the rest in terms of speed, capacity, and versatility, but in many cases a computer which might be rendered obsolete for certain types of operations might be more than

---

[1] In arithmetic addition one plus one normally equals two; in Boolean manipulation, the conjunct of two unities cannot exceed unity (one plus one equals one). It can be seen that a purely Boolean capability would be inconsistent with effective arithmetic operation and the consequent necessity for having this capability only as ancillary to the arithmetic capability.

adequate in other respects. In other words, the state of technological advance is not always a valid criterion in selecting a computer.

In selecting a computer for simulation purposes, two sets of variables should be introduced and are here restated for discussion purposes.

PROBLEM-ORIENTED VARIABLES.

1. Capacity requirement of problem
2. Timing requirement of problem
3. Versatility of operation requirements
4. Programming requirements
5. Time frame
6. Special requirements in output, special manipulations and the like.

ARTIFACTUALLY-ORIENTED VARIABLES.

1. Location of computer facility
2. Cost (based on integrating capacity, operation times as well as programming costs, etc.)
3. Availability of access to facility
4. Availability of qualified programmers.

In fact, in most cases the artifactual variables will tend to be overriding. In respect to cost it has been possible to compare two rather different simulation programs, the first having quite different operational requirements from the second, to run them on different computers and to compare the operating costs. The table below shows these relative costs; programming costs are not included.

Table 1

RELATIVE COSTS AND RUNNING TIMES

Computer	Cost/Hr	Program I		Program II	
		Time/Run	Cost/Run	Time/Run	Cost/Run
A	$ 5	72,000 min	$6000	--	--
B	$320	30 min	$ 160	52 min	$280
C	$420	42 min	$ 295	32 min	$225
D*	$760	16 min	$ 212	7 min	$ 89

*Program timed but not run; therefore figure is approximate--±10 percent.

The first computer is the facility referred to earlier; the second, third and fourth computers are "popular", high-speed digital computers, all of them scientific types[1]. The significance of the above figures lies in the fact that where one computer may better fulfill the programming requirements of a given type of problem, it may be relatively inadequate for a different problem. These differences may seem relatively slight at first glance, but consideration should be given to the fact that each of these simulations may be repeated numbers of times for changes in parameters and variables, and this can result in significant differences in costs.

In general recapitulation, simulation requirements and computer characteristics would under ideal circumstances be matched against an economic or other criterion to provide optimal computer/simulation rapport. In fact, other factors enter into the selection of a computer in most cases, these being of the artifactual quality previously described. Inasmuch as it is rare that the simulator can afford the luxury of an assortment of computers to select from, he is usually subject to the "Craps Game Restraint"[2]. Accordingly the foregoing notes are directed toward providing criteria not for optimizing the computer selection procedure, but rather for suboptimization. Perhaps the ultimate approach to selecting a computer for a simulation would be the design of a computer program which would have

[1] A scientific computer generally has different and more logical capability than a business-type computer and is designed for handling a much larger variety of problems than a business computer.

[2] If one wants to play craps and there is only one game in town, the choice of where one plays is academic.

appropriate decision rules and could evaluate the effectiveness of various computers in meeting the requirements of the given simulation. But how would one decide on the proper (and unbiased) computer to accomplish this task?

REMARKS ON STOCHASTIC SIMULATION

Gustave Rabson

1. WHY SIMULATE?

1.1. Let us consider a relatively simple problem. How fast does the population of a primitive organism increase under ideal conditions? Let us assume that reproduction is by fission: i.e., after a certain amount of time each organism is replaced by two new organisms. If we assume that each organism splits after one unit of time, we have

$$n(t) = P_o 2^{[t]} = P_o e^{[t]\ln 2}$$

$n(t)$ is the population after t units of time, $[t]$ is the greatest integer less than or equal to t, $\ln$ is the natural logarithm, P_o is the initial population and e is e.

If this seems too simple, we can make it look deeper by applying differential equations in the classical way:

$$\Delta n = kn\Delta t$$

for which we argue

$$n = P_o e^{kt} .$$

Although this approach looks deeper it really is not. The solution suffers from the same defect when we consider it as a model for the true situation. The length of time to fission is really a random variable in the first formulation and k is a random variable in the second formulation. If we want to apply this to life, we must ask the question:

Does the solution represent an expected value for the population or is it only an asymptotic value?

1.2. We may attack the problem analytically by means of some simplifying assumptions. Let us define:

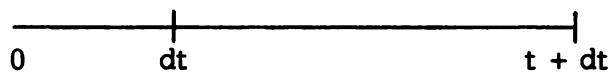
$$x(t) = \text{size of population at time } t$$

$$v_n(t) = E\left\{x(t) \mid x(0) = n\right\}$$

we then have

$$v_n(t + dt) = E\left\{x(t + dt) \mid x(0) = n\right\} .$$

Let us divide our interval $t + dt$ as follows:



and, assuming that the probability of a birth in the interval dt, given population n, is $\lambda n dt$, we have

$$v_n(t + dt) = v_{n+1}(t)\ \lambda n dt + v_n(t)\ (1 - \lambda n dt).$$

From this we obtain

$$\frac{dv_n}{dt} = \lambda n\ (v_{n+1} - v_n),\ n = 1,\ 2,\ \ldots$$

with

$$v_n(0) = n.$$

But

$$v_n(t) = nv_1(t).$$

So

$$\frac{dv_1}{dt} = \lambda(v_2 - v_1) = \lambda v_1.$$

From which we obtain

$$v_1 = v_1(0)\, e^{\lambda t} = e^{\lambda t}$$

and

$$v_n(t) = ne^{\lambda t}.$$

So we see that $ne^{\lambda t}$ is actually the expected population at time t.

It may be possible to continue this type of analysis in order to try to answer such questions as: Is there a limiting distribution of ages in the population and, if so, what is it? However the analysis becomes more and more complicated. Furthermore, the use of such analytic techniques is rather chancy. We may be able to solve the problem after a few days --but then again we may not be able to solve it at all. Furthermore, what about the assumption that the organism is equally likely to split at any time? This is certainly unrealistic. We may be able to develop a theory about when an organism is likely to split but until then we will probably want to use laboratory data.

1.3. While a solution by simulation methods is often unsatisfying, in the sense that Alexander's cutting of the Gordian knot was unsatisfying nevertheless, it is extremely powerful and can be based on experimental data. For the problem discussed above we may assume that we have laboratory data giving the distribution of ages at which splitting occurs, $P(\alpha)$ representing the fraction which split before reaching age α.

The flow chart in Figure 1 will enable us to study the growth of the population and the changes in the distribution of ages.

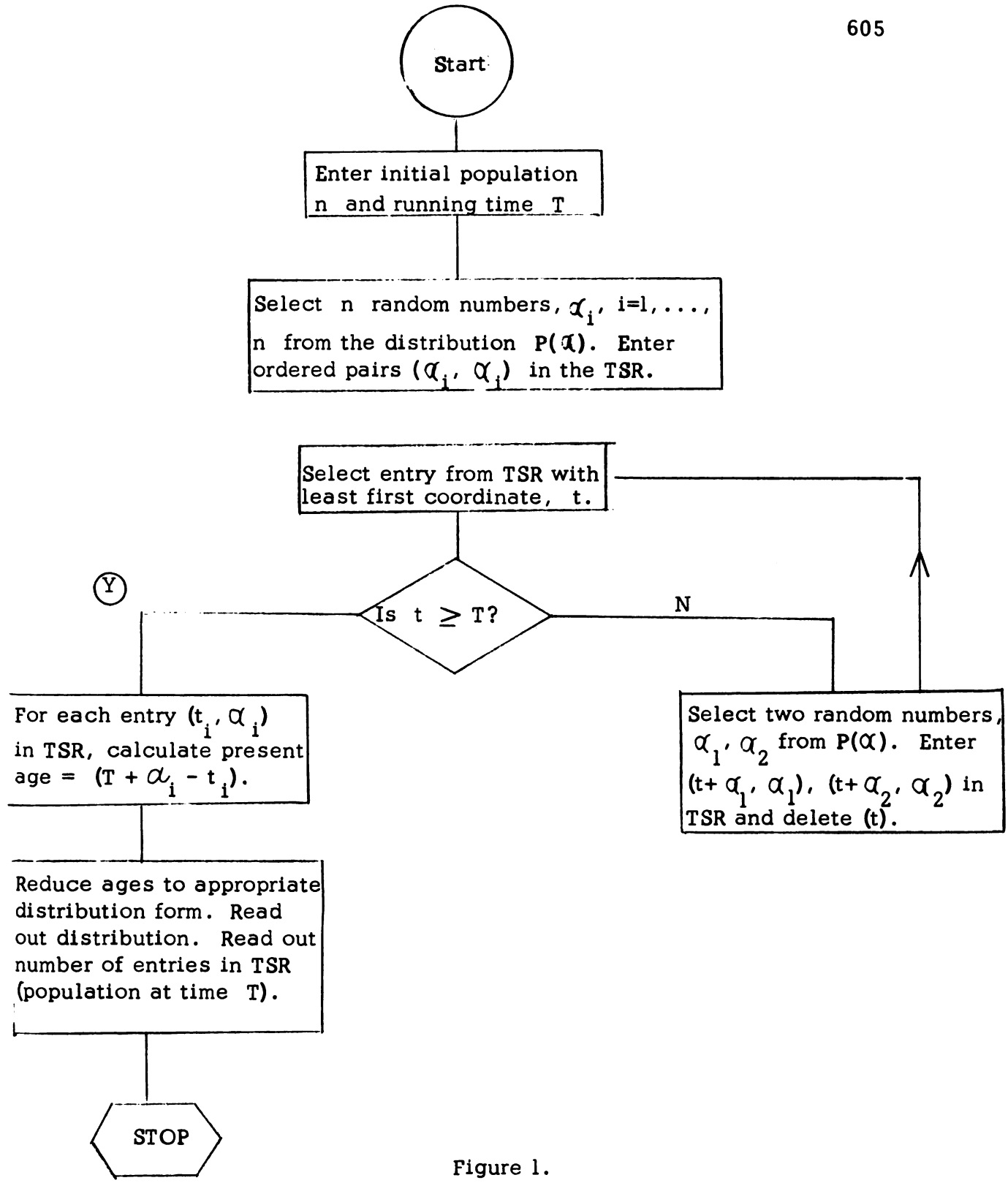


Figure 1.

1.4. The situation can easily become much more complicated. Let us consider for example the problem of growth in a limited environment. The classical approach to this problem is by means of the Verhulst-Pearl logistic curve:

$$\frac{dN}{dt} = aN\,\frac{K-N}{K}$$

where K is the maximum number of individuals that the population can support.

This seems to have the desired properties that as N approaches K the growth rate decreases and that $N = K$ is stable. However, this approach ignores the stochastic nature of the problem so that, as in the first approach to the problem of the unlimited environment we do not know how to interpret the solution. Furthermore many other functional forms have the same qualitative properties, e.g.:

$$\frac{dN}{dt} = aN\left(\frac{K-N}{K}\right)^3$$

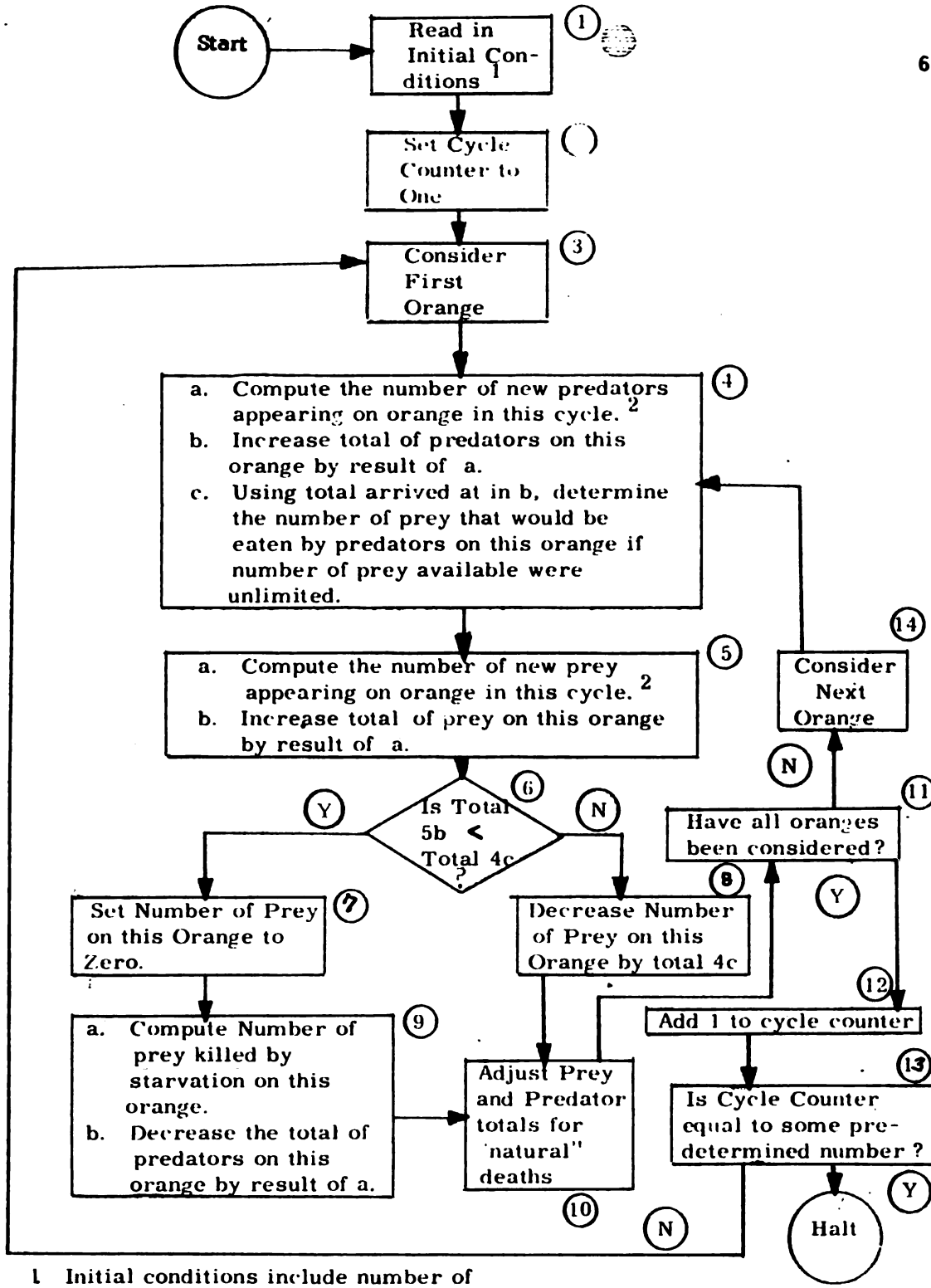
$$\frac{dN}{dt} = aN\cos\frac{\pi N}{2K} \quad .$$

1.5. Of course we have only scratched the surface of useful complexity. In order to study problems of parasitology and epidemiology we will certainly need to be able to handle at least the problem of a prey and a predator. This can be studied in a natural way by means of the following model.

A model such as this can be used to study such questions as

(1) The types of conditions under which both systems can coexist.
(2) Conditions that lead to oscillations in the populations.
(3) Distribution of ages.

This model represents a computer simulation of a laboratory experiment performed by C. B. Huffaker (Reference 2) in which he took two species of mites a prey (Eotetranychus sexmaculatus) which ate oranges and a predator (Typhodromus occidentalis) which ate the prey. Huffaker placed several oranges on a tray and separated them by barriers (vaseline lines) which were semi-permeable to the mites. Although such a laboratory experiment has the advantage that the biological constants involved are reasonable approximations to those found in nature it has the disadvantage that after a while the oranges grow moldy and that we are limited to studying only species which actually exist and can be grown in the laboratory. Experiments with a computer simulation can be validated by comparison with laboratory experiments, then used to fill in many more points on the curves showing how the population size varies with the biological parameters.



1. Initial conditions include number of oranges, number and initial distribution of prey and predators, migration routes open to mites, etc.

2. This includes births and migration,

1.6. My point, so far has been that, while analytic solutions to complicated stochastic models are highly desirable when they can be found, searching for them is somewhat chancy and they often involve assumptions which are made simply to make the mathematics more tractable. Furthermore, the solution is likely to involve gamma functions, definite integrals, etc., which we find acceptable only through familiarity. A flow chart is a function also, it is admirably suited for calculation, and deserves to be recognized as a legitimate analytic technique.

2. Technique of Modeling

One idea that is very useful in modeling and which perhaps has not been exploited as fully as it deserves is the idea of the time status record (TSR). In an air traffic control system, for example, it is possible to update time by an amount, Δt and estimate the position of each aircraft. However, the system operated only at discrete points and we will find that the computer is not really doing anything until the aircraft reaches one of the critical points. A program based on the time status record would not update time regularly but would keep a list of critical events (the time status record or TSR). The earliest event is selected from the list, processed, the list is modified on the basis of the processing and the earliest event is chosen again. I have illustrated the two methods in my two flow charts using the TSR for growth in an unlimited environment and the time update method for the prey-predator relation.

3. Purpose of the Model

In the above discussion we have introduced models for the purpose of determining desired distributions with a given degree of confidence. This implies that the model will be used with an experimental program to determine the distributions. This fact should be kept in mind throughout the course of the construction of the model since the experimental design will determine which factors must be varied (and therefore present) and which numbers are to be collected. It is obviously very embarassing to develop a model that does not include certain factors which it is desired to vary--but it is also embarassing to have a model which includes many more factors than one really wants to consider in the experimental design.

In determining the experimental design one needs to have some insight into the number of experiments that will have to be performed. It is a simple matter to estimate the number of runs required to determine

$\frac{1}{\pi}$ with 95% confidence by means of Buffon's needle experiment. We have 400 trials for 1 decimal accuracy and this goes up by a factor of 100 for each additional decimal place so 40,000 are required for 2 decimals and 4×10^6 for 3 decimals.

Several years ago I read a letter in one of the professional journals, by a well known psychologist who had tried the experiment 100 times to obtain a value of 3 for π. But no matter how many times he continued, he could not improve the estimate. He argued that this was due to lack of control of the initial conditions and that the Buffon needle should be dropped through a staw properly aligned with a plumb bob onto a sheet of paper which was flat and horizontal, etc. While he may be right in his requirement for more accurate controls, he really missed the basic point which is that the number of trials required goes up by a factor of 100 for each additional decimal point.

It appears then that Monte Carlo simulation of very simple problems may require an exorbitant number of runs. I have heard the remark made that for more complicated problems the law of large numbers can be invoked to reduce the number of runs. This may be true but I have not seen any mathematical justification of the remark. However by invoking certain variance reducing techniques, like the method of antithetic variables introduced by Hammersley and Morton [1] or the techniques discussed by Herman Kahn[3] in his report, we may be able to cut the number of runs down to something manageable. I do not believe that modelers have made as much use of these techniques as they should.

One technique that can sometimes be used to reduce the size of the experiment is to examine only the worst cases. This is sort of a minimax principle. We are not interested in winning an average war but a particular war. Consequently, perhaps we can assume that everything goes right for our enemy and wrong for us and see how badly we will do. In air traffic control, we would not be satisfied with a system that works well on the average, we want one that is adequete under the worst conditions--so perhaps we should assume Murphy's law "if anything can go wrong it will" to determine what will happen.

Sometimes a model is so large that it becomes out of the question to run any extensive experiments with it. I once asked a man who had built such a large model what its purpose was if he could not use it to perform

experiments and he answered that "he just wanted to see the kind of thing that could happen". That is to say that the purpose of the model was not to determine a distribution but to "educate". It seems to me that this is a legitimate purpose of a model but that, if this is indeed the purpose, the model should be prepared with that purpose in mind. Who is to be educated? Is it the computer, the programmer, the systems engineer or the administrator? Perhaps a semi-automatic type of model is best suited for educational purposes, that is a model in which the person to be educated plays a role in the simulation.

4. Validation

The most obvious remark about validation is that it is impractical. This is so obvious that it need not be said at all, however, if I do not say it, everybody else will.

A model is a function with an input and an output. To validate a model we would have to examine every input and show that we get the correct output for it. In a digital computer there is only a finite number but a big one.

The fact is that it is not easy to know what the "correct value" of the output should be. "Reality" can be approached only through models. Even in relatively simple cases the best that we can do is to use a statistical reduction of some simple data in lieu of reality. Consequently, we should probably never say that a model is valid but only that one model is a valid interpretation of another.

One approach that can serve to give us more confidence in our model would be to compare two conceptually different models. For example, a computer model and a field test.

Once we have obtained validation at some points by comparing the model with one that is acceptable as a standard, the problem of complete validation may be looked upon as a problem in interpolating or extrapolating the model (which is really a function). It may be that some classical results from interpolation theory would be applicable.

In a way perhaps we are taking the problem of validation too seriously. The canonical situation is that the modeler has complete confidence in his model because he understands it, but the administrator does not. The administrator doesn't admit that he doesn't understand it so he find a horde of objections about things that the modeler ignored. The modeler

duitifully runs out to add these factors to his model in such a way that they have the least effect possible. But the administrator is still not satisfied and adds more factors. Finally, the model and the modelers are so confused that nobody believes in the model anymore, and the administrator, if he is so unimaginative that he cannot think of any more factors raises the issue of validation.

Perhaps a more realistic approach would have been to do a more thorough job of transmitting the modeler's confidence to the sponsor in the first place. "Briefings" are not enough. In many cases an educational type of model in which the sponsor can actually play a role might be more convincing.

5. Some Problem Areas

I would like to close by summarizing some of the admonitions and problems on which I have touched.

5.1. When is a deterministic model, or an expected value model adequate? When do we need a stochastic model?

5.2. Do we want to use the time update method of modelling or the time status record? Are there general theorems that would enable us to make this decision? This is, of course, related to the question of estimating the time and storage capacity required.

5.3. How can we write our program in such a way that the scientist has as close contact as possible so that he can monitor it and make changes while it is running? This is related to parallel programming and time sharing.

5.4. How can variance reducing techniques be used in simulations of large scale systems? Are we really interested in studying the entire distribution or only the worst cases? Is there some sort of law of large numbers that saves us if our model is complicated enough?

5.5. Would a semi-automatic model better serve our purpose?

5.6. Another problem that should be mentioned, although I have not touched on it, is the problem of generating functions and distributions.

At the present time a great deal of storage and input of our models is devoted to this. However, various recent advances indicate that there are many clever techniques that can be employed to effect considerable reductions.

5.7. A final problem of considerable interest and importance is the problem of debugging. We must develop efficient and accurate methods for discovering codeing errors.

REFERENCES

1. Hammersley and Morton, "A New Monte Carlo Technique: Antithetic Variates", Proc. Cambr, Phil. Soc., Volume 52, pp. 449-475, 1956.

2. Huffaker, C. B., "Experimental Studies in Predation: Dispersion Factors and Predator Prey Oscillations," Hilgardia, 27, 343, 1958.

3. Kahn, Herman, "Use of Different Monte Carlo Sampling Techniques", Symposium on Monte Carlo Methods, H. A. Meyer, Ed., 1956.

4. Kahn, Herman, "Applications of Monte Carlo," Rand Corporation Research Memorandum RM-1237-AEC, 1954, Revised 1956.

PICATINNY ARSENAL TRAINING PROGRAMS IN STATISTICS

A. Bulfinch
Picatinny Arsenal

Early in our effort to further the use of statistics in the R&D work at PA, it was realized there were two major problems in attaining this end:

1. Selling the value of statistics.

It was decided that the basic approach would be one of selling, not one of executive order or decree. Since the project engineer is the responsible party for the development of hardware and the solution of problems, he should decide whether statistics is needed in any particular program. Otherwise he is reduced to a technician or an errand boy. This is not to say that the engineer should not have help. He should have. But we feel the help should be in the form of a training program. Not for the purpose of converting engineers into statisticians but for the purpose of using the training as a form of selling. A man responsible for completing an assignment in any field of endeavor will not use methods he is not familiar with and does not understand. In this sense we use training programs to give engineers an understanding of the new subject through systematic study and demonstrated application.

2. Communication between statistician and engineer.

Engineers and statisticians look at problems from entirely different viewpoints. The engineer focuses his attention on the concrete item and what it can do, while the statistician views the item in the abstract, and focuses his attention on the numerical values of measurable characteristics and what they represent. As a result engineers and statisticians have difficulty understanding each other. Our training programs are intended to alleviate this difficulty by teaching the engineer statistics rather than vice versa since the engineer holds the prime responsibility. We feel training of this kind can help the problem of communication in two ways: (a) Statistical training teaches the engineer how statistics can help him in his work and makes it possible for him to recognize situations in which statistics can be usefully applied. When situations of this kind occur the trained engineer can determine whether he needs a statistician; (b) If he calls on a statistician for advice he knows how to present his problem. This is most important since to exploit the value of experimental statistics, a close cooperative effort is required between the engineer and statistician.

Nearly a year was spent trying to find outside help to teach statistics at PA during working hours. The large number of people we planned to train made consideration of holding classes outside of PA impracticable. Holding classes after hours would add transportation and other personal problems to workload and absentee problems during hours. It is intended that every reasonable effort be made to encourage interest and participation by those people who are directly responsible for conducting experiments. We are especially interested in training first line supervisors since they are in a position to make greatest use of statistics.

Our effort to find outside help failed. In desperation we started the training program without outside help. This approach proved to be successful as shown by questionaires answered by the students and their supervisors and by the dollar savings made. In one instance, a testing program was planned without the use of statistics and then revised using experimental design techniques. As a result more information was obtained and a direct dollar saving of $360,000 was affected. This was accomplished in a testing program for only one component of a single missile system.

A training program has been started for reliability. Initially more than 25 government, educational, and industrial installations were contacted seeking help to conduct a course in reliability at PA during working hours. Again we were unsuccessful. We are now planning to conduct this training, without outside help.

We feel the need for training our engineers in the statistical aspects of reliability is more urgent than teaching the general subject of statistics. The problems in realiability are completely insurmountable without statistics. Since the technology of reliability is too new to be included in many engineering curriculums we cannot expect to hire reliability engineers. As a result we feel that PA must continue to furnish this supplementary training until more schools include this subject in their engineering courses.

A REVIEW OF A STATISTICAL WORKSHOP

Walter D. Foster
U. S. Army Chemical Corps, Fort Detrick, Frederick, Maryland
and
Theodore W. Horner
Booz-Allen Applied Research, Inc., Bethesda, Maryland

This statistical workshop was conceived and executed as a joint project by the Biomathematics Division at Fort Detrick and Booz-Allen Applied Research, Inc., Bethesda, Maryland. It was the purpose of this workshop to arouse interest in statistical methods as applied to specific problems arising from the subject matter field of Aerobiology. As such, the entire program was oriented clinically towards problem solving. The conference was arranged four afternoons in a working week, with emphasis on consecutive meetings lest the continuity from one session to the next be diminished by increasing the intervals of time between sessions.

The overriding principle underlying the presentations of each afternoon was the simplicity of approach, stressing non-technical aspects of statistical principles. In no case was any speaker allowed to become involved in an extensive derivation of statistical theory or probability underlying the methods described. Nor was the session allowed to become merely a recipe of sterile methodology with illustrated examples of application. Rather, it strove to depict the principles of statistics as applied to an experimental situation in terms both appealing and attractive as well as understanding to an audience whose formal statistical training was assummed to be near zero.

The session on the first day was opened by a message from the Chief of the Aerobiology Division at Fort Detrick, followed by a message to workshop participants by the Chief of Biomathematics Division. To set the stage for the first technical session, an experimenter from whose field of application this particular problem was taken gave a fifteen to twenty minute presentation describing the background of the problem from current results, and the need for this particular unit of experimentation, indicating objectives of this experiment in his presentation. The second speaker was a statistician. It was his job to synthesize the problem from the subject matter terminology into statistical terminology as simply and directly as possible. Ten to fifteen minutes were allowed for this.

Adhering to the criterion that no uninterrupted session should proceed more than an hour without a break, two or three coffee breaks were scheduled per session. At this point, a coffee break intervened.

The third speaker was also a statistician whose purpose was to describe in non-technical terms, the principles and basic concepts of certain applicable and appropriate statistical methodology primarily aimed to answer the questions of the second speaker who had synthesized the problem to a statistical one. This presentation of the concepts of statistical methodology was given a full hour with a question period scheduled afterwards.

After a second coffee break, the second speaker, the statistician who had synthesized the problem, returned to give the application of the statistical methodology to the specific questions raised by the experimenter and to give his interpretation of the results. The example chosen was specifically selected to illustrate the principles involved in the analysis of variance techniques. After the finish of the presentation by speaker No. 4, the floor was thrown open to discussion. This was the end of the session for the first day.

The second session was centered about linear regression as a statistical methodology, introduced again by means of a particular problem in the subject matter field. The first speaker presented a problem from the subject matter point of view in a ten minute speech. A statistician was asked to translate the problem from the subject matter field to statistical questions susceptible of specific answers. A third statistician then gave an outline of the concepts and fundamentals in the statistical methodology of linear regression for answering these questions raised by the second speaker. After the break, a third speaker again a statistician, undertook to apply the principles to the specific questions and endow his results with interpretations in terms of subject matter terminology. A surprise was in store for the audience at this point, for the data, as data often do, failed to conform to the expected behavior indicated in the initial statement of the problem and the following synthesis to a statistical question. After a coffee break, the third statistician indicated the nature of this new problem to the original synthesizer, explaining ramifications involved in the data and asking for a re-synthesis on the basis of the information provided by the data through the analysis as far as it had been carried. An additional synthesis was accorded the data by the second speaker. The synthesis then required additional methodology from the third speaker,

who then expanded his original remarks to include a prescription, again in conceptual and simple terms, for the difficulties indicated by the first inspection of the data. The opportunity for describing model building was not lost and even though some of the derivation and procedures used in model building were taken for granted by the audience, it was worthwhile re-presenting these to show that a statistician must ever be alert to create models and to use existing ones which are appropriate for a given situation.

As usual, there were periods for discussion either after each presentation, or if a speaker desired during a discussion for clarification of certain points. It was gratifying to discover a rather spirited discussion in the presentation of the problem involving the linear regression method.

Since it was the avowed aim of the entire workshop to arouse interest in the statistical methodology, the high point of the entire workshop centered around a statistical game which for a catchy title used the phrase "Ye Olde Confidence Game." A full description of this game and its results are not given here but separately by Dr. Horner in a paper with this as a title. The playing of this game and the interpretation of the results, while not occupying a full three hour session, was considered sufficient for a session without adding additional material.

The final session of the fourth afternoon involved a major presentation by a statistician who described additional statistical tools in their conceptual form as opposed to either theory or derivation. Very brief applications of each suggested tool were included even though these applications may not have been appropriate for the particular subject matter field of Aerobiology. Nevertheless, this particular presentation was very well received by the members of the audience.

A second topic for the last session involved a panel discussion by five statisticians on a series of questions originating from the subject matter field. The questions had been collected, described in a short paragraph and circulated to the panelists well before the session itself in order that they might have cogent and appropriate remarks without the need to investigate the problem during the session itself. Another panelist was "planted" in the audience to stimulate discussion from the audience during the panel discussion itself. A moderator for the panel at the close of the discussion for each problem then summarized the degree to which each problem had been answered and suggested the completeness of each answer as possibly applied to the particular situation which gave

rise to the question. The workshop was then summarized by the Chief of Biomathematics Division at the end of the fourth session.

In order to make the entire workshop as attractive as possible, each speaker came prepared with hand-outs covering the gist and extent of his own presentation. Each hand-out was given to members of the audience at the beginning of each speaker's presentation as a visual reference for the material as the speaker progressed. A notebook was provided each member of the audience with spacers included to help organize the several sessions. In this way the members of the audience were able to take away from the workshop, a notebook of some hundred pages together with their understanding of the material as it was presented.

YE OLDE CONFIDENCE GAME

Theodore W. Horner
Booz, Allen Applied Research, Inc.
Walter D. Foster
Biomathematics Division, U. S. Army Chemical Corps

A scientific investigator throughout the course of a lifetime may have occasion to examine many sets of data. In some of these, real treatment differences may be present; while in others such treatment differences may be either nonexistent or sufficiently small as to be without interest. In some sets of data the investigator may wish to form an opinion as to the existence of real treatment differences. He may do this simply on the basis of a visual examination of the data, or he may employ some form of statistical methodology. In regard to visual examination there are undoubtedly individuals who have a "feel" for numbers either natural or acquired. In order to obtain a somewhat objective measure of this ability, a game entitled "Ye Olde Confidence Game" has been designed.

A confidence game is generally the name of a game (e.g., the 3-shell game) which looks attractive to the taker but in reality is fixed to "fleece" him. The "victim" obviously feels he can beat the game, but seldom does. The title of this game was suggested by the confidence interval, which plays such an important role in statistical methodology.

The purpose of the game is to enable participants to compare intuitive conclusions drawn solely on the basis of visual examination of data items and treatment means with correct conclusions known from the way the data were constructed. Each game participant is asked to examine eighteen sets of data which are different for each player; Exhibit 1 shows a sample. Each data set has ten observations on each of two treatments, an A treatment and a B treatment. The observed means for each treatment also accompany each data set. In some sets, the true treatment difference is zero, in others there is a small difference, while in others there is a large difference. The three types (1, 2, and 3) do not occur with equal frequency. The sets examined by each player constitute a random sample of the three types. Each player has a different random sample. The player is asked to decide whether the true treatment difference is zero or non-zero and to record his conclusion for each set as $A = B$ or $A \neq B$. The participant with the highest number of correct answers is judged the winner of the game.

In addition to competing with each other, each player also competes with the statistical F-test at the five per cent level. Like the participant, the F-test for any particular set of data may indicate an erroneous conclusion. However, in those sets in which the true treatment difference is zero, the F-test should enable the correct conclusion to be drawn 95 per cent of the time. In those with a small treatment difference, the correct conclusion should be drawn about 76 per cent of the time, and in those with the large treatment difference, 100 per cent of the time. In the present game, data sheets are available for 100 players.

This game was first used at a Statistical Workshop for a department composed primarily of biologists. The percentages of correct answers from participants in the Workshop are shown below:

Type of difference	Percentages of correct answers	
	Workshop participants	Expected on basis of F-test
1	73.6	95%
2	74.4	78%
3	88.4	100%

In constructing the game, the following model was used to generate the sample observations:

$$y_{ij} = \begin{cases} u - D/2 + e_{ij} & \text{for } j = 1 \\ u + D/2 + e_{ij} & \text{for } j = 2 \end{cases}$$

where y_{ij} is the i^{th} $(i = 1, 2, \ldots n)$ observation on the j^{th} treatment, u is a constant, D is the true difference between the treatments, and e_{ij} is a random error deviation. The e's for both treatments were random observations from the same normal population, this population having a mean of zero and a variance of σ^2. The parameters u, D, and σ^2 were randomly chosen for each data set within the restriction of the type of true difference (none, small, and large) assigned at random to the data set according to certain probabilities.

The computations needed in developing the game were performed on the Remington Rand Solid State 90 Computer at Fort Detrick. The authors are indebted to Mr. Arthur J. Dukes of the Biomathematics Division at Fort Detrick for programming the calculations.

For the benefit of the reader who would like to try the game, the key for the data sets of Exhibit 1 is attached as Table 1.

The game is a simple one to administer and score, and a set of data sheets and an instruction booklet can be made available to interested groups by writing to Dr. Foster.

EXHIBIT I

1		2		3		4		5		6	
A	B	A	B	A	B	A	B	A	B	A	B
593	484	674	468	250	318	462	631	157	186	443	447
491	564	553	525	318	273	526	597	160	136	469	563
657	621	461	641	333	253	551	583	147	144	412	562
510	573	685	775	291	239	622	573	115	146	384	550
595	568	477	623	335	314	595	489	117	152	338	488
513	632	622	662	284	275	540	629	165	133	458	406
624	670	655	682	276	314	527	711	151	155	423	574
605	650	600	674	275	310	606	576	179	156	489	436
579	570	590	582	383	326	533	580	145	165	436	588
566	474	685	541	295	374	630	652	190	151	491	510
573	581	600	617	304	300	559	602	153	152	434	512

7		8		9		10		11		12	
A	B	A	B	A	B	A	B	A	B	A	B
245	125	130	196	230	209	253	299	486	372	157	330
240	181	209	161	250	321	191	278	609	562	162	300
274	192	186	217	259	282	237	312	476	461	176	291
341	244	191	207	202	230	237	296	461	397	196	329
244	200	189	184	297	219	240	252	387	554	201	321
195	203	210	203	240	224	229	259	606	443	216	283
193	188	202	162	228	157	242	282	531	387	154	216
243	307	171	170	261	238	245	216	567	551	211	304
207	193	115	186	225	234	195	294	400	272	199	266
253	184	128	162	310	243	213	277	606	409	168	256
244	202	173	185	250	236	228	276	513	441	184	289

13		14		15		16		17		18	
A	B	A	B	A	B	A	B	A	B	A	B
610	484	492	653	531	477	174	215	319	350	294	456
579	411	637	543	449	399	136	239	452	173	347	496
457	373	604	632	368	608	195	201	645	447	353	443
537	398	645	551	626	714	182	204	408	565	288	438
611	569	531	624	799	470	203	203	386	477	326	344
642	390	599	638	606	485	230	228	369	547	415	566
517	446	550	445	512	650	164	187	496	461	441	387
596	373	620	593	405	509	197	259	417	410	428	436
687	304	476	503	525	353	216	205	373	525	357	517
528	476	562	535	475	630	148	216	635	546	313	393
576	422	572	572	530	529	184	216	450	450	356	448

TABLE I

Key for Page 1 of "Ye Olde Confidence Game"

Data Set	Correct Conclusions	F test result at 0.05 Significance Level
1	A = B	Non-significant
2	A = B	Non-significant
3	A = B	Non-significant
4	A = B	Non-significant
5	A = B	Non-significant
6	A ≠ B, 2	Significant
7	A ≠ B, 2	Non-significant (Type II error)
8	A = B	Non-significant
9	A = B	Non-significant
10	A ≠ B, 2	Significant
11	A ≠ B, 2	Non-significant (Type II error)
12	A ≠ B, 3	Significant
13	A ≠ B, 2	Significant
14	A = B	Non-significant
15	A = B	Non-significant
16	A ≠ B, 2	Significant
17	A = B	Non-significant
18	A ≠ B, 2	Significant

ON LINEAR REGRESSION SYSTEMS

William S. Mallios*
Booz, Allen Applied Research, Inc.

I. SUMMARY. Methods are given for the joint estimation of the coefficients in a system of (a set of simultaneous) linear, "predictive,"[5] regression models, where the predictor (independent) variables are at least controlled but not necessarily the same from one model to the next, and the errors are correlated between, but not within, models; e.g., let there exist, say, q models which correspond to q characteristics of an experimental unit; if q response type variables are measured from each of, say, n randomly selected units, then it is assumed that the q measured variables are correlated within, but not between, experimental units.

When the correlations are known, joint or weighted least squares is applied in increasing the precision of some or all of the coefficient estimates (except the intercept terms) relative to the precision of the separate least squares estimates, separate with respect to each model. In the likely case where the correlations are unknown, an iterative estimation scheme is proposed along with a sufficient condition for convergence and an approximate covariance matrix for the resultant estimates. With unknown correlations, the precision of the estimates is possibly increased.

The foregoing is then discussed with regard to its practical implication in not only multivariate problems, but also in those of a univariate nature.

II. GENERAL ASSUMPTIONS. Suppose the mechanics of an event or phenomenon of interest are to be studied through q of its characteristics. Let $\underline{\eta}' = (\eta_1, \eta_2, \ldots, \eta_j, \ldots, \eta_q)$ denote a vector of uncontrolled, random (from one occurrence of the event to the next) variables with $\underline{\eta}$ being a true measure of the characteristics for a particular occurrence of the event. Denote by $\underline{\xi}' = (\xi_1, \xi_2, \ldots, \xi_i, \ldots, \xi_p)$ a set of controlled variables associated with the event, and suppose that only a subset of $\underline{\xi}$ need be used as a predictor for any particular η_j.

---

*Formerly at North Carolina State College

By specifying the f_j in

$$(1) \qquad \eta_j = f_j(\underline{\eta}(s_j), \underline{\xi}_j; \underline{\beta}_j) + \ddot{\epsilon}_j$$

there results a set of hypothesized structural regression models where $\underline{\eta}(s_j)$ and $\underline{\xi}_j$ are either identical to or a subset of $\underline{\eta}$ and $\underline{\xi}$ respectively; $\underline{\beta}_j$ is a vector of unknown parameters; and $\ddot{\epsilon}_j$ is the model error which is assumed additive for all models. With the models in (1) being linear, the reduced form of the models is, say,

$$(2) \qquad \eta_j = \alpha_{oj} + \sum_{j=1}^{p_j} \beta_{ij} \xi_{ij} + \dot{\epsilon}_j.$$

Example 1. If $q = 2$, $p = 3$, $\underline{\xi}' = (\xi_1, \xi_2, \xi_3) = (\xi_1, \xi_1^2, \xi_3)$, and the hypothesized structural models are

$$(3) \qquad \eta_1 = \dot{\beta}_{01} + \dot{\beta}_{11}\xi_1 + \dot{\beta}_{21}\xi_1^2 + \dot{\beta}_{31}\xi_3 + \dot{\beta}_{41}\eta_2 + \ddot{\epsilon}_1$$

$$(4) \qquad \eta_2 = \dot{\beta}_{02} + \dot{\beta}_{12}\xi_1 + \ddot{\epsilon}_2$$

then by substituting (4) into (3), the reduced models become

$$(5) \qquad \eta_1 = \alpha_{01} + \beta_{11}\xi_1 + \beta_{21}\xi_1^2 + \beta_{31}\xi_3 + \dot{\epsilon}_1$$

and (4) where $\alpha_{01} = \dot{\beta}_{01} + \dot{\beta}_{31}\dot{\beta}_{02}$, $\beta_{11} = \dot{\beta}_{11} + \dot{\beta}_{31}\dot{\beta}_{12}$, $\beta_{21} = \dot{\beta}_{21}$, $\beta_{31} = \dot{\beta}_{31}$, and $\dot{\epsilon}_1 = \ddot{\epsilon}_1 + \dot{\beta}_{31}\ddot{\epsilon}_2$.

Let the η_j be measured with error by, say, $y_j = \eta_j + \delta_j$. Since the ξ_i, measured by $\dot{x}_i$, are controlled, the difference $\underline{\dot{x}} - \underline{\xi}$ is neglected and $\underline{\dot{x}}$ is set equal to $\underline{\xi}$. Assume the availability of n independently distributed vectors $(y_{1k}, y_{2k}, \ldots, y_{jk}, \ldots, y_{qk})$, $k = 1,2, \ldots, n$, stemming from, say, n randomly selected experimental units with the corresponding $\underline{\dot{x}}$ vectors being the n row vectors of $\dot{X}(nxp) = (\underline{\dot{x}}_1 \underline{\dot{x}}_2, \ldots, \underline{\dot{x}}_i, \ldots, \underline{\dot{x}}_p)$. Setting $x_{ik} = \dot{x}_{ik} - \sum_{k=1}^{n} x_{ik/n}$ and

(6) $$X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_i, \ldots, \underline{x}_p)$$

then with $y_{jk} = \eta_{jk} + \delta_{jk}$, (2) becomes

(7) $$y_{jk} = \beta_{oj} + \sum_{j=1}^{p_j} \beta_{ij} x_{ij} + \epsilon_{jk}$$

where $\epsilon_{jk} = \dot{\epsilon}_{jk} + \delta_{jk}$, and the sample form of the models in (7) is

(8) $$y_{jk} = b_{oj} + \sum_{j=1}^{p_j} b_{ij} x_{ij} + e_{jk}.$$

Next assume that $\underline{\dot{\epsilon}}_k (1xq) = (\epsilon_{1k}, \epsilon_{2k}, \ldots, \epsilon_{jk}, \ldots, \epsilon_{qk})$ is distributed with expectation $\underline{0}$ and that

$$\text{var } \underline{\dot{\epsilon}}_k = E(\underline{\dot{\epsilon}}_k \underline{\dot{\epsilon}}'_{k'}) = \sum(qxq) = (\sigma_{jj'}) \qquad \text{for } k = k'$$

$$= 0(qxq) \qquad \text{for } k = k',$$

$j, j' = 1, 2, \ldots, q$, where it may be noted that the off diagonal elements of $\sum$ are nonzero

(i) if the measurements $(y_{1k}, y_{2k}, \ldots, y_{jk}, \ldots, y_{qk}) = \underline{\dot{y}}_k'$ stem from the same experimental unit (rendering the model errors dependent),

(ii) if the σ's or measurement errors within $\underline{\dot{y}}_k$ are correlated,

(iii) or if certain parameters such as $\dot{\beta}_{31}$ in (3) are nonzero.

With $\underline{y}_j' = (y_{j1}, y_{j2}, \ldots, y_{jk}, \ldots, y_{jn})$, it follows that

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \ldots \\ \underline{y}_j \\ \ldots \\ \underline{y}_q \end{bmatrix} = \left\{ \underline{1}^* \begin{bmatrix} \beta_{01} \\ \ldots \\ \beta_{oj} \\ \ldots \\ \beta_{oq} \end{bmatrix} + \begin{bmatrix} X_1 & & & \\ & \ddots & & \\ & & X_j & \\ & & & \ddots \\ & & & & X_q \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \ldots \\ \underline{\beta}_j \\ \ldots \\ \underline{\beta}_q \end{bmatrix}, \quad I^* \sum, \right\}$$

or

(9) $$\underline{y} : (\underline{1}^* \underline{\beta}_o + T\underline{\beta}, I^* \omega\sigma^2),$$

i.e., $\underline{y}$ is distributed with expectation $\underline{1}^* \underline{\beta}_o + T\underline{\beta}$ and variance $I^* \sum$, where $*$ denotes a direct or Kronecker product, $T(qnx \sum_{j=1}^{q} p_j)$ is a

pseudo diagonal matrix with typical element X_j (nxp_j), $\underline{\beta}$ $(\sum_{j=1}^{q} p_j x1)$

is a vector with typical pseudo element $\underline{\beta}_j(p_j x 1)$, and $\sum = \omega\sigma^2$.

III. ESTIMATION OF $\underline{\beta}$ WHEN ω IS KNOWN. Letting $A^{1/2}$ be the square root of $A = I^* \omega$, then the transformed vector

$$(10) \qquad A^{-1/2'}(\underline{y} - 1^* \underline{\beta}_o - T\underline{\beta}) = A^{-1/2'}\underline{\epsilon} : (\underline{0}, I\sigma^2)$$

where $\underline{\epsilon}'(1xnq) = (\underline{\epsilon}_1', \underline{\epsilon}_2', \ldots, \underline{\epsilon}_j', \ldots, \underline{\epsilon}_q')$ and $\underline{\epsilon}_j'(1xn) = (\epsilon_{j1}, \epsilon_{j2}, \ldots, \epsilon_{jk}, \ldots, \epsilon_{jn})$. If $\underline{e}$ is the sample form of $\underline{\epsilon}$ as in (8), then (10) directs one to minimize $\underline{e}'A^{-1}\underline{e}$ ($\underline{e} = (\underline{y} - 1^*\underline{b}_o - T\underline{b})$) with respect to $\underline{b}_o$ and $\underline{b}$. Doing so, the joint least squares estimates of the β_{oj} and $\underline{\beta}$ are $b_{oj} = \bar{y}_j$ and

$$(11) \qquad \underline{b} = (T'A^{-1}T)^{-1}T'A^{-1}\underline{y}.$$

Example 2. Let $q = 2$ and

$$\sum = \omega\sigma^2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\sigma^2.$$

Then (11) becomes

$$\begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & -\rho X_1'X_2 \\ -\rho X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'(\underline{y}_1 - \rho\underline{y}_2) \\ X_2'(\underline{y}_2 - \rho\underline{y}_1) \end{bmatrix}.$$

Example 3. Let $X_j = X$ for all j. Then (11) reduces to

$$\begin{aligned} \underline{b} &= \left[X'X * \omega^{-1}\right]^{-1} \left[X' * \omega^{-1}\right] \underline{y} \\ &= \left[(X'X)^{-1} * \omega\right] \left[X' * \omega^{-1}\right] \underline{y} \\ &= (X'X)^{-1} X' * I\underline{y} \end{aligned}$$

so that

$$\underline{b}_j = (X'X)^{-1} X' \underline{y}_j .$$

(See reference (1) for an alternative derivation.) Thus when the models contain precisely the same independent variables, the weight matrix ω offers no information in the estimation of the β_j, and the separate least squares estimation of each $\underline{\beta}_j$ is in order.

The intuitive arguments underlying the foregoing developments become apparent in the following examples.

Example 4. Consider the models

$$y_1 = \alpha_o + \epsilon_1$$

$$y_2 = \beta_o + \beta_1 x + \epsilon_2$$

under the assumptions that $E(\epsilon) = 0$, $E(\epsilon^2) = \sigma^2$, and

$$\begin{aligned} E(\epsilon_{1k} \epsilon_{2k'}) &= \rho \sigma^2 && \text{if } k = k' \\ &= 0 && \text{if } k \neq k'. \end{aligned}$$

with ρ known, the joint least squares estimate of the coefficients are $a_o = \bar{y}_1$, $b_o = \bar{y}_2$ and

$$a_1 = \sum xy_2 / \sum x^2 - \rho \sum xy_1 / \sum x^2$$

with corresponding variability

$$\text{var} \begin{bmatrix} a_o \\ b_o \\ b_1 \end{bmatrix} = \begin{bmatrix} 1/n & \rho/n & 0 \\ \rho/n & 1/n & 0 \\ 0 & 0 & (1 - \rho^2) / \sum x^2 \end{bmatrix} \sigma^2.$$

Consider Figure 1 where $\hat{y}_1 = a_o$ and $\hat{y}_2 = b_o + b_1 x$ are the best unbiased prediction lines for y_1 and y_2, $\tilde{y}_1 = \tilde{a}_1 x$ is an unbiased prediction line for y_1 which is obtained by regressing y_1 on x, and $\tilde{y}_2 = \tilde{b}_o + \tilde{b}_1 x$ is the unbiased prediction line for y_2 determined by least squares independently of y_1. But if, as in Figure 1, $\tilde{a}_1 > 0$ and, say, $\rho > 0$, then it necessarily follows that $\tilde{y}_2 = \tilde{b}_o + \tilde{b}_1 x$ is "overly steep." Consequently, estimation through joint least squares accounts for the sampling error, utilizes the information of known positive correlation and $\alpha_1 = 0$, and corrects $\tilde{b}_1$ by an amount $-\rho \tilde{a}_1$, making $b_1 = \tilde{b}_1 - \rho \tilde{a}_1$ the best unbiased estimate of β_1 under the given system of models.

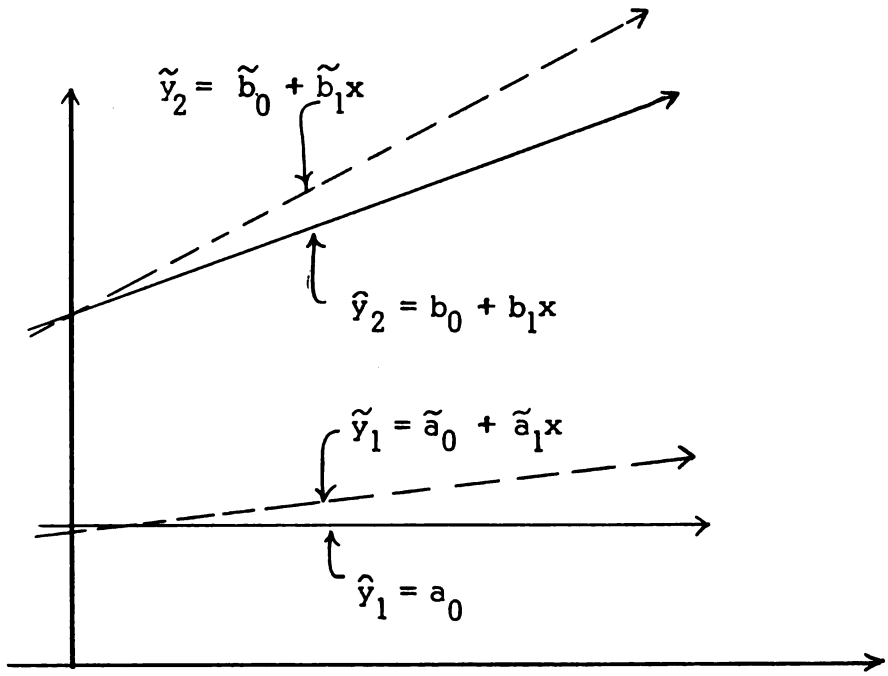


FIGURE I
Unbiased and Best Unbiased Prediction Lines

Example 5. Consider the models

$$y_1 = \alpha_o + \alpha_1 x_1 + \epsilon_1$$

$$y_2 = \beta_o + \beta_2 x_2 + \epsilon_2$$

under the same assumptions as the previous example. Let x_1 and x_2 be functionally independent with $\sum x_1 x_2 = 0$. From (11) the joint least squares estimates of the slopes are

$$a_1 = \sum x_1 y_1 / \sum x_1^2 - \rho \sum x_2 y_1 / \sum x_2^2$$

$$b_2 = \sum x_2 y_2 / \sum x_2^2 - \rho \sum x_1 y_2 / \sum x_1^2.$$

Although in the population $y_1 (y_2)$ remains constant along the $x_2 (x_1)$ dimension, in the sample, $\sum x_2 y_1$ and $\sum x_1 y_2$ will be nonzero. With a known $\rho \neq 0$, joint least squares accounts for the information supplied by the population and adjusts the slope estimates for sampling error, thus increasing their precision.

It was seen in example 3 that the correlations do not enter in least squares estimation of the coefficients. From examples 4 and 5, the correlation was used in adjusting slope estimates when a type of constraint (viz., that certain coefficients be zero in the population) was placed on one or more of parameters. However, when the models include the same independent variables, no constraints are available. Hence a knowledge of the correlations offers irrelevant information in the estimation of the coefficients.

The following theorem now becomes apparent.

Theorem: If under (9), ω is known, $X_1 \neq X_2 \neq \ldots \neq X_j \neq \ldots \neq X_q$, and $X_1 \subset X_2 \subset \ldots \subset X_j \subset \ldots \subset X_q$, i.e., X_j is contained in X_{j+1}, then $\underline{y}_j$ enters in the joint least squares estimation of some or all of the elements of the vectors $\underline{\beta}_j, \underline{\beta}_{j+1}, \ldots \underline{\beta}_q$, but does not enter in the joint least squares estimation of $\underline{\beta}_{j-1}, \underline{\beta}_{j-2}, \ldots, \underline{\beta}_1$.

Corollary: Under the assumptions of the previous theorem, the best unbiased estimate of $\underline{\beta}_1$ is $\underline{b}_1 = (X_1' X_1)^{-1} X_1' \underline{y}_1$.

(Proofs are given in reference (4).)

Finally, it is easily shown that

$$\text{(12)} \qquad \text{var } \underline{b} = (T'A^{-1}T)^{-1}\sigma^2 \qquad \text{if } X_j \neq X \text{ for some } j$$

$$\text{(13)} \qquad = (X'X)^{-1} * \sum \qquad \text{if } X_j = X \text{ for all } j.$$

((12) and (13) are derived in (4); an alternative derivation of (13) is given in (1).) In (12), σ^2 is estimated by

$$s^2 = (\underline{y} - \underline{1}*\underline{b}_o - T\underline{b})'A^{-1}(\underline{y} - 1*\underline{b}_o - T\underline{b})/(nq - \sum_{j=1}^{q} p_j - q)$$

and the element $\sigma_{jj'}$, in $\sum$ by

$$s_{jj'} = (\underline{y}_j - b_{oj}\underline{1} - X\underline{b}_j)' (\underline{y}_j - b_{oj'}\underline{1} - X\underline{b}_{j'}) /n - p - 1.$$

IV. ESTIMATION OF $\underline{\beta}$ WHEN ω IS UNKNOWN. Assume normality in (9) with no a priori knowledge of ω. If an attempt should be made to find estimates of $\underline{\beta}_o$, $\underline{\beta}$, and $\sum$ such that the likelihood function

$$L = (2\pi)^{-nq/2} |\sum|^{-n/2} \exp\left\{-1/2 \sum_{k=1}^{n} \left[\dot{\underline{y}}_k - E(\dot{\underline{y}}_k)\right]' \sum^{-1} \left[\dot{\underline{y}}_k - E(\dot{\underline{y}}_k)\right]\right\}$$

is maximized, the nonlinear system of equations resulting from equating to zero partial derivatives of L is irresolvable except possibly in simpler cases. As such, it is the intent to estimate $\underline{\beta}$ through an iterative scheme where $\omega(= \sum \sigma^{-2})$ is estimated from the data on hand.

First estimate through separate least squares each $\underline{\beta}_j$ by, say, $\underline{b}_j^{(o)}$; or there are numerous other ways by which initial estimates of the

$\underline{\beta}_j$ are attainable. Then from the $\underline{b}_j^{(o)}$ stems an estimate, say, $W^{(o)}$, of ω, where the matrix of residual sums of squares and cross products

$$(SSE_{jj'}) = \left[(\underline{y}_j - b_{oj}\underline{1} - X_j\underline{b}_j^{(o)}) \cdot (\underline{y}_{j'} - b_{oj'}\underline{1} - X_{j'}\underline{b}_{j'}^{(o)}) \right]$$

is proportional to $W^{(o)}$. But what is the proper divisor for $SSE_{jj'}$ $(j \neq j')$ if at least an unbiased estimate of $\sigma_{jj'}$, is desired?

To answer the latter question, equate (7) and (8) for the j^{th} model so that

$$\underline{e}_j^{(o)} = \phi_j(\underline{\gamma}_j - \underline{C}_j^{(o)}) + \underline{\epsilon}_j \tag{14}$$

where $\phi_j = (\underline{1} | X_j)$, $\underline{\gamma}_j' = (\beta_{oj}, \underline{\beta}_j')$ and $\underline{C}_j^{(o)'} = (b_{oj}, \underline{b}_j^{(o)})$. Since

$$\underline{C}_j^{(o)} = (\phi_j' \phi_j)^{-1} \phi_j' \underline{y}_j$$

$$= \underline{\gamma}_j + (\phi_j' A^{-1} \phi_j)^{-1} \phi_j' A^{-1} \epsilon_j, \tag{15}$$

by equating (14) and (15)

$$\underline{e}_j^{(o)} = \left[I - \phi_j (\phi_j' \phi_j)^{-1} \phi_j' \right] \underline{\epsilon}_j$$

$$= (I - U_j)\, \underline{\epsilon}_j \quad \text{(say)}$$

where $U_j(nxn)$ has rank $P_j + 1$. For the expectation of $\underline{e}_j^{(o)'}\underline{e}_{j'}^{(o)}$,

$$E(\underline{e}_j^{(o)'}\underline{e}_{j'}^{(o)}) = E\left[\underline{\epsilon}_j'(I - U_j)'(I - U_{j'})\underline{\epsilon}_{j'}\right]$$

$$= n\sigma_{jj'} - \sigma_{jj'}trU_j' - \sigma_{jj'}trU_{j'} + \sigma_{jj'}tr(U_j'U_{j'})$$

where tr denotes trace. But

$$tr\ U_j = tr\left[\phi_j(\phi_j'\phi_j)^{-1}\phi_j'\right] = tr\left[\phi_j'\phi_j(\phi_j'\phi_j)^{-1}\right] = p_j + 1$$

and similarly, $tr\ U_{j'} = p_{j'} + 1$. Also if ϕ_j is contained in $\phi_{j'}$, then $\phi_{j'}$ may be written as an augmented matrix

$$\phi_{j'}\left[nx(p_{j'} + 1)\right] = \left\{\phi_j\left[nx(p_j + 1)\right] \mid \tilde{\phi}_{j'}\left[nx(p_{j'} - p_j)\right]\right\},$$

and

$$tr\ (U_j'U_{j'}) = tr\left[\phi_j(\phi_j'\phi_j)^{-1}\phi_j'\phi_{j'}(\phi_{j'}'\phi_{j'})^{-1}\phi_{j'}'\right]$$

$$= tr\left[\phi_j(\phi_j'\phi_j)^{-1}\phi_j'(\phi_j|\tilde{\phi}_{j'})\begin{bmatrix}\phi_j'\phi_j & \phi_j'\phi_{j'} \\ \tilde{\phi}_j'\phi_j & \tilde{\phi}_j'\tilde{\phi}_{j'}\end{bmatrix}^{-1}\begin{bmatrix}\phi_j' \\ \tilde{\phi}_{j'}'\end{bmatrix}\right]$$

$$= \operatorname{tr}\left[\begin{array}{cc} \phi_j'\phi_j & \phi_j'\tilde{\phi}_{j'} \\ \tilde{\phi}_{j'}'\phi_j & \tilde{\phi}_{j'}'\phi_j(\phi_j'\phi_j)^{-1}\phi_j'\tilde{\phi}_{j'} \end{array}\right]\left[\begin{array}{cc} \phi_j'\phi_j & \phi_j'\phi_{j'} \\ \phi_j'\phi_j & \phi_{j'}'\phi_{j'} \end{array}\right]^{-1}$$

$$= \operatorname{tr}\left[\begin{array}{cc} I & O \\ (p_j + 1)^2 & (p_j + 1) \times (p_{j'} - p_j) \\ J_{jj'} & O \\ (p_{j'} - p_j) \times (p_j + 1) & (p_j - p_j)^2 \end{array}\right] = p_j + 1$$

where

$$J_{jj'} = [\tilde{\phi}_{j'}'\phi_j - \tilde{\phi}_{j'}'\phi_j(\phi_j'\phi_j)^{-1}\phi_j'\tilde{\phi}_{j'}(\tilde{\phi}_{j'}'\tilde{\phi}_{j'})^{-1}\tilde{\phi}_{j'}'\phi_j] \text{ times}$$

$$[\phi_j'\phi_j - \phi_j'\tilde{\phi}_{j'}(\tilde{\phi}_{j'}'\tilde{\phi}_{j'})^{-1}\tilde{\phi}_{j'}'\phi_j] .$$

Hence, if $\phi_j \subset \phi_{j'}$, then

$$E(\underline{e}_j^{(o)\prime}\ \underline{e}_{j'}^{(o)}) = (n - p_{j'} - 1)\ \sigma_{jj'}$$

so that an unbiased estimate of $\sigma_{jj'}$ is

$$s_{jj'} = \frac{\underline{e}_j^{(o)'}\,\underline{e}_{j'}^{(o)}}{n - p_m - 1} \qquad \text{where } p_m = \max(p_j, p_{j'})$$

$$= \frac{\underline{e}_j^{(o)'}\,\underline{e}_{j'}^{(o)}}{n - p_j - 1} \qquad \text{if } X_j = X_{j'} \text{ or } j = j'.$$

That is, to obtain an unbiased estimate of $\sigma_{jj'}$, when $X_j \neq X_{j'}$, $\rho_{jj'} \neq 0$, and $X_j \quad X_{j'}$, divide the sum of the residual cross products by n minus the larger number of coefficients occurring in the j^{th} and j'^{th} regression lines. In general

$$s_{jj'} = \frac{\underline{e}_j^{(o)'}\,\underline{e}_{j'}^{(o)}}{n - p_j - p_{j'} - 2 + \mathrm{tr}(U_j U_{j'})}$$

Having now determined a proper divisor for the residual cross products, $W^{(o)}$ is then calculated and substituted into (11) which results in a new, improved estimate of $\underline{\beta}$, say $\underline{b}^{(1)}$, improved in the sense that the off diagonal elements of $\sum$ are now being accounted for. Using $\underline{b}^{(1)}$, a new weight matrix, $W^{(1)}$, is determined and substituted into (11) resulting in $b^{(2)}$. This process is continued until $W^{(N)}$ and $W^{(N+1)}$ (and consequently $\underline{b}^{(N+1)}$ and $\underline{b}^{(N+2)}$ are identical to a desired decimal place which assumes convergence to take place at some stage of the iteration cycle.

An alternative, stepwise (from model to model) iterative process is given by the writer[4] for cases when the conditions of the theorem are fulfilled except that ω is unknown. Here, the calculations are

simplified though the slope estimates may have a larger dispersion (relative to the first iterative scheme) except for $\underline{b}_1$.

V. A SUFFICIENT CONDITION FOR CONVERGENCE AND AN APPROXIMATE COVARIANCE MATRIX OF THE RESULTING VECTOR ESTIMATE. Consider now the question of convergence when using the first iterative scheme of the last section to estimate $\underline{\beta}$.

(11) is rewritten as $\underline{b} = \underline{f}(\underline{y}; T, \omega)$ or $b_\ell = f_\ell (\underline{y}; T, \omega)$, $\ell = 1, 2, \ldots, \sum_{j=1}^{q} p_j = t$. It was seen that

$$\underline{b}^{(1)} = \underline{f}(W^{(o)}; T, \underline{y})$$
$$= \underline{h}(\underline{b}^{(o)}; T, \underline{y})$$

where the latter defined a general recurrence relationship,

$$\underline{b}^{(N+1)} = \underline{h}(\underline{b}^{(N)}) \tag{16}$$

or

$$\hat{b}_\ell^{(N+1)} = h_\ell(\hat{\underline{b}}^{(N)}),$$

which is used in solving the t nonlinear equations

$$\dot{\underline{b}} = \underline{h}(\dot{\underline{b}}) \tag{17}$$

in the same number of unknowns.

Let

$$\underline{b}^{(N)} = \underline{b}^* + \dot{\underline{\delta}}^{(N)}, \tag{18}$$

where $\underline{b}^*$ is a solution of (17) and $\dot{\underline{\delta}}^{(N)}$ is the error of the $N + 1^{st}$ iterate. Substituting (18) into (16(yields

$$(19) \qquad b_\ell^{(N+1)} = h_\ell(\underline{b}^* + \dot{\underline{\delta}}^{(N)}).$$

Expanding the right-hand side of (19) in a Taylor's series about $\underline{b}^*$, we have

$$b_\ell^{(N+1)} = h_\ell(\underline{b}^*) + \left[(\partial h_\ell / \partial \dot{\underline{b}}) \mid \dot{\underline{b}} = \underline{b}^*\right] \cdot \dot{\underline{\delta}}^{(N)} + 0(\dot{\delta}^2)$$

or generally,

$$(20) \qquad \underline{b}^{(N+1)} = \underline{h}(\underline{b}^*) + H\,\dot{\delta}^{(N)} + 0(\dot{\delta}^2),$$

where

$$H = \left[(\partial h_\ell / \partial \dot{\underline{b}}) \mid \dot{\underline{b}} = \underline{b}^*\right].$$

Noting that $\underline{b}^* = \underline{h}(\underline{b}^*)$, then (20) takes the form $\dot{\delta}^{(N+1)} = H\,\dot{\delta}^{(N)} + 0(\dot{\delta}^2)$, or, neglecting $0(\dot{\delta}^2)$, $\dot{\underline{\delta}}^{(N)} = \dot{H}\theta^N$, where the θ's are the characteristic roots of the matrix H, and $\dot{H}$ is determined from $\underline{b}^{(o)}$. Clearly, if the largest θ in modulus is less than unity, then $\dot{\underline{\delta}}^{(N)} \longrightarrow \underline{0}$ as N increases indefinitely. This, then, is a sufficient condition for the iterative process to converge to a solution $\underline{b}^*$, assuming that $0(\dot{\delta}^2)$ may be neglected. (The preceding argument is taken from reference (2).)

Next consider the spproximate variance of such an estimate. Let (18) be rewritten in the form

$$(21) \qquad \underline{b}^{(N)} = \underline{b} + \underline{\delta}^{(N)} = (\underline{\beta} + \underline{\zeta}) + \underline{\delta}^{(N)},$$

where $\underline{\delta}^{(N)}$ is the error distance of the $N + 1^{st}$ iterate from the best estimate $\underline{b}$, and $\underline{\zeta} = \underline{b} - \underline{\beta}$. Substituting the middle term of (21) into (15) and expanding in a Taylor's series about $\underline{b}$ results in

$$(22) \qquad b_\ell^{(N+1)} = h_\ell(\underline{b}) + \left[(\partial h_\ell / \partial \underline{\dot{b}}) \mid \underline{\dot{b}} = \underline{b}\right] \delta^{(N)} + 0(\delta^2).$$

Further, since $\underline{b} = \underline{\beta} + \underline{\zeta}$

$$(23) \qquad h_\ell(\underline{b}) = h_\ell(\underline{\beta} + \underline{\zeta}) = h_\ell(\underline{\beta}) + \left[(\partial h_\ell / \partial \underline{b}) \mid \underline{\dot{b}} = \underline{\beta}\right]' \underline{\zeta} + 0(\zeta^2).$$

Substituting the expansion of $h_\ell(\underline{b})$ into (22) and using (21), we have

$$\delta_\ell^{(N+1)} = h_\ell(\underline{b}) - b_\ell + \left[(\partial h_\ell / \partial \underline{\dot{b}}) \mid \underline{\dot{b}} = \underline{b}\right] \underline{\delta}^{(N)} + \left[(\partial h_\ell / \partial \underline{\dot{b}}) \mid \underline{\dot{b}} = \beta\right] \underline{\zeta} + 0(\delta^2) + 0(\zeta^2).$$

Neglecting $0(\delta^2)$ and $0(\zeta^2)$, the result in general is

$$(24) \qquad \underline{\delta}^{(N+1)} = \underline{h}(\underline{\beta}) = \underline{b} + G_1 \underline{\delta}^{(N)} + G_2 \zeta,$$

where $G_1 = \left[(\partial h_\ell / \partial \underline{\dot{b}}) \mid \underline{\dot{b}} = \underline{b}\right]$ and $G_2 = \left[(\partial h_\ell / \partial \underline{\dot{b}}) \mid \underline{\dot{b}} = \underline{\beta}\right]$.

Rewriting (24) in the form

$$\underline{\delta}^{(N)} = \underline{h}(\underline{\beta}) - \underline{b} + \dot{G}_1 \underline{\dot{\theta}}^N + G_2 \underline{\zeta},$$

where the θ's are the characteristic roots of G_1, and $\dot{G}_1$ is determined from $\underline{b}^{(o)}$, then if the largest $\dot{\theta}$ in modulus is less than unity, then $\underline{\delta}^{(N)} \longrightarrow \underline{h}(\underline{\beta}) = \underline{b} + G_2 \underline{\zeta}$ for N sufficiently large. Further, as n increases, $\underline{h}(\underline{\beta})$ rapidly approaches $\underline{b}$. (The latter statement is based on the following intuitive argument. If $W = \left[SSE_{jj'}/(df)_{jj'}\right]$, where df implies degrees of freedom, then when $\underline{\beta}$ is known, $(df)_{jj'} = n-1$ rather than $n - p_m - 1$, $p_m = \max(p_j, p_{j'})$. Clearly, as n increases,

then W should be very close to ω so that the difference between $\underline{h}(\underline{\beta})$ and $\underline{b}$ is negligible.) Then assuming convergence and n sufficiently large (which presupposes $0(\delta^2)$, $0(\zeta^2)$ negligible) we have

$$\underline{\delta}^{(N)} \cong G_2 \underline{\zeta}, \tag{25}$$

and by substituting (25) into (21),

$$\underline{b}^* = \underline{b} + \underline{\delta}^{(N)} \cong \underline{b} + G_2\underline{\zeta} = \underline{\beta} + (I + G_2)\underline{\zeta}.$$

Since $E(\underline{\zeta}) = \underline{0}$ and $E(\underline{b}^*) \cong \underline{\beta}$, then

$$\text{var } \underline{b}^* \cong (I + G_2)(T'A^{-1}T)^{-1}(I + G_2)'\sigma^2. \tag{26}$$

Assuming convergence, it is proposed that $\underline{b}^*$ is approximately $N(\underline{\beta}, \text{var } \underline{b}^*)$ where var $\underline{b}^*$ is estimated by substituting $\underline{b}^*$ for $\underline{\beta}$ and $W^{(N-1)} = W$ for ω (with convergence occurring at the N^{th} iterate) in (26).

It should be noted that we have neglected in var $(\underline{b}^*)$ the variability of $\underline{b}_0$ which arises in repeated sampling. As such, var $(\underline{b}^*)$ in (26) will tend to underestimate the true small sample variance of the iterated estimate $\underline{b}^*$.

VI. GENERAL COMMENTS ON ESTIMATION IN THE PREDICTIVE REGRESSION SYSTEM. Unless the off diagonal elements of $\sum$ are not "nearly zero," one should go from the restricted regression system (the system in which the independent variables are not the same from model to model) to the unrestricted regression system, i.e., $X_j = X$ for all j. After regressing each response type on all the independent variables, the anticipated result is that the coefficients not appearing in the restricted regression system have near zero estimates in the unrestricted system. A test of significance can be made to determine whether the matrix,

$B'_U(pxq) = (\underline{b}_1, \ldots, \underline{b}_j, \ldots, \underline{b}_q)$ (the estimated coefficient matrix of the unrestricted regression system), differs significantly from B_R, which is the matrix B_U with zeros replacing those b_U's not appearing in the restricted system. Unfortunately, methods such as the likelihood ratio test used for testing $H_0: B_U = B_R$ are not usually not sensitive to small departures from the null hypothesis whereas we are particularly interested in detecting small departures.

If B_R is "sufficiently close" to B_U, then the coefficients in the models of the restricted regression system may be estimated by separate least squares and from these estimates, say, $\underline{b}_j^{**}$, the matrix $(SSE_{jj'})$ may be determined. The $\underline{b}_j^{**}$ will be unbiased, consistent, and efficient enough not to warrant a search for alternative estimates.

If those coefficients not appearing in the restrictive regression system have "decidedly" nonzero estimates in the unrestricted system, then one should either

(i) question the models of the restricted regression system, or
(ii) make adjustments by using some iterative scheme to estimate $\underline{\beta}$.

With the models having strong theoretical justification, there may be no other choice than to make adjustments through iteration. Then assuming convergence, there is the formidable task of evaluating (26). The likely resolution will be the use of the estimated asymptotic variance of $\underline{b}^*$, namely,

$$(27) \qquad \left[T'(I*W)^{-1}T\right]^{-1}s^2 ,$$

even though the variances of the b*'s may be considerably underestimated, particularly if there are many parameters to estimate through iteration. For more simple systems, it is seen in the forthcoming numetical example that (27) is very close to the estimate of (26). This is due possibly to the fact that iteration is required in the estimation of only one parameter, and that the estimate of the zero population parameter is not significantly different from zero.

In general, if there are very few parameters which must be estimated through iteration, the estimate of (26(and the estimated asymptotic

covariance matrix (27) may be close, particularly if those coefficients in the restricted regression system have nonsignificant (from zero) estimates in the unrestricted regression system. If there exist many parameters and if adjustments need by made, then one should be well aware that the variances of such estimates increase, possibly by very large amounts.

Example 6. To illustrate a numerical example of example 4, the data were as follows:

$\underline{y}_1'$:	8.8,	9.6,	10.6,	8.3,	10.4,	10.1
$\underline{y}_2'$:	3.7,	12.5,	19.8,	24.4,	31.6,	40.1
$\underline{x}'$:	-5 ,	-3 ,	-1 ,	1 ,	3 ,	5

Intercept estimates were $a_0 = 9.6333$ and $b_0 = 22.0167$. Even though $\tilde{a}_1 = .0943$ was not significant, an adjustment was made by means of

$$b_1^{(N+1)} = \sum xy_2 / \sum x^2 - \rho^{(N)} \sum xy_1 / \sum x^2$$

where

$$\rho^{(N)} = (SSE_{12}^{(N-1)}/4)/(SSE_p^{(N-1)}/9)$$

$$= (9/4)\frac{\sum (y_1 - a_0)(y_2 - b_0 - b_1^{(N-1)} x)}{\sum (y_1 - a_0)^2 + \sum (y_2 - b_0 - b_1^{(N-1)} x)^2}$$

since $\sigma_{y_1}^2 = \sigma_{y_2}^2$. Three iterations were required to converge to the third decimal place of b_1^* and second decimal place of ρ^*:

$$b^{(0)} = 3.4843,\quad b^{(1)} = 3.4190,\quad b^{(2)} = 3.4119,\quad b^{(3)} = 3.4114;$$

$$\rho^{(0)} = 0.6930,\quad \rho^{(1)} = 0.7677,\quad \rho^{(2)} = 0.7729,\quad \rho^{(3)} = 0.7703.$$

Then

$$\partial b_1^{(N+1)} / \partial \dot{b}_1 = \left[-\left(\sum xy_1 / \sum x^2 \right) \left(\partial \rho^{(N)} / \partial \dot{b}_1 \right) \Big| \dot{b}_1 = \dot{b}_1^* \right] = 0.0639$$

so that the approximate variance of b_1^* is $(1.0639)^2 = 1.1319$ times the asymptotic variance of b_1^*. Hence

$$\operatorname{var} \begin{bmatrix} a_0 \\ b_0 \\ b_1^* \end{bmatrix} \cong \begin{bmatrix} 1/n & \rho^*/n & 0 \\ \rho^*/n & 1/n & 0 \\ 0 & 0 & (1.1319)(1-\rho^{*2}) / \sum x^2 \end{bmatrix} s^2 = \begin{bmatrix} .222 & .170 & 0 \\ .170 & .222 & 0 \\ 0 & 0 & .008 \end{bmatrix}.$$

Note that the zeros in the preceding matrix are fictitious, since b_1^* is dependent on a_0 and b_0, i.e., b_1^* depends on ρ^* which in turn depends on a_0 and b_0.

VII. THE PRACTICAL IMPLICATIONS OF LINEAR REGRESSION SYSTEMS. In practice, linear regression systems may arise from two rather distinct considerations: (i) the multivariate problem and (ii) the univariate problem

(i) The multivariate problem. When an attempt is made to study relationships between different types of dependent responses, one is usually, whether directly or indirectly, investigating the mechanics of a particular phenomenon or event of interest. More often than not, this should lead to some consideration of the events causal network (at least with regard to the characteristics of interest), i.e., a network inherent

to the event which states whether or not one characteristic causes or is instrumental in the occurrence of another characteristic. As such, an initial step is to hypothetically sketch the workings of the event in a path diagram. Such diagrams may be described hypothetically in terms of linear and/or nonlinear regression models which are deduced possibly from one or more differential equations. Thereafter comes experimentation designed to either support or discount the hypothesized models, and finally, inferences are drawn. The process of hypothesized models--designed experimentation--inferences describes, in general, the scientific cycle of investigation. The inferences gained in the first cycle then dictate changes in the structural models so that the entire process is repeated with the revised models. The cycle is continued until a structural regression system is determined which adequately describes the mechanics of the phenomenon and hence adequately predicts the event. Obviously, a lesser effort need be exerted if interest is centered only in prediction.

If estimable, the structural parameters at each cycle can be estimated through the reduced models (or the predictive system) such as in example 1. When the structural parameters are over-identified[3],[7], a recourse is the application of the Hood-Koopmans method of limited information[3]. For the case of under-identification, the statistician is forced to postulate whatever the required number of relationships from physical considerations, provided he has a sufficient knowledge of the subject matter and/or the aid of a mathematically inclined experimentor.

Though an adequate structural system is obviously preferable to an adequate predictive system, the current tendency to neglect the latter in favor of the former is particularly dangerous. Firstly, any one of many structural systems may lead to the same predictive system or approximately so; e.g., the structural models

$$\eta_1 = \ddot{\beta}_{01} + \ddot{\beta}_{21}\xi_1^2 + \ddot{\beta}_{31}\xi_2 + \ddot{\beta}_{41}\eta_2 + \epsilon_1^* \tag{28}$$

and (4) lead to the same predictive system as do models (3) and (4); further for the models given by (3) and (4), the structural parameters are under-identified with there being six equations in seven unknowns; in (28) and (4) there is "just identification" with six equations in six unknowns. And while the hypothesized structural system (hypothesized,

say, in the first or early cycles of investigation) may be totally inadequate, the predictive system may lead to valid inferences for certain problems. Further, the experimentor is often reluctant to carry through many cycles, thus possibly preventing the statistician from ascertaining the adequacy of an hypothesized structural system. Secondly what is to be done if there exists under-identification and no further relationships are known ? Thirdly, a predictive system may be sufficient to answer many oft-occurring questions regarding relationships between dependent response types, e.g., problems of extrema and constrained extrema[4]. And finally a predictive system of one cycle may offer invaluable aid and/or insight in the revision of the models to be used in the next cycle of investigation.

Consider problems where the predictive system is sufficient. Here it is often the case that structural models are not or cannot realistically be hypothesized, whereupon a system of linear predictive models is hypothesized. Due to the nature of the errors, the latter are assumed correlated between but not within models. However, much of the literature voids itself of practical use by assuming each model to contain precisely the same independent variables. In the latter circumstance, it has been shown (example 3) that no estimation difficulties are encountered. It was originally on the premise that the models did not necessarily contain the same independent variables that this research was instigated.

(ii) The univariate problem. When interest is centered on the prediction of one uncontrolled variable, η_1, by means of a set of fixed or controlled variables (say $\underline{x}_1$), the parameter estimates can be easily and most efficiently attained, at least for linear models. But as was seen in examples 4 and 5, if another (several) uncontrolled variable η_2 is found which can be adequately predicted by $\underline{x}_2 \neq \underline{x}_1$, then it may be possible to increase the precision of the predicted variable of interest.

The foregoing sheds a somewhat different light on the usual regression prediction. Any experimental unit has associated many (possibly an infinity of) measurable characteristics. If the intent is the prediction of one of these characteristics, it is intuitively evident that there exists a second (a third, a fourth, ...) characteristic which is "correlated" with the first and which can be adequately predicted

(a) by a set of controlled variables containing at least one variable which is different from all those variables used in the prediction

of the first characteristic, or

(b) by a set of controlled variables which is a subset of that set used in the prediction of the first characteristic.

Note further that while η_1 and η_2 may be independent (stemming, say, from different experimental units), a correlation between y_1 and y_2 may be induced through the measurement device.

With the correlations being known, then the prediction precision is definitely increased. If not, the precision is possibly increased. In the latter case, the loss in introducing an additional, uncontrolled variable (to be predicted) is the entry of a nuisance parameter or the unknown correlation. However, it is possible that the loss is more than offset by a gain which is not only an increased prediction precision, but also an automatic future detection of large sampling error assuming the system of linear predictive models is adequate.

VIII. CONCLUDING REMARKS. These developments are applicable to wider ranges of problems than hitherto mentioned. For example, not only does the approximate variance given by (26) apply to the variability of estimates obtained through iteration in certain nonlinear models, e.g., the rational regression model [6], as is shown in reference (4), but also, one or more linear models, introduced simultaneously with a nonlinear model, may act to increase the precision of the predicted nonlinear response.

In a sense, the methods here described are somewhat analogous to the estimation of parameters in functional or structural relationships (with independent variables uncontrolled) by introducing instrumental variables [8]. Further retrospect reveals the possibility of estimating these same parameters through the joint prediction of the functional or structural variables, though in this case, the parameters are often overidentified. The writer hopes to illustrate these facets at a future date.

IX. ACKNOWLEDGMENTS. The writer is grateful to R. L. Anderson for his ideas and aid and to E. J. Williams who first introduced the writer to this topic.

1. Anderson, T. W., 1958. An Introduction to Multivariate Statistical Analysis, J. Wiley and Sons, New York.

2. Bodmer, W. F., 1960. On the convergence of iterative processes for the solution of simultaneous equations in several variables. Pro. Camb. Phil. Soc. 56:286-289.

3. Hood, W. C., and Koopmans, T. C., 1953. Studies in Econometric Method, Cowles Commission Monograph No. 14. J. Wiley and Sons, New York.

4. Mallios, W. S., 1961. Some aspects of linear regression systems. North Carolina State College Mimeo Series No. 298.

5. Tukey, J. W., 1954. Statistics and Mathematics in Biology, Ch. 3. Iowa State College Press, Ames, Iowa.

6. Turner, M. E., Monroe, R. J., and Lucas, H. L., 1961. Generalized asymptotic regression and nonlinear path analysis. Biometrics, 17: 120-143.

7. Turner, M. E., and Stevens, C. D., 1959. The regression analysis of causal paths. Biometrics 15: 236-258.

8. Williams, E. J., 1959. Regression Analysis. J. Wiley and Sons, New York.

SOME RESULTS CONCERNING THE REDUCTION OF PRODUCT VARIABILITY THROUGH THE USE OF VARIANCE COMPONENTS

R. R. Prairie*
Sandia Corporation

In many situations in which a result from an experimental unit is subject to several sources of variation, the problem is to identify the sources and to estimate the variation associated with each. Generally, estimation and identification are accomplished through the use of experimental procedures. While considerable effort has been devoted to the efficient design of experiments for the estimation of treatment contrasts and regression parameters, little has been done on designing experiments for the estimation of variance components. Crump (1954) proposed designs for the two-stage nested classification that provides optimal estimates of the between-class variance component, the within-class variance component, or the ratio of the two components. Gaylor (1960) presented optimal designs for the estimation of components of variance and certain functions of the components for the two-way crossed classification. Anderson and Bancroft (1952) and Anderson (1960) introduced a staggered design for the estimation of parameters from a three- (or more) stage nested classification. No known work has been done on the problem of designing experiments for the estimation of variance components with the objective of using the estimates for the efficient planning of future experiments or courses of action.

This investigation is concerned with the design of experiments for the efficient estimation of functions of the components of variance with the objective of reducing product, or process, variability. Only the case in which a result is subject to two sources of variation will be considered. It will be assumed that a single observation can be presented by

$$(1) \qquad x_{ij} = m + A_i + B_{ij} \qquad \begin{matrix} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n_i \\ \sum n_i = N \end{matrix}$$

Where m is a constant, A_i and B_{ij} are both normally independently

*Formerly at North Carolina State College, Raleigh, North Carolina.

distributed (NID) with means zero, and variances, σ_A^2 and σ_B^2, respectively. Under this model, the product variance, i.e., the variance of an observation, may be represented by

$$\sigma_T^2 = \sigma_A^2 + \sigma_B^2 \quad .$$

An analysis of variance for data obeying (Equation 1) is given in Table I.

TABLE I

An Analysis of Variance for a Two-Stage Nested Classification			
Source of Variation	Degrees of Freedom	Mean Square	Expectation of* Mean Square
Classes	a - 1	M_A	$\sigma_B^2 + K\sigma_a^2$
Within classes	N - a	M_B	σ_B^2
Total	N - 1		

$$*K = \left(N - \frac{\sum n_i^2}{N}\right) / (a - 1)$$

If one is given the task of reducing σ_T^2, he could quite conceivably proceed by affecting individually the sources that give rise to σ_A^2 and σ_B^2. In the usual circumstance, only a limited amount of funds would be available for reducing σ_T^2, and one would attempt to devise a program that would yield the greatest reduction in σ_T^2 for a given expenditure of funds. The crucial matter here is the proper apportionment of funds toward reduction of the two sources of variation.

Two factors would determine the specific apportionment of funds. They are: (1) the relative rate at which σ_A^2 can be reduced compared to σ_B^2, and (2) the magnitude of $\rho = \sigma_A^2 / \sigma_B^2$. In many experimental situations, the relative costs of reducing σ_A^2 and σ_B^2 may be adequately known, but ρ may not be. Often the estimate of ρ is obtained by performing an experiment with a fixed sample size, with units arranged in a specified design.

The primary purpose of this work will be to investigate, under a given cost model, the influence of the design of the experiment on the apportionment of funds and, therefore, on the effectiveness of the reduction of the total variance. It will be assumed that the cost of sampling for the experiment is directly proportional to the sample size N.

In order to arrive at an explicit solution for the optimal apportionment of funds, it is necessary to adopt a model which relates funds expended to reduction in variance.

There are several different models that could be used to express reduction in variance as a function of funds expended. One of the simplest is the additive model. Under this model the variance obtained after expending d units is equal to the original variance minus an amount proportional to d. For the situation at hand, the reduced total variance, symbolized by $\sigma_T'^2$, attained after expending d_A units on source A and d_B units on source B, is given by

$$\sigma_T'^2 = \sigma_B^2 - K_B d_B + \sigma_A^2 - K_A d_A ,$$

where K_A and K_B are constants of proportionality.

Such a model is quite unrealistic for depicting an actual real-world situation. According to this model, the amount that a variance is reduced does not depend on the magnitude of the variance. Generally, more effort would be required to reduce a small variance a given amount than to reduce

a large variance by the same amount. Also, the model indicates that if enough funds are made available, a variance can be reduced to zero and even made negative, a circumstance which is quite absurd.

The basic model that is proposed here is of the power-function type. Under this model it is assumed that the expenditure of one unit of funds on source A reduces σ_A^2 by 100 A' percent and, on source B, reduces σ_B^2 by 100 B' percent. Hence, the reduced total variance attained, after expending d_A units on source A and d_B units on source B, is

$$\sigma_T'^2 = \sigma_B^2 (1 - B')^{d_B} + \sigma_A^2 (1 - A')^{d_A} . \tag{2}$$

In order to eliminate one parameter, the model proposed above will be slightly modified. If it is assumed that one unit of funds expended reduces σ_B^2 by one percent and reduces σ_A^2 by 100A percent, the reduced total variance may be written as

$$\sigma_T'^2 = \sigma_B^2 (0.99)^{d_B} + \sigma_A^2 (1 - A)^{d_A} . \tag{3}$$

Because it is more convenient to work with an exponential than with a power function, $\sigma_T'^2$ will be expressed in exponential form. By setting $k_1 = -\ln(0.99)$ and $k_2 = -\ln(1 - A)$ we have

$$\sigma_T'^2 = \sigma_B^2 e^{-k_1 d_B} + \sigma_A^2 e^{-k_2 d_A} . \tag{4}$$

The model, as given by Equation 4, shall be the model used throughout.

When the total amount of funds is fixed at, say, D units, $\sigma_T'^2$ may be written as

$$\text{(5a)} \qquad \sigma_B^2 e^{-k_1 D} + \sigma_A^2, \qquad d_A = 0 (d_B = D)$$

$$\text{(5b)} \qquad \sigma_T'^2 = \sigma_B^2 e^{-k_1 (D - d_A)} + \sigma_A^2 e^{-k_2 d_A}, \quad 0 \leq d_A \leq D$$

$$\text{(5c)} \qquad \sigma_B^2 + \sigma_A^2 e^{-k_2 D}, \qquad d_A = D (d_B = 0)$$

In determining the optimal allocation of funds to minimize $\sigma_T'^2$, we first find that value of d_A which minimizes Equation 5b. This value is

$$\text{(6)} \qquad d_1 = \frac{Dk_1 + \ln k_2 - \ln k_1 + \ln \rho}{k_1 + k_2}$$

$$\text{(7)} \qquad = C_1 + \frac{\ln \rho}{C_2} .$$

If $d_1 < 0$, we set $d_A = 0$; and, if $d_1 > 0$, we set $d_A = D$. Hence, the complete solution is

$$\text{(8)} \qquad d_1 = \begin{cases} 0, & 0 \leq \rho \leq \rho_1 \\ C_1 + \frac{\ln \rho}{C_2}, & \rho_1 \leq \rho \leq \rho_2 \\ D, & \rho \geq \rho_2 \end{cases}$$

where

$$\rho_1 = e^{-C_1 C_2} = \frac{k_1}{k_2} e^{-k_1 D}$$

and

$$\rho_2 = e^{C_2(D-C_1)} = \frac{k_1}{k_2} e^{k_2 D} .$$

For this work, it is more convenient to use the ratio $\sigma'^2_R = \sigma'^2_T / \sigma^2_B$, rather than σ'^2_T. Since d_1, as given by Equation 8, also minimizes σ'^2_R, and since the statistical properties of both are the same, working with σ'^2_R is equivalent to working with σ'^2_T, and σ'^2_R will be used in all future development. The quantity σ'^2_R will be referred to as the reduced total variance ratio.

As was mentioned previously, $\underline{A}$ is often known, but an estimate of ρ is required before the allocation of funds can be completely specified. If an estimator of ρ, say $\hat{\rho}$, is substituted into (3.7) for ρ, then as an estimator of d_1, say $\hat{d}_1$, we have

$$(9) \qquad \hat{d}_1 = \begin{cases} 0, & L \leq \hat{\rho} \leq \rho_1 \\ C_1 + \dfrac{\ln \hat{\rho}}{C_2}, & \rho_1 \leq \hat{\rho} \leq \rho_2 \\ D, & \hat{\rho} \geq \rho_2 \end{cases}$$

where L, which may be negative, depends on the particular estimator of ρ used.

Therefore, the true reduced total variance ratio that exists, after expending $D - \hat{d}_1$ on source B and $\hat{d}_1$ on source A, is

$$(10)\quad \sigma'^2_R \Big| \hat{d}_1 = \begin{cases} e^{-k_1 D} + \rho, & L \le \rho \le \rho_1 \\ e^{-k_1(D-C_1)\hat{\rho}^{k_1/C_2}} & \\ & \rho_1 \le \hat{\rho} \le \rho_2 \\ + \rho e^{-k_2 C_1 \hat{\rho}^{-k_2/C_2}}, & \\ 1 + \rho e^{-k_2 D}, & \hat{\rho} \ge \rho_2 \end{cases}$$

It is important to realize that $\sigma'^2_R \Big| \hat{d}_1$, where d_1 is obtained by experimental methods, may be considered as either a random variable or as a constant, according as $\hat{d}_1$ is considered as a random variable or a constant. The reduced total variance ratio is a random variable in the sense that, if we were to perform a large number of experiments, they would yield differing values of $\hat{\rho}$ and, hence, possibly differing values of $\sigma'^2_R \Big| \hat{d}_1$. Conversely, once the experiment has been performed, the allocation determined, and the funds actually expended on the two sources contributing to σ_B^2 and σ_A^2, $\sigma'^2_R \Big| \hat{d}_1$ is a constant associated with the product or process. We shall be concerned with $\sigma'^2_R \Big| \hat{d}_1$ as a random variable.

The situation shall be considered in which ρ is to be estimated from an experiment where N units may be arranged in a two-stage nested classification with a classes. The model that represents the data collected according to such a scheme is taken to be of the form given by

Equation 1. As previously stated, it will be assumed that the cost of sampling the units is proportional to N regardless of the type of design. Further, it will be assumed that for any a the N units are allocated as equally as possible to the a classes. Specifically, for $N/a = p + s/a (0 \leq s < a)$, there are $p + 1$ units assigned to each of s classes (referred to as A_1) and p units to each of $a - s$ classes (referred to as A_2).

In Table II an analysis of variance is given for such data. It should be apparent that the partition of the class sum of squares as given in Table II is relevant only for situations in which $0 < s < a$.

TABLE II

An Analysis of Variance for a Two-Stage Nested Classification with $p + 1$ Units in Each of s Classes and p Units in Each of $a - s$ Classes

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expectation of* Mean Square
Classes	$a - 1$	S_A	M_A	$\sigma_1^2 = \sigma_B^2\left[1 + K\rho\right]$
A_1	$s - 1$	S_{A_1}	M_{A_1}	$\sigma_{11}^2 = \sigma_B^2\left[1 + (p+1)\rho\right]$
A_2	$a - s - 1$	S_{A_2}	M_{A_2}	$\sigma_{12}^2 = \sigma_B^2\left[1 + p\rho\right]$
A_1 vs. A_2	1	S_{A_3}	M_{A_3}	$\sigma_{13}^2 = \sigma_B^2\left[1 + \frac{ap(p+1)}{N}\rho\right]$
Within	$N - a$	S_B	M_B	σ_B^2
Total	$N - 1$			

*$K = \left[N(N - 2p - 1) + ap(p + 1)\right] / N(a - 1)$

It may be noted, for this particular allocation of units to classes, that,

$$N = (p + 1)s + p(a - s) = s + pa.$$

Also, when $s = 0$; then $a = N/p$, $ap(p + 1)/N = p + 1$ and the expectation of the mean square for classes becomes $E(M_A) = \sigma_B^2(1 + p\rho)$.

From the manner in which the units are assigned to the classes, i.e., $p + 1$ units are assigned to each of s classes and p units to each of $a - s$ classes, it is realized that designating $\underline{a}$ completely specifies the design. Therefore, varying the design would be accomplished by varying $\underline{a}$.

As was previously stated, the concern of this work is where ρ and hence d_1 must be estimated. The estimator of ρ that we shall consider is

$$\hat{\rho} = (M_A - M_B)/KM_B. \tag{11}$$

This estimator will be considered because it is the one that is used most extensively in practice.

From Equation 11 it is noted that

$$\hat{\rho} = \frac{1}{K}\left[\frac{M_A - M_B}{M_B}\right] = \frac{1}{K}\left[\frac{M_A}{M_B} - 1\right],$$

and therefore a study of the properties of $\hat{\rho}$ becomes a study of M_A/M_B.

Now, M_B is distributed as $\sigma_B^2 \chi^2_{(N-a)}/(N-a)$, where $\chi^2_{(v)}$ is distributed as chi square with v degrees of freedom. If the design is ballanced, M_A is distributed as $\sigma_B^2\left[1 + \frac{N}{a}\rho\right]\chi^2_{(a-1)}/(a-1)$. Hence, if N/a is an integer,

$$F = \frac{M_A}{M_B(1 + \frac{N}{a}\rho)}$$

follows Snedecor's F distribution, and the distribution of $\hat{\rho}$ can be linearly transformed to that of F so that all the known properties of F can be utilized. However, this work will not be limited to considering only balanced designs. When N/a is not an integer, M_A is not distributed as $\sigma_B^2\left[1+\frac{N}{a}\rho\right]\chi^2_{(a-1)}/(a-1)$, and it is not possible to make a simple transformation from the distribution of $\hat{\rho}$ to that of F. To circumvent this difficulty, use was made of an approximate distribution of a linear function of chi-square variables [studied by Satterthwaite (1946)]. The approximate distribution is set up such that its mean and variance are equal to the mean and variance of the linear function of chi-square variables.

Consider the weighted sum of independent chi-square variables

$$Q = \sum\lambda_i \chi^2_{(v_i)} .$$

Let Q be represented by $Z = g\chi^2_{(h)}$ where g and h are chosen such that E(Q) = E(Z) and var(Q) = var(Z). It is easily seen that $E(Q) = \sum\lambda_i v_i$, $E(Z) = gh$, $var(Q) = 2\sum\lambda_i^2 v_i$, and $var(Z) = 2g^2h$. Setting E(Q) = E(Z) and var(Q) = var(Z), we obtain

$$g = \sum\lambda_i^2 v_i / \sum\lambda_i v_i$$

$$h = (\sum\lambda_i v_i)^2 / \sum\lambda_i^2 v_i .$$

Hence, $Q = \sum\lambda_i \chi^2_{(v_i)}$ is approximately distributed as $g\chi^2_{(h)}$ where g and h are given above.

From Table II we see that $\hat{\rho}$ may be written as

$$\hat{\rho} = \frac{1}{K}\left[\frac{M_A}{M_B} - 1\right]$$

$$= \frac{1}{K}\left[\frac{S_{A_1} + S_{A_2} + S_{A_3}/(a-1)}{M_B} - 1\right] \tag{12}$$

$$= \frac{1}{K}\left[\frac{Q_1/(a-1)}{M_B} - 1\right].$$

From $0 < s < a$, it is observed that

$$S_{A_1} = \sigma_{11}^2 \chi^2_{(s-1)},\ S_{A_2} = \sigma_{12}^2 \chi^2_{(a-s-1)},\ S_{A_3} = \sigma_{13}^2 \chi^2_{(1)}$$

and

$$M_B = \sigma_B^2 \chi^2_{(N-a)}/(N-a),$$

where

$$\sigma_{11}^2 = \sigma_B^2\left[1 + (p+1)\rho\right],\ \sigma_{12}^2 = \sigma_B^2\left[1 + p\rho\right]$$

and

$$\sigma_{13}^2 = \sigma_B^2 \left[1 + \frac{ap(p+1)}{N}\rho\right] .$$

Therefore, Q_1 in Equation 12 can be written as a linear function of chi-square variables,

$$Q_1 = \sigma_{11}^2 \chi^2_{(s-1)} + \sigma_{12}^2 \chi^2_{(a-s-1)} + \sigma_{13}^2 \chi^2_{(1)} .$$

Using the result concerning the approximate distribution of a linear function of chi-square variables, it is seen that Q_1 is approximately distributed as $\sigma_B^2 g_1 \chi^2_{(h_1)}$ where

$$g_1 = \frac{(s-1)\left[\sigma_{11}^2\right]^2 + (a-s-1)\left[\sigma_{12}^2\right]^2 + \left[\sigma_{13}^2\right]^2}{\left[(s-1)\sigma_{11}^2 + (a-s-1)\sigma_{12}^2 + \sigma_{13}^2\right]\sigma_B^2}$$

$$(13) \quad = \frac{(s-1)\left[1+(p+1)\rho\right]^2 + (a-s-1)\left[1+p\rho\right]^2 + \left[1+\frac{ap(p+1)}{N}\rho\right]^2}{(s-1)\left[1+(p+1)\rho\right] + (a-s-1)\left[1+p\rho\right] + \left[1+\frac{ap(p+1)}{N}\rho\right]} \equiv \frac{T_1}{R_1}$$

$$(14) \quad h_1 = R_1^2/T_1 .$$

The substitution of $\sigma_B^2 g_1 \chi^2_{(h_1)}$ for Q_1 and $\sigma_B^2 \chi^2_{(N-a)}/(N-a)$ for M_B in Equation 12 gives

$$\hat{\rho} \doteq \frac{1}{K}\left[\frac{g_1 \chi^2_{(h_1)}/(a-1)}{\chi^2_{(N-a)}/(N-a)} - 1\right] \tag{15}$$

$$= \frac{1}{K}\left[\frac{g_1 h_1 \chi^2_{(h_1)}/(a-1)h_1}{\chi^2_{(N-a)}/(N-a)} - 1\right]$$

where $\doteq$ means approximately equal. Recall that

$$F_1 = \frac{\chi^2_{(v_1)}/v_1}{\chi^2_{(v_2)}/v_2}$$

follows Snedecor's F distribution with v_1 and v_2 degrees of freedom. Let

$$F_1 = \frac{\chi^2_{(h_1)}/h_1}{\chi^2_{(N-a)}/(N-a)}, \tag{16}$$

which is distributed as F with h_1 and $N - a$ degrees of freedom. Hence,

$$\hat{\rho} \doteq \frac{1}{K}\left[\frac{g_1 h_1 F_1}{a - 1} - 1\right]. \tag{17}$$

It may be noted that when $N/a = p = K$ is an integer, $h_1 = a - 1$, $g_1 = (1 + \frac{N}{a}\rho)$, and

$$\hat{\rho} = \frac{a}{N}\left[\left(1 + \frac{N}{a}\rho\right) F_1 - 1\right].$$

The reduced total variance ratio, $\sigma'^2_R \mid \hat{d}_1$, has a distribution whose lower limit is the minimum attainable variance ratio; namely,

$$\min\left(\sigma'^2_R \mid \hat{d}_1\right) = e^{-k_1(D-d_1)} + \rho\, e^{-k_2 d_1}. \tag{18}$$

Therefore, it is reasonable that one would seek a design or estimator which would generate a distribution for $\sigma'^2_R \mid \hat{d}_1$ that has the largest possible density near $\min\left(\sigma'^2_R \mid \hat{d}_1\right)$. The first criterion that is suggested concerns the expectation of $\sigma'^2_R \mid \hat{d}_1$. Ostensibly, the best design or estimator would be one that yields the smallest expectation of $\sigma'^2_R \mid \hat{d}_1$. Therefore, an expression will be developed for the expectation of $\sigma'^2_R \mid \hat{d}_1$.

From consideration of $\sigma'^2_R \mid \hat{d}_1$, as given by Equations 10 and 11, it is apparent that the expectation of $\sigma'^2_R \mid \hat{d}_1$ is

$$E\left(\sigma_R^{\prime 2} \mid \hat{d}_1\right) = \left(e^{-k_1 D} + \rho\right) \int_{-\frac{1}{K}}^{\rho_1} f(\hat{\rho}) d\hat{\rho}$$

$$+ \int_{\rho_1}^{\rho_2} \left[e^{-k_1 (D - C_1)} \hat{\rho}^{k_1/C_2} + \rho\, e^{-k_2 C_1}\, \hat{\rho}^{-k_2/C_2} \right] f(\hat{\rho}) d\hat{\rho} \tag{19}$$

$$+ \left(1 + \rho\, e^{-k_2 D}\right) \int_{\rho_2}^{\infty} f(\hat{\rho}) d\hat{\rho} \ .$$

Making the linear transformation

$$\hat{\rho} = \frac{1}{K}\left[\frac{g_1 h_1 F_1}{\hat{a} - 1} - 1\right] ,$$

the expectation may be expressed as

$$E(\sigma_R'^2 \mid \hat{d}_1) \doteq (e^{-k_1 D} + \rho) \int_0^{\ell_1} g(F_1)dF_1$$

$$+ \int_{\ell_1}^{\ell_2} \left\{ e^{-k_1(D-C_1)} \left[\frac{1}{K} \left(\frac{g_1 h_1 F_1}{a-1} - 1 \right) \right]^{k_1/C_2} \right.$$

(20)

$$\left. + \rho\, e^{-k_2 C_1} \left[\frac{1}{K} \left(\frac{g_1 h_1 F_1}{a-1} - 1 \right) \right]^{k_2/C_2} \right\} g(F_1)dF_1$$

$$+ \left(1 + \rho\, e^{-k_2 D}\right) \int_{\ell_2}^{\infty} g(F_1)dF_1 \,,$$

where

$$\ell_1 = \frac{a-1}{g_1 h_1} (K\rho_1 + 1)$$

$$\ell_2 = \frac{a-1}{g_1 h_1} (K\rho_2 + 1)$$

and $g(F_1)$ is the density function for the F statistic with h_1 and $N-a$ degrees of freedom.

Ideally, one would carry out the integration required in Equation 20 and then specify numerical values for ρ, A, D, and N. The effect of the design would be determined by substituting various values of $\underline{a}$ and observing the numerical results of $E(\sigma'^2_R \mid \hat{d}_1)$. A major difficulty with using this procedure concerned the evaluation of the second integral on the RHS of Equation 20. Numerical integration could have been used; however, a simpler criterion was selected which may be even more useful than that involving the average variance ratio.

This criterion concerns the quantity

$$P = \text{prob}\left[(\sigma'^2_R \mid \hat{d}_1) \leq \beta\right] , \tag{21}$$

where β is a fixed number between $\min(\sigma'^2_R \mid \hat{d}_1)$ and $\max(\sigma'^2_R \mid \hat{d}_1)$. By careful selection of β, such a criterion should do well to pinpoint the effect of design and estimator. Obviously, extreme values of β, either small or large, would yield P quite insensitive to variation in design or estimator, regardless of their effects on the distribution of $\sigma'^2_R \mid \hat{d}_1$. An expression for P is now developed.

First recourse is made to investigation $\sigma'^2_R \mid \hat{d}_1$ as a function of $\hat{\rho}$. Consider the function

$$V_o(\hat{\rho}) = \sigma'^2_R \mid \hat{d}_1 .$$

Referring to Equations 10 and 11, it is seen that V_o takes on three forms,

$$(22)\qquad V_o = \begin{cases} V_1 = e^{-k_1 D} + \rho\,, & -\frac{1}{K} \le \rho \le \rho_1 \\ V_2 = e^{-k_1(D-C_1)}\hat{\rho}^{k_1/C_2} & \\ \qquad + \rho e^{-k_2 C_1}\hat{\rho}^{-k_2/C_2}\,, & \rho_1 \le \hat{\rho} \le \rho_2 \\ V_3 = 1 + \rho e^{-k_2 D}\,, & \hat{\rho} \ge \rho_2 \end{cases}$$

where

$$\rho_1 = e^{-C_1 C_2} = \frac{k_1}{k_2}\, e^{-k_1 D}$$

and

$$\rho_2 = e^{C_2(D-C_1)} = \frac{k_1}{k_2}\, e^{k_2 D}\,.$$

Let

$$\rho_o = \left(1 - e^{-k_1 D}\right) /\!/ \left(1 - e^{-k_2 D}\right)\,.$$

An inspection of V_1 and V_3 shows that

$$(23)\qquad V_1 \underset{\geqq}{\leqq} V_3 \quad \text{if} \quad \rho \underset{\geqq}{\leqq} \rho_o\,.$$

In addition, we see that

$$V_2(\rho_1) = V_1;\ V_2(\rho_2) = V_3 \quad (24)$$

Also, it is apparent that V_2 is continuous in $[\rho_1, \rho_2]$. The continuity of V_2 plus the equalities given by Equation 24 establish that V_o is continuous over its entire range.

Next, consider the behavior of V_2 in $[\rho_1, \rho_2]$. First, consider V_2 at its end points. To do this, the first derivative of V_2 with respect to $\hat{\rho}$ is needed. The derivative is

$$V_2'(\hat{\rho}) = \frac{k_1}{C_2} e^{-k_1(D-C_1)} \hat{\rho}^{(k_1-C_2)/C_2} - \frac{k_2}{C_2} \rho e^{-k_2 C_1} \hat{\rho}^{-(k_2+C_2)/C_2} \quad (25)$$

From Equation 25 it is easily verified that

$$V_2'(\rho_1) \gtrless 0 \text{ if } \rho \lessgtr \rho_1;\ V_2'(\rho_2) \lessgtr 0 \text{ if } \rho \gtrless \rho_2.$$

It is well known that a necessary, but not sufficient, condition for a function f to have a relative extremum at a point x_o is that $f'(x_o) = 0$. Using this result for Equation 24, we note that $V_2'(\hat{\rho})$ is zero when

$$\hat{\rho}^{(k_1+k_2)/C_2} = \rho \frac{k_2}{k_1} e^{k_1 D - C_1(k_1+k_2)}.$$

Since

$$C_1 = \frac{Dk_1 + \ln k_2 - \ln k_1}{k_1 + k_2} ,$$

and

$$C_2 = k_1 + k_2 ,$$

this reduces to

$$\hat{\rho} = \rho . \tag{26}$$

Hence, $V_2(\hat{\rho})$ has at most one extremum; namely, $V_2(\rho)$.

The second derivative at $\hat{\rho} = \rho$ is

$$V_2''(\rho) = -\frac{k_1 k_2}{C_2^2} e^{-k_1 D + k_1 C_1} \rho^{-(2C_2 - k_1)/C_2}$$

$$+ \frac{k_2(k_2 + C_2)}{C_2^2} e^{-k_2 C_1} \rho^{-(k_2 + C_2)/C_2} \tag{27}$$

$$= \frac{k_1 k_2}{C_2^2} \rho^{(k_1 - 2C_2)/C_2} e^{k_1(C_1 - D)} \left[\frac{k_2 + C_2}{k_2} - 1 \right].$$

The second derivative will always be positive because k_1, $k_2 > 0$ and $\left[(k_2 + C_2)/k_2\right] - 1 = C_2/k_2 = 1 + k_1/k_2 > 0$; hence, $f'(x_o) = 0$ is a sufficient as well as necessary condition and the extremum will be a minimum.

Collecting these results, we see that V_o can be characterized by three different profiles, depending on the value of ρ, as shown in Figures 1, 2, and 3. These three ways in which V_o behaves are described below:

1. When $\rho \leq \rho_1$, V_o increases from a minimum of $V_o = V_1$ to a maximum of $V_0 = V_3$.
2. When $\rho \geq \rho_2$, V_o decreases from a maximum of $V_o = V_1$ to a minimum of $V_o = V_3$.
3. When $\rho_1 < \rho < \rho_2$, V_o decreases from a value of $V_o = V_1$ to a minimum of $V_o = V_2(\rho)$ and then increases to a value of $V_o = V_3$.

In Figure 3 it is noted that $V_1 > V_3$ if $\rho > \rho_o$, $V_1 < V_3$ if $\rho < \rho_o$; regardless of whichever V_1 or V_3 is the smaller, V_2 reaches a minimum value which is less than the smaller of V_1 and V_3. The minimum of V_2 equals V_1 or V_3 only when $\rho = \rho_1$ or $\rho = \rho_2$.

Therefore, if V_o were graphed and then a horizontal line were drawn at $V = \beta$, $\min\left[V_o(\hat{\rho})\right] < \beta < \max\left[V_o(\hat{\rho})\right]$, the line would intersect the graph of V_o at either one or two places. In Figures 1 and 2 there would be exactly one point of intersection. In Figure 3 the line would intersect the graph of V_o at exactly two places if both $V_1 > \beta$ and $V_3 > \beta$, and would intersect at exactly one place if $V_1 > \beta$ and $V_3 \leq \beta$ or $V_1 \leq \beta$ and $V_3 > \beta$. It should be observed that, in Figure 3, if $V_1 = \beta$ and $V_3 > \beta$, the lower point of intersection is taken to be $\hat{\rho} = -1/K_1$; and, if $V_1 > \beta$ and $V_3 = \beta$, the upper point of the intersection is taken to be at positive infinity.

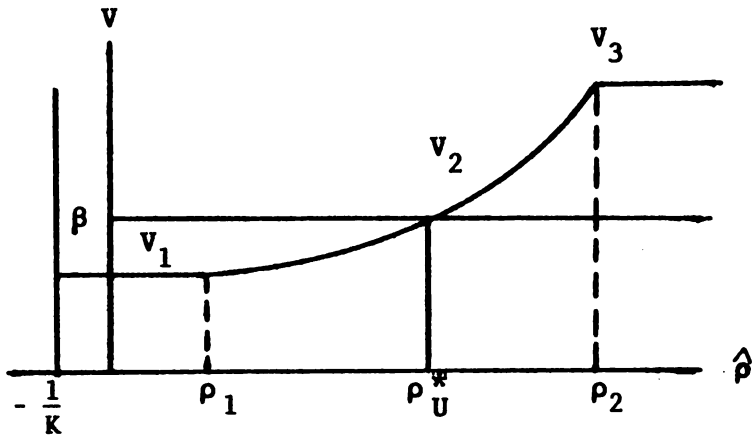


Figure 1. Profile of V_o when $\rho \leq \rho_1$

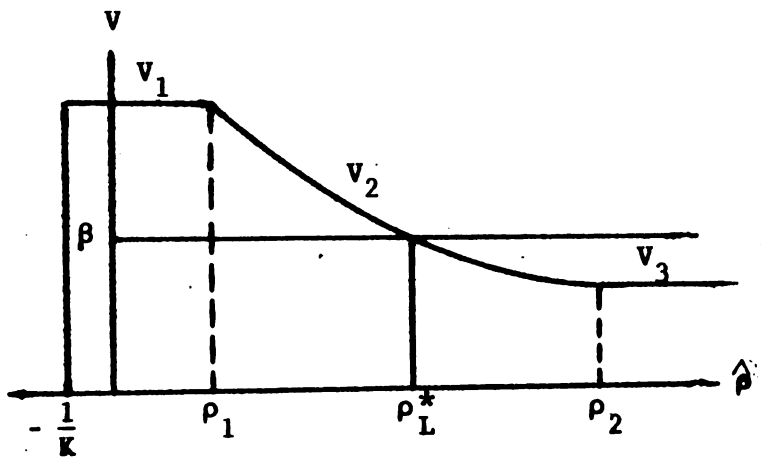


Figure 2. Profile of V_o when $\rho \geq \rho_2$

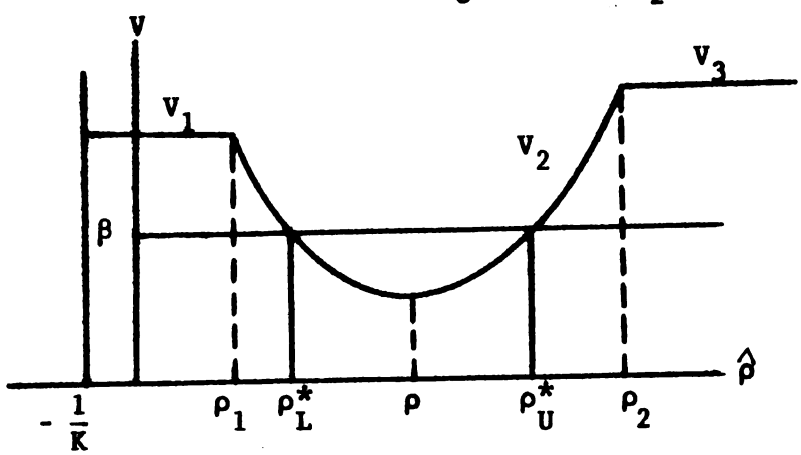


Figure 3. Profile of V_o when $\rho_1 < \rho < \rho_2$

Also, in each of the figures (1 through 3), the line $V = \beta$ is shown, and the values of $\hat{\rho}$, ρ_L^*, and ρ_U^*, at which $V_o(\hat{\rho}) = \beta$ are indicated.

Using the information regarding the behavior of $V_o(\hat{\rho}) = \sigma_R'^2 \left| \hat{d}_1 \right.$, it is found that P may be expressed as

$$P = \begin{cases} P_1 = \text{prob}\ (\hat{\rho} \leq \rho_U^*) \text{ if } \rho \leq \rho_1 \text{ or if } \rho_1 < \rho < \rho_2 \text{ and } V_1 \leq \beta,\ V_3 > \beta \\ P_2 = \text{prob}\ (\hat{\rho} \geq \rho_L^*) \text{ if } \rho \geq \rho_2 \text{ or if } \rho_1 < \rho < \rho_2 \text{ and } V_1 > \beta,\ V_3 \leq \beta \\ P_3 = \text{prob}\ (\rho \leq \hat{\rho} \leq \rho_U^*) \text{ if } \rho_1 < \rho < \rho_2 \text{ and } V_1 > \beta,\ V_3 > \beta \end{cases}$$

where ρ_L^* and ρ_U^* are the smallest and largest roots, respectively, of the equation

(28) $$V_2(\hat{\rho}) = \beta\ .$$

P is expressed in integral form as

$$P = \begin{cases} P_1 = \int_{-\frac{1}{K}}^{\rho_U^*} f(\hat{\rho})\, d\hat{\rho} & (29) \\ P_2 = \int_{\rho_L^*}^{\infty} f(\hat{\rho})\, d\hat{\rho} & (30) \\ P_3 = \int_{\rho_L^*}^{\rho_U^*} f(\hat{\rho})\, d\hat{\rho} & (31) \end{cases}$$

To evaluate the integrals (equations 29 through 31), use is made of the relation given by Equation 17. Then, using the density function of F, the general form for P can be approximated by

$$(32) \qquad P \doteq \int_{R^*} \frac{\frac{a-1}{(N-a)g_1}(K\hat{\rho}+1)^{(h_1-2)/2}\,\frac{(a-1)K}{(N-a)g_1}}{B\left(\frac{h_1}{2},\frac{N-a}{2}\right)\left[1+\frac{a-1}{(N-a)g_1}(K\hat{\rho}+1)\right]^{(h_1+N-a)/2}}\,d\hat{\rho}$$

where R* is the appropriate domain of integration. Note that P, as given by Equation 32, is exact when N/a is an integer. By making the transformation of variable

$$\frac{a-1}{(N-a)g_1}(K\hat{\rho}+1) = \frac{x}{1-x},$$

Equation 32 is transformed into the incomplete beta form

$$(33) \qquad P \doteq \int_{R} \frac{x^{\frac{h_1}{2}-1}(1-x)^{\frac{N-a}{2}-1}}{B\left(\frac{h_1}{2},\frac{N-a}{2}\right)}\,dx,$$

where R is the transformed domain of integration.

Each of the integrals (Equations 29 through 31) can thus be approximated by incomplete beta functions, and P can be expressed as

$$P \doteq \begin{cases} P_1 = I_U\left(\frac{h_1}{2}, \frac{N-a}{2}\right) & (34) \\ P_2 = 1 - I_L\left(\frac{h_1}{2}, \frac{N-a}{2}\right) & (35) \\ P_3 = I_U\left(\frac{h_1}{2}, \frac{N-a}{2}\right) - I_L\left(\frac{h_1}{2}, \frac{N-a}{2}\right) & (36) \end{cases}$$

where

$$U = \frac{\frac{a-1}{g_1(N-a)}(K\rho_U^* + 1)}{1 + \frac{a-1}{g_1(N-a)}(K\rho_U^* + 1)} \quad (37)$$

$$L = \frac{\frac{a-1}{g_1(N-a)}(K\rho_L^* + 1)}{1 + \frac{a-1}{g_1(N-a)}(K\rho_L^* + 1)} \quad (38)$$

$$I_x(n, n) = \int_0^x \frac{x^{m-1}(1-x)^{n-1}}{B(m, n)}\, dx \quad . \quad (39)$$

After we have an estimator and a value for β, and have selected a set of values for the system parameters D, A, and ρ, the influence of the design can then be ascertained by varying the number of classes $\underline{a}$ for fixed N, over a specified range and computing for each value of $\underline{a}$ the quantity P as given by use of the forms (Equations 34 through 36).

Hence, as experimenter could determine for himself which design he should use for his particular situation, i.e., for a specific D, A, ρ, and sample of N. To make the results of this report more useful, it was decided to present some numerical results.

Using the results concerning the behavior of $\sigma'^{2}_{R}\left|\hat{d}_{1}\right.$, a series of tables of value of P were prepared. It was hoped that consideration of some numerical results would aid in understanding the effect of design and estimator on the specific objective considered, viz., reducing product variability.

To compute P, it is necessary first to specify the system parameters ρ, 100A, and D; the design perameters N and $\underline{a}$; and the fixed number β. Unfortunately, one is not able to examine all the combinations of system and design parameters that he would like. A set of parameters were chosen that should represent many situations encountered in actual practice. The values of ρ selected were 1/10, 1/4, 1/2, 1, 2, 4, and 10. The values of 100A selected were 1/10, 1/4, 1/2, 1, 2, 4, and 10. Note, for example, that $\rho = 10$ means that σ_{A}^{2} is 10 times larger then σ_{B}^{2}, and $100A = 10$ means that σ_{A}^{2} is reduced at 10 times the rate at which σ_{B}^{2} is reduced. For D, the value $D = 100$ was selected.

The sample size considered was $N = 72$. The values of $\underline{a}$ used were 2, 4, 8, 12, 18, 24, 36, 48, and 60.

The value of β used was

$$\beta(0.90) = \left[9 \min\left(\sigma'^{2}_{R}\left|\hat{d}_{1}\right.\right) + \max\left(\sigma'^{2}_{R}\left|\hat{d}_{1}\right.\right)\right]/10.$$

The value of $\beta(0.90)$ represents a reduction in total variance of 90 percent of the maximum attainable amount.

For the 49 combinations of ρ and 100A, the values of P are presented in Tables III through IX.

TABLE III

Values of $P = \text{Prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for 100A = 1/10

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	1.00	1.00	1.00	0.96	0.87	0.76	0.39
4	1.00	1.00	1.00	0.99	0.91	0.76	0.53
8	1.00	1.00	1.00	1.00	0.96	0.79	0.63
12	1.00	1.00	1.00	1.00	0.98	0.81	0.68
18	1.00	1.00	1.00	1.00	0.98	0.83	0.73
24	1.00	1.00	1.00	1.00	0.99	0.84	0.76
36	1.00	1.00	1.00	1.00	0.98	0.83	0.79
48	1.00	1.00	1.00	1.00	0.96	0.79	0.78
60	1.00	1.00	0.99	0.98	0.89	0.71	0.76

TABLE IV

Values of $P = \text{Prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for 100A = 1/4

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	1.00	0.99	0.94	0.83	0.75	0.40	0.52
4	1.00	1.00	0.98	0.87	0.75	0.53	0.73
8	1.00	1.00	0.99	0.91	0.76	0.63	0.88
12	1.00	1.00	1.00	0.93	0.78	0.68	0.93
18	1.00	1.00	1.00	0.95	0.79	0.73	0.97
24	1.00	1.00	1.00	0.95	0.80	0.76	0.98
36	1.00	1.00	0.99	0.93	0.78	0.78	0.99
48	1.00	0.99	0.98	0.89	0.74	0.77	0.99
60	0.97	0.95	0.91	0.79	0.66	0.74	0.97

TABLE V

Values of P = Prob $\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for 100A = 1/2

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	0.99	0.94	0.84	0.77	0.14	0.45	0.61
4	1.00	0.98	0.88	0.77	0.26	0.63	0.85
8	1.00	0.99	0.92	0.80	0.38	0.76	0.96
12	1.00	0.99	0.93	0.81	0.45	0.83	0.99
18	1.00	0.99	0.94	0.82	0.50	0.87	1.00
24	1.00	0.99	0.93	0.82	0.52	0.90	1.00
36	0.99	0.98	0.90	0.80	0.52	0.92	1.00
48	0.97	0.94	0.85	0.75	0.47	0.90	1.00
60	0.89	0.83	0.74	0.66	0.36	0.86	1.00

TABLE VI

Values of P = Prob $\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for 100A = 1

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	0.96	0.86	0.80	0.15	0.45	0.52	0.66
4	0.99	0.90	0.82	0.27	0.63	0.74	0.90
8	0.99	0.93	0.85	0.38	0.75	0.88	0.98
12	0.99	0.93	0.86	0.44	0.81	0.93	1.00
18	0.99	0.93	0.87	0.48	0.86	0.96	1.00
24	0.98	0.91	0.86	0.49	0.88	0.98	1.00
36	0.94	0.87	0.82	0.47	0.89	0.98	1.00
48	0.88	0.76	0.76	0.41	0.87	0.98	1.00
60	0.76	0.69	0.66	0.30	0.82	0.95	1.00

TABLE VII

Values of $P = \text{Prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for $100A = 2$

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	0.91	0.85	0.26	0.28	0.52	0.59	0.69
4	0.94	0.88	0.45	0.48	0.73	0.82	0.92
8	0.95	0.91	0.59	0.65	0.87	0.94	0.99
12	0.95	0.92	0.65	0.72	0.92	0.98	1.00
18	0.93	0.91	0.67	0.77	0.95	0.99	1.00
24	0.91	0.90	0.67	0.78	0.96	1.00	1.00
36	0.85	0.85	0.62	0.76	0.97	1.00	1.00
48	0.77	0.78	0.52	0.69	0.95	1.00	1.00
60	0.66	0.68	0.38	0.53	0.91	0.98	1.00

TABLE VIII

Values of $P = \text{Prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for $100A = 4$

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	0.92	0.40	0.37	0.44	0.58	0.67	0.75
4	0.96	0.62	0.60	0.70	0.83	0.90	0.96
8	0.97	0.73	0.75	0.86	0.95	0.98	1.00
12	0.96	0.76	0.80	0.91	0.98	1.00	1.00
18	0.95	0.76	0.82	0.93	0.99	1.00	1.00
24	0.93	0.74	0.82	0.94	0.99	1.00	1.00
36	0.88	0.68	0.78	0.93	0.99	1.00	1.00
48	0.80	0.58	0.69	0.89	0.99	1.00	1.00
60	0.68	0.43	0.53	0.78	0.96	1.00	1.00

TABLE IX

Values of $P = \text{Prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right]$ for 100A = 10

	ρ						
a	1/10	1/4	1/2	1	2	4	10
2	0.99	0.56	0.57	0.63	0.70	0.77	0.84
4	1.00	0.76	0.79	0.86	0.93	0.96	0.99
8	1.00	0.84	0.89	0.95	0.99	1.00	1.00
12	1.00	0.86	0.92	0.98	1.00	1.00	1.00
18	1.00	0.85	0.93	0.99	1.00	1.00	1.00
24	1.00	0.84	0.93	0.99	1.00	1.00	1.00
36	1.00	0.80	0.91	0.98	1.00	1.00	1.00
48	0.98	0.75	0.86	0.97	1.00	1.00	1.00
60	0.90	0.67	0.78	0.92	0.99	1.00	1.00

It is also obvious that the magnitude of P depends on the specific combination of ρ and 100A.

For situations where 100A and ρ are both very large, or both very small, the values of P are near one. The reason for this may be understood by considering a specific case. Consider the combination ρ = 1/10 and 100A = 1/10. With 100A = 1/10, we would want to allocate all our funds to source B unless ρ were quite large. This condition is reflected in the value of ρ_1, which is 3.67; i.e., we would want to allocate all our funds to source B unless ρ were greater than 3.67. Hence, $\text{prob}\left[\left(\sigma_R'^2 \mid \hat{d}_1\right) \leq \beta\right] = \text{prob}\left[(\hat{\rho} \leq 4.39) \mid (\rho = 1/10)\right]$ is very large, no matter what design or estimator is used.

From general consideration of the manner in which P varies as a function of $\underline{a}$, it appears that, if one uses a design that is moderately near the optimal, he will do quite well in achieving his objective of reducing total variance.

The results seem to indicate that, for most situations, an intermediate value of $\underline{a}$, say $N/4 \leq a \leq N/2$, will give results that are quite close to the optimal. If one value of $\underline{a}$ were to be recommended, it would be $a = N/3$.

LIST OF REFERENCES

Anderson, R. L., "Uses of Variance Component Analysis in the Interpretation of Biological Experiments," Bullètin Int. Statist. Instit., 37:3:71-90, 1960.

Anderson, R. L., and Bancroft, T. A., Statistical Theory in Research, McGraw-Hill Book Co., Inc., New York, 1952.

Crump, P. P., Optimal Designs to Estimate the Parameters of a Variance Component Model, unpublished Ph.D. thesis, North Carolina State College, Raleigh, 1954.

Gaylor, D. W., The Construction and Evaluation of Some Designs for the Estimation of Parameters in Random Models, unpublished Ph.D. thesis, North Carolina State College, Raleigh, Inst. of Stat. Mimeo Series No. 256, 1960.

Satterthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components," Biometrics Bulletin, Volume 2, pp. 110-114, 1946.

USE OF STATISTICAL DESIGN IN LABORATORY ENVIRONMENTAL TESTING OF ADAPTION KITS

D. J. Taravella
Picatinny Arsenal

1. This presentation will attempt to give you an idea of the complexity, cost, and problems associated with conducting a laboratory environmental test program for a large ballistic missile adaption kit. The use of a modified fractional factorial design is a method being utilized to assess the reliability of an expensive and complex item with some degree of assurance.

2. The primary function of Picatinny Arsenal in the special weapons area is the development of adaption kits for use in nuclear applications. An adaption kit is the equipment required to adapt a nuclear bomb for use with a missile or a projectile. This equipment furnishes the arming, safing, and fuzing signals to the warhead, and includes the necessary ground support equipment for handling and checkout of the warhead section. The type of fuzing used includes electronic, contact, inertial, timers, and altitude sensing devices. The safing and arming devices are generally a collection of "g" weight systems, switches, timers, and solenoids. The ground handling equipment must be capable of handling warheads and warhead sections weighing up to several thousand pounds. The checkout equipment must be capable of checking simple continuity, to setting and checking of any of the complicated fuzing systems. The reliability requirements, as stated in the Military Characteristics, and Stockpile-to-Target Sequences, are usually 99% with certain safety or functioning requirements as high as 99.99%. The environmental requirements are normally high and low temperature (+160° and -65°F), sand and dust, humidity, ozone, salt atmosphere, fungus, rain, sunshine, thermal shock, transportation and handling (all forms of transportation, such as airplane, railroad, truck, tracked vehicle, helicopter, and boat) plus the flight or firing environments of acceleration, shock, temperature-altitude, and vibration.

3. The following photographs* are examples of a ballistic missile adaption kit hardware which will give you an idea of the size, complexity, and cost of an adaption kit and its required ground support equipment.

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

Photograph 1* - This electronic fuze is a pulsed radar fuze which has a dual altitude sensing system that is commonly referred to as a Target Detecting Device. Each Target Detecting Device has a separate power supply. The outputs of both systems are cross-coupled to both arm-safe devices. The system utilizes a common transmitting and receiving antenna. The fuze contains the following type of circuits or parts: RF amplifier, klystron and converter, AFC, magnetron, synchronizer, modulator, IF amplifier, and firing circuit. This fuze weighs 400 lbs.and costs approximately $30,000, including a self-contained power supply.

Photograph 2- This photograph shows the location and arrangement of the impact fuzing system. This system consists of two electrically independent fuze sets. Each fuze set consists of an impact crystal, T-connectors, and cabling.

Photograph 3-The impact crystal consists mainly of the piezo-electric ceramic elements, a weight, a spring, and a coaxial connector. The impact fuzing system costs approximately $1,200.

Photograph 4- This safing and arming device is a programming mechanism which consists of the following parts:

a. Three "g" sensing weights, accelerometer, "g" weight, and decelerometer.
b. Four micro switches.
c. Two modified T3 Timers; a 140-second and a 4-second clock.
d. Three banks of wafer switches.
e. Two solenoids.
f. Arming cam, spring loaded.

Photograph 5- All of the above are housed within a cylindrical aluminum housing with a connector to the fuze in front and the cable at the rear end. This item costs approximately $2,000; two of these items are required.

Photograph 6- This cabling connects the safing and arming devices to the warhead. It is approximately 276 inches in length. This cabling costs approximately $200.

---

*Photographs have been placed at the end of this article.

Photograph 7- Two mechanical support structures are used to support and mate the nuclear warhead to the missile skin structure. This is the forward payload support.

Photograph 8- This is the aft payload support. These structures cost approximately $2,500.

Photograph 9- This is the bleeder tube assembly which is used to vent moisture and gas from the warhead. This assembly costs approximately $325. All of these components are the flight components of the adaption kit.

4. The following are the ground handling equipment:

Photograph 10- This is the warhead stand which is used to assemble the warhead into the missile structure.

Photograph 11- This shows the fuze cone handling tool which is used to assemble the fuze cone assembly (400 pounds) to the warhead section. Photograph 1 shows this tool with the fuze cone assembly.

Photograph 12- This shows the fuze cone stand which is used during checkout of the fuze cone assembly prior to assembly.

Photograph 13- This shows the warhead section sling which is used to handling the warhead section which weighs over 7,000 pounds.

The ground handling equipment costs from $250 to $3,000 for each item.

5. Six testers are used in the checkout of the adaption kit components and the warhead section. They are as follows:

Photograph 14- XT-4000 Fuze Tester is a GO NO GO tester which programs the air burst fuze through a simulated flight. Certain test points such as magnetron, IF, crystal current, noise, AFC, and continuity are checked.

Photograph 15- XT-4004 Tester - This tester checks the capacitance of the networks and contains a device which is used to check the T53 crystal.

Photograph 16- XT-4005 Cable Tester - This tester is used to check the continuity of the cable.

Photograph 17- XT-4006 Assembly Tester - Checks the various circuits of the warhead section to assure continuity and to ascertain that certain switches are in the prescribed position.

Photograph 18- XT-4007 Control and Monitor Panel - This tester is part of the missile Prefire Control and Test Truck. The burst option is controlled by this tester. It checks certain continuity circuits and activates a safing switch within the warhead.

Photograph 19- Warhead Installation Simulator - This tester is used to assure that the T-4006 and T-4007 Testers are in proper working order. This tester can also be used to introduce simulated malfunctions for troop training. These testers cost approximately $300 to $12,000 each.

6. Due to the high cost and complexity of special weapons hardware and the high cost of flight testing, considerable time and effort are being devoted to the laboratory testing activity. One group within the Long Range Atomic Warheads Laboratory, of which I head, has been established with the principle function to organize, establish, and conduct the laboratory environmental portion of the qualification program for certain adaption kits. The term qualification applies to the demonstration to all concerned that the adaption kit is capable of being exposed to the Stockpile-to-Target Sequence environments and still perform satisfactorily. The normal storage life requirement is five years. These environmental tests are considered to be accelerated aging tests.

7. Within the Special Weapons Development Division at Picatinny Arsenal, the environments to be encountered during the Stockpile-to-Target Sequence have been standardized and separated into three categories: Transportation and Handling, Field and Storage, and Flight or Firing. These, in turn, are divided for convenience as follows:

a. Transportation and Handling

(1) Vibration

(2) Packaged Drop

b. Field and Storage

(1) Field Vibration

(2) Shock (impact)

(3) High Temperature

(4) Low Temperature

(5) Temperature-Shock

(6) Temperature-Altitude

(7) Sand and Dust

(8) Salt Spray

(9) Immersion

(10) Rain

(11) Sunshine

(12) Fungus

(13) Ozone

(14) Humidity

c. Flight Environments

(1) Temperature-Altitude

(2) Shock

(3) Vibration

(4) Acceleration

8. Laboratory simulation of the flight environments is often difficult to achieve. The most significant flight environments are Temperature, Altitude, Mechanical Shock, Acceleration or Deceleration, and Vibration. Equipment for the simulation of the combined environments of temperature, altitude, and vibration is readily available. The combination of temperature, altitude, vibration, and acceleration is available for small test specimens. True simulation of these combined environments is not available with present-day equipment. Of all of the most significant environments, vibration is the most difficult to reproduce. Most vibrators simulate sinusoidal vibration; some others can simulate random vibration. However, the vibration experienced during missile flight is usually a combination of both intermixed with mechanical shocks. Valid flight vibration data are the most difficult to obtain. Some of the contributing factors in not being able to secure valid flight information are: capabilities of the telemetering systems and transducers, priorities of desired flight information, time, and economics.

9. The REDSTONE Missile was the first large ballistic missile developed by the United States Army. Hence, this adaption kit for the warhead section was the first ever to be laboratory tested. In the planning and organization of the testing program, economics and availability of hardware must be considered. Except for the safing and arming devices, only a limited amount of test samples were available for the laboratory testing program. The test program was conducted with a limited number of test samples being exposed to a large number of environmental exposures. Only four of the twenty samples of safing and arming devices were subjected to the same testing

sequence. Analysis of the resulting data was often difficult and inconclusive.

10. Mindful of this experience in the first ballistic missile programs, more consideration was given to hardware allocations for environmental testing. This permitted the allocation, scheduling, and justification of a larger number of test samples. After consultation with our statisticians at the Arsenal, the use of bi-level fractional factorial design was recommended for our use in the laboratory test program. This statistical design was modified in order to obtain maximum engineering information. This approach is being used for the PERSHING and NIKE ZEUS Adaption Kits.

Photograph 20 shows this application for the flight components for PERSHING Adaption Kit. Originally, sixteen adaption kits were scheduled for the program. However, due to a major design change, the allocation was reduced. By another re-allocation, a group of twelve units of the same design will be used.

11. The principal modification is the use of an upper and lower level of severity in place of an absence of an exposure. These upper and lower levels are as follows:

Test	Lower (0)	Upper (X)
7 Temperature Shock	-40 to + 135°F	-65 to + 160°F
16 Humidity	5 cycles	10 cycles
17 Temperature-Altitude	-40 to 135°F. at 50,000 ft.	-65 to 160°F, at 100,000 ft.
18 Shock	20 Gs	30 Gs
19 Vibration	10 Gs	15 Gs
20 Acceleration	10 Gs/5Gs	15 Gs/7.5Gs

Checkouts will be conducted after exposure to the Temperature-Shock and Humidity environments, and during the flight environments of Temperature-Altitude, Shock, Vibration, and Acceleration. The

Flight Vibration environments will be a combined environment of vibration and temperature; 8 at ambient (+77°F), 4 at -65°F, and 4 at + 160°F.

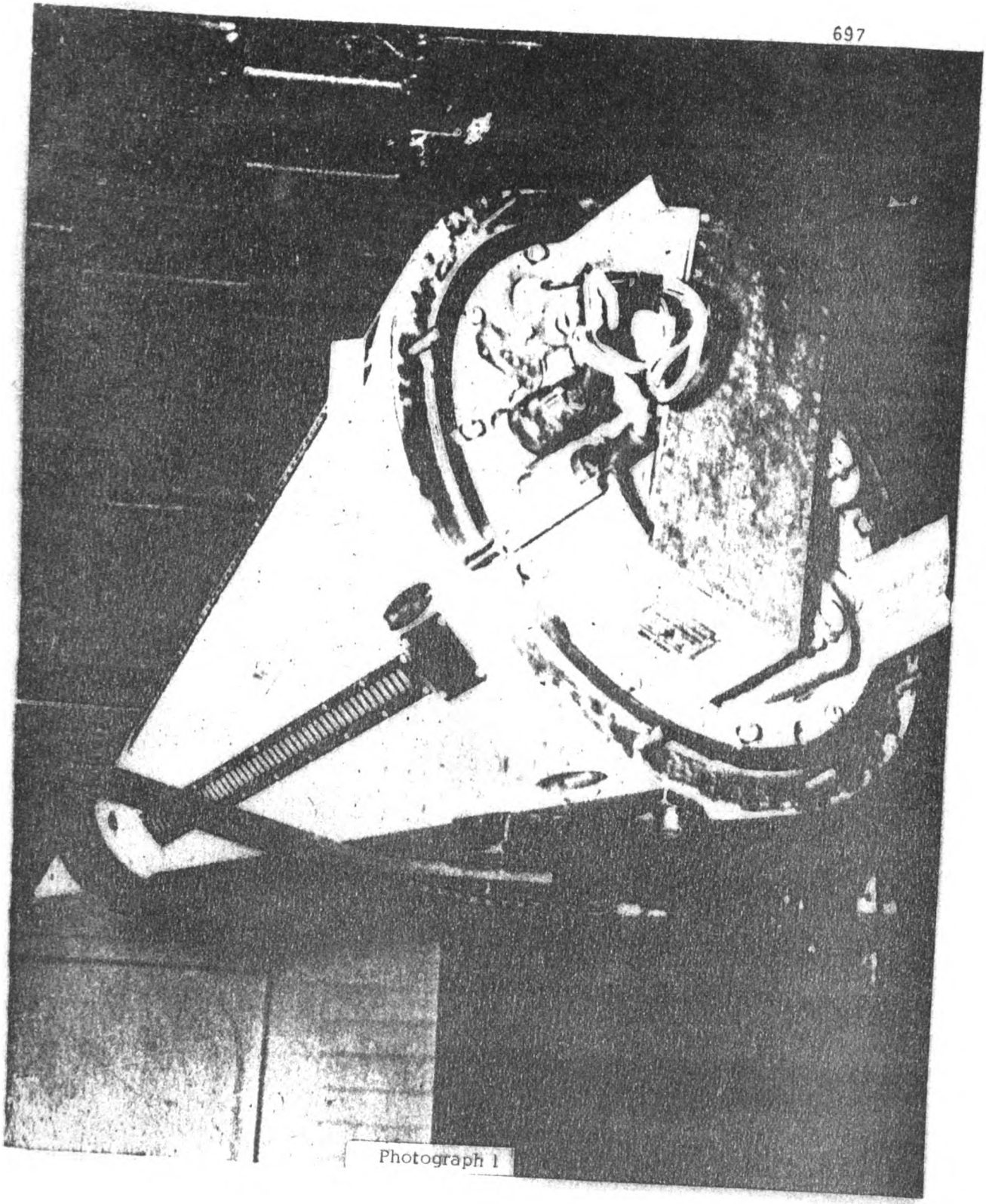
The upper limits of the firing environments of shock, vibration, and acceleration are approximately 50% above the required levels. So, this is somewhat of an overtest.

12. The cost of the hardware and testing is $100,000 per adaption kit; hence, the cost of this design is $1.2 million.

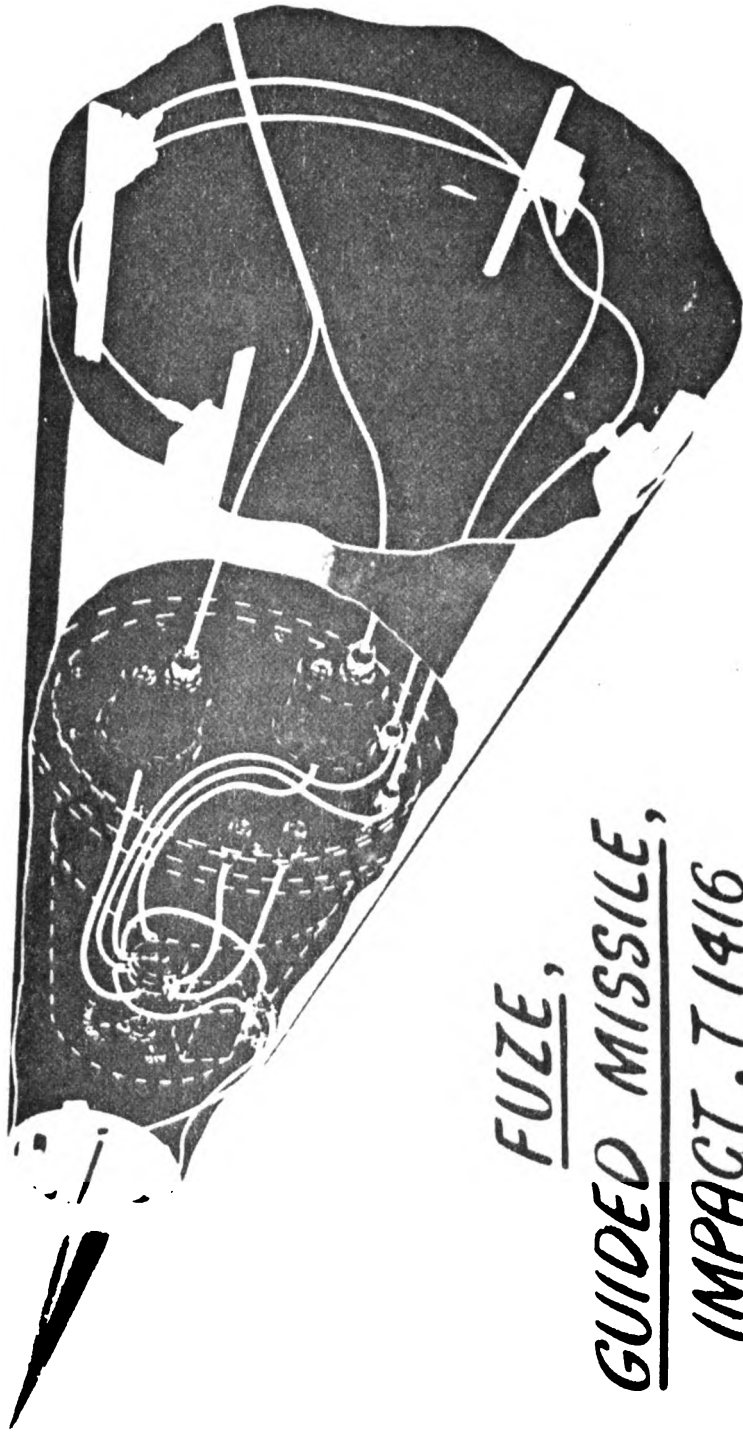
Photograph 21 shows this same application for the NIKE ZEUS Adaption Kit. The upper and lower levels are similar.

13. From an engineering point of view, the plan will feature a "larger" sample size undergoing the same and the most significant or severe environments to be encountered in the Stockpile-to-Target Sequence. The effect of each environmental exposure will be more readily discernible. This also applies to the combination of exposures. The desired goal of this application is to be able to assess the reliability of an adaption kit with some degree of assurance. One shortcoming of the plan is flexibility.

14. In summary, this presentation has attempted to give you an idea of the cost, complexity, and problems associated with conducting a laboratory environmental test program for an adaption kit. With this as a background, the use of a modified fractional factorial design is a method being utilized to assess the reliability of an expensive and complex item with some degree of assurance.



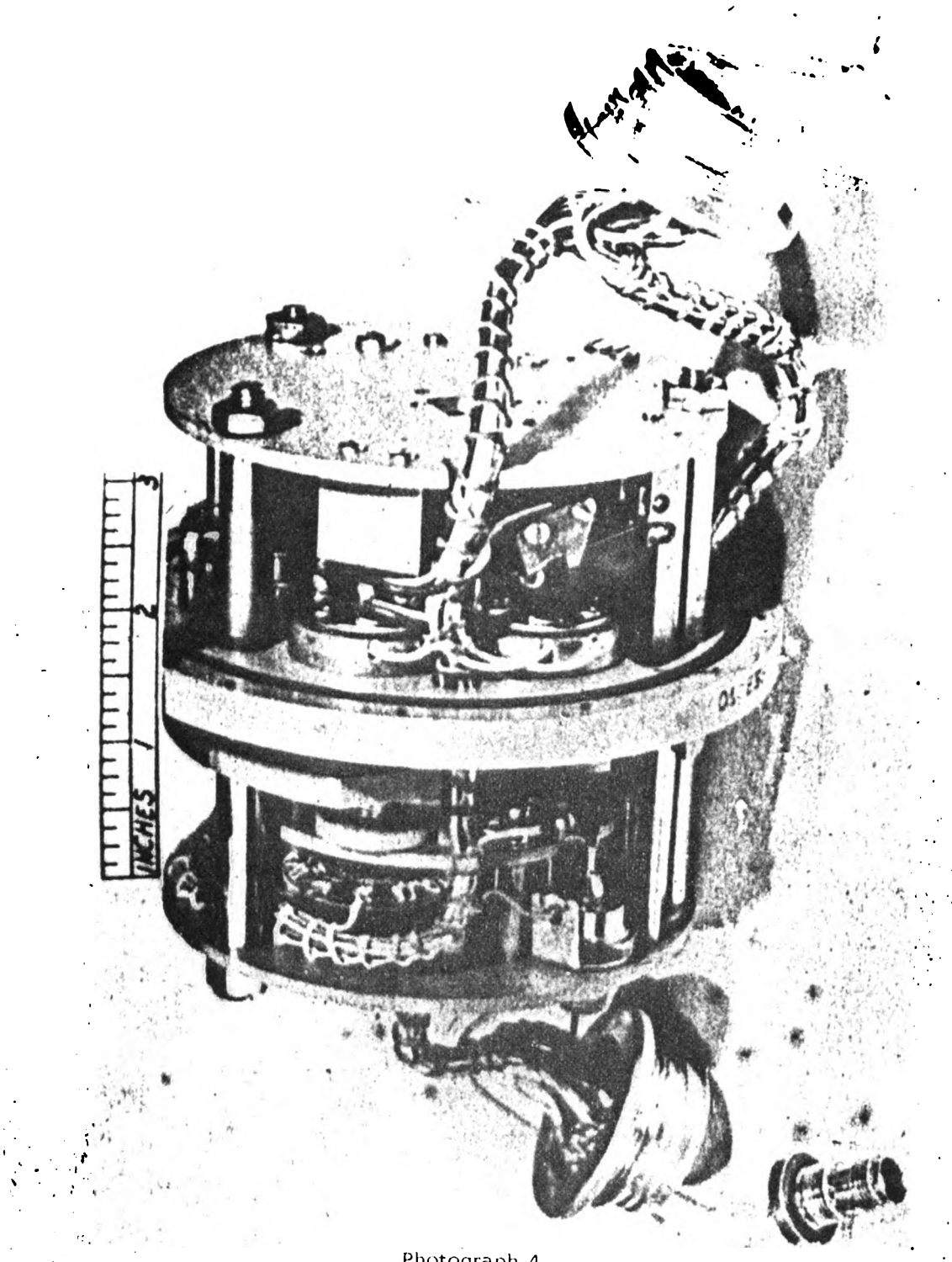
Photograph 1



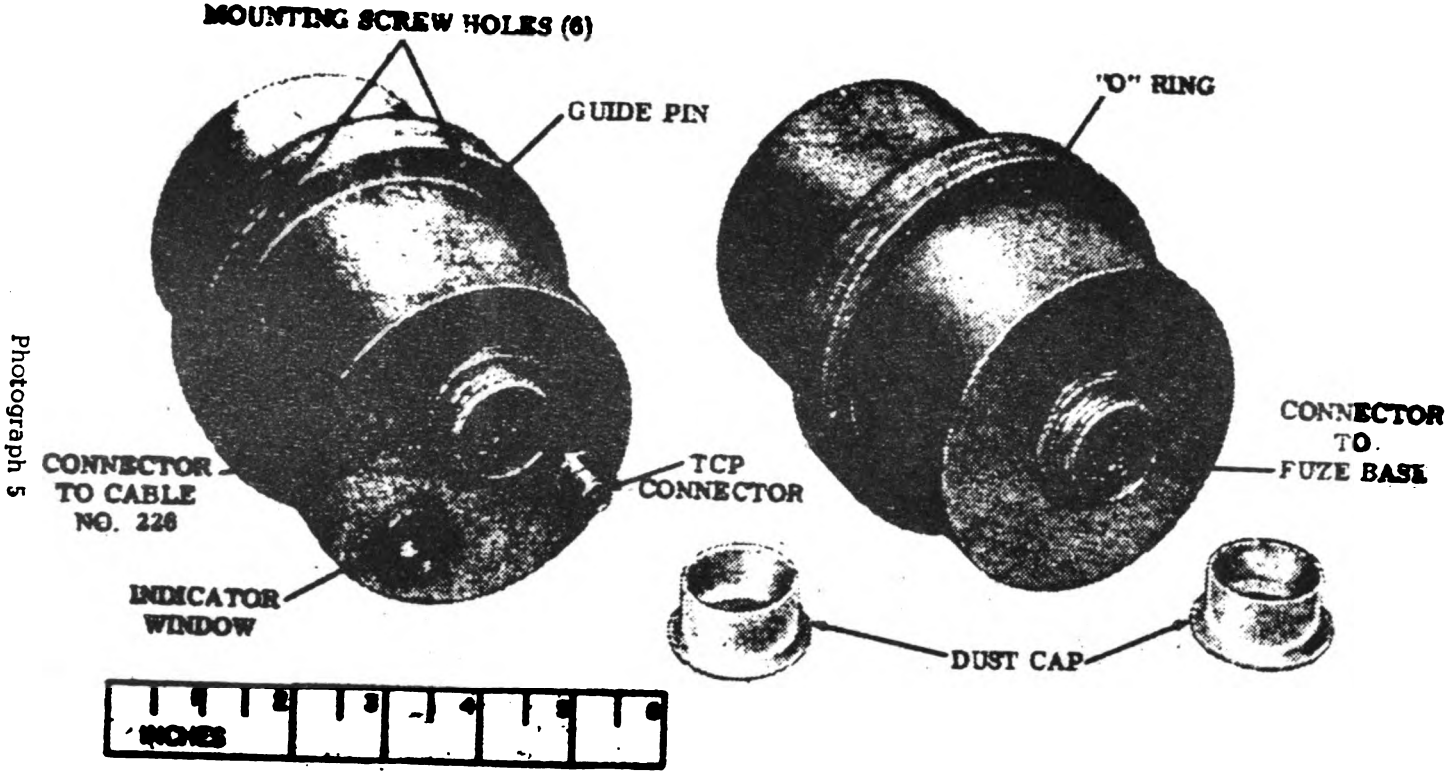
FUZE, GUIDED MISSILE, IMPACT, T 1416

Photograph 2

Photograph 3

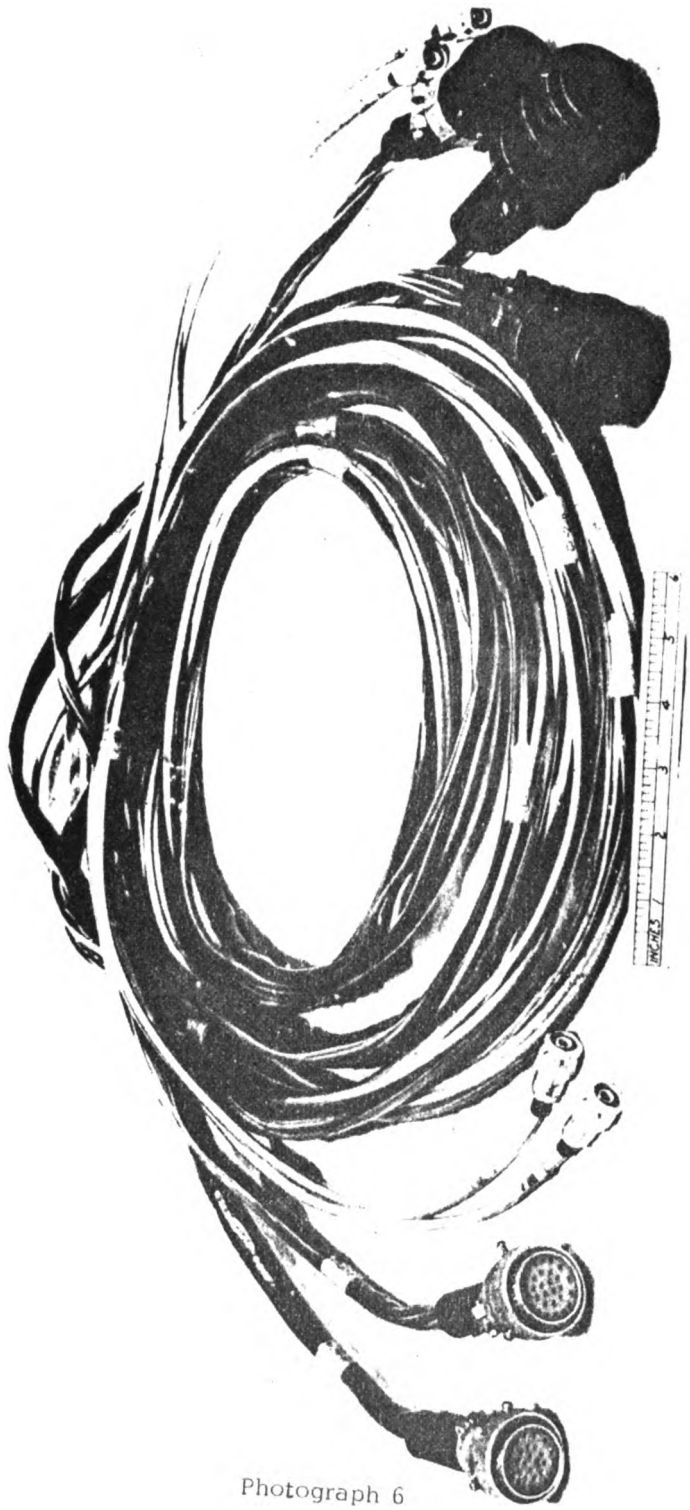


Photograph 4

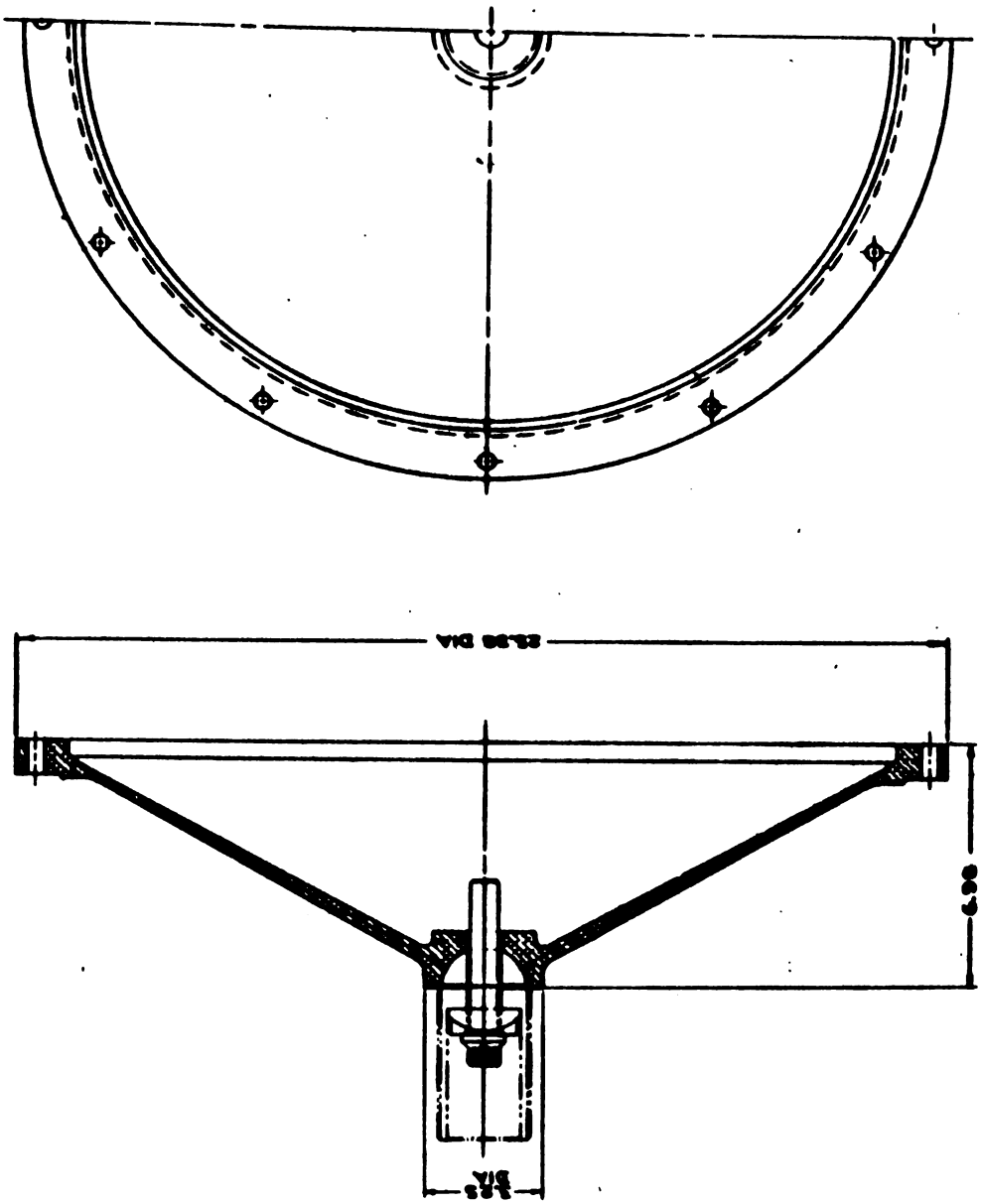


Photograph 5

T107 SAFE AND ARMING MECHANISM

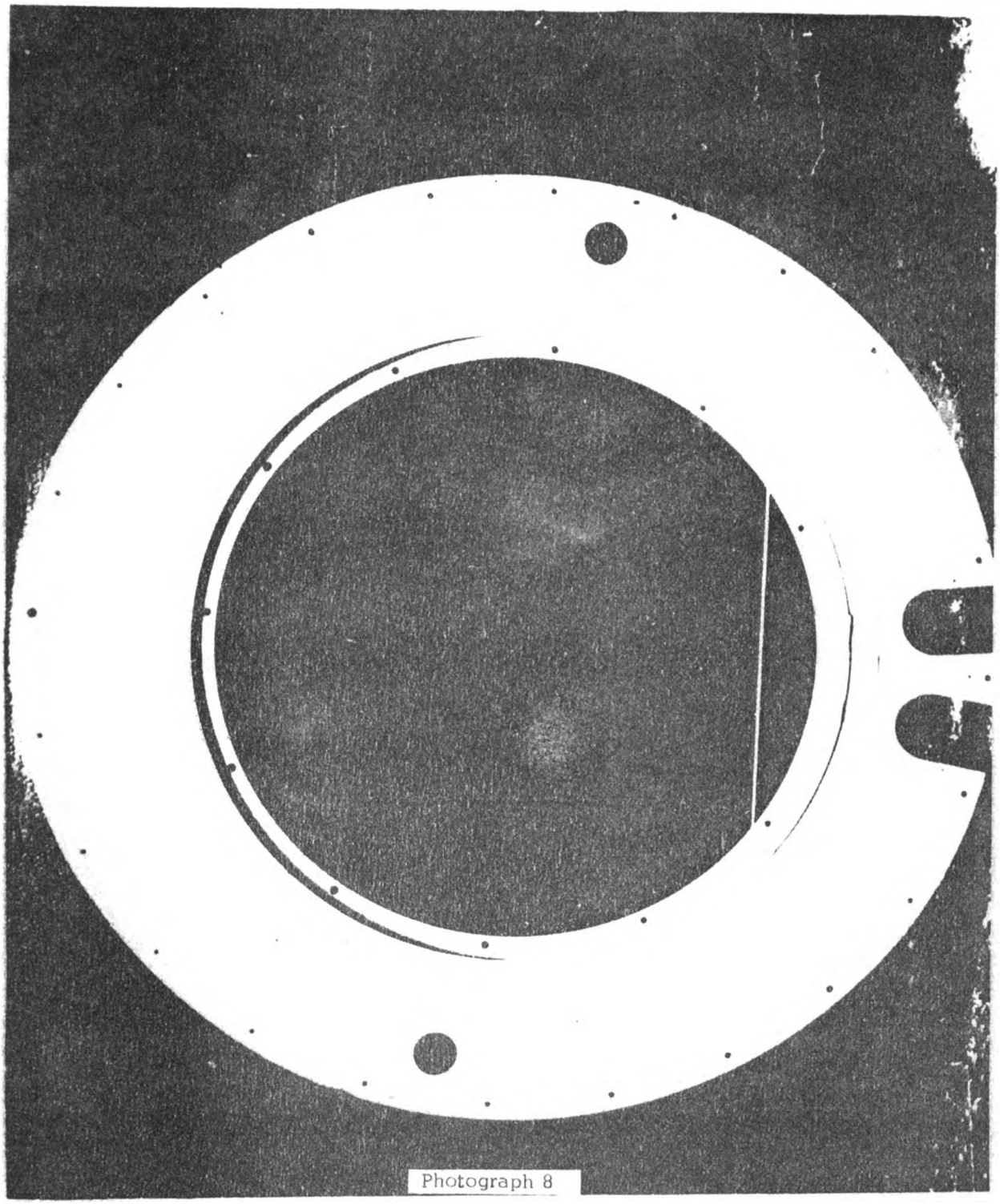


Photograph 6

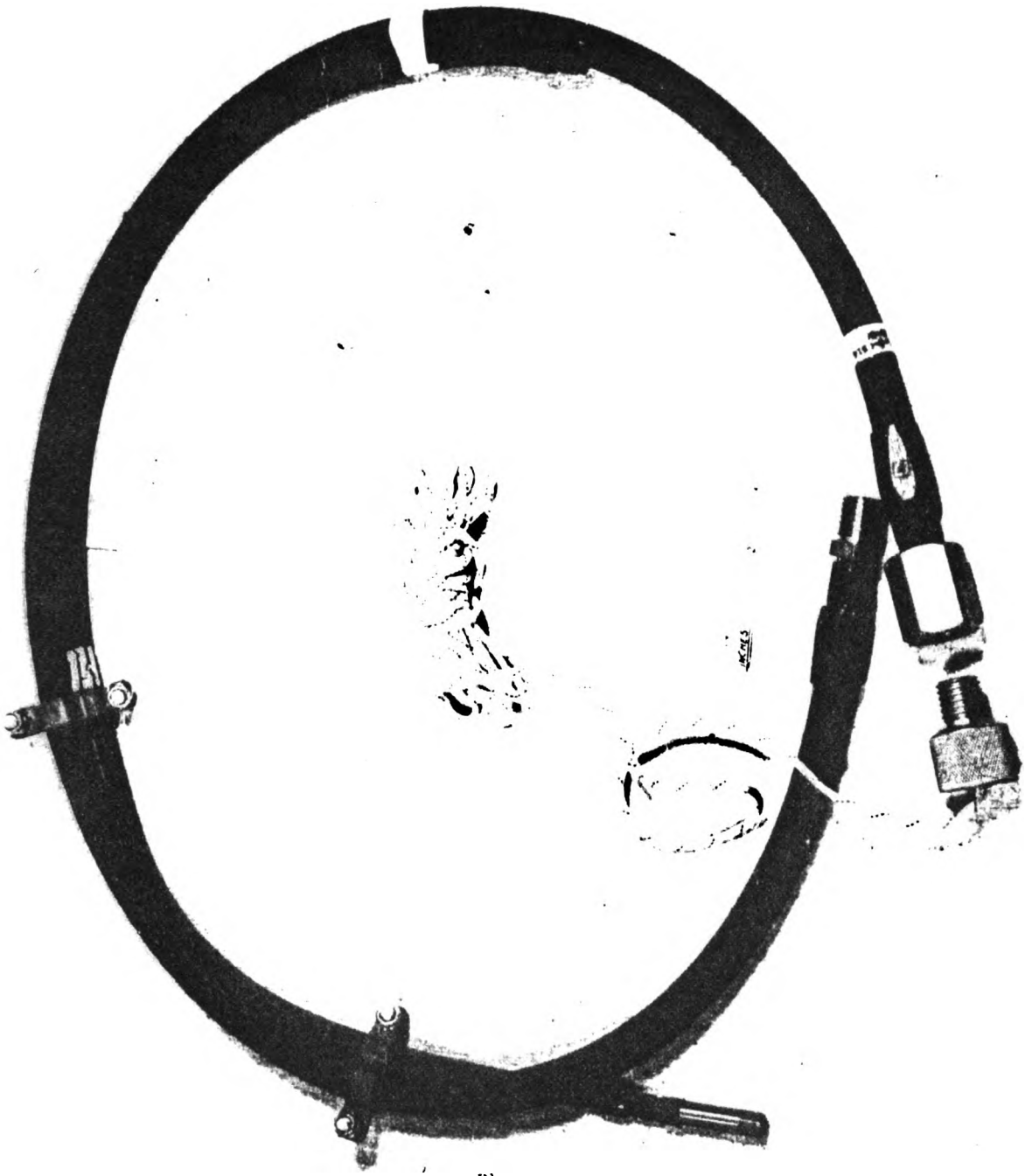


Forward Payload Adapter

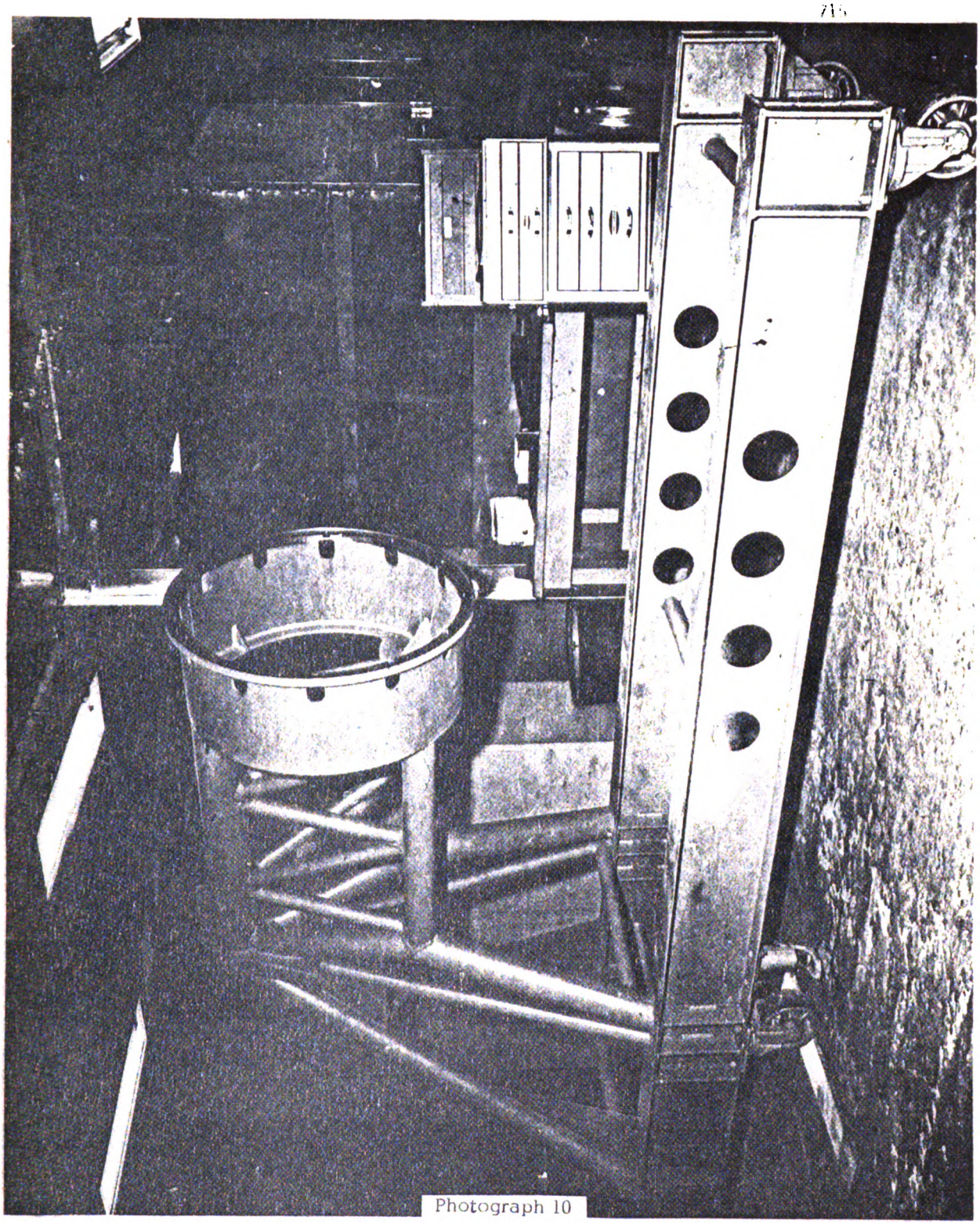
Photograph 7



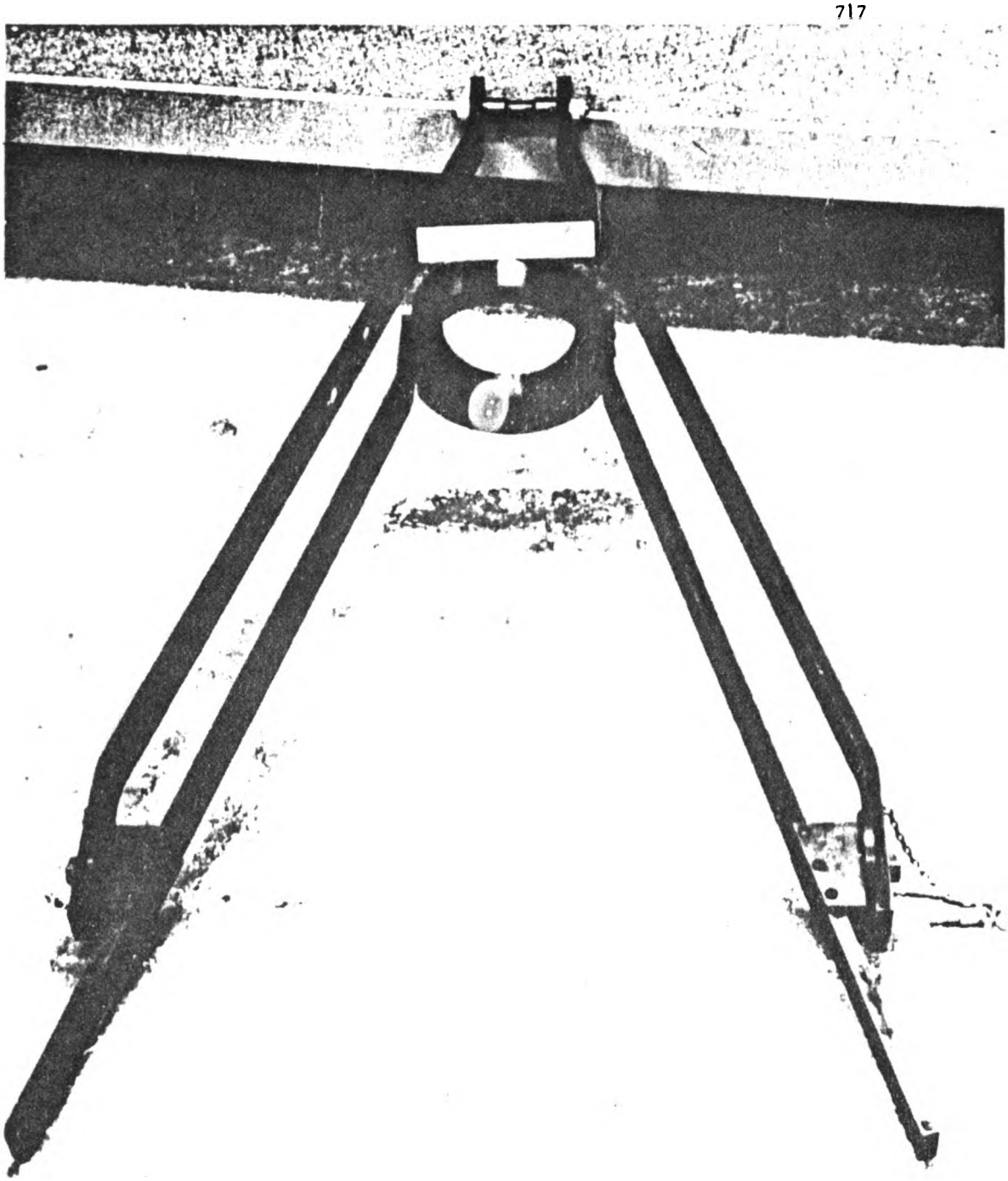
Photograph 8



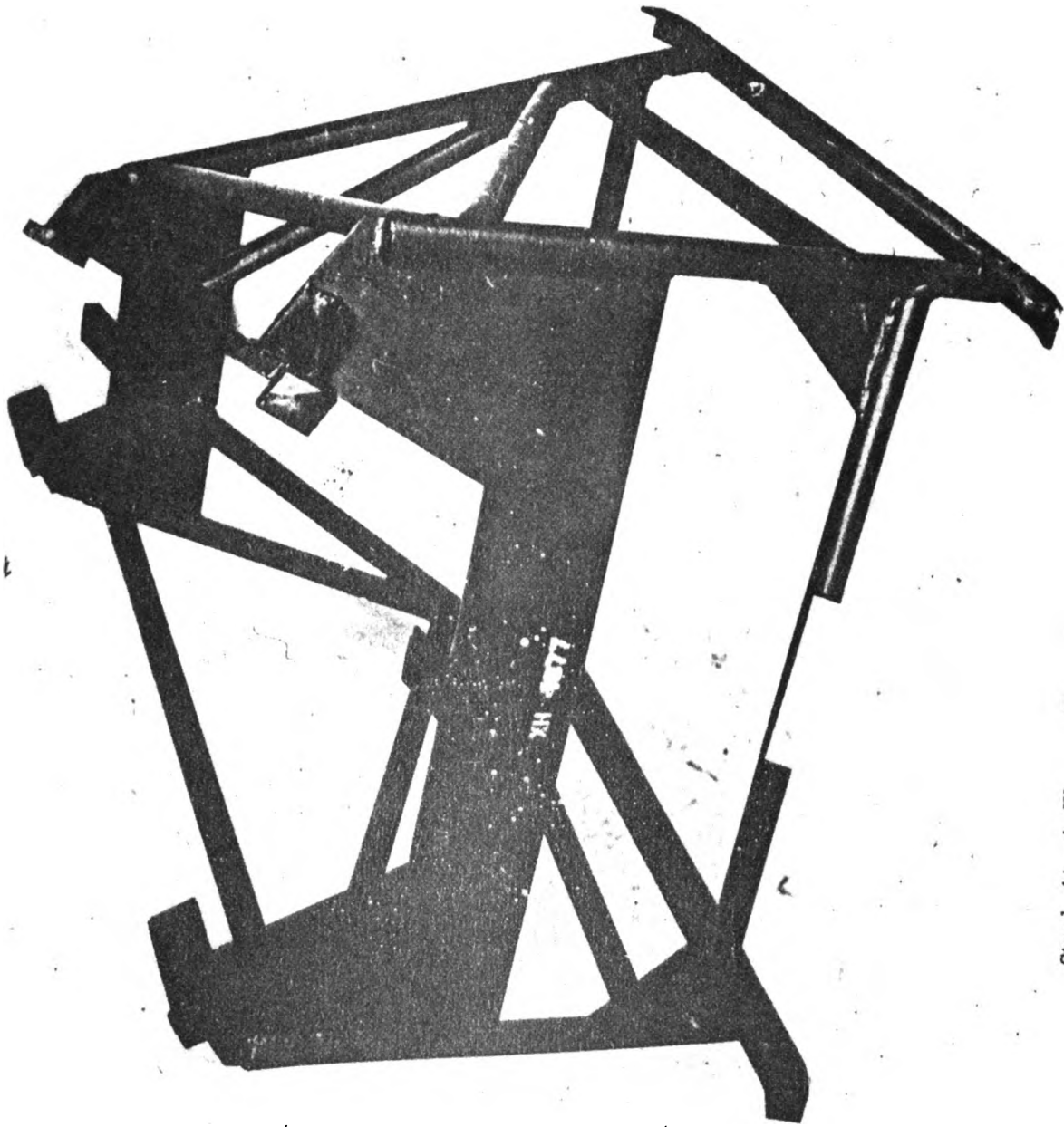
Photograph 9



Photograph 10

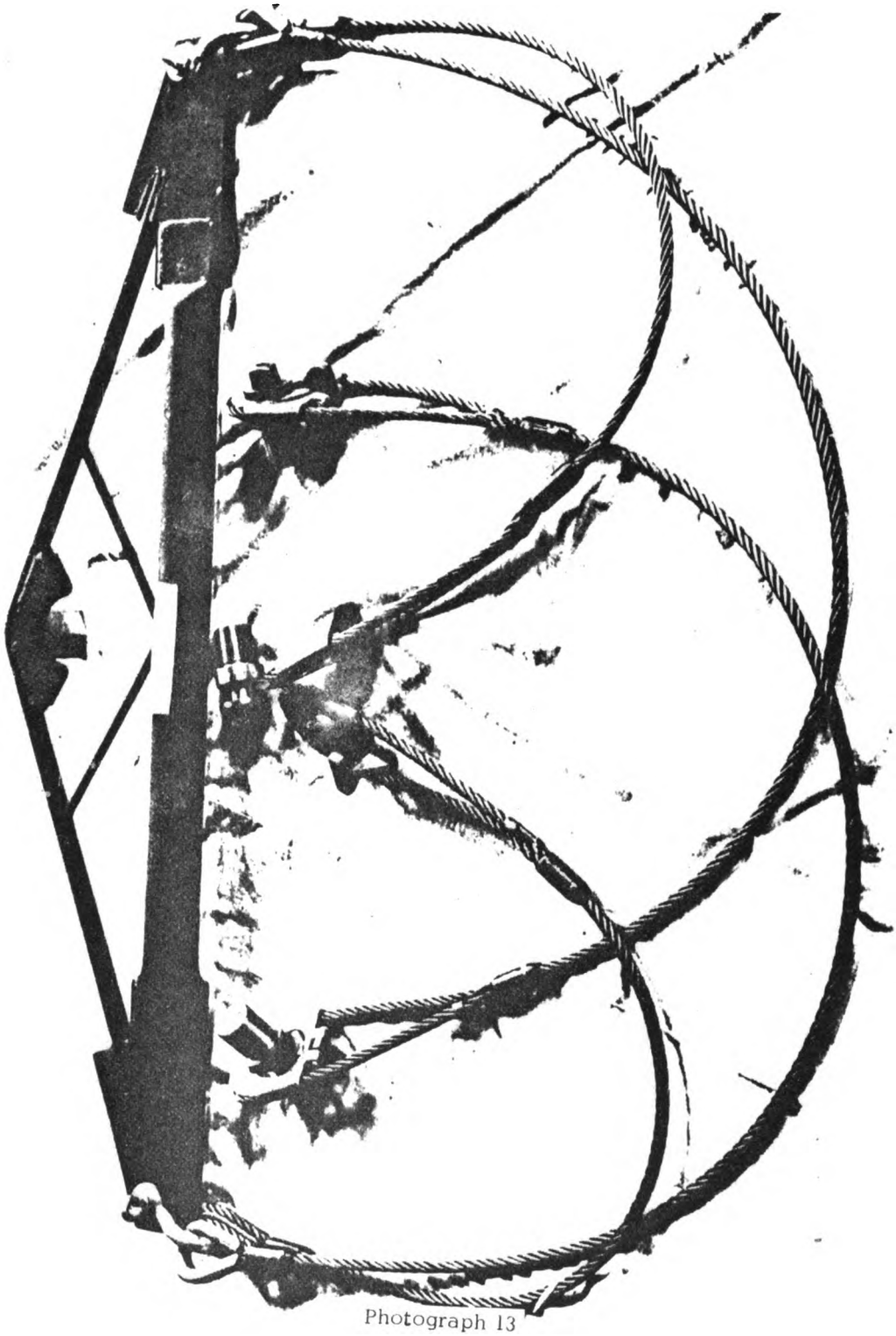


Handling Tool, Fuse Cone, XH-4057 (Lifting Bail)
Photograph 11



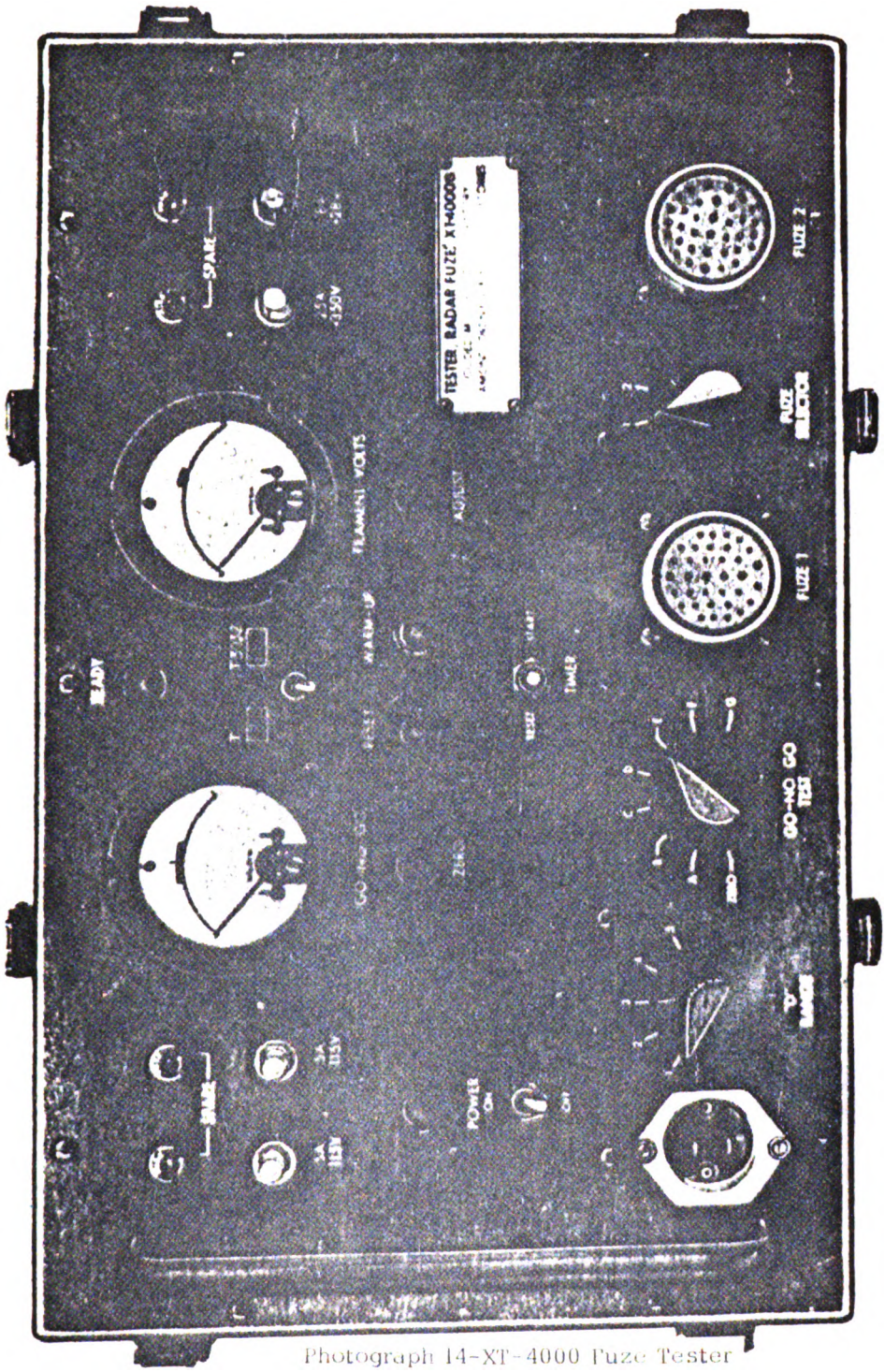
Stand, Atomic Weapon, Fuze, Cone: XH-4077

Photograph 12



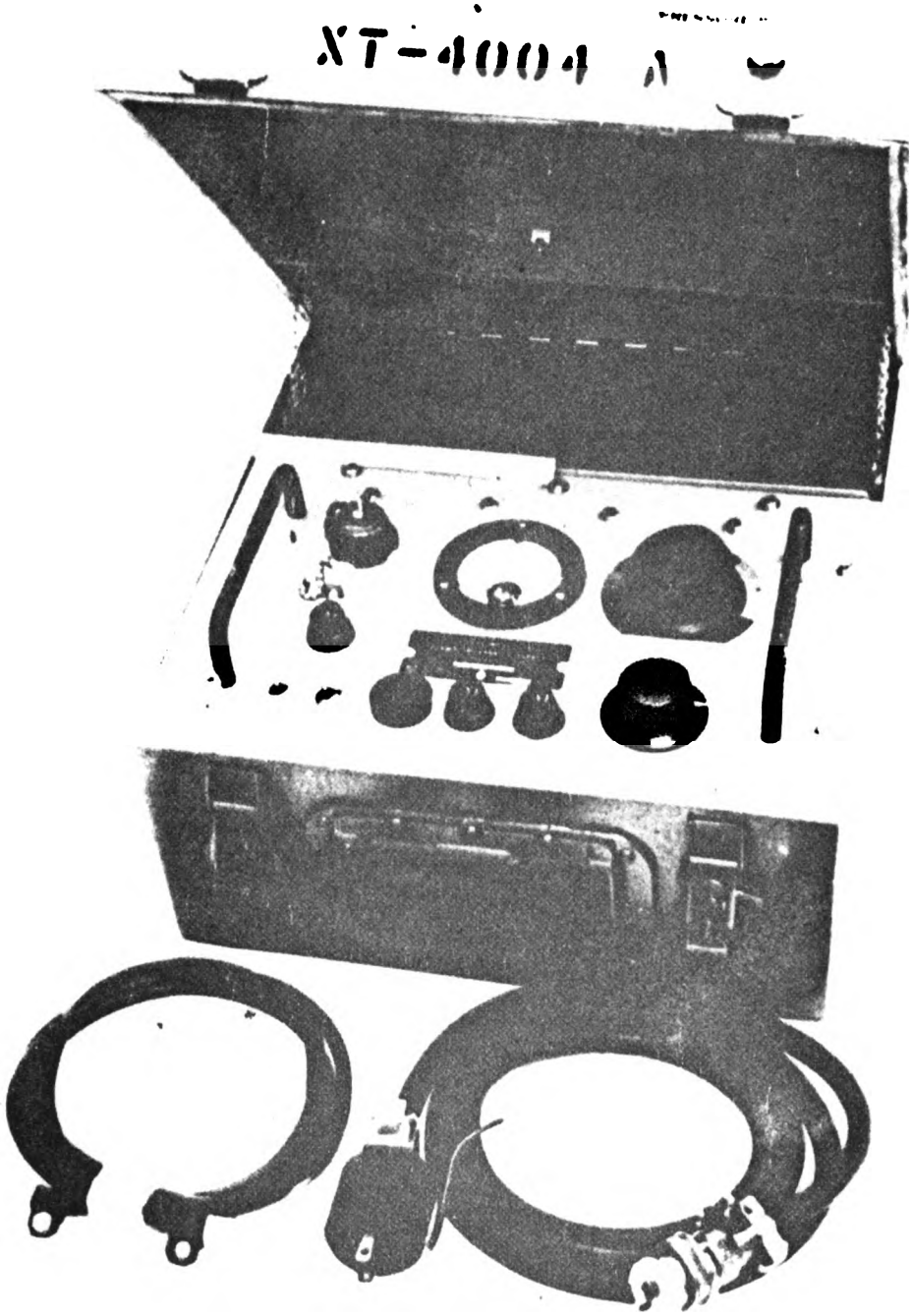
Photograph 13

Sling, Multiple Leg; XH-4060 (Nose Sling)



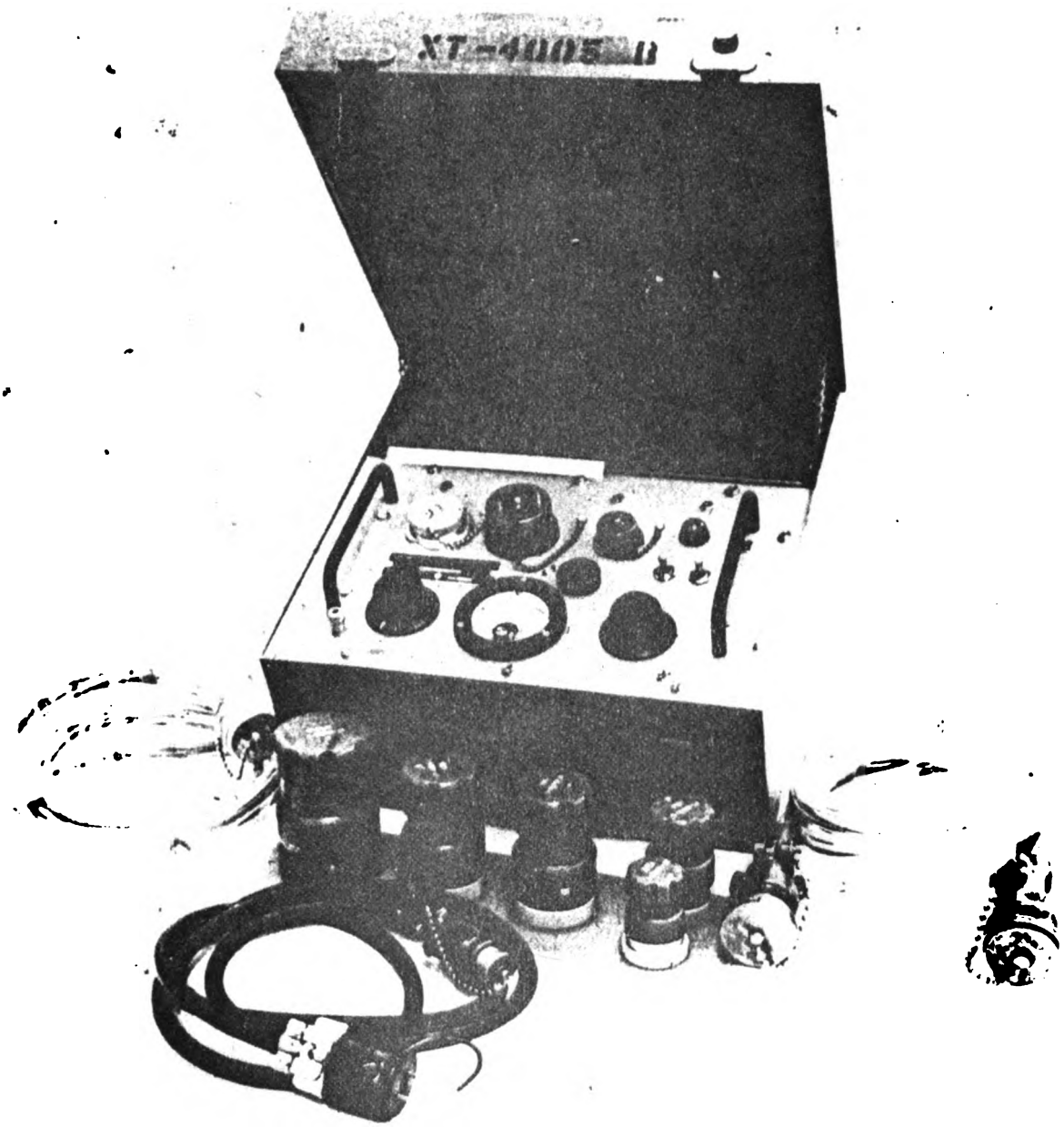
Photograph 14-XT-4000 Fuze Tester

RADAR FUZE TESTER, XT4000B



Photograph 15-XT-4004 Tester

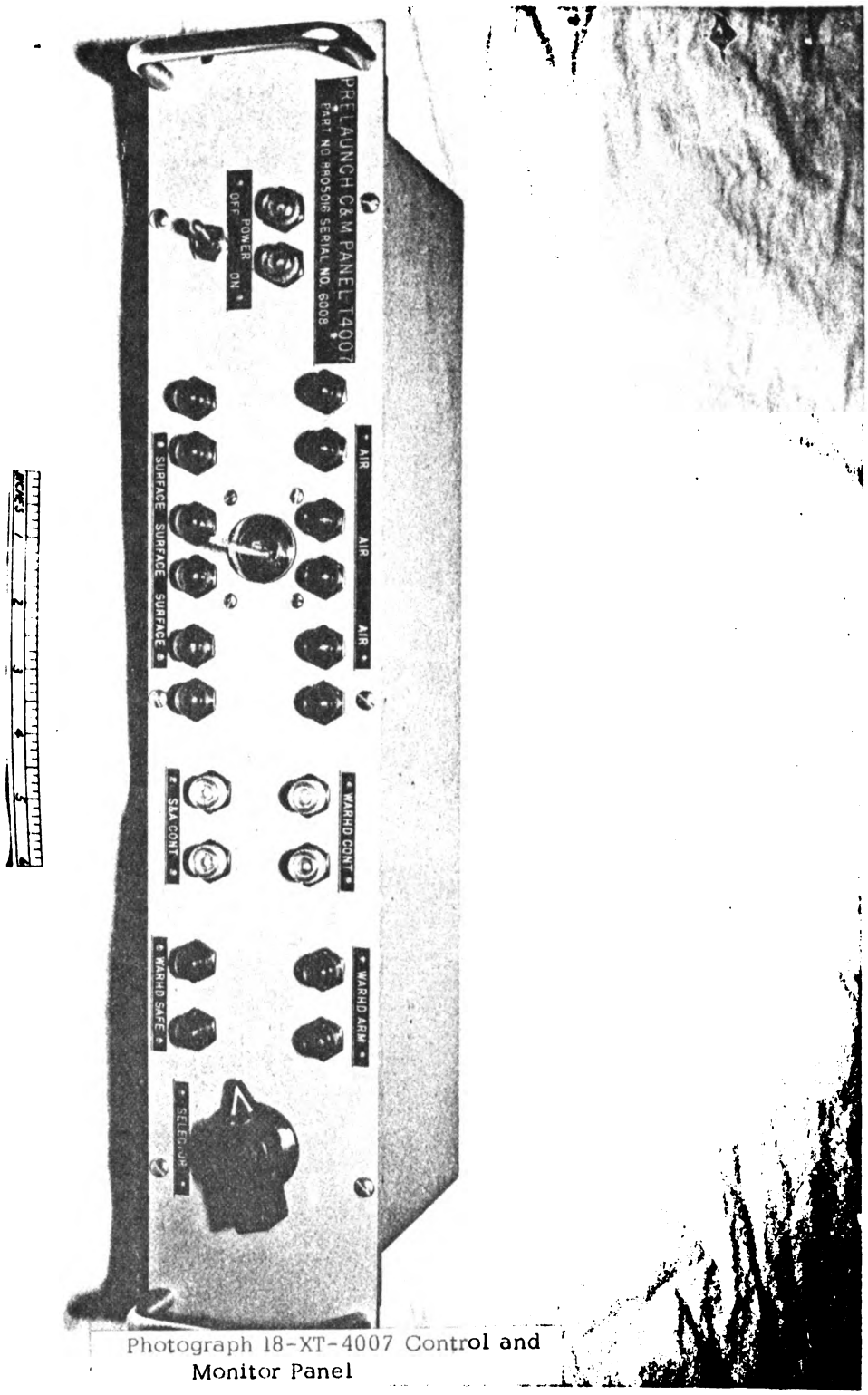
Test Set, Impact Fuze, Atomic Weapon, XT-4004A



Test Set, Electrical Cable: XT-4005B
Photograph 16-XT-4005 Cable Tester

Photograph 17-XT-4006 Assembly Tester

Test Set, Atomic Weapon: XT-4006B



Photograph 18-XT-4007 Control and Monitor Panel

Simulator, Warhead Installation, Atomic Weapon XT-400-B
Photograph 19-Warhead Installation
Simulator

QUALIFICATION AND RELIABILITY TEST PROGRAM
PERSHING ADAPTION KIT
FLIGHT-TEST COMPONENTS

TESTS	STATISTICAL EXPERIMENT ASSEMBLIES											
	1	2	3	4	5	6	7	8	9	10	11	12
A. TRANSPORTATION & HANDLING												
1. VIBRATION												
a. FOR ISOLATION SYSTEMS		X	X			X	X		X		X	
B. FIELD & STORAGE												
2. TEMP.-SHOCK (-65°F TO +160°F)	O	X	O	O	O	O	X	X	O	X	X	X
3. HUMIDITY (AMBIENT TO +160°F)	O	O	O	O	X	X	X	O	X	X	O	X
C. FLIGHT ENVIRONMENT												
4. TEMPERATURE-ALTITUDE	O	X	O	X	X	O	O	O	X	O	X	X
5. SHOCK	O	X	X	X	X	O	X	O	O	X	O	O
6. VIBRATION (2000 CPS)	O	O	X	X	O	O	X	X	X	O	O	X
7. ACCELERATION	O	X	X	O	X	X	O	X	O	O	O	X

Photograph 20

QUALIFICATION AND RELIABILITY TEST PROGRAM
NIKE ZEUS ADAPTION KIT

TESTS	QUALIFICATION TESTS															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A. TRANSPORTATION & HANDLING																
1. VIBRATION	X	X	X	X	O	O	O	O	X	X	X	X	O	O	O	O
B. FIELD & STORAGE																
2. TEMP. SHOCK (-65°F TO +160°F)	X	X	O	O	X	X	O	O	X	X	O	O	X	X	O	O
3. HUMIDITY (AMBIENT TO +160°F)	X	O	X	O	X	O	X	O	X	O	X	O	X	O	X	O
C. FLIGHT ENVIRONMENT																
4. SHOCK	X	X	X	X	O	O	O	O	X	X	X	X	O	O	O	O
5. ACCELERATION	X	X	O	O	X	X	O	O	X	X	O	O	X	X	O	O
6. VIBRATION (2000 CPS)	X	O	O	X	O	X	X	O	X	O	O	X	O	X	X	O

Photograph 21

A SERIES OF TWO-PHASE EXPERIMENTS*

Emil H. Jebe**
The University of Michigan

In pursuing a program of modeling and simulation, it is desirable to obtain from time to time some "real world" data on the performance of systems being studied in order to check and validate mathematical models used in the simulations. If properly designed experiments are carried out during the conduct of engineering tests, sound and useful interpretations may be obtained for validating a model. Among the systems studied have been various combat surveillance sensor subsystems. For one of these sensors, (hereafter referred to as sensor ANZ) which was to be simulated, it appeared that a program could be proposed for conducting engineering tests and obtaining data for sensor model validation. In exploring this problem, a series of two-phase experiments was designed. A description of these experiments and some explanation of their design and analysis provides the motivation for this paper.

For the investigation of a situation of interest to combat surveillance, sensor ANZ provides only a film record. When the film record has been interpreted, information is available for assessing the situation under surveillance. Instead of going directly to the design and film interpretation problems it will be useful to give some background of the sensor responses to be studied, parameters to be varied for the engineering evaluation tests and the environmental circumstances to be investigated.

---

*This paper is based upon work conducted for the Information Processing Task, Project Michigan, under Department of the Army Contract DA-36-039 SC-78801, administered by the U. S. Army Signal Corps.

**Research Mathematician, Operations Research Department, Institute of Science and Technology, The University of Michigan.

The author wishes to acknowledge the aid, the stimulation of and the interchange of ideas with Professor I. Copi, Research Logician, Dr. Thomas S. Lough, Research Associate and John H. Wyman, Research Assistant, Institute of Science and Technology, The University of Michigan, which helped develop this series of two-phase experiments.

From sensor responses to be observed during engineering tests a total of 13 characteristics could be evaluated for describing its performance. Examples of these 13 items are:

R_1: Probability of show (proportion of total number of moving targets detected in area of search),

R_2: Number of false targets reported in areas where targets might be expected,

R_3: Proportion of total number of parked vehicles detected in area mapped,

I: Proportion of total number of parked vehicles correctly identified in area mapped,

D: Distance errors (distances between reported and actual locations of targets),

PQ: Photographic qualities of the radar map.

Examples of parameters that are subject to experimental control in tests are the following:

r: slant range between aircraft and target,

TT: target type, e.g., jeep, truck or tank,

V_t: velocity of targets,

d_t: distance between targets,

V_a: velocity of aircraft,

h_a: altitude of aircraft,

R_s: sensor range setting.

Environmental conditions which may affect the sensor outputs are:

Foliage density,

Density of targets when uncontrolled as to speed and distance separation,

Turbulence of the air,

Interference from other radars,

Variation associated with the experimental units, e.g., an aircraft flight over a given path or a group of such flights.

It may be of interest to present a list of tests specified for investigating this complex of responses, parameters, and conditions and the list of corresponding experiments which were designed to implement such tests. The list and the experiments are given in Tables 1 and 2.

With this general picture of the test program before us we may return to the sensor's output, the film record and its interpretation. At this stage, we may ask, "What are two-phase experiments ?" A definition that has been given by Curnow is, "These are experiments in which the effects of the various treatments on the experimental units cannot be assessed directly and a further experiment is needed to estimate them" (Reference 3, page 60). McIntyre (Reference 1) has given examples of various combinations of First Phase and Second Phase experimental patterns. In our case, for any single experiment a film strip is obtained as output from the sensor for the observation of each experimental unit. If a machine could be used to interpret such a set of films or if a single highly reliable and accurate film interpreter not subject to fatigue could be used to interpret the film strips then a second phase experiment would not be required. Since a number of interpreters would be used to interpret the film outputs from the seven experiments shown in Table 2, it seemed wise to set up a series of carefully arranged film interpretation schedules. Problems of fatigue and unknown accuracy and reliability may then be avoided or assessed. Each combination of a field experiment with its film interpretation schedule then comprises a two-phase experiment; the First Phase is the field program and the Second Phase is the schedule of film assessment which is to be carried out in a Photo Interpretation Section.

TABLE 1

LIST OF TESTS

Test No	
1.	Probability of show as a function of traffic density See Experiment A
2.	Probability of show as a function of target speed See Experiment C
3.	Radial resolution - Included in Experiment D
4.	Lateral resolution - Included in Experiment D
5.	Minimum range determination and range calibration Included in Experiment D
6.	Comparison of radar map with conventional photography See Experiment E
7.	Vehicle Identification - See Experiment A
8.	Coordinate accuracy of radar photo map Included in Experiment D
9.	Interference from other radars - See Experiment F
10.	Accuracy of navigation radar compared with beacon or skin tracking - Included in Experiment D if needed equipment becomes available
11.	Effect of roll and yaw stabilization Included in Experiments A and C
12.	Foliage penetration - See Experiment G
13.	Human Factors Evaluation - See Experiment H and **
14.	Vulnerability to countermeasures - Refer ***
15.	Maintenance evaluation - Refer **

TABLE 2

LIST OF EXPERIMENTS

A	Stationary and Uncontrolled Moving Vehicle Experiment - Implements Tests 1, 7, and 11
B	Altitude Experiment - Tests Optimal Altitude Hypotheses
C	Controlled Vehicle Experiment - Implements Test 2 and is relevant to Tests 3, 4, and 11
D	Resolution, Fidelity, and Accuracy Experiment Implements Tests 3, 4, 5, and 8. If a sufficiently precise tracking radar is available test 10 may also be implemented.
E	Radar Mapping Experiment - For comparison with Photography. Implements Test 6
F	Radar Interference Experiment - Implements Test 9
G	Foliage Experiment - Implements Test 12
H	Photo Interpretation Experiments - Needed for implementation of most of the tests.

**Experiment Log Book will be maintained during the entire period of experimentation by the Project Engineer and the officer in charge of each day's activities. This record will be supported by a maintenance log and reports from personnel involved in the program. Implements Tests 13 and 15.

***Vulnerability Experiment
It is understood that implementation of Test 14 will be the responsibility of another department.

The literature on two-phase experiments is limited. Only three references are available (References 1, 2, 3). It is curious to note that all of these papers are by non-U. S. workers. Perhaps, this situation may be partially accounted for by the lack of recognition of the problem or the incomplete description of some complex research problems. Professor Kempthorne, Iowa State University, first called my attention to the two-phase problem in my consulting work there. I found that I had actually designed (without so labeling them) a number of two-phase programs at Iowa State (a report on one of these has been published, Reference 4).

In discussing the problem, McIntyre set forth a number of principles to be followed in designing the two-phase experiments. These are paraphrased as follows:

a. Replication in the first phase is needed to measure consistency of performance and provide a basis for a valid test of significance,

b. The results from each experimental unit of the First Phase should be separately evaluated in the Second Phase,

c. Replication in the Second Phase is not necessary but is highly desirable in order to cope with
   (1) Large uncontrollable variation in the Second Phase relative to expected First Phase treatment effects,
   (2) Analytical mistakes,
   (3) Expected small treatment effects in the First Phase. For this situation, McIntyre also suggested a high degree of "local control" in the Second Phase,

d. Results from experimental units of the Second Phase shall permit direct analysis in terms of the design of the First Phase experiment.

e. Replication in the Second Phase should be the same with respect to all elements of the First Phase so that errors of measurement for all experimental units of the First Phase will be uniform.

I have no criticism of McIntyre's points. They are sound principles in the design of experiments. Most statisticians would try to carry out these concepts in designing experiments of this degree of complexity. Curnow's paper (Reference 3) advances no new principles but makes the contribution of showing that McIntyre's analysis of a complicated example was incomplete. The total variation was not broken down into all of its component parts. This point is highly relevant and in designing this series of two-phase experiments for sensor ANZ I have aimed at securing information on the interesting components of the total variation.

Thus, in following McIntyre's and Curnow's lead it seemed desirable to set up the structure of these two-phase experiments so that:

a. The numerical results of the Second Phase can be subjected to a direct statistical analysis in terms of the design of the First Phase experiment,

b. Analysis will provide unbiased estimates of the sources of variation in both the First Phase and the Second Phase experiments,

c. Estimates of experimental error can be obtained for comparing interpreters,

d. Estimates of the repeatability of photo interpreters in reading the same film may be computed.

It was not found practicable, however, to set up the structure of all 7 experiments so that a, b, c, and d could be completed. As a minimum, item a can be carried out for all although in some cases a modified analysis has been set up. Items c and d were built into some of the designs.

Essentially, the material presented so far has been an extended introduction. We now come to the actual presentation of the experimental designs as they were worked out. In this paper I shall try to describe several of these structures. Experiment A in the First Phase is a randomized complete block in two replicates for three factors each at two levels in factorial arrangement on the whole plots of the experiment. In addition, a split-plot feature was introduced with one factor at four levels. Thus, this experiment is designed to produce 64 film strips from 16 aircraft missions. If the film strips provided data for immediate analysis, the schematic analysis of variance shown in Table 3 could be

obtained. A review of the symbols in this table will help us understand the nature of the experiment. V_a, is the speed of the aircraft carrying the sensor, R_s is a range control setting of the sensor itself, and D_t is the density of the moving targets on a segment of highway to be observed. These three provide the three factors each at two levels to be imposed in their eight treatment combinations on the whole plots. It should be noted that the factor D_t could not be randomized since ordinary vehicular traffic in and out of a military installation was to be observed at different times of the day when high and low densities would occur. The split-plot feature is the Flight Path treatment which actually gives four sensor-target aspect angles by means of the aircraft carrying the sensor flying at four different angles with the road segment to be observed.

With this pattern for the First Phase of experiment A, it was not easy to set up a good pattern for the Second Phase. For these photo interpretation experiments it was assumed that a minimum of four interpreters or interpreter teams consisting of two men working together would always be available. This is not exactly a modest assumption when one is dealing with military personnel who may be absent on any given day for one of a myriad of reasons. Since it was desired to obtain considerable information on film interpretation from this first experiment choice of an experimental pattern was approached from this point of view. The number four for interpreters seemed to suggest some form of Latin square arrangement. A single 4×4 square would permit the reading of only 16 film strips. Thus, four such Latin squares would be required as a minimum. How to distribute the 64 films among the four squares was a real problem. After some struggle, a pattern was obtained, but its analysis is not in accord with the analysis of Table 3. The films from one replicate of the First Phase experiment were artitrarily separated into two groups and each group assigned to one 4×4 square. The arbitrary division was made by confounding the three-factor or second order interaction with square. If an entire square is thought of as a block, the procedure for making this sub-division is described in detail by Professor O. Kempthorne (Reference 5) and I shall not go into it here.

In the Photo Interpretation Section a set of four stations would need to be set up for carrying on the film interpretation work. Then the interpreter teams are randomly assigned to the stations (columns of the Latin Square) and the four whole-plot treatment combinations to the rows. Within the cells of each square the split-plot treatments, i.e., the four

TABLE 3

ANALYSIS OF VARIANCE FOR OBSERVATIONS
ON UNCONTROLLED VEHICLE MOVEMENTS

Source of Variation	Degrees of Freedom
TOTAL	64
General Mean	1
Replications	1
V_a	1
R_s	1
D_t	1
$V_a \times R_s$	1
$V_a \times D_t$	1
$R_s \times D_t$	1
$V_a \times R_s \times D_t$	1
Experimental Error (a)	7
Flight Aspect Angles, L	3
Flight Aspects x Main Plot Treatments	21
Error (b)	24

Flight Paths, were then randomized. Finally, one of the four squares was duplicated, or a total of five squares used, in order to obtain a measure of repeatability of the same observer reading the same film strip twice.

Anyone of the five 4 x 4 Latin Squares could be analyzed, of course, by itself but this would provide only information for comparing Flight Paths. The two duplicate squares could be combined to give the analysis presented in Table 4. By omitting data for one of the duplicate squares an analysis of data read from the 64 film strips may be made as shown in Table 5.

After struggling with the complexity of experiment A we turn to a smaller experiment designed to test a simple hypothesis, "What is the effect of aircraft altitude on the output of the airborne sensor?" It was believed that there might be an optimal altitude for employment of the sensor at each of its range settings. Eight altitudes were selected for study and these have been arranged in two randomized complete blocks with R_s (the range setting) completely confounded with blocks since there was no interest in the interaction of altitude with R_s. Such an arrangement does not provide a proper estimate of experimental error so four of the eight altitudes in each block were duplicated. In a sense then, two estimates of error are provided by the selected pattern, an estimate based on deviations from regression and an estimate based upon the duplicates (refer Table 6, below).

The introduction of the duplicates makes each block rather large, i.e., 12 passes or flights over the same flight path are required to be completed within one aircraft mission (mission = take-off, flying the required passes and returning for landing). With experimental equipment and modified aircraft, it is often difficult to secure a desired set of successful passes in one mission. For this simple experiment B , it was assumed that the 12 passes could be obtained during one aircraft mission.

For this experiment B a convoy of nine military target vehicles was to operate on a selected straight road segment at a specified fixed speed and fixed spacing. Aircraft speed, V_a, was also to be controlled with only the altitude to be varied. The altitudes selected were 500, 1000, 1500, 2000, 2500, 3000, 4000 and 5000 feet above the ground level. It may be asked, "Why this peculiar choice of unequal spacings?"

TABLE 4

ANALYSIS OF VARIANCE OF TWO LATIN SQUARES FOR EXPERIMENT A
(Analysis of Duplicate Film Interpretations)

Source of Variation	Degrees of Freedom
Total	32
Mean	1
Squares	1
Stations (Columns)	3 (Equals interpreters)
Rows	3 (Sub-set of treatments)
Flight Paths	3
Remainder	6
Duplicates	15

TABLE 5

ANALYSIS OF VARIANCE OF FIRST AND SECOND PHASE PATTERNS FOR EXPERIMENT A

Source of Variation	Degrees of Freedom	
Total	64	
Mean	1	
Sessions (based on 4 Latin Squares)	(3)	
Blocks of Phase 1 Experiment		1
Squares within Blocks		2*
Stations	3	
Interpreters	3	
Remainder (for comparing interpreters)	6	
Rows in Sessions	(12)	
Treatment Combinations		6
Treatment Combinations by Blocks		6
Columns in Sessions	(12)	
Interpreters (given above)		
Remainder (confounded with other effects)		
Flight Paths	3	
Flight Paths by Blocks	3	
Flight Paths by Sessions in Blocks	6	
Remainder (within square error for Flight Paths)	24	

*Includes the three-factor interaction used for forming the squares by dividing the main treatment combinations into two groups.

TABLE 6

ANALYSIS FOR EXPERIMENT B

INVESTIGATION OF ALTITUDE EFFECT FOR EACH RANGE SETTING*

Source of Variation	Degrees of Freedom
Total	24
Mean	1
R_s (equals block)	1
Altitudes	7
500' Spacings--Linear	1
Quad.	1
Rem.	3
1000' Spacings--Linear	1
Quad.	1
Rem.	2
Altitudes by Range Setting*	7
Duplicates in Range Settings	8

*Alternate Analysis

If Altitudes by Range Settings is a large effect (which is to be expected) then the spacings analyses should be carried out separately for each Range Setting.

TABLE 7

ANALYSIS OF VARIANCE FOR STUDY OF INTERPRETER TEAM VARIABILITY USING EXPERIMENT B FILMS

Source of Variation	Degrees of Freedom
Total	32
Mean	1
Squares	1
Altitudes	3*
Altitudes by Squares	3
Rows and Rows by Squares	6
Interpreter Teams	3
I. Teams by Squares	3**
Exp'tal Error in Squares	12**

* May be separated into Linear Quadratic and Remainder effects

** May be combined for assessing Interpreter Teams.

Again, these choices are a compromise. You will note that the set of eight altitudes contains two sequences, one at 500' spacings and one at 1000' spacings. The closer spacing was considered relevant to one range setting and the wider spacing for the other range setting although some information on the complete range of altitudes was desired for both range settings. The altitudes selected for duplication were 1000, 2000, 3000, and 4000.

Analysis of the 24 film strips to be obtained from experiment B would be as set forth in Table 6 if no Second Phase experiment were required. For the Second Phase of B two designs were suggested. The simpler pattern merely assigned one photo interpreter team to each block with film interpretation to be carried out in the same order as the randomized field arrangement of the First Phase. Thus, interpreters are completely confounded with the blocks and range settings. Analysis would then follow exactly the scheme of Table 6.

The purpose of the additional design was to obtain more information on interpreter performance. The four altitudes duplicated in each block make it convenient to employ two 4 x 4 Latin Squares and four teams. Altitudes are randomly assigned to the columns in each square. Rows are formed by arbitrarily setting up two groups of four films containing each of the four altitudes within each block of the First Phase. The total of four such groups are randomly assigned to rows in each square. The interpreter teams are assigned to the cells in each square by a separate randomization. Analysis of this added design then would follow the scheme of Table 7. I do not consider this last pattern a wholly satisfactory design but it does provide more information on photo interpreters. At present I am not aware that adequate statistical studies have been made of the photo interpretation process.[7] Even if such studies were available, I should note that film strips produced by a sensor such as ANZ are sufficiently different from the usual aerial photographs interpreted in a Photo Interpretation Section that some studies of the interpretation of these film strips should be conducted.

It will be of interest to take a look at another of the experiments in the series. Our experiment C has some interesting features. First, the Phase One experiment is a fractional plan. The pattern is a one-half replicate of a 2^6 factorial arranged in four blocks, each block comprising eight passes by the sensor aircraft at a constant altitude along flight paths described as near (in range) to a road segment to be observed. Within each block requires 10 passes and the whole experiment a total of 40 passes. This pattern may seem like a rather large experiment. It does require four days, one day for each block, but it is expected that only four successful aircraft missions

will be required. In terms of the amount of information to be obtained from the experiment, the plan does not seem large to the statistician.

The complexity of the First Phase plan and the desire to keep the Second Phase program from becoming too large led to selection of a simplified film interpretation schedule. Each block of the First Phase experiment was assigned to one photo interpretation team so that teams are confounded with blocks. Such a program still permits analysis of the interpretations directly in terms of the First Phase design. For convenience, two interpretation schedules were set up, one for the one-half replicate of the 2^6 factorial, and one for a sub-set of the experimental units in each block not involving the near range setting of the sensor. This last schedule uses film strips obtained using the far flight paths already mentioned and two additional passes from each block. The total of sixteen treatment combinations from the four blocks provide data for securing partial information on near vs. far flight paths and two other factors.

Analysis of the fractional replicate is presented in Table 8. It is seen that the six factors are Target Speed, Target Spacing, Aircraft Speed, Sensor Range Setting, and Aircraft-Vehicle Path Angle which make only five, but the last factor occurs at four levels so it is sub-divided into 2 factors each at two levels (Refer 6, Sec. 6A 33, pp. 273-4 for explanation of this procedure). Thus, E and F are designated the pseudo-factors for the single factor at four levels.

For this experiment C, the aircraft with sensor is to be flown at a constant altitude and the targets consist of a convoy of nine military vehicles whose speed and spacing is controlled. Since the convoy is restricted to a single straight road segment, the aircraft Flight Paths are varied with respect to the road.

For the second interpretation schedule, analysis of results is outlined in Table 9. Again, interpreters are confounded with blocks and block assignment for the teams is the same as for the fractional replicate. How Table 9 is obtained is not immediately obvious. It was actually worked out from first principles after careful choice of the treatment combinations on the far flight paths to be added to each block. A larger First Phase experiment was not considered feasible so the information obtained here is a sort of minimum compromise.

TABLE 8

PRIMARY ANALYSIS OF VARIANCE FOR A SINGLE RESPONSE FROM EXPERIMENT C INVOLVING CONTROLLED VEHICLE MOVEMENTS

Source of Variation	Degrees of Freedom	
Total	32	
Mean	1	
Blocks (includes BC)	3	
Factors (main effects)	7*	
A Target Speed		1
B Range Setting (of the sensor)		1
C A/C Speed		1
D Target Spacing		1
E&F A/C-Vehicle Path Angle		3*
Two-factor Interactions	13**	
Remainder	8	

*Explanation of the Pseudo-factors E and F:

	Flight Path Angle			
Factor	0°	30°	60°	90°
	Factor Level			
E	E_1	E_2	E_1	E_2
F	F_1	F_1	F_2	F_2
Contrast	Contrast Vector (Here + and - indicate +1 and -1)			
C(E)	-	+	-	+
C(F)	-	-	+	+
C(EF)	+	-	-	+

**There are 15 degrees of freedom for the two-factor interactions but BC is confounded with Blocks and EF is a part of the Vehicle Path Angles effect; hence, only 13 remain here.

It may be noted from what I have just described that some of the film strips, eight in total, will be interpreted twice. That is, each team will have made duplicate interpretations of two films. These data can be analyzed as shown in Table 10. Thus, a little more information is secured on the film interpretation process.

It would be of further interest to describe the two-phase experiments set up to investigate the effect of foliage on the film strips produced by the sensor and the comparison of the sensor film strips with information furnished by standard aerial photographs over different types of terrain conditions. For these experiments some special procedures had to be worked out for the film interpretation. In order to limit this paper no details on these experiments are given at this time.

It is feared that the account given thus far is rather sketchy. There are too many things that have been glossed over, not mentioned or inadequately conveyed to the reader. The reader can understand, however, that a great many details on organization of the field test program, instrumentation of test sites and road segments, programs of the aircraft flights and vehicle convoys, film interpretation schedules and many other aspects of experiments of this type have not been mentioned at all or merely noted. All this office work and planning and much more is required to complete experiments of this kind. Even so, revisions may have to be made after the field program and/or film interpretation work gets under way. If the First Phase and Second Phase experiments and associated procedures have been carefully worked out in advance of any testing, then it will be clear what effects necessary changes or modifications will have and how they should be made.

No discussion of the power or the sensitivity of these experiments for detecting differences of specified magnitudes has been given. Nor are any educated guesses available on the magnitudes of the experimental errors that will be obtained from experiments that have been described. The seven experiments are considered to be minimal in size for their intended objectives. Only two of the experiments may be considered really large, A, producing 64 film strips and C, a total of 40 film strips. Until experimental results from such trials become available, it has not seemed worthwhile to give further attention to the sensitivity of the experiments. In development work it may often be best to complete a number of small experiments that provide some information and possible guides to further investigation.

TABLE 9

SECONDARY ANALYSIS OF VARIANCE FOR A SINGLE RESPONSE FROM EXPERIMENT C INVOLVING CONTROLLED VEHICLE MOVEMENTS FOR STUDY OF NEAR VS. FAR FLIGHT PATH AT THE FAR RANGE SETTING

Source of Variation	Degrees of Freedom
Total	16
Mean*	1
Blocks	3
Factors* (4)	
G Near vs. Far Flight Path	1
A Target Speed	1
D Target Spacings	1
E and F (combined): Vehicle Path Angles	1*
Interactions (3)	
GA	1
GD	1
AD	1
Error Components (5)**	
G by Blocks	3
A by Blocks	2***

*C is combined with blocks E and F confounded or not separable.

B at the upper level equals the mean vector.

**These two components may be combined to form a five degree of freedom error team.

***One of the A by blocks and all three of the D by blocks are not orthogonal to E and F.

TABLE 10

ANALYSIS OF VARIANCE FOR INTERPRETER PERFORMANCE ON A SUB-SET OF EXPERIMENT C RESULTS

Source of Variation	Degrees of Freedom
Total	16
Mean	1
Blocks (equals Interpreters)	3
Trial 1 vs. Trial 2	1
Trial by Interpreters	3
Duplicates	8

In closing, appreciation is expressed for this opportunity to present the subject of two-phase experiments. Some principles for the design of these combined patterns have been reviewed and problems encountered in setting up a series of such experiments have been described. Comments, criticisms and suggestions from readers will be most welcome.

REFERENCES

1. G. A. McIntyre, "Design and Analysis of Two-Phase Experiments," Biometrics 11, 324 (1955).

2. G. A. McIntyre, "Query 123", Biometrics 12, 527 (1956).

3. R. N. Curnow, "The Analysis of a Two-Phase Experiment", Biometrics 15, 60 (1959).

4. E. H. Jebe, et. al., "How to Study Consumer Satisfaction with Children's Clothing Made of Different Textile Fibers", Journal Home Economics, 50, No. 3, 213.

5. O. Kempthorne, "The Design and Analysis of Experiments", New York: J. Wiley and Sons, (1952).

6. W. G. Cochran and G. M. Cox, "Experimental Designs", New York: J. Wiley and Sons, (1957 , Second Edition).

7. W. T. Pollock, "Photo Interpreter Performance with Different Conditions of Tactical Aerial Photography" (U), Report Number 2900-9-T, The University of Michigan, Willow Run Laboratories, Report of Project MICHIGAN, May 1959. This report describes a large photo interpretation experiment, but no analysis of results in terms of sources of variability is presented. This report includes one reference in the area of interpretation of aerial photographs.

FITTING THE MODIFIED EXPONENTIAL FUNCTION BY THE METHODS OF MULTIPLE REGRESSION

Willis L. Hasty
U. S. Army Chemical Corps
Fort Detrick, Frederick, Maryland

The function

$$(1) \qquad y = A\left[1 - e^{-B(X+C)}\right]$$

is variously known as the modified exponential function, Mitscherlich's equation and the law of diminishing returns. It has a rather long history as a proposed model in economics and agronomy, and has been a source of frustration to the applied mathematician who was responsible for estimating its parameters by a least-squares fitting to data.

The obvious source of difficulty is the non-linearity of (1) in its parameters and the consequent non-linearity of the system of normal equations of least-squares. This situation cannot be avoided, as it can in the case of the simple exponential function, by the logarithmic transformation. An early modification did, indeed, admit the logarithmic transformation: one made an independent estimate of A and fitted the function

$$(2) \qquad -\ln (1 - y/A) = Bx + C.$$

This, however, requires the exclusion of observations of greater magnitude than A.

The general problem of fitting sums of exponentials has been discussed extensively; in this note, however, we limit ourselves to consideration of (1) and to a method of fitting which uses the technique of multiple regression analysis.

If (1) is written in the form $y = A - be^{-cx}$ and expanded in Maclaurin's series, terminating with the kth power of x, we have

$$y = A - b + bc\,x - bc^2x^2/2! + \ldots + (-1)^{k+1}\,bc^kx^k/k! \,. \tag{3}$$

By identifying x^i with x_i and proceeding as in multiple regression analysis, we can obtain least-squares estimates of the regression coefficients b_i $(i = 0, 1, \ldots, k)$ which can be identified with the parameters in (1) as follows:

$$b_0 = A - b = A(1-e^{-BC})$$

$$b_1 = bc = ABe^{-BC} \tag{4}$$

$$\vdots$$

$$b_k = (-1)^{k+1}bc^k/k! = (-1)^{k+1}AB^ke^{-BC}/k!.$$

It appears that the ratios of regression coefficients give estimates of powers of B according to the expression

$$b_j/b_i = (-1)^{j-i}(i!/j!)\,B^{j-i} \qquad (i < j,\ j = 2, 3, \ldots, k) \,. \tag{5}$$

Two difficulties stand in the way of using (5):

a. The b's are in general correlated and have unequal variances.
b. We have estimates of powers of B, whereas we require B itself.

The second of these obstacles does not in itself appear insurmountable. We have the relationship

$$E(x^n) = \mu_n - \sum_1^n (-1)^i \binom{n}{i} \alpha_1^i E(x^{n-i})$$

(6) $$\alpha_1 = E(x)$$

$$\mu_k = E(x - \alpha_1)^k$$

which, used recursively, allows us to exhibit $E(x^n)$ as a polynomial in α_1, with the central moments figuring in the coefficients. The $k + 1$ regression coefficients generate $k(k - 1)/2$ estimates of the powers of B, so there are $k(k - 1)/2$ equations and $k - 1$ quantities to be determined (α_1 and the μ_i, $i = 2, 3, \ldots, k - 1$).

Thus the system of equations for determining α_1, (and the μ_i) is overdetermined and the equations of the system are non-linear and do not have equal weights.

These difficulties may be eliminated by forming the weighted sums of the groups of equations which estimate the same power of α_1 ($= B$). For example α_1 is estimated by each of the $(k - 1)$ ratios $-ib_i/b_{i-1}$, $i = 2, \ldots, k$. So the weighted sum provides a single estimate of α_1:

$$\alpha_1 = -\sum_2^k iW_i \; b_i/b_i = 1$$

where the W_i's are functions of the variances and covariances of the b_{i-1} and $\sum_2^k iW_i = 1$. Continuing this procedure with the $k - 2$ equations which estimate $\mu_2 + \alpha_1^2$, and so on, we have finally the system shown below, in which there are $k - 1$ equation and $k - 1$ unknown quanties: α_1 and the μ_i, $i = 2, 3, \ldots, k - 1$. The first

column shows the b's from which the equations in the second column are derived:

$$b_i,\ b_{i-1} \qquad \alpha_1 = -\sum_{2}^{k} i \cdot b_i W_i / b_{i-1}$$

$$i = 2, \ldots, k$$

(7)
$$b_i,\ b_{i-2} \qquad \mu_2 + \alpha_1^2 = \sum_{3}^{k} i\,(i-1) W_i\ b_i / b_{i-2}$$

$$i = 3, \ldots, k$$

$$\vdots$$

$$b_k,\ b_1 \qquad \mu_{k-1} + \sum_{1}^{k-1} (-1)^{k-1} \binom{k-1}{i} \mu_{k-i-1} \alpha_1^i = (-1)^{k-1} k!\ W_k\ b_k / b_1.$$

The system is such that α_1, and the μ's can be recovered in succession: the first equation gives α_1, whereupon the second gives μ_2 and so on. The value of α_1, as determined from the first equation, is an estimate of B in (1). This estimate permits (1) to be linearized in the form

(8) $$y = A - DZ$$

where

$$D = Ae^{-BC}, \quad Z = e^{-BX}$$

The least-squares solution of this equation gives A, and $C = (1/B) \ln D/A$.

The μ's, which are available from the last $k - 2$ equations of (7), are formally identified with the central moments of the distribution of B. No investigation of the properties of these estimates has been made.

The weights appearing in (7) are to be determined such that the variance of the weighted sum of estimates of B shall be a minimum. Let $b_i/b_{i-1} = r_i$; denote a Lagrange multiplier by λ and the mean squared deviation from (3) by S^2. Then the function

$$(9)\quad \sum_{2}^{k} i^2 W_i \operatorname{var}(r_i) + 2\sum_{i=2}^{k-1} iW_i \sum_{j=i+1}^{k} jWj \operatorname{cov}(r_i, r_j) + \lambda\left(S^2 \sum_{2}^{k} iW_i - S^2\right)$$

is to be minimized with respect to W_i.

The system obtained by setting the partial derivatives equal to zero is

$$\begin{array}{ccccccccc}
\underline{W_2} & & \underline{W_3} & & \underline{W_4} & & \underline{W_k} & & \underline{\lambda} \\
4V_2 & + & 6C_{23} & + & 8C_{24} & +\ldots+ & 2KC_{2K} & + & 2 = 0 \\
6C_{23} & + & 9V_3 & + & 12C_{34} & +\ldots+ & 3KC_{3K} & + & 3 = 0 \\
 & & \vdots & & & & & & \\
2KC_{2K} & + & 3KC_{3K} & + & 4KC_{4K} & +\ldots+ & K^2V_K & + & K = 0 \\
2 & + & 3 & + & 4 & +\ldots+ & K & & = 1
\end{array} \qquad (10)$$

where $V_i = \operatorname{var}(b_i/b_{i-1})$, $C_{ij} = \operatorname{cov}(b_i/b_{i-1}, b_j/b_{j-1})$.

Using a well-known approximation for the variance of the ratio of random variables

(11)
$$V_i = \frac{b_i}{b_{i-1}}\left[\frac{c_{ii}}{b_i^2} + \frac{c_{i-1,\,i-1}}{b_{i-1}^2} - \frac{2c_{i,\,i-1}}{b_i\, b_{i-1}}\right] \cdot s^2$$

where c_{ii} is the element in the ith row and ith column of the inverse of the matrix of the system of normal equations of least-squares corresponding to (3). In a similar manner one may develop the expression for the covariance of two ratios b_i/b_{i-1}, b_j/b_{j-1}:

(12)
$$C_{ij} = \frac{b_i b_j}{b_{i-1} b_{j-1}}\left[\frac{c_{ij}}{b_i b_j} - \frac{c_{i-1,j}}{b_{i-1} b_j} - \frac{c_{i,j-1}}{b_i b_{j-1}} + \frac{c_{i-1,j-1}}{b_{i-1} b_{j-1}}\right] \cdot s^2$$

It seems clear that the applicability of this method would be limited to cases in which a large body of data was available for the estimation of the parameters. This, however, is the case of interest here, since the use of an electronic digital computer is assumed.

PROBLEMS RELATED TO A BIO-ASSAY FOR SPORE-GERMINATION INHIBITORS ASSOCIATED WITH UREDOSPORES

K. R. Bromfield and Marian W. Jones
U. S. Army Chemical Corps Biological Laboratories
Fort Detrick, Frederick, Maryland

I. INTRODUCTION. The uredospores of Puccinia graminis tritici, the organism causing stem rust of wheat, have associated with them a water-soluble substance that at certain concentrations inhibits their germination. Since the chemical identity of the inhibitor has not yet been established, we are compelled to rely solely on a bio-assay (the response of test indicator spores) to indicate the presence and relative amount of inhibitor.

The bio-assay procedure is as follows: Two hundred mg of uredospores are weighed out and agitated for 5 minutes in 20 ml of distilled water in a Potter-type tissue-homogenizer. The liquid containing the inhibitor (crude extract) is filtered from the spores and diluted with distilled water to give a series of concentrations. Each portion of diluted crude extract is mixed with an equal volume of liquefied 2% water agar, distributed among 3 micro-dishes, and permitted to gel.

The solidified test substrates are then inoculated with "indicator spores" and placed under standardized conditions conducive to spore germination.

At the conclusion of the germination period the indicator spores are exposed to formaldehyde to prevent further development. Individual plates are then observed under the microscope and the number of germinating indicator spores in each of 2 groups of 100 spores is recorded.

With the procedure outlined briefly above we are attempting to attain the following objectives:

1. Find the relationship between inhibitor concentration and per cent inhibition of indicator spores.

2. Make quantitative estimates of inhibitor content of various spore lots.

3. Generate variances of estimates of inhibitor content.

4. Investigate the behavior of estimates of inhibitor content in response to various experimental treatments.

II. GENERAL METHOD OF ANALYSIS. The tests to be discussed were made to determine assay reproducibility. Test substrates were prepared from random spore samples from the same lot of "donor spores" and care was taken to use comparable "indicator spores."

In preliminary experiments it was found that relative concentrations of crude extract in the range from 0.1 to 3.0 covered the effective range of per cent germination of indicator spores. The relationship had the general form shown in Figure 1.

However, because the inhibitor effect is the factor of interest, it is desirable to measure response in terms of per cent inhibition rather than per cent germination. For each level of inhibitor concentration, per cent inhibition is computed as follows:

$$\% \text{ Inhibition} = 100 \left(1 - \frac{\% \text{ germination of Indicator Spores on test substrate}}{\% \text{ germination of Indicator Spores on control plates}} \right).$$

The search for a linearizing transformation of the response of per cent inhibition to relative concentration led to the tentative conclusion that the log-probit transformations meet the requirement. It will be noted that per cent inhibition is identical with Finney's adjustment for natural mortality using Abbott's formula. Therefore the usual probit with adjustments for natural mortality seems appropriate. See Figure 2.

If this method of analysis is valid it will 1) establish the relationship between inhibitor concentration and per cent inhibition of indicator spores, 2) permit quantitative estimates of inhibitor content to be made from parameters of the probit regression, 3) permit the calculation of internal and between-trial variances of parameters, and 4) permit treatment comparisons to be made using one or more of the parameters as an index of inhibitor content.

III. PROBLEMS ENCOUNTERED. To date three major problems have arisen in using the log-probit method of analysis.

The first is concerned with the appropriateness of the model. Are we justified in assuming a linear model in the transformed data over the entire range ?

Plots resulting from the current series of tests are shown in Figure 3. Plots 2, 5, 7, and 8 hint that curvature still exists in the transformed data. In all plots except 3 and 5, responses at the lowest concentration lie above the probit line. It is tentatively assumed that this deviation may be associated with some lack of precision in the laboratory procedure employed for obtaining low doses.

The second problem concerns the selection of a parameter indicating the best quantitative measure of inhibition. Initially the ED_{50} seemed to be the logical parameter to use as an index of inhibition. However, this parameter is positively associated with per cent germination of indicator spores on control plates as shown in Figure 4. If this association is real how can we then make treatment comparisons among tests conducted on different days ? Would some other parameter, perhaps slope, be a better index?

The last, and perhaps most vexing problem, is locally termed the question of the "bouncing" controls. Although care is taken to use comparable "indicator spores" in a given series of tests and to maintain standard germination procedures, the level of germination of indicator spores on the control plates varies more from day to day than one would expect from sampling variation alone. Apparently some uncontrolled, and as yet unidentified, factor is operative.

The necessity for a control value with stability can be appreciated from examination of the expression for percentage inhibition given previously. Any factor operating to depress or increase the check values could insert a bias that might be mistaken for experimental error. With low control values, especially, the adjustment currently used seems to have the effect of overestimating percentage inhibition . Specifically then, would some other adjustment for check values be more accurate?

IV. SUMMARY. Although work is underway on the chemical characterization of the germination inhibitor, its identity has not yet been established. When the chemical is identified we will then be able to assay for inhibitor by chemical techniques. Until that time, however, we must rely on the behavior of indicator spores to indicate the presence and relative amount of inhibitor.

Per cent germination as a function of concentration of a water-soluble germination-inhibitor obtained from uredospores.

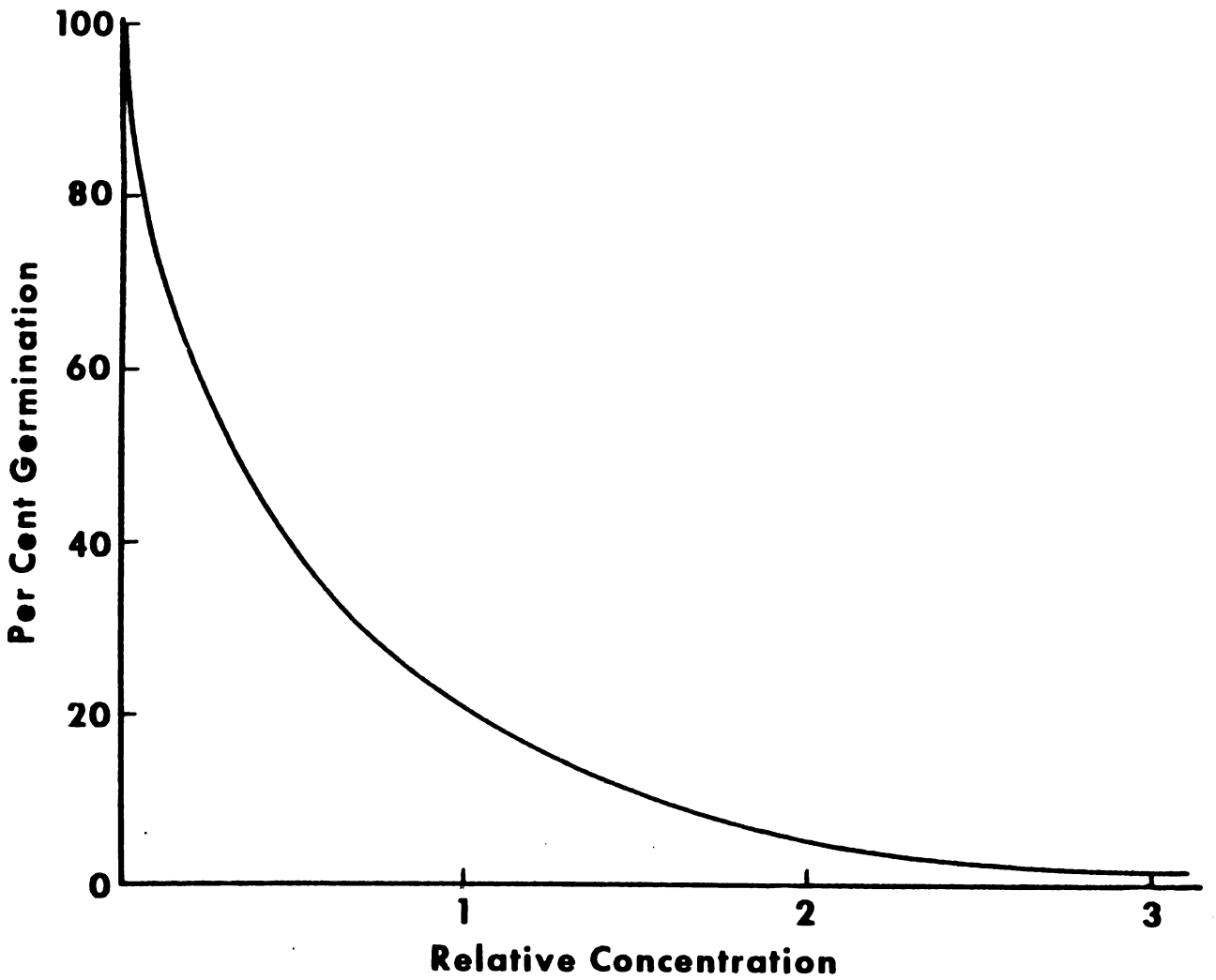


Figure 1

Probit of per cent inhibition of spore germination as a function of inhibitor concentration.

Figure 2

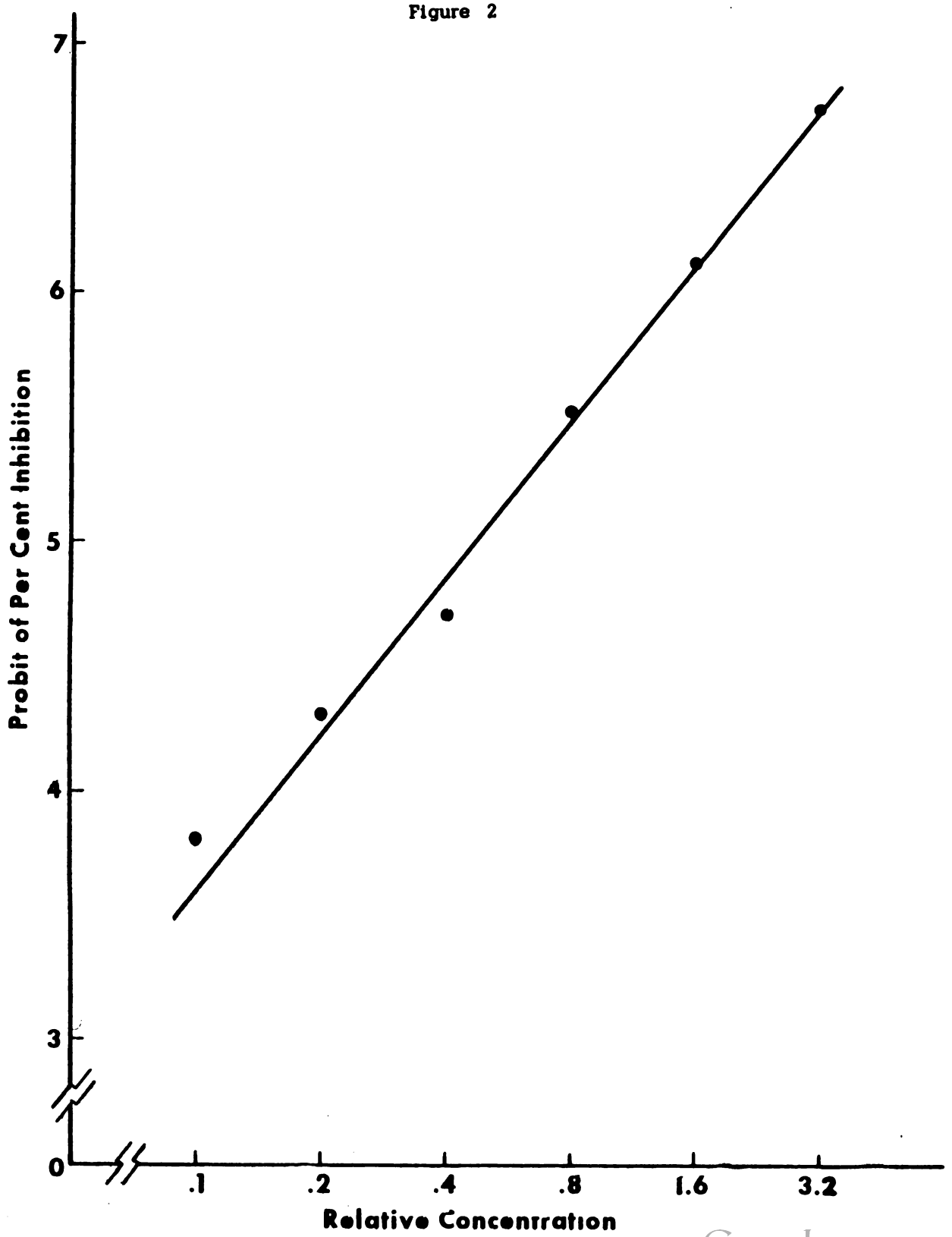
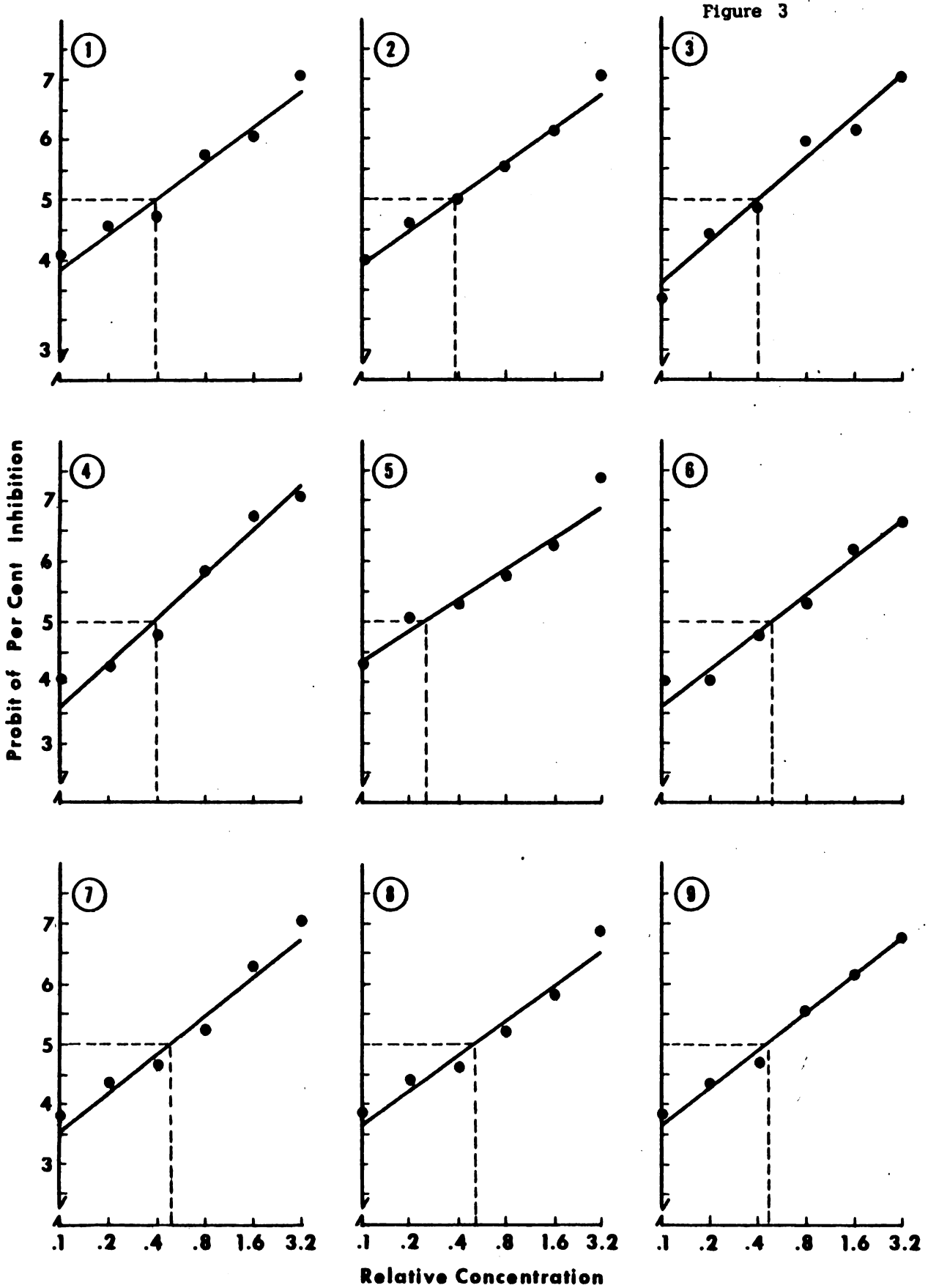


Figure 3



Comparison of 9 experiments. Probit of per cent inhibition of spore germination as a function of inhibitor concentration.

Regression of ED_{50} on per cent germination of spores on check plates.

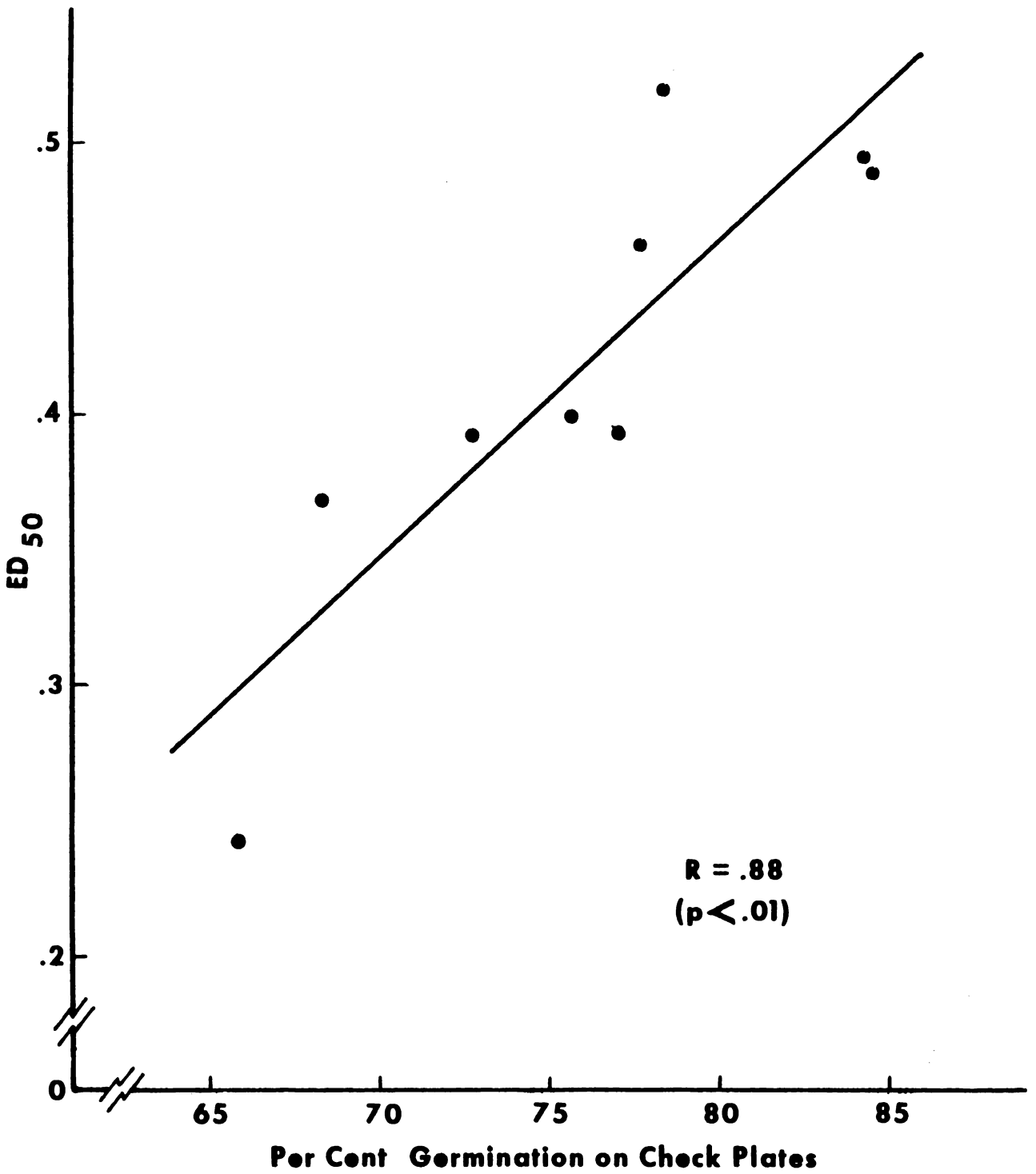


Figure 4

The present technique is to germinate indicator spores on agar substrates containing a range of concentrations of crude inhibitor extracted from donor spores. The per cent inhibition of germination of indicator spores is calculated and plotted over relative concentration of crude inhibitor extract to give a dosage-response curve. A log-probit transformation appears to linearize the response of per cent inhibition to relative concentration.

The following questions or problems have arisen and must be answered before we can rely on the bio-assay procedure:

1. Is the log-probit transformation appropriate?

2. Is the ED_{50} the best parameter to use as an index of inhibition?

3. What adjustment can be made for day-to-day variations in level of germination of indicator spores on control plates?

DESIGNS FOR ESTIMATING VARIANCE COMPONENTS

R. L. Anderson
Institute of Statistics
North Carolina State College

1. INTRODUCTION. The science of statistics is essentially the application of the mathematical theory of probability to the study of variation in experimental and operational data. In a paper prepared for the 1958 meeting of the International Statistical Institute in Brussels, I discussed the "Uses of Variance Component Analysis in the Interpretation of Biological Experiments." Some of this discussion is repeated here.

The identification of the sources of variation and development of methods of estimating the separate variances and testing hypotheses concerning them is one of the statistician's important tasks. In most experiments, some variation is imposed by the experimenter, in the form of different treatments, varieties, or practices. Often inferences are desired concerning only the particular treatments used in the experiment; it has been the custom to label these as fixed sources of variability. In other cases, it is assumed that the variability among the sources used in the experiment is representative of the variability in a much larger population; these are usually labeled as random sources of variability. Admittedly the particular representatives used in the experiment sometimes are not selected in an accepted random manner; however, the experimenter can often justify the assumption that order of selection is of little importance. Most experiments have a combination of fixed and random sources of variability; the mixed category. Often an experimenter will want to examine a given variable from both the random and fixed point of views.

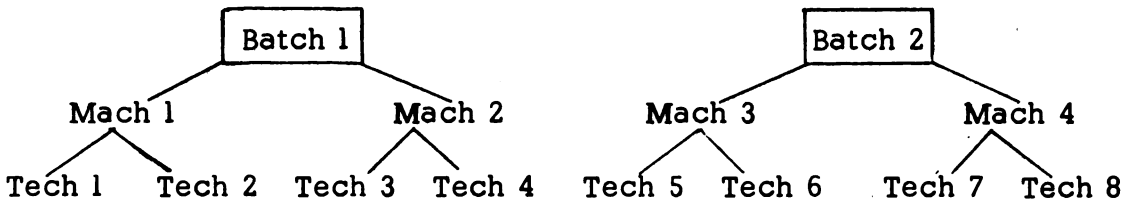
In this paper I will discuss only situations in which all sources of variation are essentially random. Some examples are presented on the next page. There are two experimental and operational procedures to be considered: the nested or hierarchical type and the classification type. To illustrate the basic differences between these two types, suppose we have a large number (b) of batches of some material. These are to be processed by one of a large number of more or less identical machines. Suppose material from the first batch is subdivided into m_1 sub-batches and each sub-batch processed by a different machine; the second batch is processed by m_2 machines, all different from the first m_1; and so forth for all batches. Hence there are $m = \sum_{i=1}^{b} m_i$ machines used

for processing. The processed material from each machine is tested, assayed or inspected by a corps of technicians, a different corps for each machine. The material is subdivided into samples, each sample handled by one technician. Let us assume there are t_{ij} technicians for the j^{th} machine processing the i^{th} batch, so that the total number of technicians is $t = \sum_{i,j} t_{ij}$. In the following example of two batches, we have assumed that each $m_i = 2$ and each $t_{ij} = 2$.

Now suppose that there are only, say 4 processing machines and, say 2 technicians at a given plant, but that these machines and technicians are "representative" of large populations of machines and technicians. Let us further assume that every batch is subdivided into 4 sub-batches, one for each machine. Then the product from each machine is divided into 2 parts, one part for each technician. This operation for two batches is also exemplified on the following page.

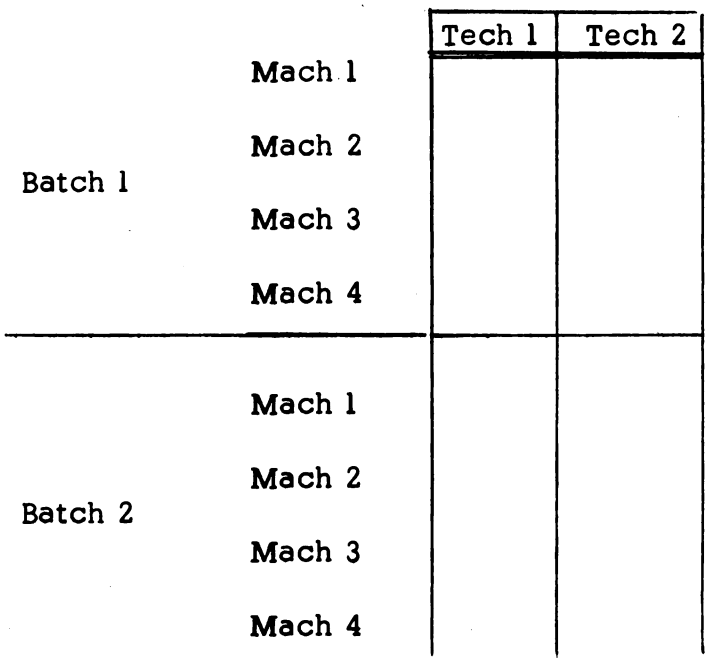
In general, the first operation will be a combination of the two types, since subsequent batches will be allocated to machines 1 and 2, 3 and 4, etc.

EXAMPLES OF RANDOM OPERATIONS



Nested or Hierarchical-Type Operation

* *



Classification-Type Operation

* *

Streptomycin Experiment

Initial Incubation → Primary Inoculation → Secondary Inoculation → Fermentation → Assay

* *

Rubber Sampling

Producer's Estate → Day at Estate → Bale on Given Day → Sheet of Rubber from Bale → Sample from Sheet

A better example of the strictly nested-type operation might be one I used for the International Statistical Institute paper: A pilot study was considered to assess the various sources of variability in the production and assay for streptomycin before conducting an experiment on the efficacy of various molds. There were five stages in this process: an initial incubation in a test tube; a primary inoculation period in a petrie dish; a secondary inoculation period in another petrie dish; a fermentation period in a bath; and the final assay of the amount of streptomycin produced. Newton, et al. (1951) in a study of the variability of rubber considered the following sources of variation: producer's estates, days at a given estate, bales on a given day, sheets from a given bale, and samples from each sheet, Technicians then took measurements on the samples. In this paper I shall discuss only purely nested and classification-type operations but the same methods are applicable to the combination.

Returning to our batch problem, the final measurement made by the technician is designated as the process yield, Y. We are here interested in the variability of the Y's. Such variation is a result of variable batches, variable batch sampling and subsequent machine processing and variable product sampling and measurement. The variation in machine processing results from machine-to-machine differences and failure of a given machine to perform in the same manner on different batches, and similarly for the variation in the measuring process. The failure of a given machine to perform similarly on different batches is called a machine-batch interaction and similarly for the technician-batch and technician-machine interactions. We are assuming that only two-factor interactions need be considered; however, one might also be interested in a technician-machine-batch interaction. Interactions can be detected only with the classification-type operation.

It is assumed that the sources of variability act independently and additively; hence, one can write the mathematical model for the final yield as follows:

$$Y_j = \mu + \sum_{i=1}^{V} e_{ij}, \tag{1}$$

where there are v sources of variability, the $\{e_{ij}\}$ assumed independently distributed random variables with zero means and constant variances, σ_i^2; hence, μ is the product average. The variance of the final produce is $\sigma^2 = \sum_{i=1}^{v} \sigma_i^2$. It is assumed that the experimenter or manufactorer is interested in reducing the magnitude of σ^2. In order to do this, a preliminary experiment is to be set up to estimate the individual variances, σ_i^2.

The estimates of the σ_i^2 will be designated as $\hat{\sigma}_i^2$. In studying the adequacy of these estimates, it will be assumed that the random variables (the e's) are normally distributed. Hence the observed yields are assumed normally distributed with the same mean μ and variance $\sigma^2 = \sum \sigma_i^2$. However, if there is more than one source of variation, the Y's (in general) will be correlated. Consider the nested process with 2 batches, 2 machines per batch and 2 technicians per machine. One might represent the measurements as follows:

$$Y_{111} = \mu + b_1 + m_1 + t_1, \quad Y_{112} = \mu + b_1 + m_1 + t_2,$$

(2)

$$Y_{123} = \mu + b_1 + m_2 + t_3, \ \ldots, \ Y_{248} = \mu + b_2 + m_4 + t_8.$$

The usual short cut notation is to renumber the technicians within each machine, the machines within each batch, etc,; i.e.,

$$Y_{111} = \mu + b_1 + m_{11} + t_{111}, \quad Y_{112} = \mu + b_1 + m_{11} + t_{112},$$

(3)

$$Y_{121} = \mu + b_1 + m_{12} + t_{121}, \ \ldots, \ Y_{222} = \mu + b_2 + m_{22} + t_{222}.$$

The latter notation is much simpler for complicated designs and will be used here. In any case, it is easy to see that the covariance between Y_{111} and Y_{112} is $\sigma_b^2 + \sigma_m^2$; between Y_{111} and Y_{121} is σ_b^2; and between Y_{111} and Y_{222} is 0. Hence it is possible to write the likelihood of the sample in the familiar multivariate normal form, with a constant mean (μ) and variance ($\sigma^2 = \sigma_b^2 + \sigma_m^2 + \sigma_t^2$) but with correlations of either $(\sigma_b^2 + \sigma_m^2)/\sigma^2$, σ_b^2/σ^2 or 0. The method of estimation would be the familiar maximum likelihood (ML).

The model for the classification-type process might be

$$(4) \quad Y_{ijk} = \mu + b_i + m_j + (bm)_{ij} + t_k + (bt)_{ik} + (mt)_{jk} + e_{ijk},$$

where e may be a three-factor interaction effect or technique error or a combination of the two. In this case we have variance components:

$$\sigma_b^2, \sigma_m^2, \sigma_{bm}^2, \sigma_t^2, \sigma_{bt}^2, \sigma_{mt}^2, \sigma_e^2.$$

Again if the components are assumed to be normally distributed, one can set up the normal multivariate likelihood function, which will be considerably more complicated than the nested one.

The usual estimating procedure for variance components is based on the familiar, I hope, Analysis of Variance table. For balanced designs, the analysis of variance (ANOV) and maximum likelihood (ML) estimators are essentially the same; I will not take the time to discuss this difference here. For non-balanced designs the ML equations are complicated functions of the estimators; iterative methods are usually needed to solve them. There is no unique ANOV estimating procedure for non-balanced designs. The usual procedure is based on a pooling procedure first presented by Ganguli (1941) and summarized in Anderson and Bancroft (1952). This procedure produces unbiased estimators, but of uncertain efficiency. The total sum of squares can be subdivided into orthogonal sources but

there remains the problem of how to best combine these sources. One probably would prefer unbiased estimators which are linear combinations of these mean squares in the ANOV table. Hence one solution to the estimation problem would be to select that unbiased linear estimator which has the smallest variance. In all discussion which follows, a linear estimator of a variance component will refer to a linear function of the mean squares; obviously this is a quadratic function of the observations. Much theoretical work has been done on the construction of best quadratic estimators for balanced designs, but the results have been rather sterile for non-balanced designs.

2. SOME SIMPLE EXAMPLES. To illustrate the complexity of the problem, let us consider a two-stage nested process with only 8 final measurements, coming from $a = 2, 3, \ldots,$ or 7 classes, with the model

$$(5) \qquad Y_{ij} = \mu + a_i + b_{ij}; \quad i = 1, 2, \ldots, a.$$

The object of the investigation is to obtain a minimum variance unbiased linear estimator of σ_a^2. One possible design has $a = 4$ classes and 2 samples per class. The analysis of variance would be as follows

($\rho = \sigma_a^2 / \sigma_b^2$) 2/

Table 1. Analysis of Variance for 4 Classes (N=8)

Source of Variation (SV)	Degrees of Freedom (DF)	Mean Square (MS)	Expected Value of Mean Square (EMS)
Classes (A)	3	MSA	$\sigma_b^2 + 2\sigma_a^2 = \sigma_b^2(1+2\rho)$
Samples (B)	4	MSB	σ_b^2

2/ It is assumed that the reader is familiar with ANOV procedures. If not, reference is made to Anderson and Bancroft (1952, Chapter 22).

The ANOV estimates of σ_a^2 and σ_b^2 are

$$(6) \qquad \hat{\sigma}_a^2 = \frac{MSA - MSB}{2}; \quad \hat{\sigma}_b^2 = MSB.$$

Under the normality assumption, MSA and MSB are distributed as multiples of independent χ^2-variates. In fact $MSA = \chi_1^2(\sigma_b^2 + 2\sigma_a^2)/3$ and $MSB = \chi_2^2 \sigma_b^2/4$, where χ_1^2 and χ_2^2 are independent χ^2-variates with 3 and 4 respective degrees of freedom. Hence

$$(7) \qquad \mathrm{Var}(\hat{\sigma}_a^2) = \frac{2}{4}\left[\frac{(\sigma_b^2 + 2\sigma_a^2)^2}{3} + \frac{\sigma_b^4}{4}\right] = 2\sigma_b^4\left[7/48 + \rho/3 + \rho^2/3\right].$$

In comparing designs and estimators, we will consider

$$(8) \qquad V_a = \frac{\mathrm{Var}(\hat{\sigma}_a^2)}{2\sigma_b^4} = 7/48 + \rho/3 + \rho^2/3.$$

$$(9) \qquad \mathrm{Var}(\hat{\sigma}_b^2) = 2\sigma_b^4/4; \quad V_b = 1/4.$$

Another plan would have $a = 2$ classes and 4 samples per class, with the results given in Table 2 and equations (10). A comparison of these two designs shows that the one with 2 classes will be superior (in estimating σ_a^2) only when ρ is quite small. But we could consider 3, 5, 6 or 7 classes. Unfortunately all of these will involve non-balanced designs. As an example, consider the design with 7 classes, one sample per class for 6 classes and 2 samples from the seventh class; the ANOV is in Table 3.

A_1 represents the variation among the 6 classes with one sample per class. A_2 represents the comparison of the sum of these first 6 observations and 3 times the sum of the two observations from the seventh class. In estimating σ_a^2, one must decide how to weight MSA_1 and MSA_2. If these are weighted directly as their DF, we have the results in equations (11).

Using estimators of this type, the values of V_a and V_b have been computed when $a = 2, 3, \ldots,$ or 7 classes are used, for $\rho = .05, .1, .2, .5, 1, 2, 5$ and 10 (Table 4). As expected, one requires more classes to estimate σ_a^2 efficiently as ρ increases; at the same time, the variance of $\hat{\sigma}_b^2$ increases. Hence for large ρ, if one wishes good estimates of both σ_a^2 and σ_b^2, a compromise is necessary. This will be discussed in more detail later.

Table 2. Analysis of Variance for 2 Classes

SV	DF	MS	EMS
A	1	MSA	$\sigma_b^2 + 4\sigma_a^2 = \sigma_b^2(1 + 4\rho)$
B	6	MSB	σ_b^2

(10) $\hat{\sigma}_a^2 = \frac{MSA - MSB}{4}$; $\hat{\sigma}_b^2 = MSB$; $V_a = \frac{7}{96} + \frac{\rho}{2} + \rho^2$; $V_b = \frac{1}{6}$.

Table 3. Analysis of Variance for 7 Classes

SV	DF	MS	EMS
A_1	5	MSA_1	$\sigma_b^2 + \sigma_a^2 = \sigma_b^2(1+\rho)$
A_2	1	MSA_2	$\sigma_b^2 + 1.75\,\sigma_a^2 = \sigma_b^2(1+1.75\rho)$
B	1	MSB	σ_b^2

(11) $\hat{\sigma}_a^2 = \dfrac{5\,MSA_1 + MSA_2 - 6\,MSB}{6.75}$; $V_a = .922 + .2963\rho + .17695\rho^2$; $V_b = 1/(8-a)$.

* * * * * * * * *

Table 4. Values of V_a and V_b for Various Designs, N = 8 [a]

No. Classes (a)	V_a for given ρ .05	.10	.2	.5	1.0	2.0	5.0	10.0	V_b
2	.100	.133	.213	.573	1.57	5.07	27.6	105	.167
3	.121	.145	.198	.419	.993	2.90	14.8	54.9	.200
4	.163	.182	.226	.396	.812	2.15	10.1	36.8	.250
5	.255	.274	.314	.467	.829	1.96	8.60	30.5	.333
6	.430	.447	.485	.623	.941	1.90	7.43	25.4	.500
7	.937	.953	.988	1.11	1.40	2.22	6.83	21.6	1.000

[a] $Var(\hat{\sigma}_a^2) = 2V_a\sigma_b^4$; $Var(\hat{\sigma}_b^2) = 2V_b\sigma_b^4$. Underlined values are minima for a given $\rho = \sigma_a^2/\sigma_b^2$.

3. Selection of Best "Linear" Estimators. As stated earlier, for non-balanced designs the choice of even the best linear unbiased estimator is not obvious. We might use an iterative weighted least squares procedure, using the mean squares as the dependent variable and the variance components as regression coefficients. Since the mean squares are multiples of χ^2-variates, they have unequal variances, both due to EMS and DF; i.e.,

$$\text{Var (MS)} = (\text{EMS})^2/\text{DF}.$$

For the above 7-class problem the least squares equations are given in equations (12), with limiting values for $\rho = 0$ and $\rho \to \infty$ in equations (13) and (15). In the latter two cases, explicit estimators of σ_a^2 and σ_b^2 are available. These and their variances are presented in equations (14) and (16). For the general case (12), a suggested procedure is to use a preliminary estimate of ρ, such as $\hat{\sigma}_a^2/\hat{\sigma}_b^2$ from the pooled ANOV results; then resolve for $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ using the least squares equations; continue the iteration until two successive sets of estimates agree. Unfortunately, it is almost impossible to obtain the variance of the final estimate of $\hat{\sigma}_a^2$; hence, a comparison of designs becomes very difficult. Very little research has been pursued along these lines; the ideas are merely tossed out here for consideration.

A comparison of the limiting estimators with those of Table 4 is presented in Table 5 for $\rho = .05, .1, 1$ and 10. The first two estimators are about the same for all ρ but the third one is good only for very small ρ. It appears that one does not lose much information by use of MSB to estimate σ_b^2 and the pooled MSA_i to estimate σ_a^2, except possibly when ρ is very small. In subsequent discussions, it will be assumed that the estimators will be of the type used in Section 2.

Least Squares equations for 7-class problem.

$$\left[\frac{5}{(1+\rho)^2}+\frac{(1.75)^2}{(1+1.75\rho)^2}\right]\hat{\sigma}_a^2+\left[\frac{5}{(1+\rho)^2}+\frac{1.75}{(1+1.75\rho)^2}\right]\hat{\sigma}_b^2=\frac{5A_1}{(1+\rho)^2}+\frac{1.75A_2}{(1+1.75\rho)^2}$$

(12)

$$\left[\frac{5}{(1+\rho)^2}+\frac{1.75}{(1+1.75\rho)^2}\right]\hat{\sigma}_a^2+\left[\frac{5}{(1+\rho)^2}+\frac{1}{(1+1.75\rho)^2}+1\right]\hat{\sigma}_b^2=\frac{5A_1}{(1+\rho)^2}+\frac{A_2}{(1+1.75\rho)^2}+B.$$

$\rho = 0$

$$\left[5+(1.75)^2\right]\hat{\sigma}_a^2+6.75\,\hat{\sigma}_b^2=5A_1+1.75A_2$$

(13)

$$6.75\,\hat{\sigma}_a^2+7\,\hat{\sigma}_b^2=5A_1+A_2+B$$

$$\hat{\sigma}_{ao}^2=\frac{10A_1+44A_2-54B}{87};\quad V_{ao}=\frac{20(1+\rho)^2+1936(1+1.75\rho)^2+2916}{7569}$$

(14)

$$\hat{\sigma}_{bo}^2=\frac{35A_1-20A_2+43B}{58};\quad V_{bo}=\frac{245(1+\rho)^2+400(1+1.75\rho)^2+1849}{3364}$$

$$\rho = \infty$$

(15) $$\hat{\sigma}^2_{b,\infty} = B = \hat{\sigma}^2_b ; \; 6\hat{\sigma}^2_a + (5 + \frac{1}{1.75})\hat{\sigma}^2_b = 5A_1 + \frac{A_2}{1.75} ;$$

(16) $$\hat{\sigma}^2_{a,\infty} = \frac{35A_1 + 4A_2 - 39B}{42} ; \; V_{a,\infty} = \frac{245(1+\rho)^2 + 16(1+1.75\rho)^2 + 1521}{1764} .$$

Table 5. Comparison of Variances of Estimators (N=8)

	ρ			
V	.05	.5	1	10
V_a	.94	1.11	1.40	21.6
$V_{a,\infty}$	1.03	1.21	1.49	20.8
$V_{a,o}$	.69	1.29	2.32	88.2
V_b	1.00	1.00	1.00	1.00
$V_{b,o}$	.77	.80	.87	1.13

These methods could be extended to any design with any number of variance components, provided independent mean squares can be computed. In general let M_h represent the h^{th} mean square, with n_h D.F. and $E\ M_h = \sum_{i=1}^{v} k_{hi}\, \sigma^2_i$, where there are v variance components (some of the k_{hi} will be zero; none will be negative). The i^{th} normal equation will be

(17) $$\sum_{j=1}^{v} S_h [w_h k_{hi} k_{hj}] \cdot \hat{\sigma}^2_j = S_h [w_h k_{hi} M_h] ,$$

where $w_h = n_h/(EM_h)^2$. One can simplify the weights (w_h) by using ratios of variance components, as above.

4. OPTIMAL DESIGNS FOR A TWO-STAGE NESTED PROCESS. P. P. Crump (1954) considered these two problems for the two-stage nested process:

(1) Given N, the total sample size, find the best design for use in estimating μ, σ_b^2, σ_a^2 or $\rho = \sigma_a^2/\sigma_b^2$, using the pooled ANOV considered in Section 2.

(2) For certain selected designs, compare some alternative estimators of μ and σ_a^2.

Only (1) will be discussed at this time.

The best design for estimating μ is a = N classes with 1 observation per class and for σ_b^2 is a = 1 class of N observations; the results for σ_a^2 and ρ are intermediate. Given a classes, Crump proved that for both σ_a^2 and ρ, the optimum allocation of N was to have r classes with p+1 observations per class and a - r classes with p observations per class, where N = ap + r, r (an integer)<a. However, the optimal a would be different for estimating σ_a^2 and ρ. It could not be proven rigorously but it was conjectured (and shown to be correct for all computed examples) that the following procedure should be followed to determine the optimal a:

(1) If a did not have to be an integer, it has been shown by several authors that the optimal a (the one giving an unbiased minimum variance estimator based on the pooled ANOV) is

i) for σ_a^2: $a = a_1 = \dfrac{N(N\rho+2)}{N(\rho+1)+1}$;

(18)

ii) for ρ: $a = a_2 = 1 + \dfrac{(N-5)(N\rho+1)}{2N\rho + N - 3}$.

(2) Since a_1 and a_2 will not, in general, be integers, select the integers just larger and just smaller than a_1 (or a_2); compute the exact variance for each; then whichever variance is smaller, compute the variance for the next adjacent integral value of $\underline{a}$; continue this process until a smallest variance is obtained. In all examples computed, this was found for one of the two original integers. These calculations usually can be omitted since the variance profile is quite flat near the optimum; hence, I suggest using for $\underline{a}$ the integer closest to a_1 (or a_2).

(3) The true variances of $\hat{\sigma}_a^2$ and $\hat{\rho}$ for a given $\underline{a}$ ($N = ap + r$) are:

$$\mathrm{Var}(\hat{\sigma}_a^2) = 2\sigma_b^4(A\rho^2 + B\rho + C),$$

$$\mathrm{Var}(\hat{\rho}) = 2(A'\rho^2 + B'\rho + C').$$

(19)
$$A = \frac{N^2S_2 - 2NS_3 + S_2^2}{(N^2 - S_2)^2}\ ;\ B = \frac{2N}{N^2 - S_2};\ C = \frac{N^2(N-1)(a-1)}{(N-a)(N^2 - S_2)^2};$$

$$A' = \frac{(N-a-2)A+1}{N-a-4};\ B' = \frac{(N-3)B}{N-a-4};\ C' = \frac{(N-a)(N-3)}{(N-1)(N-a-4)}C;$$

$$S_2 = Np + r(p+1)\ ;\ S_3 = Np^2 + r(p+1)(2p+1).$$

Crump presents values of A, B, C, A', B', and C' for selected values of $\underline{a}$ for N = 10, 20, 30, and 100.

Crump compared the minimum variances for $\hat{\sigma}_a^2$ and $\hat{\rho}$ following the above procedure with the theoretical minimum disregarding the fact that a and the number of observations per class must be integers. Some of these efficiency ratios of variances (E_o) are given in Table 6 for $\hat{\sigma}_a^2$.

Table 6. Values of $E_o(\hat{\sigma}_a^2)$ [a]

	ρ			
N	0	.50	1.00	2.00
10	.988	.981	.980	.940
20	.997	.987	.995	.943
100	1.000	.998	1.000	.947

[a] $E_o = \frac{\text{min. hypothetical variance}}{\text{min. realizable variance}}$

These results indicate that the procedure outlined above must be quite good.

Recognizing that the allocation which produces a good estimator of one parameter may be poor for estimating some other parameter, Crump computed the efficiency ratio (E_f) for all four parameters, for various allocation plans. These are presented in Table 7. This table presents (for each parameter) the ratio of the variance for the optimum allocation to estimate that parameter to the variance for the given allocation. For example with N = 10 and $\rho = 1$, the best allocation plans for μ, σ_b^2, σ_a^2 and ρ are, respectively, a = 10, 1, 5, and 3. The minimum respective variances are $\sigma_b^2/5$; $2\sigma_b^4/9$; $2\sigma_b^4(.6125)$; $2(1.98775)$. The respective variances for a = 4 are $\sigma_b^2(.36)$; $2\sigma_b^4/6$; $2\sigma_b^4(.6950)$; $2(2.32285)$. The ratios of the minimum variances to those for a = 4 give the respective efficiency ratios, E_f: 5/9 = .56; 2/3 = .67; .6125/.6950 = .88; 1.98775/2.32285 = .86.

Table 7. Efficiency Ratios (E_f) for $\hat{\mu}$, $\hat{\sigma}_b^2$, $\hat{\sigma}_a^2$, $\hat{\rho}$. [1]

N	ρ	a	$E_f(\hat{\mu})$	$E_f(\hat{\sigma}_b^2)$	$E_f(\hat{\sigma}_a^2)$	$E_f(\hat{\rho})$
10	0.5	2	.43	.89	.61	.88
		3	.56	.78	.89	1.00
		4	.65	.67	.99	.78
		5	.75	.56	1.00	.43
	1.0	2	.33	.89	.42	.79
		3	.45	.78	.71	1.00
		4	.56	.67	.88	.86
		5	.67	.56	1.00	.50
20	0.5	5	.50	.79	.90	1.00
		7	.61	.68	1.00	.95
		8	.65	.63	.99	.87
		10	.75	.53	.96	.69
	1.0	5	.40	.79	.70	.96
		7	.51	.68	.87	1.00
		8	.56	.63	.92	.96
		10	.67	.53	1.00	.82
		11	.70	.47	.98	.69
	2.0	5	.33	.79	.54	.90
		7	.44	.68	.72	1.00
		10	.60	.53	.95	.89
		11	.63	.47	.97	.82
		14	.71	.32	1.00	.36
100	0.5	10	.25	.91	.58	.71
		20	.43	.81	.88	.97
		25	.50	.76	.96	1.00
		33	.60	.68	1.00	.96
		35	.61	.66	.99	.94
		50	.75	.51	.91	.74
	1.0	25	.40	.76	.77	.96
		33	.49	.68	.89	1.00
		40	.56	.61	.94	.97
		50	.67	.51	1.00	.88
		60	.71	.40	.95	.72
	2.0	25	.33	.76	.61	.90
		38	.47	.63	.81	1.00
		50	.60	.51	.97	.97
		66	.69	.34	1.00	.72
		75	.82	.25	.98	.54

[1] E_f = Variance for optimum allocation at optimum a / Variance for optimum allocation at given a

An examination of Table 7 indicates that large gains can be achieved in the efficiency of estimates of σ_a^2 and ρ by suitable choice of the sampling plan, for fixed total number of observations, N. For $N = 20$, $\rho = 1.0$, one should use 7, 8, 9, or 10 classes depending on the relative importance of the two parameters, assuming μ and σ_b^2 were of secondary importance. Similarly if $\rho = 2.0$, one should use between 7 and 14 classes. And if $\rho = 0.5$, use 5, 6, or 7 classes. It is important to note that if one thought that $0.5 < \rho < 2.0$, he might logically choose 7, 8, or 9 classes without departing too far from the optimum. For $N = 100$ and ρ somewhere between 0.5 and 2.0, he could choose around 40 classes without much risk of departing far from the optimum.

One can use the results in Table 7 to study the effects of using the incorrect ρ in making his allocations. Some of these comparisons plus others regarding the estimation of σ_a^2 have been made in Table 8. In general it appears that if the true ρ is 1 or less, one will not lose more than 10% in efficiency even if he uses 2ρ or $\rho/2$. The loss is considerably less when ρ is as large as 2.

5. OPTIMAL DESIGNS FOR A TWO-WAY CLASSIFICATION MODEL.

5.1 Introduction. D. W. Gaylor (1960) considered methods of sampling, under certain restrictions, to minimize the variance of certain estimators of variance components for a two-way crossed classification

$$(20) \qquad Y_{ijk} = \mu + r_i + c_j + (rc)_{ij} + s_{ijk} \; ; \; \begin{matrix} i = 1, 2, \ldots, r < N \\ j = 1, 2, \ldots, c < N, \end{matrix}$$

where the effects are assumed normally and independently distributed with zero means and respective variances, σ_r^2, σ_c^2, σ_{rc}^2 and σ_s^2 and n_{ij} observations are obtained in the (ij) cell. It is assumed that a total of $N = \sum\sum n_{ij}$ samples are allocated to the rc cells. The problem is to make the allocation so that the variances of the estimates of the variance components will be as small as possible. This again poses the problem of

Table 8. Effect of Incorrect Estimate of ρ on the Efficiency of Estimates of σ_a^2

N	ρ	Optimal No. Classes (a)	Estimated ρ	No. Classes Used (a)	Efficiency
20	0.5	7	0.22	5	.903
			0.31	6	.964
			0.68	9	.985
			0.85	10	.960
			1.06	11	.897
	1.0	10	0.41	7	.869
			0.53	8	.918
			1.32	12	.957
			1.66	13	.921
			2.12	14	.869
	2.0	14	0.41	7	.725
			0.68	9	.868
			0.85	10	.947
			1.06	11	.969
			3.70	16	.951
100	0.5	33	0.14	14	.739
			0.26	22	.911
			0.59	38	.986
			0.97	50	.912
			1.46	60	.775
	1.0	50	0.31	25	.772
			0.47	33	.893
			0.64	40	.944
			1.46	60	.946
			1.90	66	.894
	2.0	66	0.64	40	.832
			0.97	50	.969
			2.95	75	.976
			3.94	80	.935

deciding which parameters are most important. Gaylor studied optimal allocations for individual components and for certain combinations

Given any quadratic estimator

(21) $$Q_i = \underset{\sim}{y}' M_i \underset{\sim}{y}$$

where $\underset{\sim}{y}$ is the (N x 1) vector of observations and $M_i = M_i'$ an N x N matrix of coefficients. In vector form, we can write

(22) $$\underset{\sim}{y}(N \times 1) = \underset{\sim}{\mu}(N \times 1) + \underset{\sim}{e}(N \times 1),$$

where $\underset{\sim}{e}$ is the error or random effects vector. The variance-covariance matrix of the e's is

(23) $$V = E(\underset{\sim}{e}\,\underset{\sim}{e}').$$

The diagonal elements of V will be

(24) $$\sigma_r^2 + \sigma_c^2 + \sigma_{rc}^2 + \sigma_s^2.$$

The off-diagonal elements will be

(25) $$\sigma_r^2,\ \sigma_c^2,\ \sigma_r^2 + \sigma_c^2 + \sigma_{rc}^2 \text{ or } 0,$$

depending on whether the two corresponding observations are in the same row but different columns, the same column but different rows, the same

row and column or in different rows and columns. It is easy to show that, for unbiased estimators,

$$(26) \qquad E(Q_i) = \mathrm{tr}(VM_i) = \sigma_i^2; \quad \mathrm{Var}(Q_i) = 2\mathrm{tr}(VM_i)^2,$$

where σ_i^2 is the variance component to be estimated and "tr" refers to the trace of the matrix. The lower bound to the variance is

$$(27) \qquad \text{L.B.Var}(Q_i) = 2\,\sigma_i^4/(N-1).$$

For the following cases, this lower bound can be realized; for others, it usually is not possible.

Function estimated	r	c	n_{ij}
σ_s^2	1	1	N
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2$	N	1	1
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_c^2$	1	N	1
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2 + \sigma_c^2$	N	N	1 or 0*

*each observation from a different row and column

The estimator in each case is $S(y - \bar{y})^2/(N-1)$. The last design is also the optimal design for estimating μ. The variance of $\bar{y}$ is then

$$(29) \qquad (\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2 + \sigma_c^2)/N.$$

One special case is of importance. If $\sigma_{rc}^2 = 0$, the optimal design for estimating σ_r^2 or $\rho_r = \sigma_r^2/\sigma_s^2$ consists of one column; in this case the results by Crump apply. Similarly, if only σ_c^2 or ρ_c is of interest, use one row and apply Crump's results.

5.2 Estimation of σ_r^2 or ρ_r when $\sigma_{rc}^2 > 0$. It was not possible to develop a complete general class of designs to estimate σ_r^2 or $\rho_r = \sigma_r^2/\sigma^2$, where $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$, which could be proven optimal for all situations. It was first shown that if the design were limited to a class in which each $n_{ij} = 0$ or n(an integer) observations per cell, n should equal 1. Hence each cell should either be empty or contain one observation. The sums of squares used for estimating σ_r^2 and ρ_r were those given in an ANOV table based on the method of fitting constants (Table 9). The estimators of σ_r^2 and ρ_r and the variances of these estimators are given in equations (30) and (31). These results are based on the fact that (r-1)R* can be subdivided into r-1 independent quantities.

Table 9. Analysis of Variance for 2-Way Classification Data

SV	DF	MS	EMS
Columns	c - 1	C	$\sigma_s^2 + \sigma_{rc}^2 + c_1\sigma_r^2 + r_o\sigma_c^2$
Rows (adjusted for cols)	r - 1	R*	$\sigma_s^2 + \sigma_{rc}^2 + c_o\sigma_r^2$
Interaction (adjusted for rows and cols.)	N-r-c+1	I*	$\sigma_s^2 + \sigma_{rc}^2$
Total	N - 1		

(30) $$\hat{\sigma}_r^2 = (R^* - I^*)/c_o\,; \quad \hat{\rho}_r = \left[\frac{N-r-c-1}{N-r-c+1}F' - 1\right]/c_o,$$

where $c_o = (N-c)/(r-1)$ and $F' = R^*/I^*$.

(31a) $$\frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \frac{r-1}{m(N-c)} + \frac{2\rho_r}{N-c} + \frac{\rho_r^2}{r-1} + \frac{\rho_r^2}{(N-c)^2}\sum_{i=1}^{r-1}(d_i - \bar{d})^2;$$

(31b) $$\frac{\mathrm{Var}(\hat{\rho}_r)}{2} = \frac{(r-1)(1+c_o\rho_r)^2(m+r-3) + (m-2)\rho_r^2\sum(d_i - \bar{d})^2}{c_o^2(r-1)^2(m-4)}$$

where $m = N-r-c+1$ (>4 for $\hat{\rho}_r$), $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$, $\rho_r = \sigma_r^2/\sigma^2$ and $(r-1)R^* = \sigma^2\sum_{i=1}^{r-1}(1 + d_i\rho_r)\chi_i^2$.

If $N = rc$,

(32a) $$\mathrm{Var}(\hat{\sigma}_r^2) = \frac{2\sigma^4}{r-1}\left[\frac{1}{c(c-1)} + \frac{2\rho_r}{c} + \rho_r^2\right];$$

(32b) $$\mathrm{Var}(\hat{\rho}_r) = 2(1+c\rho_r)^2(N-c-2)/c^2(r-1)(N-r-c-3).$$

All statements refer to the above estimators. Admittedly other estimators might be more efficient for a given design.

If for a given N and r, N is an integral multiple of r, N = rc, then the best design with $n_{ij} = 0$ or n to estimate σ_r^2 or ρ_r consists of r rows and c columns with one observation per cell. Graybill and Wortham (1956) show that the resultant $\hat{\sigma}_r^2$ is a minimum variance unbiased quadratic estimator. In this case in Table 9, $c_1 = 0$, $c_o = c$ and $r_o = r$, and $d_i = \bar{d} = c$. The variances of the estimators are given in equations (32).

In general N/r will not be an integer. Suppose $N = r(k-1)+s$, $0<s<r$. For a given r, a design of the type $n_{ij} = 0$ or 1 which minimizes $\mathrm{Var}(\hat{\sigma}_r^2)$ based on the above ANOV table consists of r rows by k-1 columns (using only N-s observations) or r rows by k-1 columns plus one column with s of the r rows. The variances are given in equations (33). This leads to the interesting result that a balanced design may lead to a smaller variance than an unbalanced design with more observations; this situation arises when ρ_r is large and indicates that presumably the estimators used are not the best for the unbalanced situation.

In order to find the optimal value of r, one should minimize (33a) or (33b). This is quite complicated because k and s are functions of r. Because of the near balance of the design, a good approximation to (33) is available by replacing each d_i in (31) by $\bar{d} = c_o = (N-c)/(r-1)$ and c by $c_o + 1/2$ (since $c_o \leq c \leq c_o + 1$). These approximate variances are given in equations (34). Setting the derivative of (34a) with respect to c_o equal to 0, we have

$$(35) \qquad c_o \doteq \frac{\rho_r^2(N-1/2) + 2\rho_r - 1 + \sqrt{\rho_r^2(N-1/2)(N-3/2) + 2\rho_r(N-3/2) + 1}}{\rho_r^2(N-1/2) + 2\rho_r}$$

This is approximated by (36a) with the approximate value of $\underline{r}$ in (37a). Corresponding results for $\hat{\rho}_r$ are presented in (36b) and (37b).

If $N = r(k-1) + s, \quad 0 < s < r$

$$(33a)\quad \frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \begin{cases} \left[\rho_r^2 + 2\rho_r/(k-1) + 1/(k-1)(k-2)\right]/(r-1), \text{ using } k-1 \text{ cols.} \\ \dfrac{\rho_r^2}{r-1} + \dfrac{\rho_r^2(s-1)(r-s)}{(N-k)^2(r-1)} + \dfrac{2\rho_r}{N-k} + \dfrac{r-1}{(N-k)(N-k-r+1)}, \text{ using } k \text{ cols.} \end{cases}$$

$$(33b)\quad \frac{\mathrm{Var}(\hat{\rho}_r)}{2} = \begin{cases} \left[1 + (k-1)\rho_r\right]^2 (rk-r-k-1)/(k-1)^2(r-1)(rk-2r-k-2), \text{ using } k-1 \text{ cols.} \\ \dfrac{(N-r-k-1)(s-1)(r-s)\rho_r^2 + \left[(r-1) + (N-k)\rho_r\right]^2 (N-k-2)}{(N-k)^2(r-1)(N-r-k-3)}, \text{ using } k \text{ cols.} \end{cases}$$

Using all N observations,

$$(34a)\quad \frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} \doteq \frac{1 + 2c_o\rho_r + c_o^2\rho_r^2 - 2\rho_r - c_o\rho_r^2}{(c_o-1)(N-c_o-1/2)}.$$

$$(34b)\quad \frac{\mathrm{Var}(\hat{\rho}_r)}{2} \doteq \frac{(1 + c_o\rho_r)^2}{(c_o-1)(N-c_o-1/2)}.$$

For $\hat{\sigma}_r^2$,

(36a)
$$c_o \doteq \tilde{\tilde{c}}_o = \frac{\rho_r(N-1/2) + (N-1/2) + 1}{\rho_r(N-1/2) + 2} \to 1 + 1/\rho_r \text{ for large } N,$$

and

(37a) $$r \doteq \tilde{\tilde{r}} = (N - 1/2)/\tilde{\tilde{c}}_o.$$

For $\hat{\rho}_r$,

(36b)
$$c_o \doteq \tilde{c}_o = \frac{2\rho_r(N-1/2) + (N-1/2) + 1}{\rho_r(N-1/2) + \rho_r + 2} \to 2 + 1/\rho_r \text{ for large } N,$$

(37b) $$r \doteq \tilde{r} = (N - 1/2)/\tilde{c}_o.$$

Again one can consider integers above and below $\tilde{\tilde{r}}$ or $\tilde{r}$ as the optimal r, testing by insertion in the true variance formula (33). As in Crump's case, the variance profile is so flat that the integer closest to $\underline{r}$ will probably be sufficient. Once $\underline{r}$ is determined, the number of columns will be found from the fact that $N = kr$ or $N = (k-1)r+s$, where $s < r$. In the latter case, the experimenter presumably should check whether to use a balanced design with $(k-1)r$ observations or an unbalanced one with $\underline{s}$ rows in the k^{th} column; since it requires a very large ρ_r to prefer the smaller balanced design, the unbalanced design usually will be better. In the latter case, the experimenter might wish to consider the use of enough more observations to have a (kxr) balanced design; matters of this kind need further investigation, such as maximizing the information per observation rather than the total information.

Gaylor also investigated the loss of information due to an incorrect estimate of ρ_r, as given in Table 10. These results are of the same order of magnitude as found by Crump. As before, losses due to use of the incorrect ρ are less serious for large ρ. Gaylor checked some of these results against those found by use of exact variances; the agreement on the efficiency factor was very good; e.g., $E_{f2} = .907$ instead of .909 for $N = 30$, $\rho_r = 1.0$, $\rho_r' = 2.0$. It should be noted that Gaylor considered ratios of 4:1 for $\hat{\rho}_r$ and only 2:1 for $\hat{\sigma}_r^2$; hence, some of these efficiencies are lower than for $\hat{\sigma}_r^2$.

Table 11 presents the efficiency of various designs used to obtain information on $\hat{\sigma}_r^2$ and ρ_r for $N = 30$. This table shows that extreme departures from the optimal design results in a considerable loss of efficiency; however, moderate departures usually result in small losses. Where ρ_r is small, a balanced design in the neighborhood of the optimal design has high efficiency for estimating both parameters. However, even for $\rho_r = 1$, the efficiency of each estimator is less than 90% when the optimal design for the other is used. On the other hand, one is able to find an intermediate design which is quite good for estimating both parameters when ρ_r is no larger than 2.

Table 10. Approximate Relative Efficiencies, E_f of the Restricted Optimal Designs for $\hat{\sigma}_r^2$ and $\hat{\rho}_r$ Based on Incorrect Values of ρ.

Results for $\hat{\sigma}_r^2$

N	Design Based on True ρ_r		Design Based on Incorrect ρ'_r					
	ρ_r	$\approx c_o$	ρ'_{r-1}	$\approx c'_{o1}$	E_{f1}	ρ'_{r-2}	$\approx c'_{o2}$	E_{f2}
30	.25	4.04	.125	6.01	.928	.50	2.70	.905
	1.0	1.90	.50	2.70	.920	2.00	1.47	.909
	4.0	1.24	2.00	1.47	.943	8.00	1.12	.939
100	.25	4.67	.125	7.82	.907	.50	2.90	.900
	1.0	1.97	.50	2.90	.912	2.00	1.49	.909
	4.0	1.25	2.00	1.49	.942	8.00	1.12	.936

Results for $\hat{\rho}_r$

True ρ_r	ρ'_r Used	E_f N = 30	E_f N = 100
0.25	1.00	.849	.829
1.00	0.25	.847	.823
0.50	2.00	.900	.891
2.00	0.50	.907	.896
2.00	8.00	.970	.976
8.00	2.00	.993	.983

Table 11. Efficiency of Some Designs for Estimating σ_r^2 and ρ_r, N = 30 a/

ρ_r	r	c	s	$E_f(\hat{\sigma}_r^2)$	$E_f(\hat{\rho}_r)$
.25	2	15	0	.422	.446
	3	10	0	.687	.739
	5	6	0	.948	.975
	6	5	0	.999	1.000
	7	5	2	.984	.947
	7*	4*	0*	.941*	.889*
	8	4	6	1.000	.943
	9	4	3	.959	.871
	10	3	0	.965	.862
	15	2	0	.732	.557
.50	5	6	0	---	.889
	6	5	0	.833	.964
	7	5	2	.882	.968
	7*	4*	0*	.857*	.922*
	8	4	6	.938	1.000
	9	4	3	.954	.967
	10	3	0	1.000	.986
	11	3	8	.986	.927
	15	2	0	.933	.732
1.0	6	5	0	---	.852
	7*	4*	0*	---	.862*
	9	4	3	---	.956
	10	3	0	.877	1.000
	11	3	8	.899	.968
	14	3	2	.957	.833
	15	2	0	1.000	.844
	16	2	14	.984	---
	19	2	11	.907	---
2.0	6	5	0	---	.776
	7*	4*	0*	---	.814*
	9	4	3	---	.938
	10	3	0	.740	1.000
	11	3	8	---	.987
	12	3	6	.822	.969
	15	2	0	.974	.931
	18	2	12	.997	---
	19	2	11	1.000	.631
	20	2	10	.998	---
	24	2	6	.902	---
4.0	6	5	0	---	.735
	7*	4*	0*	---	.788*
	9	4	3	---	.927
	10	3	0	.586	1.000
	11	3	8	---	1.000
	12	3	6	---	.994
	15	2	0	.837	.991
	20	2	10	.943	.527
	23	2	7	.994	---
	24	2	6	1.000	---
	25	2	5	.993	---
	28	2	2	.658	---

$E_f(\hat{\rho}_r)$ only for ρ_r = 8 and 16

r	c	s	ρ_r = 8	ρ_r = 16
10	3	0	.974	.955
11	3	8	.981	.965
12	3	6	.982	.970
13	3	4	.977	.969
14	3	2	.963	.960
15	2	0	1.000	1.000
16	2	14	.945	.948
20	2	10	.634	.646

a/ r = No. rows ; c = No. columns. If s = o, N = rc; if s >o, N = r(c-1) + s;
* designs have only 28 observations.

The two sets of restricted optimal designs begin to diverge considerably as ρ_r increases beyond 2. One notes that the limiting value of $\underline{c}$ is $1 + 1/\rho_r$ for $\hat{\sigma}_r^2$ and $2 + 1/\rho_r$ for $\hat{\rho}_r$. For $\rho_r = 1$, the optimal designs have 2 and 3 columns, respectively. However, for $\rho > 1$, the optimal design for $\hat{\sigma}_r^2$ is nonbalanced with 2 columns, whereas that for $\hat{\rho}_r$ approaches a balanced design with 2 columns.

The particular case of $c = 2$ has been investigated more thoroughly for $\hat{\sigma}_r^2$. Consider a two-column restricted optimal design with $N-r'$ rows in one column and $\underline{r'}$ of these same rows occupied in the second column; i.e., $s = 0$ if $r' = N/2$ or $2 \leq r' = s < N/2$. From (33a),

$$(38) \qquad \frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \left[\frac{N + 2r' - 4}{N-2}\rho_r^2 + 2\rho_r + \frac{N - r' - 1}{r' - 1}\right]/(N-2),$$

If $r' = N/2$, (38) simplified to $(\rho_r^2 + \rho_r + 0.5)/(r'-1)$. The value of $\underline{r'}$ which minimizes (38) is

$$(39) \qquad r' = \begin{cases} 2 & \rho_r \geq (N-2)/\sqrt{2} \\ 1 + (N-2)/\rho_r\sqrt{2} & \sqrt{2} \leq \rho_r \leq (N-2)/\sqrt{2} \\ N/2 & \rho_r \leq \sqrt{2} \end{cases}$$

Hence when $\rho_r > \sqrt{2}$, a two-column design with one column shorter than the other is better than two columns of equal length.

These results suggest the following procedure for finding a nearly optimal design for $\hat{\sigma}_r^2$:

(i) If $\rho_r > \sqrt{2}$, use one column with $r = N - r'$ rows and a second column with r' of these rows, where r' is the integer (≥ 2) which is nearest $1 + (N-2)/\rho_r\sqrt{2}$.

(ii) When $\rho_r \leq \sqrt{2}$, use a balanced design with $\underline{c}$ as the integer above or below

$$\tilde{\tilde{c}}_o = \left[\rho_r(N-1/2) + N + 1/2\right] / \left[\rho_r(N-1/2) + 2\right],$$

which minimizes

$$V(c) = \frac{1}{(N-c)(c-1)} + \frac{2\rho_r}{N-c} + \frac{c\rho_r^2}{N-c} \tag{40}$$

and where $\underline{r}$ was set equal to N/c. Presumably little information is lost by simply using as $\underline{c}$ the integer closer to $\tilde{\tilde{c}}_o$.

Another approach to (ii) above is to determine intervals of ρ_r for which $c = 3$ is best, $c = 4$ is best, ... This is accomplished by determining $\rho_r = \rho(c)$ in the equation

$$V(c) = V(c-1). \tag{41}$$

When $\rho_r < \rho(c)$, $\underline{c}$ is preferred to $c-1$ columns. The solution to (41) is

$$\rho_1(c) = \frac{1}{\sqrt{(c-1)(c-2)}} - \frac{1}{N} - \frac{1}{N}\sqrt{\frac{c-1}{c-2}}\left[1 + 1/2N + 0(1/N^2)\right] \tag{42}$$

For $N = 30$, the following are the first three separation points:

$$\rho_1(3) = .62, \quad \rho_1(4) = .33, \quad \rho_1(5) = .22.$$

Hence

$$c = 2 \text{ for } .62 < \rho_r < \sqrt{2};$$
$$c = 3 \text{ for } .33 < \rho_r < .62$$
$$c = 4 \text{ for } .22 < \rho_r < .33.$$

Note that for $\rho_r = .25$ in Table 11, $c = 4$ is slightly preferred to $c = 5$.

Table 11 indicates that for $\hat{\rho}_r$ there is a considerable range of ρ_r for which the best allocation plan will have 3 columns and N/3 rows. If N is not divisible by 3, some gain may be obtained by using an extra row with one column not filled. It appears that the best plan would be to either increase N or reduce N by enough to have it divisible by 3. As ρ_r increases, only 2 columns should be used with N/2 rows. Again add or subtract one to make N even.

The following general procedure is suggested for $\hat{\rho}_r$:

(i) Use a balanced design with $\underline{c}$ as the integer above or below (36b),

$$\tilde{c}_o = (2\rho_r N + N - \rho_r + 1/2)/(\rho_r N + \rho_r/2 + 2),$$

which minimizes (32b):

$$\frac{\text{Var}(\hat{\rho}_r)}{2} = (1 + c\rho_r)^2(N-c-2)/(N-c)(Nc-N-c^2-3c). \tag{43}$$

(ii) Choose r so that $rc \doteq N$.

The same procedure of finding separation points on ρ_r could be used on (43) as was done above. The equation comparable to (41) is considerably more complicated; hence, all I am reporting here is the asymptotic result (for large N). In this case the solution is

$$\rho_2(c) = \frac{1 + \sqrt{(c-1)(c-2)}}{c^2 - 3c + 1} \tag{44}$$

Some of the separation points are:

$$\rho_2(3) = 1 + \sqrt{2} = 2.4, \quad \rho_2(4) = (1 + \sqrt{6})/5 = .7$$

$$\rho_2(5) = (1 + 2\sqrt{3})/11 = .4, \quad \rho_2(6) = (1 + 2\sqrt{5})/19 = .28.$$

Hence

$$c = 3 \text{ for } .7 < \rho_r < 2.4$$

$$c = 4 \text{ for } .4 < \rho_r < .7$$

$$c = 5 \text{ for } .28 < \rho_r < .4$$

Note that in Table 11, $c = 5$ is slightly preferred to $c = 6$ for $\rho = .25$. The above separation points would be altered slightly if the adjustments of $0(1/N)$ were made.

5.3 Estimation of σ_c^2 or $\rho_c = \sigma_c^2/\sigma^2$. Use the results in Section 5.2 by reversing the roles of r and c. In Table 9, one computes mean squares, R and C*, for Rows and Columns (adjusted for rows). The latter can be computed quickly by use of the identity:

$$(45) \qquad (r-1)R^* + (c-1)C = (r-1)R + (c-1)C^*.$$

5.4 Simultaneous Estimation of σ_r^2 and σ_c^2. Gaylor obtained only some tentative results for this case but they are indicative of the problem. Since efficient estimation of σ_r^2 requires many rows and efficient estimation of σ_c^2 requires many columns, simultaneous estimation requires many rows and columns. Gaylor considered two special designs, called the L-design and the Balanced Disjoint Rectangles (BDR)-design, and compared them for the case $\sigma_r^2 = \sigma_c^2$.

The L-design is as follows:

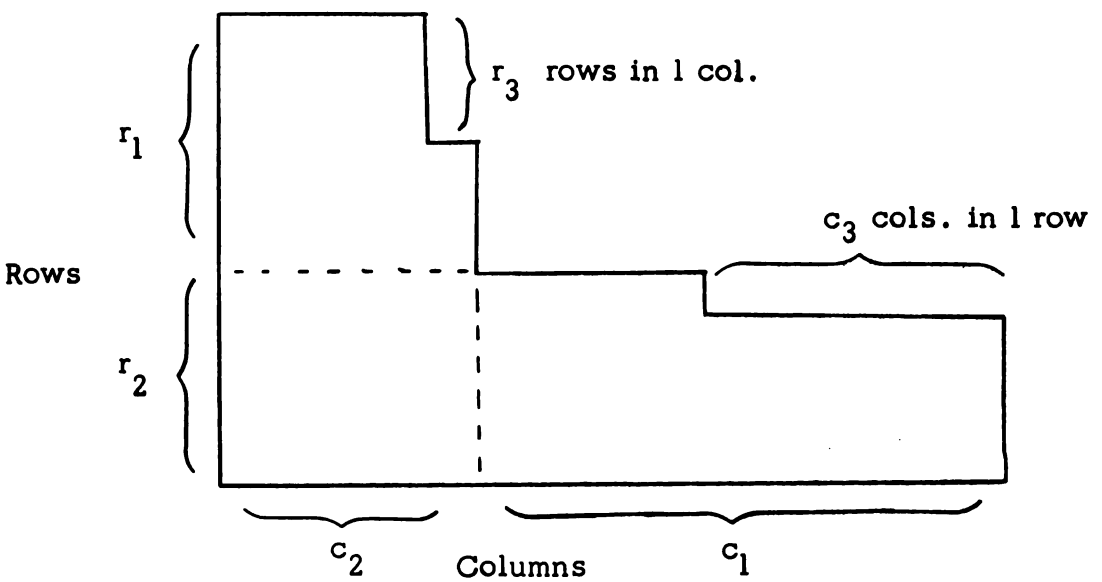


Figure 1. The L-Design

For $\sigma_r^2 = \sigma_c^2$, one should use $r_1 = c_1$, $r_2 = c_2$, $r_3 = c_3$. The total rows and columns used are $r = r_1 + r_2$ and $c = c_1 + c_2$. $N_1 = rc_2 - r_3$ observations contain one column with $r_3 (0 \leq r_3 < r_1)$ empty cells and $N_2 = cr_2 - c_3$ observations contain one row with $c_3 (0 \leq c_3 < c_1)$ empty cells. The N_1 observations are of the form of a restricted optimal design for $\hat{\sigma}_r^2$ and the N_2 observations for $\hat{\sigma}_c^2$. In general c_2 and r_2 will be small.

The BDR design consists of g distinct rectangles, each consisting of r rows by c columns, with one observation per cell. The rectangles are distinct in that each rectangle samples a different set of r rows and c columns. The ANOV for this design is given in Table 12.

If the component of variance for the rectangle effects is zero, $\sigma_g^2 = 0$, then a unique solution for the estimators of the components of variance is not obtained by equating the mean squares to their expected values. Then, there are five mean squares (equations) with which to estimate four components of variance. Since the best procedure in this case is unknown, the case will not be considered here, but this certainly warrants future investigation.

If G is neglected, the estimators and their variances are quite simple, as given in equations (46).

The two designs have been compared for the special case $\sigma_r^2 = \sigma_c^2$ and $\text{Var}(\hat{\sigma}_r^2) = \text{Var}(\hat{\sigma}_c^2)$. Some specific examples are given in Table 13 for $\rho = \sigma_r^2 / \sigma^2 = \sigma_c^2 / \sigma^2$, where $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$. The total number of observations, N, was allowed to vary slightly in order to obtain balance if desired. Thus comparisons were made on the basis of $N\text{Var}(\sigma_r^2) = N\text{Var}(\sigma_c^2)$ which is the reciprocal of the amount of information per observation. The estimator for σ_r^2 for the BDR-design is given in (46) and for the L-design in (30). The variances in the last column of Table 13, $N\dot{V}$, are based on the estimator, $\dot{\sigma}_r^2$, in which the last c_1 columns in

Table 12. ANOV for a Balanced Disjoint Rectangles Design

SV	DF	MS	EMS
Rectangles	g-1	G	$\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2 + r\sigma_c^2 + rc\sigma_g^2$
Rows	g(r-1)	R	$\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2$
Columns	g(c-1)	C	$\sigma_s^2 + \sigma_{rc}^2 + r\sigma_c^2$
Interaction	g(r-1)(c-1)	I	$\sigma_s^2 + \sigma_{rc}^2$

(46)

$$\hat{\sigma}_r^2 = (R-I)/c; \quad \sigma_c^2 = (C-I)/r;$$

$$\mathrm{Var}(\hat{\sigma}_r^2) = \frac{2r}{N(r-1)c}\left[(\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2)^2 + \frac{(\sigma_s^2 + \sigma_{rc}^2)^2}{c-1}\right];$$

$$\mathrm{Var}(\hat{\sigma}_c^2) = \frac{2c}{N(c-1)r}\left[(\sigma_s^2 + \sigma_{rc}^2 + r\sigma_c^2)^2 + \frac{(\sigma_s^2 + \sigma_{rc}^2)^2}{r-1}\right].$$

Table 13. Comparison of BDR and L-Designs with $\mathrm{Var}(\hat{\sigma}_r^2) = \mathrm{Var}(\hat{\sigma}_c^2)$, Where $\sigma_r^2 = \sigma_c^2$.

	BDR-Design				L-Design					
ρ	g	r=c	N	NV[a]	r=c	$r_2=c_2$	$r_3=c_3$	N	NV	$\dot{NV}$
.5	4	3	36	3.38	10	2	0	36	4.24	5.00
.5	2	4	32	3.11	7	3	0	33	3.29	4.12
.5	1	6	36	3.24	6	6	0	36	3.24	3.24
1	9	2	36	10.00	10	2	0	36	10.89	10.00
1	4	3	36	8.25	7	3	0	33	9.46	10.08
1	2	4	32	8.44	6	6	0	36	9.84	9.84
2	9	2	36	26.0	14	2	8	36	26.8	25.6
2	4	3	36	24.8	10	2	0	36	34.8	26.0
2	2	4	32	27.1	7	3	0	33	31.6	30.2
4	9	2	36	82.0	16	2	12	36	72	79.8
4	4	3	36	84.8	10	2	0	36	125	82.0

[a] $V = \mathrm{Var}(\hat{\sigma}_r^2)/2\sigma^4$ and $\dot{V} = \mathrm{Var}(\dot{\sigma}_r^2)/2\sigma^4$.

Figure 1 for the L-design are ignored. The estimator, σ_r^2, could be improved slightly by pooling the error sum of squares from both legs of the L-design.

Several observations can be made from Table 13. When ρ is small, BDR-designs are better than the L-designs. When ρ is large, the L-design is better. For those designs where $\rho^2 \sum (d_i - \bar{d})^2$ is large, see (31a), the estimator, $\dot{\sigma}_r^2$, which discards the last c_1 columns is often better than the estimator, $\hat{\sigma}_r^2$, which uses all of the observations. This is a result found previously where it was shown desirable to keep the numbers of observations per row nearly equal. Also, as was shown in Section 5.2, the L-design is desirable which has less than two complete columns in the N_1 observations and less than two complete rows in the N_2 observations when $\rho > \sqrt{2}$.

For an L-design with $r_2 = c_2 = 2$, $r = c$, and $r_3 = c_3 = 0$,

$$\text{Var}(\dot{\sigma}_r^2) = 2\sigma^4(2 + 4\rho_r + 4\rho_r^2)/N, \tag{47}$$

where $\rho_r = \sigma_r^2/\sigma^2$. For a BDR-design with $r = c = 2$ and $g = N/4$, from (46), $\text{Var}(\hat{\sigma}_r^2)$ is the same as (47). This result is borne out in Table 13. Similarly, $\text{Var}(\dot{\sigma}_c^2) = \text{Var}(\hat{\sigma}_c^2)$. No restriction that $\sigma_r^2 = \sigma_c^2$ has been imposed here.

When $\rho \geq 2.87$, the best BDR-design, has $r = c = 2$ and $g = N/4$. Since $\rho \geq \sqrt{2}$, the L-design with $r_2 = c_2 = 2$ can be improved upon by letting $r_3 = c_3 > 0$. Hence, when $\rho \geq 2.87$, for the L-design $\text{Var}(\dot{\sigma}_r^2) \leq \text{Var}(\hat{\sigma}_r^2)$ for a BDR-design. $\text{Var}(\dot{\sigma}_r^2)$ could be reduced still further by using the observations in the last c_1 columns to estimate σ^2 and pooling this with the estimate of σ^2 from the first c_2 columns.

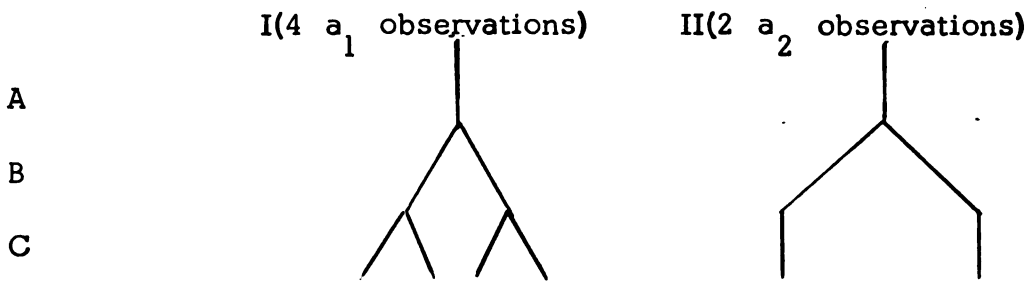
Also, when $r = c$ is small for a BDR-design, the number of rectangles, g, becomes large. If $\sigma_g^2 = 0$, considerable information may be lost here by not making use of the mean squares among rectangles.

A few tentative observations can be made. In general, as ρ increases, the L-design is favored over disjoint rectangles if $\sigma_g^2 > 0$. Also, the analysis which ignores the observations in the other leg of the L-design in order to achieve balance sometimes reduces the variance.

6. ADDITIONAL COMMENTS AND SUGGESTIONS FOR FUTURE RESEARCH.

6.1 Extensions of Nested Sampling to More than Two Stages. One important research problem is to extend Crump's results to more than two stages of sampling. In Section 1, two five-stage problems were mentioned. In the Anderson-Bancroft book and in my International Statistical Institute paper, I proposed a so-called staggered design, which consists of a number of balanced designs but with incomplete sampling at some stages.

For three stages, two different groups (I and II) would be used with a_1 first-stage samples for I and a_2 for II, as follows:



ANALYSIS OF VARIANCE

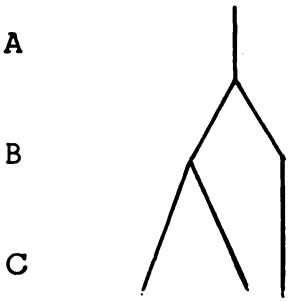
SV	DF	MS	EMS
Groups	1	MSA_3	$\sigma_c^2 + k_b\sigma_b^2 + k_a\sigma_a^2$
A_1	a_1-1	MSA_1	$\sigma_c^2 + 2\sigma_b^2 + 4\sigma_a^2$
A_2	a_2-1	MSA_2	$\sigma_c^2 + \sigma_b^2 + 2\sigma_a^2$
B_1	a_1	MSB_1	$\sigma_c^2 + 2\sigma_b^2$
B_2	a_2	MSB_2	$\sigma_c^2 + \sigma_b^2$
C	$2a_1$	MSC	σ_c^2

where $MSA_3 = (T_1/4a_1 - T_2/2a_2)^2/(1/4a_1 + 1/2a_2)$, $k_b = \dfrac{2(a_1 + a_2)}{2a_1 + a_2}$,

$k_a = \dfrac{4(a_1 + a_2)}{2a_1 + a_2}$, $T_1 = SY_I$ and $T_2 = SY_{II}$.

One might consider using an additional third plan with a_3 first-stage samples, only one second-stage sample per first-stage and one third-stage sample per second-stage, giving a_3 observations.

R. R. Prairie is studying another plan, which would attempt to equalize the total number of observations at each stage. In this case there would be a first-stage samples, with each followed by the following second- and third-stages:



ANALYSIS OF VARIANCE

SV	DF	MS	EMS
A	a-1	MSA	$\sigma_c^2 + \frac{5}{3}\sigma_b^2 + 3\sigma_a^2$
B	a	MSB	$\sigma_c^2 + \frac{4}{3}\sigma_b^2$
C	a	MSC	σ_c^2

The variance of $\hat{\sigma}_a^2$ for this design is complicated by the fact that MSA and MSB are positively correlated. This is a feature of non-balanced designs which we were not anticipating. It indicates a fact which many of us often forget; namely, that the usually accepted concepts regarding the analysis of variance are based on a regression and not a variance components model. Naturally if one uses the ML approach, he does not start with such preconceived ideas. Also one never has any trouble when the design is balanced. Note that my staggered design is balanced within groups; the major difficulty there is how to use the between-groups mean square.

6.2 Other Suggestions.

1) (a) Develop numerical methods, suitable for high-speed computers, to obtain, under normal theory, the ML estimates of variance components from unbalanced data.

(b) Study the small sample properties of the ML estimators by empirical means, and, where possible, by analytic means.

(c) It might be possible to find approximate formulas to estimate the biases and variances of these ML estimates. A comparison of asymptotic variances and empirically determined small sample variances would be useful.

(d) Compare the small sample properties of the ML estimates with those of more convenient estimates, e.g., ANOV or iterated least squares.

2) Study the effect of non-normality on estimating procedures.

3) Introduce unequal cost factors into the problem.

4) Develop suitable criteria for use in simultaneous estimation of several parameters.

5) Develop a sequential estimation procedure in which the first stage (or stages) would be used to estimate ρ (or various ρ's) needed to design the remaining stages.

REFERENCES

Anderson, R. L. 1960. Uses of variance component analysis in the interpretation of biological experiments. Bulletin Int. Statist. Instit. 37:3:71-90.

Anderson, R. L. and T. A. Bancroft. 1952. Statistical Theory in Research. McGraw-Hill Book Company, Inc., New York.

Crump, P. P. 1954. Optimal designs to estimate the parameters of a variance component model. Unpublished Ph.D. Thesis, North Carolina State College, Raleigh, N. C.

Ganguli, M. 1941. A note on nested sampling. Sankhya 5:449-452.

Gaylor, D. W. 1960. The construction and evaluation of some designs for the estimation of parameters in random models. Unpublished Ph.D. Thesis, North Carolina State College, Raleigh. Institute of Statistics Mimeo Series No. 256.

Graybill, F. A. and A. W. Wortham. 1956. A note on uniformly best unbiased estimators for variance components. J. Amer. Stat. Assoc. 51:266-268.

Newton, R. G., M. W. Philpott, H. Fairfield-Smith and W. G. Wren. 1951. Variability of Malayan rubber. Industrial and Engineering Chemistry 43:329-334.

LECTURE ON HAZARD ANALYSIS

G. S. Watson
University of Toronto

INTRODUCTION. Suppose that the time X to failure of some article, randomly selected from a large batch of supposedly identical items, has a probability density and distribution functions $f(x)$, $F(x)$ with $f = F'$. Naturally $X \geq 0$. Then the time rate of failure of items on test at time x is

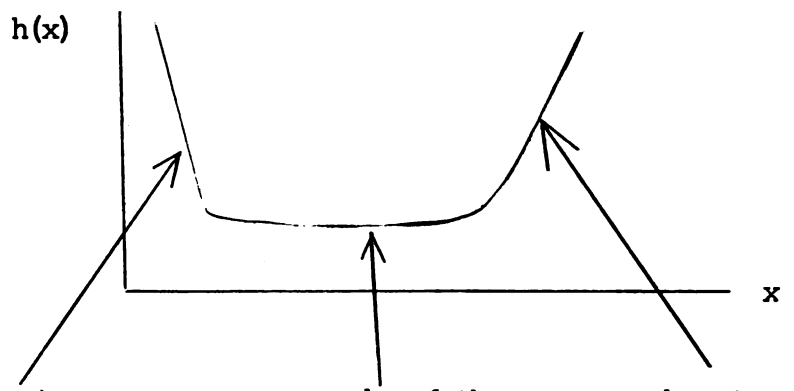
$$\lim_{\Delta \to 0} \left\{ \frac{\text{Prob}\ \{x < X \leq x + \Delta / X > x\}}{\Delta} = \frac{\text{Prob}\ \{x < X \leq x + \Delta\}}{\Delta\ \text{Prob}\ (X > x)} \right\}$$

$$= f(x) / (1 - F(x))$$

$$\equiv h(x), \text{ the Hazard Function}$$

For $f = \lambda e^{-\lambda x}$, $h(x) = \dfrac{\lambda e^{-\lambda x}}{1 - (1 - e^{-\lambda x})} = \lambda \qquad (0 \leq x \leq \infty)$.

In this case failures occur at the same rate, no matter how long the test is run. More typically in practice



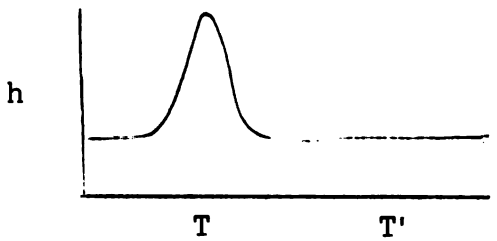
To understand the interpretation of hazard curves better, suppose that an item fails from either of two independent causes so that

$$X = \min (Y_1, Y_2)$$

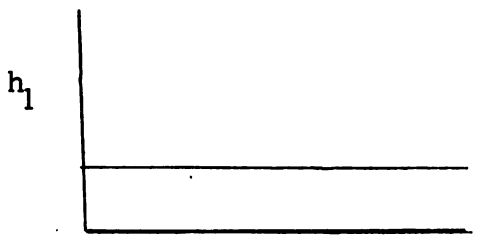
where Y_1 has hazard h_1 and Y_2 has hazard h_2. Then it is easily proved that

$$h(x) = h_1(x) + h_2(x).$$

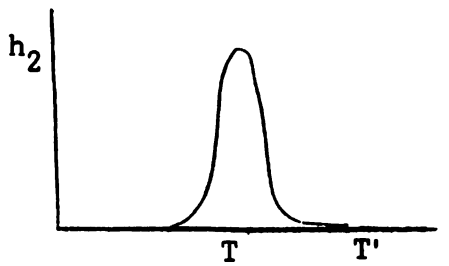
Thus, if in practice, we observed the curve



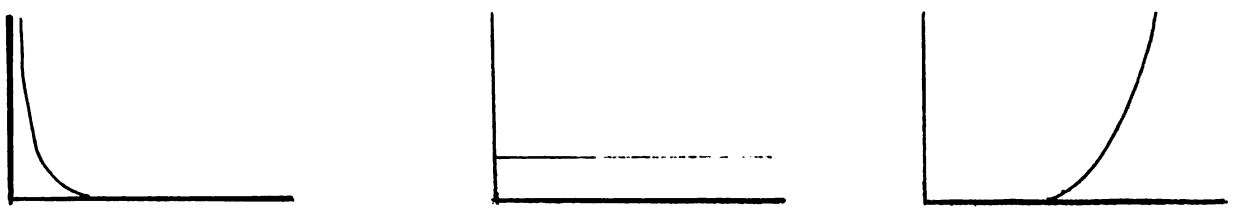
we would be tempted to suppose that it is the sum of



and



An examination of the items failing near time T might then show a different failure mode than those at, say T'. Further examination of all failures might substantiate this guess. Again the "typical" curve drawn above may well be composed of



and examination may reveal three failure modes.

Of course these decompositions are not mathematically necessary, since to <u>any</u> $h(x)$ there is a unique distribution given by

$$F(x) = 1 - \exp\left(-\int_0^x h(t)dt\right)$$

since $h(x) = -\frac{d}{dx}\left[\log_e (1 - F(x))\right]$.

They are merely sensible in the light of what may be expected in practice - different failure modes at different times. For this reason the hazard function seems to be fundamental for reliability-studies made in the development of a device. There are occasions when the hazard itself is required. For an item that is only required to operate successfully for a short time (like a rocket motor), it may be wise to run it on a test stand until the hazard has reached a minimum.

Thus we have made a practical case for estimating the hazard function. The usual approach in life-testing is to assume, sometimes without sound reasons, a specific form $f(x, \theta)$ for the density of x and merely to fit the parameters θ. If the hazard is then required it would begin as

$$\frac{f(x, \hat{\theta})}{1 - F(x, \hat{\theta})} .$$

This is particularly unsound when dealing with a new item. Our approach is to estimate, from a random sample of lives, $h(x)$, without making any prior assumptions about its form or the form of the density $f(x)$. We will refer to this approach and its ramifications as "Hazard Analysis".

There is an analogy in the history of the statistical analysis of stationary time series. Once the procedure was to assume specific models (such autoregressive and moving average processes) and fit the parameters. Now most writers press first for a <u>spectral analysis</u> of the data, i.e. they seek, without further assumptions, to estimate a function

$$s(w) \geq 0, \quad 0 \leq w \leq \pi, \quad 2\int_0^{\pi} s(w)dw = \sigma^2. \quad \text{(called the}$$

spectral density function) with the property that

$$E(X_i X_{i+j}) = \int_0^{\pi} s(w) \cos jw \, dw$$

where X_i and X_{i+j} are the ith and $(i+j)$th observations. Clearly $s(w)$ determines the covariance matrix of the process and conversely, so are mathematically equivalent (like h and f above). But $s(w)$ is more suggestive in practice because one may write

$$x_t = \int_0^{\pi} \cos tw \, du(w) + \int_0^{\pi} \sin tw \, dv(w)$$

where $\text{var}(du(w)) = \text{var}(dv(w)) = s(w)dw$, $\text{cov}(du(w), dv(w)) = 0$.

In other words, $2s(w)dw$ represents the variance of x_t in the frequency range $(w, w + dw)$. Thus peaks in $s(w)$ should correspond to favored frequences in the process generating x_t and so throw light on its nature--just as peaks in $h(x)$ throw light on the failure mechanism.

The two problems are also mathematically similar in one very basic respect - in both cases we are trying to estimate a function. In statistics, we usually try to estimate a point in a finite dimensional space. Here it is a point in a function space. There are other such problems, the most related one being the estimation of a probability density. This latter problem is mathematically a fundamental one and by studying it, we will see how to solve our original problem, the estimation of the hazard function.

The remarks of the last few minutes show us the place of one problem in mathematical statistics and give us a theoretical interest in its solution, regardless of the practical utility of our answer.

The work I will now report was carried out with O.N.R. support by M. R. Leadbetter of R.T.I. and myself.

2. MATHEMATICAL CONSIDERATIONS. Let $X_1 \ldots X_n$ be a random sample from a density $f(x)$ then for any function $q(\,.\,)$,

$$(1) \qquad E\left\{\frac{\sum_1^n q(X_i)}{n}\right\} = \int_{-\infty}^{\infty} q(x)\ f(x)\ dx.$$

Suppose that we want to know the value of f at y, $f(y)$. The most primitive method is to form a histogram from the sample and if I_y is the interval containing y, and n_y the number of $X_i \in I_y$, and to use

$$(2) \qquad \frac{n_y}{n\,||I_y||} \qquad \text{as dst } f(y)$$

the expectation of this estimator is

$$\frac{1}{||I_y||}\int_{I_y} f(x)dx \approx f(y) \text{ if } f(x) \text{ is "smooth" in } I_y \text{ for } I_y$$

small. This is the average of f in I_y. Thus (2) is an example of (1) with

$$q(x) = 1/||I_y||, \quad x \in I_y,$$

$$= 0 \qquad\quad , \quad x \notin I_y.$$

In order that this estimator (2) should be consistent it is necessary to have $||I_y|| \longrightarrow 0$ at the right rate as $n \longrightarrow \infty$.

It is well known to anyone who has ever made a histogram of some data that

too many intervals gives a histogram that is too rough or variable,

too few intervals " " " " " " smooth or lacking in interest.

The vaguely expressed balance that is required in practice can be expressed here mathematically.

If, instead of (1), we write

$$E\left(\hat{f}_n(y) = \frac{\sum_{i=1}^{n} \delta_n(y_i - x_i)}{n}\right) = \int_{-\infty}^{\infty} \delta_n(y-x) f(x)dx,$$

(a "smoothing" of f)

(3) $$\longrightarrow f(y),$$

if f is continuous at y and $f_n(\,.\,)$ tends to the Delta-Function. Now

(4) $$\operatorname{var}(\hat{f}_n(y)) = \frac{1}{n}\int_{-\infty}^{\infty} \delta_n^{\,2}(y - x)f(x)dx - \frac{1}{n}\left[\int \delta_n(y-x)f\ dx\right]^2.$$

The last term tends to zero as $n \longrightarrow \infty$ if (3) is true so that

$\operatorname{var}(\hat{f}_n) \longrightarrow 0$ provided

(5) $$\int \delta_n^{\,2}(y - x)\, f\,(x)dx = \sigma(n).$$

Thus $\delta_n(\ )$ must not tend too quickly to $\delta(\ )$ e.g. L.H.S. of (5) is, for the usual forms of $\delta_n(\ )$, less than $\delta_n(0)$ f(y) approximately so that $\delta_n(0) = \alpha(n)$ will suffice. This still leaves a large class of admissable sequences $\delta_n(\ .\ )$. An attractive criterion that leads to an "optimum" sequence is

$$J_n \equiv E\left(\int (f_n(y) - f(y))^2 dy\right)$$

(6)

$$= \int |\phi_{\hat{f}_n}(t) - \phi_f(t)|^2 dt,$$

by Parseval where

$$\phi_f = \int e^{iyt} f(y)dy, \quad \phi_{\hat{f}_n} = \int e^{iyt} \hat{f}_n(y)\, dy.$$

It may be shown that the characteristic function of the $\delta_n(\ )$ which minimizes J_n is given by

$$\phi_{\delta_n}(t) = \frac{|\phi_f(t)|^2}{\frac{1}{n} + \frac{(n-1)}{n}|\phi_f(t)|^2} \tag{7}$$

Thus $\delta_n(\ .\ )$ is symmetric about zero. Examples show that $\delta_n(\ .\ )$ is not always positive. The important point is that the optimum $\delta_n(\ .\ )$ depends on the unknown f. However, for $n \longrightarrow \infty$, the optimum $\delta_n(\ .\ )$ depends only on the rate at which $|\phi_f(t)| \longrightarrow 0$ as $|t| \longrightarrow \infty$. Our results in this direction are analogous to those of Parzen (1958) for the spectral density. To see how spectral analysis has a similar form mathematically, we have only to take the relation I mentioned before

$$E(X_i\ X_{i+j}) = \int \cos\ wj \cdot s(w)dw$$

and form

$$E(Q = \sum_i \sum_j a_j X_i X_{i+j}) =$$

$$\int \left[na_0 \cos\ w0 + (n-1)a_1 \cos\ w + \ldots + a_{n-1} \cos\ w(n-1)\right] s(w)dw$$

$$= \int q_n\ (w)s(w)dw,$$

where the coefficients a_j will be chosen to make the trigonometric polynomial $q_n(w)$ resemble $\delta(w_o - w)$ in order to estimate $s(w_o)$. In this problem the choice $\delta_n(\ .\ )$ is very restricted compared with ones previously considered.

Familiarity with the results above enables us to proceed quickly to

3. The estimation of the Hazard Function. Referring to the equations (1) and (3), we want a function $q(\ .\ )$ such that

$$(8) \qquad E(\frac{1}{n} \sum_{i=1}^{n} q(X_i)) = \int_{-\infty}^{\infty} \delta_n(y-x) \frac{f(x)}{1-F(x)}\ dx$$

which implies that

$$(9) \qquad q(x) = \frac{\delta_n(y-x)}{1-F(x)}$$

which is no good because of the presence of $F(x)$. However at $X_{r|n}$, the rth smallest observation, an estimate of $F(x)$ is r/n and so

$\frac{n-r}{n}$ is an estimate of $1 - F(x)$. The latter occurs in the denominator and may be zero so we suggest

$$\tilde{h}_n(y) = \frac{1}{n} \sum_{r=1}^{n} \frac{\delta_n(y - X_{r|n})}{\frac{n-r+1}{n}} ,$$

i.e.

$$\tilde{h}_n(y) = \sum_{r=1}^{n} \frac{\delta_n(y - X_{r|n})}{n-r+1} . \tag{10}$$

This then is a most natural estimator, following along the path of Section 2.

Since the probability density of $X_{r|n}$ is given by

$$\frac{n!}{(r-1)!\,(n-r)!} \; F_x^{r-1} \; (1 - F_x)^{n-r} \; f(x),$$

$$E(\tilde{h}_n(y)) = \sum_{r=1}^{n} \frac{n!}{(r-1)!\,(n-r+1)!} \int_0^\infty \delta_n(y-x) F_x^{r-1} (1-F_x)^{n-r} f(x)\,dx,$$

$$= \int_0^\infty \frac{\delta_n(x-y)}{1 - F(y)} \left\{1 - F(y)^n\right\} f(y)\,dy,$$

$$\longrightarrow h(y) ,$$

with some mild restrictions on $\delta_n(\,.\,)$ and $f(\,.\,)$. Several more mild conditions ensure that

$$\text{var}\ (\tilde{h}_n(y)) \longrightarrow 0 \quad , \quad \text{as} \quad n \longrightarrow \infty .$$

The arguments and conditions are rather lengthy and I will omit them. The initial simplicity of our problem is seen to be deceptive - it seems to be due to the fact that we are dealing with a ratio.

It is again true that many sequences $\delta_n(\,.\,)$ will suffice, and, for theoretical completeness, one would like to see a criterion that would select an optimum sequence. So far however none of the obvious tactics have yielded a solution. One cannot use

$$\min_{\delta_n} \mathbf{E} \int_0^\infty (\hat{h}_n(y) - h(y))^2 dy$$

because the integral will not converge in cases of interest, in contrast to the probability and spectral density problems. A convergence factor could be inserted but this means more arbitrariness; however a solution along these lines is surely possible. Work in this direction has been temporarily suspended in order to examine in more detail estimators of this and other types. For practical applications this seems more useful since, as we have seen in § 2 for the probability density, the optimum $\delta_n(\,.\,)$, for any criterion, will depend strongly on the unknown $h(x)$, at least in small samples and that is what we always have in practice.

The most primitive estimator of h is again that formed from the histogram. Let $(0,\infty) = I_1 + I_2 + \ldots + I_k$ and

$$\begin{cases} n_j = \text{number of } X_i \in I_j, & j = 1, \ldots, k \\ \sum n_j = n, \end{cases}$$

and define

$$(12) \qquad h^*_{j,n} = \frac{1}{\| I_j \|} \cdot \frac{n_j}{n - n_1 - \cdots - n_{j-1} + 1}.$$

Then if $\| I_j \| \longrightarrow 0$ as $n \longrightarrow \infty$, $E(h^*_{j,n}) \longrightarrow h$.

However this estimator can be closely matched by one of the form $\tilde{h}_n$. If

$$(13) \qquad \begin{cases} I_j = \left\{ x \mid x \in \left(y - \frac{1}{2\delta}, \; y + \frac{1}{2\delta} \right) \right\}, \\ \delta_n(z) = \begin{cases} \delta & \text{if } -\frac{1}{2\delta} \le z \le \frac{1}{2\delta} \\ 0 & \text{otherwise} \end{cases}, \end{cases}$$

then

$$(14) \qquad \tilde{h}_n = \delta \sum_{r \in R} \frac{1}{n - r + 1}$$

with

$$R = \left\{ r \text{ ranks of } X_i \mid y - \frac{1}{2\delta} \le X_r \le y + \frac{1}{2\delta} \right\}.$$

However if

$$\begin{cases} r_1 = \text{least member of } R \\ r_2 = \text{greatest member of } R \end{cases}$$

(14) may be rewritten as

$$\tilde{h}_n = \delta \sum_{r_1}^{r_2} \frac{1}{n - r + 1}$$

Approximating the sum by an integral we find

$$(15) \qquad \delta \log_e \frac{n - r_1 + 1}{n - r_2} > \tilde{h}_n > \delta \log_e \frac{n - r_1 + 2}{n - r_2 + 1} .$$

These bounds could also be used. In this way a number of estimators, connected by inequalities, can be devised that are based on data that are contained in the histogram.

A graphical estimation method can be based on the relation

$$h(x) = -\frac{d}{dx} \left\{ \log_e (1 - F(x)) \right\}.$$

For let $F_n(x)$ be the sample distribution function and plot $\log_e \left\{1 - F_n(x)\right\}$ against x. If a smooth curve is drawn through these points and its slope determined, perhaps just with a straight edge, we obtain an estimator that might be called

$$(16) \qquad \hat{h}_g(x).$$

This estimator, by its construction, does not come with formulae for its mean and variance.

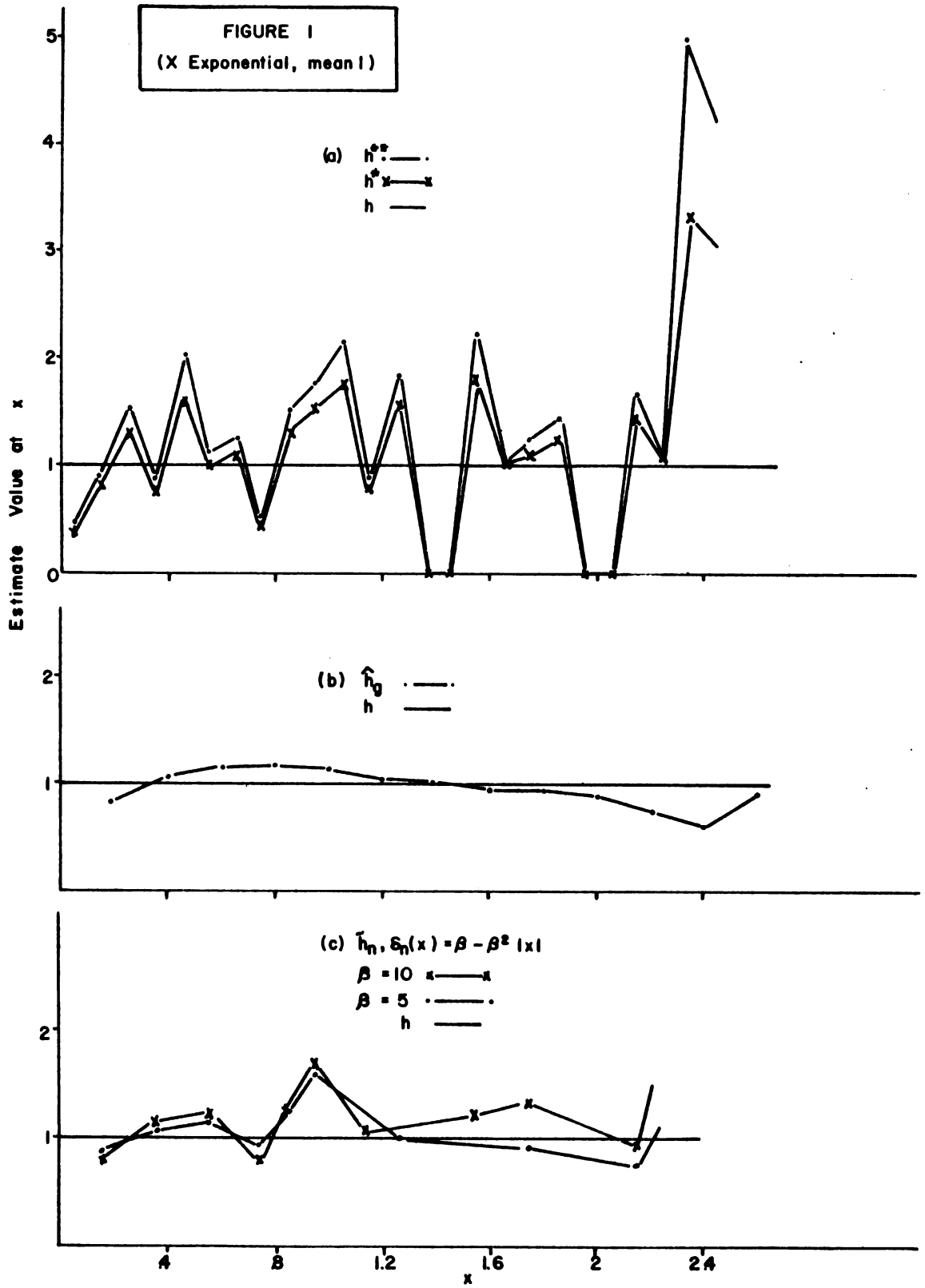
4. NUMERICAL EXAMPLES. To test out the various methods, samples of 50 were drawn from 4 populations with respectively, constant, never increasing and (2 cases) never decreasing hazards. While in practice we will often be concerned with the detection of peaks, the hazards here are all smooth. Since the class intervals (histogram type estimates) and $\delta_n(\,.\,)$, which we call the "hazard window" are at our disposal, it is possible here to vary them so that we get estimated functions that are as close as possible to the known functions. In particular since we know there are no peaks, we can use broader intervals and hazard windows than one might dare to use in practice.

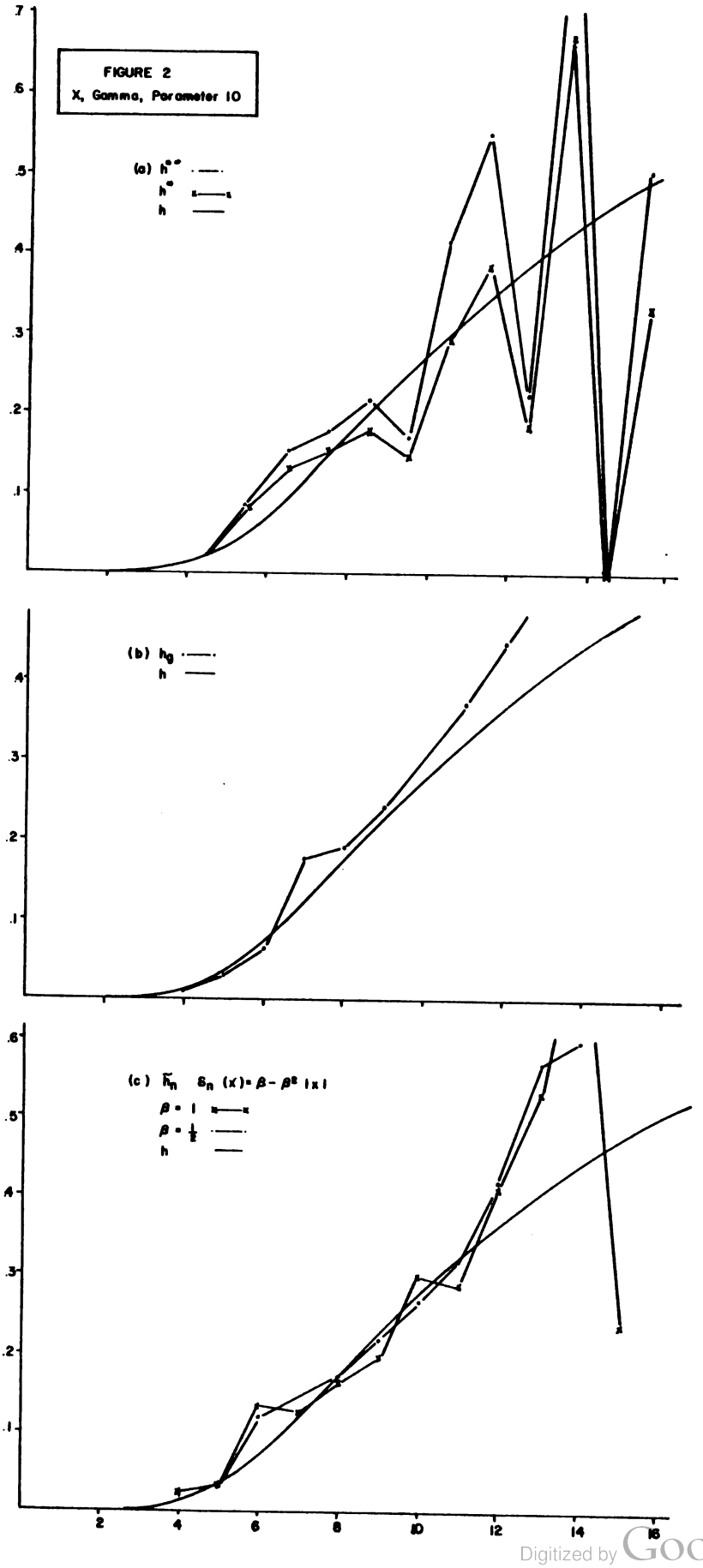
On looking at the graphs† of these experiments, it is clear that the simplest histogram estimators $h^*_{j,n}$, $h^{**}_{j,n}$ are the worst. The graphical estimator $\hat{h}_g$ does very well because it involves heavy smoothing (and we have smooth hazards). $\tilde{h}_n$ was computed using a triangular window, since we had no reason for choosing anything more complicated. It too does quite well and has the advantage of being treatable mathematically and amenable to automatic machine calculation. The results suggested that $\tilde{h}_n$ was unreliable at any point depending on too few observations. We are therefore now considering a more general form of $\tilde{h}_n$ where $\delta_n(\,)$'s form depends to some extent on the observations.

In conclusion it is evident that we have just begun this investigation. We are a long way from the reliable detection of peaks and from confidence bands for $h(x)$. The long investigation of probability density estimation was a necessary preliminary.

---

† Graphs are on the following pages.





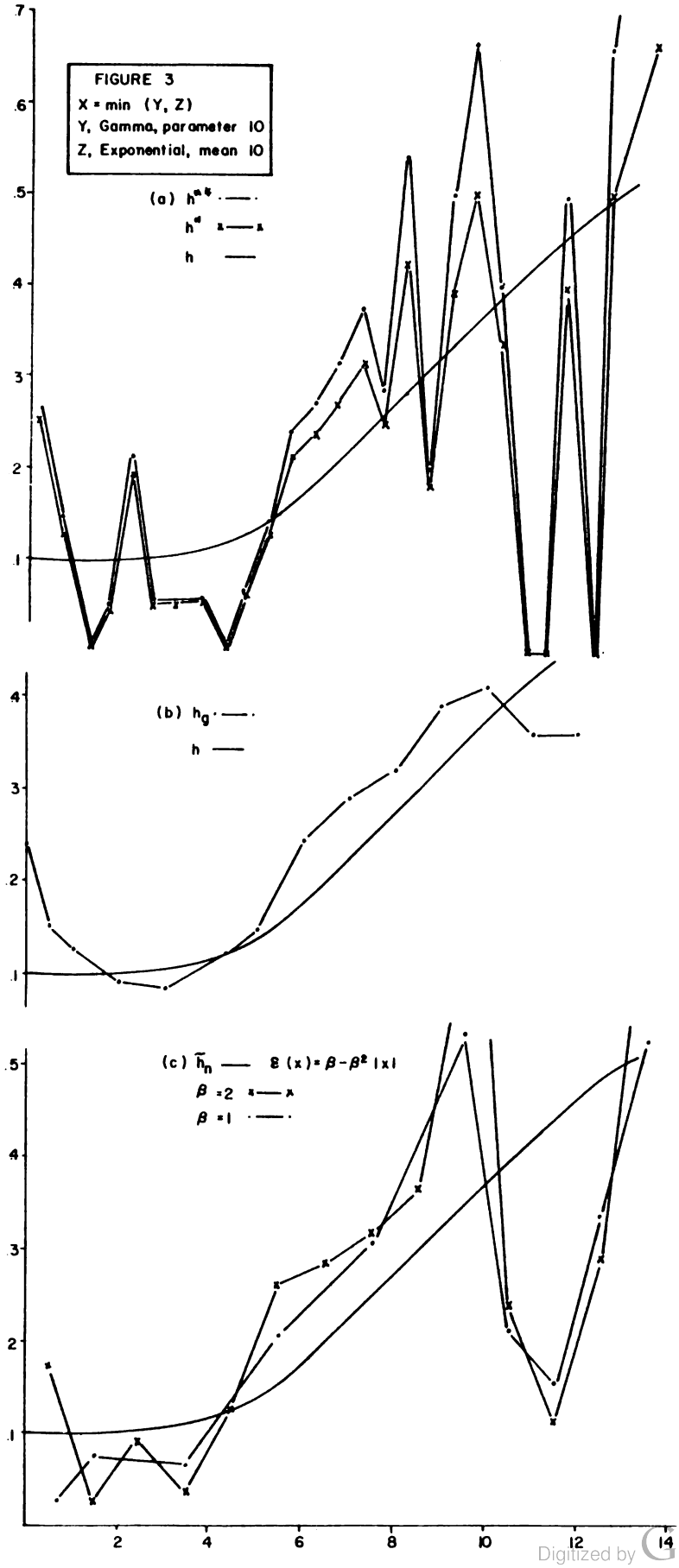
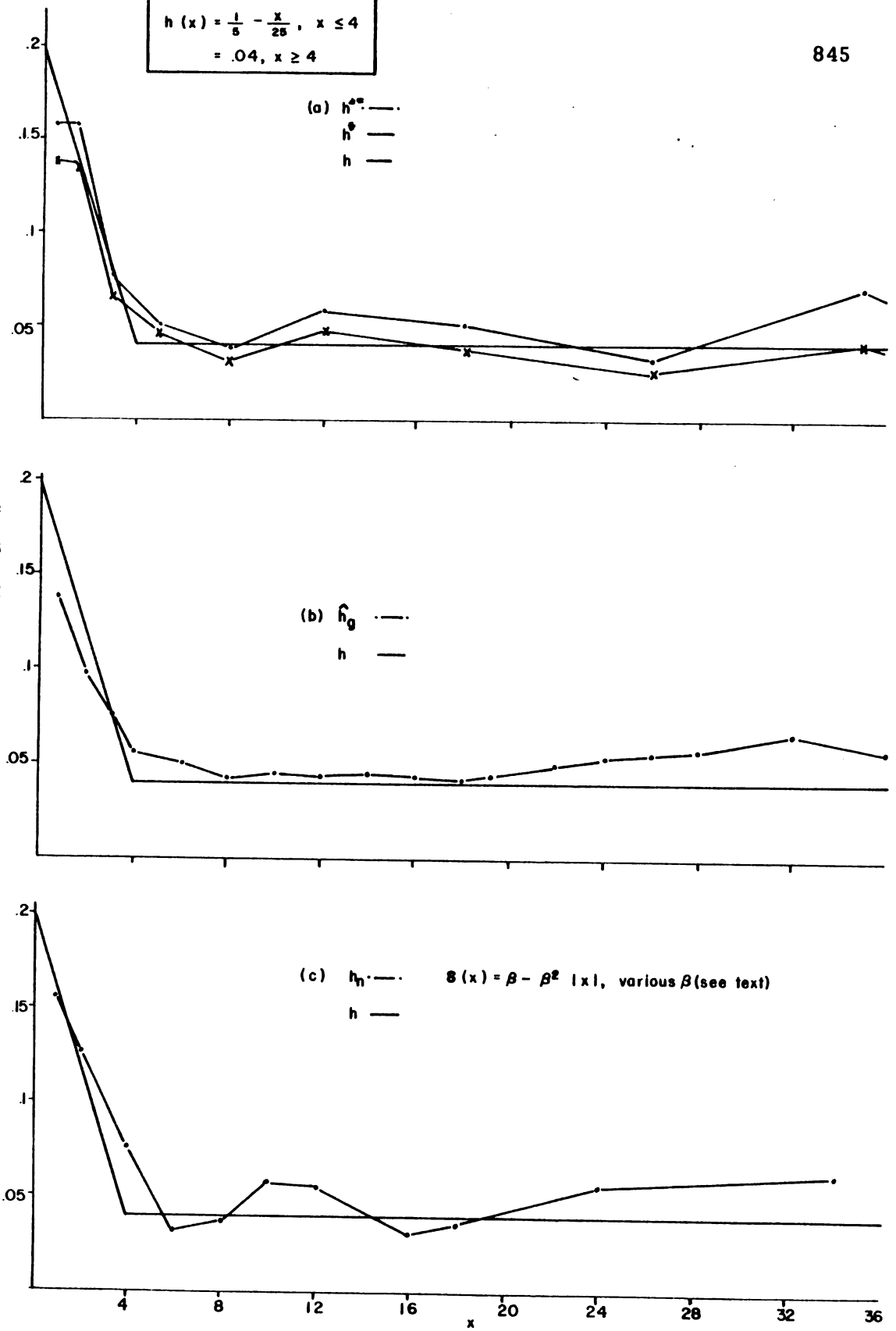


FIGURE 4

$h(x) = \frac{1}{5} - \frac{x}{25}$, $x \le 4$

$= .04$, $x \ge 4$



ATTENDEES TO THE SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

18-20 Oct. 1961

Abraham, John K.	BRL, Aberdeen Proving Ground, Maryland
Addelman, Dr. Sidney	Research Triangle Institute, Durham, N. C.
Anderson, R. L.	North Carolina State College, Raleigh, N. C.
Anscombe, F. J.	Princeton University, Princeton, N. J.
Armstrong, J.	USASRDL, Ft. Monmouth, N. J.
Beaube, Capt. George P.	USASADEA, Ft. Meade, Maryland
Bechhofer, Robert	Cornell University, Ithaca, N. Y.
Berger, Dr. Agnes	Columbia University, New York, N. Y.
Biser, Dr. Erwin	USASRDL (Systems), Ft. Monmouth, N. J.
Bombara, E. L.	USARGMA, Redstone Arsenal, Huntsville, Ala.
Boyle, Douglas G.	Hercules Powder Co., Salt Lake, Utah
Bromfield, Dr. K. R.	U. S. Army Biological Labs., Ft. Detrick, Md.
Brown, Ralph E.	Frankford Arsenal, Philadelphia, Pa.
Brown, William A.	Dugway Proving Grounds, Dugway, Utah.
Bruno, O. P.	Aberdeen Proving Ground (BRL), Md.
Bulfinch, A.	Picatinny Arsenal, Dover, N. J.
Chlupsa, J. Matt	Picatinny Arsenal, Dover, N. J.
Clark, David	USASRDL, Ft. Monmouth, N. J.
Cohen, A. C., Jr.	University of Georgia, Athens, Georgia
Cowie, William S.	QMR&E Command, Natick, Massachusetts
DeArmon, Ira A., Jr.	U. S. Army Chemical Corps, ACC, Md.
DeQuoy, Alfred W.	Strat. Tactics & Analysis Group, Bethesda, Md.
Donaldson, John L.	Research Analysis Corporation, Bethesda, Md.
Doner, R. D.	ARGMA, Redstone Arsenal, Huntsville, Ala.
Dressel, F. G.	ARO, Durham, N. C.
Eckard, Cecil O.	U. S. Army Biological Warfare Lab., Ft. Detrick
Eckstein, A.	USASRDL, Ft. Monmouth, N. J.
Einbinder, Seymour K.	Picatinny Arsenal, Dover, N. J.
Eissner, Robert M.	Aberdeen Proving Ground, Md.
Farnochi, A.	USASRDL, Ft. Monmouth, N. J.
Feyder, F. O.	RCA Electron Tube Division, Harr., N. J.
Fischer, M.	USNTDC, Port Washington, N. Y.

Fisher, Lt. Col. Paul A.	Strat. & Tact. Analysis Group, Bethesda, Md.
Fitch, Dr. David J.	TAGO, Washington D. C. (Research & Dev. Com)
Floyd, William B.	Litton Systems, Inc., Waltham, Mass.
Freiberger, Professor W.	Brown University, Providence,R. I.
Frishman, Fred	ARO, Washington D. C.
Fulton, C. L.	Frankford Arsenal, Philadelphia, Pa.
Galbraith, Allen S.	ARO, Durham, N. C.
Gergen, J. J.	Duke University, Durham, N. C.
Gilford, Mrs. Dorothy M.	Office of Naval Research, Washington D. C.
Geohner, G. H.	Tung Sol Electric, Inc., Bloomfield, N. J.
Goroff, L.	USASRDL
Grant, Roy L.	U. S. Bureau of Mines, Pittsburgh, Pa.
Green, Lyle D.	University of Michigan, Ann Arbor, Mich.
Grubbs, Dr. F. E.	Aberdeen Proving Ground, Md. (BRL)
Gurland, John	Mathematics Research Center, Madison, Wis.
Haines, Bertram W.	Ft. Detrick, Frederick, Md.
Hamblin, James M.	Office Chief of Ordnance, Washington, D. C.
Hammer, Alexander	Springfield Armory, Springfield, Mass.
Hammersley, John M.	Princeton, University, Princeton, N. J.
Harshbarger, Boyd	Virginia Polytechnic Institute, Blacksburg, Va.
Hasty, Willis L.	Ft. Detrick, Md. U. S. Army Biol. Warfare Lab.
Hecht, Edward C.	Picatinny Arsenal, Dover, N. J.
Heifetz, Sam	Ballistic Research Lab. Md.
Hobbs, R. H.	Research Analysis Corp., Bethesda, Md.
Hooke, Dr. Robert	Westinghouse Research Lab., Pittsburgh, Pa.
Horner, T. W.	Booz, Allen Applied Research, Bethesda, Md.
Howard, B. A .,Jr.	Ordnance Weapons Command, Rock Island, Ill.
Howes, David R.	U. S. Army Chem. Center, Md.
Hunter, J. S.	Princeton University, Princeton, New Jersey
Inselmann, E.	Frankford Arsenal, Philadelphia, Pa.
Isaac, Gerhard	U. S. Army Medical Research & Nutrition Lab., Denver, Colo.
Jebe, Emil H.	University of Michigan, Ann Arbor, Mich.
Johnson, Cecil D.	TAGO, Washington, D. C.
Johnson, J. R.	BRL, Aberdeen Proving Ground, Md.
Kahn, Lt. Robert I.	TRGCOM, Ft. Eustis, Va.
Kaplan, L. L.	USASRDL, Ft. Monmouth, N. J.
Karp, A. E.	Aberdeen Proving Ground, Md.

Keefer, John H.	Ballistic Research Laboratory
Kingsley, O. L.	White Sands Missile Range, N. Mexico
Krane, Scott A.	CEIR Inc., Dugway, Utah
Kurkjian, B.	DOFL, Washington D. C.
Laderman, Jack	ONR, New York, N. Y.
Lehnigh, Siegfried H.	ARGMA, Redstone Arsenal, Huntsville, Ala.
Lelivelt, John B.	USASRDL, Ft. Monmouth, N. J.
Leskowitz, Leon	USASRDL, Ft. Monmouth, N. J.
Levin, S. G.	DOFL, Washington D. C.
Liller, P. R.	RCA Electron Tube Division, Lancaster, Pa.
Ling, Dr. Rufus C.	STAG, Bethesda, Md.
Love , Benjamin T.	USAEPG, Ft. Huachuca, Arizona
Lucas, H. L.	North Carolina State Collefe, Raleigh, N. C.
Lyon, D. A.	Suffield Expt'l Station, Ralston,Alberta,Canada
Mallios, W. S.	Institute of Stat. of N. Car., Raleigh, N. C.
Maloney, Clifford J.	Ft. Detrick , Frederick, Md.
Mantel, Nathan	Nat'l Cancer Institute, Bethesda, Md.
Marcous, Lawrence	U. S. Army Naval Eng. Exp. Station, Annapolis Maryland
Matteis, Richard J.	Research Analysis Corp., Bethesda, Md.
McAreavy, John	OMETA, OWC, Rock Island, Ill.
McCormick, Garth	Research Analysis Corp., Bethesda, Md.
McKeague, Jr., Robert I.	Ordnance Ammunition Command, Joliet, Ill.
Meehan, J. J.	RCA Semiconductor Division, N. J.
Millman, George	USASRDL Systems Division, Asbury Park, N. J.
Mindes, A. S.	Picatinny Arsenal, Dover, N. J.
Mitchel, A. V.	Tung-Sol Electric, Bloomfield, N. J.
Moss, J. H.	RAC, Washington, D. C.
Motyka, Peter	Picatinny Arsenal, Dover, N. J.
Neal, Donald M.	Watertown Arsenal, Watertown, Mass.
Neyman, J.	University of California, Berkeley, Cal.
Norden, Monroe L.	New York University, College of Engineering, New York, N. Y.
Olson, K. W.	USMA, West Point, N. Y.
O'Neil, James F.	Springfield Armory, Springfield, Mass.
Orleans, Beatrice	Bureau of Ships R & D, Washington, D. C.
Ostrofsky, Dr. Morris	Westinghouse Research Lab., Pittsburgh, Pa.
Parks, A. B.	DOFL, Washington D. C.
Parrish, Gene B.	ARO, Durham, N. C.

Pascual, Michael	Watervliet Arsenal, Watervliet, N. Y.
Perlman, Sol	USAS Radio Propagation Agency, Ft. Mon., N. J.
Pethel, Frank C.	USASRDL, Ft. Monmouth, N. J.
Prairie, R. R.	North Carolina State College, Raleigh, N. C.
Purtell, John P.	Watervliet Arsenal, Watervliet, N. Y.
Rabson, Gustave	University of Michigan, Ann Arbor, Mich.
Rappaport, Donald C.	Ammo Group, Picatinny Arsenal, Dover, N.J.
Reynolds, George W.	USAEPG, Ft. Huachuca, Arizona
Rhodes, Richard R.	Frankford Arsenal, Philadelphia, Pa.
Rouschex, William V.	Tung-Sol Electric, Newark, N. J.
Rush, Harold R.	QMR&E, Ft. Lee, Virginia
Saelens, R. G.	USASRDL, Ft. Monmouth, N. J.
Santone, L. C.	Frankford Arsenal, Philadelphia, Pa.
Scharber, Walter C.	Electron Tube Dir., E. C. Department, Belmount, N. J. (U.S. Army S&R Dev. Lab.)
Scott, E. L.	University of California, Berkeley, Cal.
Selig, Seymour M.	Office Naval Research, Washington, D. C.
Sewell, Walter E.	ARO, Durham, N. C.
Shaw, Thomas R.	Operations Research Inc., Silver Springs, Md.
Sherman, H.	Tung-Sol Electric, Inc., Bloomfield, N. J.
Sirota, Milton	Picatinny Arsenal, Dover, N. J.
Sloane, Harry S.	Dugway Proving Ground, Dugway, Utah.
Smith, Jr., Harry G.	Picatinny Arsenal, Dover, N.J.
Stanley, Jerry	USASRDL, Ft. Monmouth, N. J.
Starr, Selig	Army Research Office, Washington D. C.
Steiger, B. F.	Tung-Sol Electric, Livingston, N. J.
Stephanides, Mrs. Agatha	U. S. Army Strat & Tact Analysis Group, Bethesda, Md.
Suhler, William C.	Research Analysis Corp., Bethesda, Md.
Taravella, D. J.	Picatinny Arsenal FREL, Dover, N. J.
Taylor, George W.	ARO, Durham, North Carolina
Taylor, W. Bruce	Research Analysis Corp., Bethesda, Md.
Thompson, Jr., W. A.	University of Delaware, Newark, Del.
Thrall, Robert M.	University of Michigan, Ann Arbor, Mich.
Trigg, Col. Thomas K.	USASRDL, Ft. Monmouth, N. J.
Vivona, Stefano (Lt. Col.)	Walter Reed Army Institute of Research Washington D. C.

Wadley, F. M.	U. S. Army Biological Warfare Lab., Ft. Detrick, Md.
Walner, Arthur H.	N. Y. Naval Shipyard, Material Lab., New York,
Watson, G. S.	University of Toronto, Toronto, Ontario Canada
Watterson, Geoffrey	Virginia Polytechnic Institute, Blacksburg, Va.
Weinstein, Joseph	USASRDL, Ft. Monmouth, N. J.
Weintraub, Gertrude	Picatinny Arsenal, Dover, N. J.
West, C. A.	RCA Electron Tube Division, Harrington, N. J.
White, John E.	Bridgewater College, Bridgewater, Va.
Wilks, S. S.	Princeton University, Princeton, N. J.
Wood, Dr. Walter F.	QMR & E Center, Natick, Mass. U. S. Army R & E Command.
Zelen, Marvin	MRC, Madison 6, Wisconsin
Zoellner, John A.	Armour Research Foundation, Annapolis, Md.