AMC PAMPHLET

AMCP 706-110

BEST
AVAILABLE COPY

ENGINEERING DESIGN HANDBOOK

EXPERIMENTAL STATISTICS

SECTION 1

BASIC CONCEPTS AND ANALYSIS OF MEASUREMENT DATA

HEADQUARTERS, U.S. ARMY MATERIEL COMMAND

DECEMBER 1969

HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D.C. 20315

16 December 1969

AMC PAMPHLET
No. 706-110*

ENGINEERING DESIGN HANDBOOK

EXPERIMENTAL STATISTICS (SEC 1)

Paragraph *Page*

CHAPTER 1

SOME BASIC STATISTICAL CONCEPTS AND PRELIMINARY CONSIDERATIONS

CHAPTER 2

CHARACTERIZING THE MEASURED PERFORMANCE OF A MATERIAL, PRODUCT, OR PROCESS

*This pamphlet supersedes AMCP 706-110, 31 July 1963.

Paragraph *Page*

CHAPTER 2 (Cont)

CHAPTER 3

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO AVERAGE PERFORMANCE

Paragraph *Page*

CHAPTER 3 (Cont)

CHAPTER 4

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO VARIABILITY OF PERFORMANCE

Paragraph *Page*

CHAPTER 4 (Cont)

CHAPTER 5

CHARACTERIZING LINEAR RELATIONSHIPS BETWEEN TWO VARIABLES

Paragraph *Page*

CHAPTER 5 (Cont)

CHAPTER 6

POLYNOMIAL AND MULTIVARIABLE RELATIONSHIPS ANALYSIS BY THE METHOD OF LEAST SQUARES

LIST OF ILLUSTRATIONS

LIST OF TABLES

FOREWORD

INTRODUCTION

This is one of a group of handbooks covering the engineering information and quantitative data needed in the design, development, construction, and test of military equipment which (as a group) constitute the Army Materiel Command Engineering Design Handbook.

PURPOSE OF HANDBOOK

The Handbook on Experimental Statistics has been prepared as an aid to scientists and engineers engaged in Army research and development programs, and especially as a guide and ready reference for military and civilian personnel who have responsibility for the planning and interpretation of experiments and tests relating to the performance of Army equipment in the design and developmental stages of production.

SCOPE AND USE OF HANDBOOK

This Handbook is a collection of statistical procedures and tables. It is presented in five sections, viz:

AMCP 706-110, Section 1, Basic Concepts and Analysis of Measurement Data (Chapters 1-6)

AMCP 706-111, Section 2, Analysis of Enumerative and Classificatory Data (Chapters 7-10)

AMCP 706-112, Section 3, Planning and Analysis of Comparative Experiments (Chapters 11-14)

AMCP 706-113, Section 4, Special Topics (Chapters 15-23)

AMCP 706-114, Section 5, Tables

Section 1 provides an elementary introduction to basic statistical concepts and furnishes full details on standard statistical techniques for the analysis and interpretation of measurement data. Section 2 provides detailed procedures for the analysis and interpretation of enumerative and classificatory data. Section 3 has to do with the planning and analysis of comparative experiments. Section 4 is devoted to consideration and exemplification of a number of important but as yet non-standard statistical techniques, and to discussion of various other special topics. An index for the material in all four sections is placed at the end of Section 4. Section 5 contains all the mathematical tables needed for application of the procedures given in Sections 1 through 4.

An understanding of a few basic statistical concepts, as given in Chapter 1, is necesssary; otherwise each of the first four sections is largely independent of the others. Each procedure, test, and technique described is illustrated by means of a worked example. A list of authoritative references is included, where appropriate, at the end of each chapter. Step-by-step instructions are given for attaining a stated goal, and the conditions under which a particular procedure is strictly valid are stated explicitly. An attempt is made to indicate the extent to which results obtained by a given procedure are valid to a good approximation when these conditions are not fully met. Alternative procedures are given for handling cases where the more standard procedures cannot be trusted to yield reliable results.

The Handbook is intended for the user with an engineering background who, although he has an occasional need for statistical techniques, does not have the time or inclination to become an expert on statistical theory and methodology.

The Handbook has been written with three types of users in mind. The first is the person who has had a course or two in statistics, and who may even have had some practical experience in applying statistical methods in the past, but who does not have statistical ideas and techniques at his fingertips. For him, the Handbook will provide a ready reference source of once familiar ideas and techniques. The second is the person who feels, or has been advised, that some particular problem can be solved by means of fairly simple statistical techniques, and is in need of a book that will enable him to obtain the solution to his problem with a minimum of outside assistance. The Handbook should enable such a person to become familiar with the statistical ideas, and reasonably adept at the techniques, that are most fruitful in his particular line of research and development work. Finally, there is the individual who, as the head of, or as a member of a service group, has responsibility for analyzing and interpreting experimental and test data brought in by scientists and engineers engaged in Army research and development work. This individual needs a ready source of model work sheets and worked examples corresponding to the more common applications of statistics, to free him from the need of translating textbook discussions into step-by-step procedures that can be followed by individuals having little or no previous experience with statistical methods.

It is with this last need in mind that some of the procedures included in the Handbook have been explained and illustrated in detail twice: once for the case where the important question is whether the performance of a new material, product, or process exceeds an established standard; and again for the case where the important question is whether its performance is not up to the specified standards. Small but serious errors are often made in changing "greater than" procedures into "less than" procedures.

AUTHORSHIP AND ACKNOWLEDGMENTS

The Handbook on Experimental Statistics was prepared in the Statistical Engineering Laboratory, National Bureau of Standards, under a contract with the Department of Army. The project was under the general guidance of Churchill Eisenhart, Chief, Statistical Engineering Laboratory.

Most of the present text is by Mary G. Natrella, who had overall responsibility for the completion of the final version of the Handbook. The original plans for coverage, a first draft of the text, and some original tables were prepared by Paul N. Somerville. Chapter 6 is by Joseph M. Cameron; most of Chapter 1 and all of Chapters 20 and 23 are by Churchill Eisenhart; and Chapter 10 is based on a nearly-final draft by Mary L. Epling.

Other members of the staff of the Statistical Engineering Laboratory have aided in various ways through the years, and the assistance of all who helped is gratefully acknowledged. Particular mention should be made of Norman C. Severo, for assistance with Section 2, and of Shirley Young Lehman for help in the collection and computation of examples.

Editorial assistance and art preparation were provided by John I. Thompson & Company, Washington, D. C. Final preparation and arrangement for publication of the Handbook were performed by the Engineering Handbook Office, Duke University.

Appreciation is expressed for the generous cooperation of publishers and authors in granting permission for the use of their source material. References for tables and other material, taken wholly or in part, from published works, are given on the respective first pages.

Elements of the U. S. Army Materiel Command having need for handbooks may submit requisitions or official requests directly to the Publications and Reproduction Agency, Letterkenny Army Depot, Chambersburg, Pennsylvania 17201. Contractors should submit such requisitions or requests to their contracting officers.

Comments and suggestions on this handbook are welcome and should be addressed to Army Research Office-Durham, Box CM, Duke Station, Durham, North Carolina 27706.

PREFACE

This listing is a guide to the Section and Chapter subject coverage in all Sections of the Handbook on Experimental Statistics.

Chapter No. — *Title*

CHAPTER 1

SOME BASIC STATISTICAL CONCEPTS AND PRELIMINARY CONSIDERATIONS

1-1 INTRODUCTION

Statistics deals with the collection, analysis, interpretation, and presentation of numerical data. Statistical methods may be divided into two classes—descriptive and inductive. Descriptive statistical methods are those which are used to summarize or describe data. They are the kind we see used everyday in the newspapers and magazines. Inductive statistical methods are used when we wish to generalize from a small body of data to a larger system of similar data. The generalizations usually are in the form of estimates or predictions. In this handbook we are mainly concerned with inductive statistical methods.

1-2 POPULATIONS, SAMPLES, AND DISTRIBUTIONS

The concepts of a *population* and a *sample* are basic to inductive statistical methods. Equally important is the concept of a *distribution.*

Any finite or infinite collection of individual things—objects or events—constitutes a *population.* A population (also known as a universe) is thought of not as just a heap of things specified by enumerating them one after another, but rather as an aggregate determined by some property that distinguishes between things that do and things that do not belong. Thus, the term *population* carries with it the connotation of completeness. In contrast, a *sample,* defined as a portion of a population, has the connotation of incompleteness.

Examples of populations are:

(a) The corporals in the Marines on July 1, 1956.

(b) A production lot of fuzes.

(c) The rounds of ammunition produced by a particular production process.

(d) Fridays the 13th.

(e) Repeated weighings of the powder charge of a particular round of ammunition.

(f) Firings of rounds from a given production lot.

In examples (a), (b), and (c), the "individuals" comprising the population are material objects (corporals, fuzes, rounds); in (d) they are periods of time of a very restricted type; and in (e) and (f) they are physical operations. Populations (a) and (b) are clearly finite, and their constituents are determined by the official records of the Marine Corps and the appropriate production records, respectively. Populations (c), (d), and (e) are conceptually infinite. Offhand, the population example (f) would seem to be finite, because firing is a destructive operation; but in order to allow for variation in quality among "firings" performed in accordance with the same general procedure it is sometimes useful, by analogy with repetitive weighings, to regard an actual firing as a sample of size one from a conceptually infinite population of "possible" firings, any one of which might have been associated with the particular round conceived. In this connection, note that in examples (e) and (f) the populations involved are not completely defined until the weighing and firing procedures concerned have been fully specified.

Attention to some characteristic of the individuals of a population that is not the same for every individual leads immediately to recognition of the *distribution* of this characteristic in the population. Thus, the heights of the corporals in the Marines on July 1, 1956, the burning times of a production lot of fuzes, and the outcomes of successive weighings of a powder charge ("observed weights" of the charge) are examples of distributions. The presence or absence of an attribute is a characteristic of an individual in a population, such as "tatooed" or "not tatooed" for the privates in the Marines. This kind of characteristic has a particularly simple type of distribution in the population.

Attention to one, two, three, or more characteristics for each individual leads to a univariate, bivariate, trivariate, or multivariate distribution in the population. The examples of populations given previously were examples of univariate distributions. Simultaneous consideration of the muzzle velocities and weights of powder charges of rounds of ammunition from a given production process determines a bivariate distribution of these characteristics in the population. Simultaneous recognition of the frequencies of each of a variety of different types of accidents on Friday the 13th leads to a multivariate distribution. In connection with these examples, note that, as a general principle, the distribution of a characteristic or a group of characteristics in a population is not completely defined until the method or methods of measurement or enumeration involved are fully specified.

The distribution of some particular property of the individuals in a population is a collective property of the population; and so, also, are the average and other characteristics of the distribution. The methods of inductive statistics enable us to learn about such population characteristics from a study of samples.

An example will illustrate an important class of derived distributions. Suppose we select 10 rounds of ammunition from a given lot and measure their muzzle velocities when the rounds are fired in a given test weapon. Let $\bar{X}$ be the average muzzle velocity of the 10 rounds. If the lot is large, there will be many different sets of 10 rounds which could have been obtained from the lot. For each such sample of 10 rounds, there will correspond an average muzzle velocity $\bar{X}_i$. These averages, from all possible samples of 10, themselves form a distribution of sample averages. This kind of distribution is called the *sampling distribution of* $\bar{X}$ *for samples of size 10* from the population concerned. Similarly, we may determine the *range* R of muzzle velocities (i.e., the difference between the largest and the smallest) for each of all possible samples of 10 rounds each. These ranges R_i $(i = 1, 2, \ldots)$ collectively determine the *sampling distribution of the range* of muzzle velocities *in samples of size 10* from the population concerned. The methods of inductive statistics are based upon the mathematical properties of sampling distributions of sample *statistics* such as $\bar{X}$ and R.

Let us summarize: A population in Statistics corresponds to what in Logic is termed the "universe of discourse"—it's what we are talking about. By the methods of inductive statistics we can learn, from a study of samples, only about population characteristics—only about *collective* properties of the populations represented by the individuals in the samples—not about characteristics of specific individuals with unique idiosyncrasies. The population studied may be large or small, but there must be a population; and it should be well defined. The characteristic of interest must be a collective property of the population.

1-3 STATISTICAL INFERENCES AND SAMPLING

1-3.1 STATISTICAL INFERENCES

If we were willing or able to examine an entire population, our task would be merely that of describing that population, using whatever numbers, figures, or charts we cared to use. Since it is ordinarily inconvenient or impossible to observe every item in the population, we take a sample—a portion of the population. Our task is now to generalize from our observations on this portion (which usually is small) to the population. Such generalizations about characteristics of a population from a study of one or more samples from the population are termed *statistical inferences.*

Statistical inferences take two forms: *estimates* of the magnitudes of population characteristics, and *tests of hypotheses* regarding population characteristics. Both are useful for determining which among two or more courses of action to follow in practice when the "correct" course is determined by some particular but unknown characteristic of the population.

Statistical inferences all involve reaching conclusions about population characteristics (or at least acting as if one had reached such conclusions) from a study of samples which are known or assumed to be portions of the population concerned. Statistical inferences are basically predictions of what would be found to be the case if the parent populations could be and were fully analyzed with respect to the relevant characteristic or characteristics.

A simple example will serve to bring out a number of essential features of statistical inferences and the methods of inductive statistics. Suppose that four cards have been drawn from a deck of cards and have been found to be the Ace of Hearts, the Five of Diamonds, the Three of Clubs, and the Jack of Clubs. The specific methods discussed in the following paragraphs will be illustrated from this example.

First of all, from the example, we can clearly conclude at once that the deck contained at least one Heart, at least one Diamond, and at least two Clubs. We also can conclude from the presence of the Five and the Three that the deck is definitely not a pinochle deck. These are perhaps trivial inferences, but their validity is above question and does not depend in any way on the *modus operandi* of drawing the four cards.

In order to be able to make inferences of a more substantial character, we must know the nature of the sampling operation that yielded the sample of four cards actually obtained. Suppose, for example, that the sampling procedure was as follows: The cards were drawn in the order listed, each card being selected *at random* from all the cards present in the deck when the card was drawn. This defines a hypothetical population of drawings. By using an appropriate technique of inductive statistics—essentially, a "catalog" of all possible samples of four, showing for each sample the conclusion to be adopted whenever that sample occurs—we can make statistical inferences about properties of this population of drawings. The statistical inferences made will be rigorous if, and only if, the inductive technique used is appropriate to the sampling procedure actually employed.

Thus, by taking the observed proportion of Clubs as an estimate of the proportion of Clubs in the abstract population of drawings, we may assert: the proportion of Clubs is 50%. Since random sampling of the type assumed assures that the proportion of Clubs in the population of drawings is the same as the proportion of Clubs in the deck, we may assert with equal validity: the proportion of Clubs in the deck is 50%. If the deck concerned actually was a standard bridge deck, then in the present instance our estimate is wrong in spite of being the best single estimate available.

We know from experience that with samples of four we cannot expect to "hit the nail on the head" every time. If instead of attempting to make a single-number estimate we had chosen to refer to a "catalog" of *interval estimates* (see, for example, Table A-22*), we would have concluded that the proportion of Clubs is between 14% and 86% inclusive, with an expectation of being correct 9 times out of 10. If the deck was in fact a standard bridge deck, then our conclusion is correct in this instance, but its validity depends on whether the sampling procedure employed in drawing the four cards corresponds to the sampling procedure assumed in the preparation of the "catalog" of answers.

It is important to notice, moreover, that strictly we have a right to make statistical inferences only with respect to the hypothetical population of drawings defined by the sampling operation concerned. In the present instance, as we shall see, the sampling operation was so chosen that the parameters (i.e., the proportions of Hearts, Clubs, and Diamonds) of the hypothetical population of drawings coincide with the corresponding parameters of the deck. Hence, in the present case, inferences about the parameters of the population of drawings may be interpreted as inferences about the composition of the deck. This emphasizes the importance of selecting and employing a sampling procedure such that the relevant parameters of the population of drawings bear a known relation to the corresponding parameters of the real-life situation. Otherwise, statistical inferences with respect to the population of drawings carried over to the real-life population will be lacking in rigor, even though by luck they may sometimes be correct.

* The A-Tables referenced in this handbook are contained in Section 5, AMCP 706-114.

1-3.2 RANDOM SAMPLING

In order to make valid nontrivial generalizations from samples about characteristics of the populations from which they came, the samples must have been obtained by a sampling scheme which insures two conditions:

(a) Relevant characteristics of the populations sampled must bear a known relation to the corresponding characteristics of the population of all possible samples associated with the sampling scheme.

(b) Generalizations may be drawn from such samples in accordance with a given "book of rules" whose validity rests on the mathematical theory of probability.

If a sampling scheme is to meet these two requirements, it is necessary that the selection of the individuals to be included in a sample involve some type of *random selection,* that is, each possible sample must have a fixed and determinate probability of selection. (For a very readable expository discussion of the general principles of sampling, with examples of some of the more common procedures, see the article by Cochran, Mosteller, and Tukey[1]. For fuller details see, for example, Cochran's book[2].

The most widely useful type of random selection is *simple* (or *unrestricted*) *random sampling*. This type of sampling is defined by the requirement that each individual in the population has an equal chance of being the first member of the sample; after the first member is selected, each of the remaining individuals in the population has an equal chance of being the second member of the sample; and so forth. For a sampling scheme to qualify as simple random sampling, it is not sufficient that "each individual in the population have an equal chance of appearing in the sample," as is sometimes said, but it is sufficient that "each possible sample have an equal chance of being selected." Throughout this handbook, we shall assume that all samples are random samples in the sense of having been obtained by simple random sampling.

It cannot be overemphasized that the *randomness* of a sample is inherent in the sampling scheme employed to obtain the sample and not an intrinsic property of the sample itself. Experience teaches that it is not safe to assume that a sample selected haphazardly, without any conscious plan, can be regarded as if it had been obtained by simple random sampling. Nor does it seem to be possible to consciously draw a sample *at random*. As stated by Cochran, Mosteller, and Tukey[1],

> We insist on some semblance of mechanical (dice, coins, random number tables, etc.) randomization before we treat a sample from an existent population as if it were random. We realize that if someone just "grabs a handful," the individuals in the handful almost always resemble one another (on the average) more than do the members of a simple random sample. Even if the "grabs" are randomly spread around so that every individual has an equal chance of entering the sample, there are difficulties. Since the individuals of grab samples resemble one another *more* than do individuals of random samples, it follows (by a simple mathematical argument) that the means of grab samples resemble one another *less* than the means of random samples of the same size. From a grab sample, therefore, we tend to *under*estimate the variability in the population, although we should have to *over*estimate it in order to obtain valid estimates of variability of grab sample means by substituting such an estimate into the formula for the variability of means of simple random samples. Thus, using simple random sample formulas for grab sample means introduces a double bias, both parts of which lead to an unwarranted appearance of higher stability.

Instructions for formally drawing a sample at random from a particular population are given in Paragraph 1-4.

Finally, it needs to be noticed that a particular sample often qualifies as "a sample" from any one of several populations. For example, a sample of n rounds from a single carton is a sample from that carton, from the production lot of which the rounds in that carton are a portion, and from the production process concerned. By drawing these rounds from the carton in accordance with a simple random sampling scheme, we can insure that they are a (simple) random sample from the carton, not from the production lot or the production process. Only if the production process is in a "state of statistical control" may our sample also be considered to be a simple random sample from the production lot and the production process. In a similar fashion, a sample of repeated weighings can validly be considered to be a random sample from the conceptually infinite population of repeated weighings by the same procedure only if the weighing procedure is in a state of statistical control (see Chapter 18, in Section 4, AMCP 706-113).

It is therefore important in practice to know from which of several possible "parent" populations a sample was obtained *by simple random sampling*. This population is termed the *sampled population,* and may be quite different from the population of interest, termed the *target population*, to which we would like our conclusions to be applicable. In practice, they are rarely identical, though the difference is often small. A sample from the target population of rounds of ammunition produced by a particular production process will actually be a sample from one or more production lots (sampled population), and the difference between sampled and target populations will be smaller if the sampled population comprises a larger number of production lots. The further the sampled population is removed from the target population, the more the burden of validity of conclusions is shifted from the shoulders of the statistician to those of the subject matter expert, who must place greater and greater (and perhaps unwarranted) reliance on "other considerations."

1-4 SELECTION OF A RANDOM SAMPLE

As has been brought out previously, the method of choosing a sample is an all-important factor in determining what use can be made of it. In order for the techniques described in this handbook to be valid as bases for making statements from samples about populations, we must have unrestricted random samples from these populations. In practice, it is not always easy to obtain a random sample from a given population. Unconscious selections and biases tend to enter. For this reason, it is advisable to use a table of random numbers as an aid in selecting the sample. Two tables of random numbers which are recommended are by L. H. C. Tippett(3) and The Rand Corporation(4). These tables contain detailed instructions for their use. An excerpt from one of these tables(4) is given in Table A-36. This sample is included for illustration only; a larger table should be used in any actual problem. Repeated use of the same portion of a table of random numbers will not satisfy the requirements of randomness.

An illustration of the method of use of tables of random numbers follows. Suppose the population consists of 87 items, and we wish to select a random sample of 10. Assign to each individual a separate two-digit number between 00 and 86. In a table of random numbers, pick an arbitrary starting place and decide upon the direction of reading the numbers. Any direction may be used, provided the rule is fixed in advance and is independent of the numbers occurring. Read two-digit numbers from the table, and select for the sample those individuals whose numbers occur until 10 individuals have been selected. For example, in Table A-36, start with the second page of the Table (p. T-83), column 20, line 6, and read down. The 10 items picked for the sample would thus be numbers 38, 44, 13, 73, 39, 41, 35, 07, 14, and 47.

The method described is applicable for obtaining simple random samples from any sampled population consisting of a finite set of individuals. In the case of an infinite sampled population, these procedures do not apply. Thus, we might think of the sampled population for the target population of weighings as comprising all weighings which might conceptually have been made during the time while weighing was done. We cannot by mechanical randomization draw a random sample from this population, and so must recognize that we have a random sample only *by assumption.* This assumption will be warranted if previous data indicate that the weighing procedure is in a state of statistical control; unwarranted if the contrary is indicated; and a leap in the dark if no previous data are available.

1-5 SOME PROPERTIES OF DISTRIBUTIONS

Although it is unusual to examine populations in their entirety, the examination of a large sample or of many small samples from a population can give us much information about the general nature of the population's characteristics.

One device for revealing the general nature of a population distribution is a histogram. Suppose we have a large number of observed items and a numerical measurement for each item, such as, for example, a Rockwell hardness reading for each of 5,000 specimens. We first make a table showing the numerical measurement and the number of times (i.e., frequency) this measurement was recorded.

Rockwell Hardness Number	Frequency
55	1
56	17
57	135
58	503
59	1,110
60	1,470
61	1,120
62	490
63	125
64	26
65	3

Data taken, by permission, from *Sampling Inspection by Variables* by A. H. Bowker and H. P. Goode, Copyright, 1952, McGraw-Hill Book Company, Inc.

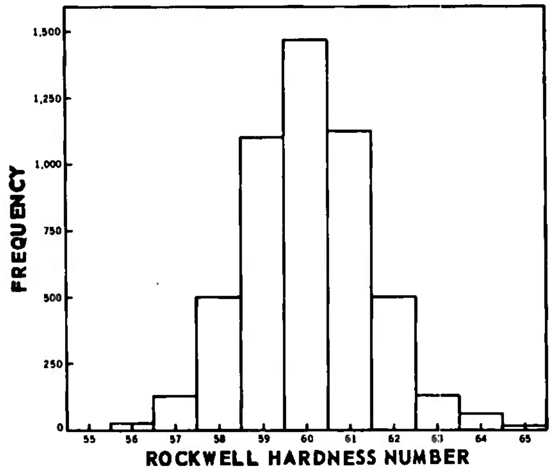

Figure 1-1. Histogram representing the distribution of 5,000 Rockwell hardness readings.

Reproduced by permission from *Sampling Inspection by Variables* by A. H. Bowker and H. P. Goode, Copyright, 1952, McGraw-Hill Book Company, Inc.

From this frequency table we can make the histogram as shown in Figure 1-1. The height of the rectangle for any hardness range is determined by the number of items in that hardness range. The rectangle is centered at the tabulated hardness value. If we take the sum of all the rectangular areas to be one square unit, then the area of an individual rectangle is equal to the *proportion* of items in the sample that have hardness values in the corresponding range. When the sample is large, as in the present instance, the histogram may be taken to exemplify the general nature of the corresponding distribution in the population.

If it were possible to measure hardness in finer intervals, we would be able to draw a larger number of rectangles, smaller in width than before. For a sufficiently large sample and a sufficiently fine "mesh," we would be justified in blending the tops of the rectangles into a continuous curve, such as that shown in Figure 1-2, which we could expect to more nearly represent the underlying population distribution.

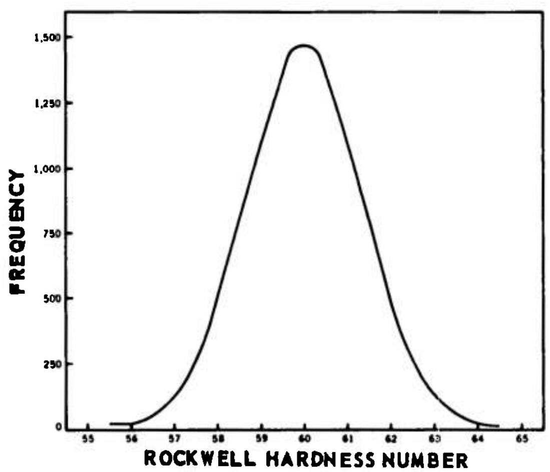

Figure 1-2. Normal curve fitted to the distribution of 5,000 Rockwell hardness readings.

Reproduced by permission from *Sampling Inspection by Variables* by A. H. Bowker and H. P. Goode, Copyright, 1952, McGraw-Hill Book Company, Inc.

If we were to carry out this sort of scheme on a large number of populations, we would find that many different curves would arise, as illustrated in Figure 1-3. Possibly, the majority of them would resemble the class of symmetrical bell-shaped curves called "normal" or "Gaussian" distributions, an example of which is shown in the center of Figure 1-3. A normal distribution is unimodal, i.e., has only a single highest point or *mode,* as also are the two asymmetrical curves in the lower left and upper right of Figure 1-3.

A "normal" distribution is completely determined by two parameters: *m,* the arithmetic mean (or simply "the mean") of the distribution, and σ, the standard deviation (often termed the "population mean" and "population standard deviation"). The *variance* of the distribution is σ^2. Since a normal curve is both unimodal and symmetrical,

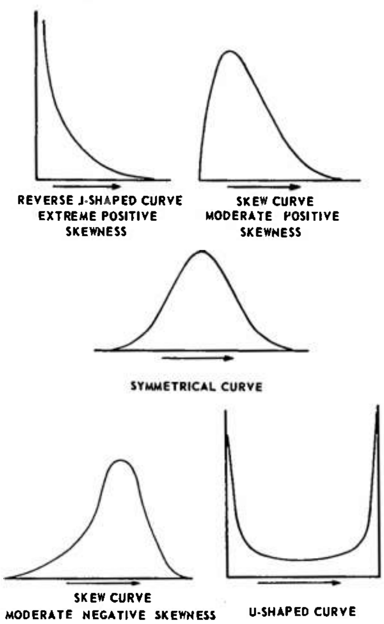

Figure 1-3. Frequency distributions of various shapes.

Adapted with permission from *Elements of Statistical Reasoning* by A. E. Treloar, Copyright, 1939, John Wiley & Sons, Inc.

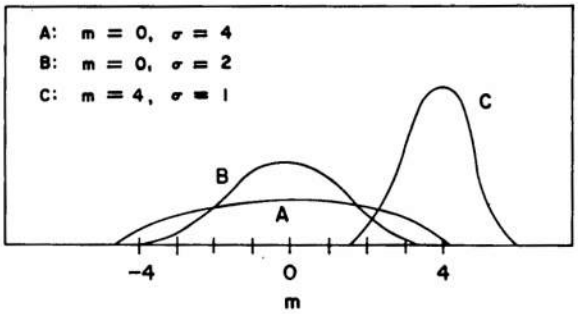

Figure 1-4. Three different normal distributions.

m is also the *mode* and the value which divides the area under the curve in half, i.e., the *median.* It is useful to remember that σ is the distance from *m* to either of the two inflection points on the curve. (The inflection point is the point at which the curve changes from concave upward to concave downward.) This is a special property of the normal distribution. More generally, the mean of a distribution *m* is the "center of gravity" of the distribution; σ is the "radius of gyration" of the distribution about *m,* in the language of mechanics; and σ^2 is the second moment about *m.*

The parameter *m* is the *location parameter* of a normal distribution, while σ is a measure of its spread, scatter, or dispersion. Thus, a change in *m* merely slides the curve right or left without changing its profile, while a change in σ widens or narrows the curve without changing the location of its center. Three different normal curves are shown in Figure 1-4. (All normal curves in this section are drawn so that the area under the curve is equal to one, which is a standard convention.)

Figure 1-5 shows the percentage of elements of the population contained in various intervals of a normal distribution. z is the distance from the population mean in units of the standard deviation and is computed using the formula $z = (X-m)/\sigma$, where X represents any value in the population. Using z to enter Table A-1, we find P, the proportion of elements in the population which have values of z smaller than any given z. Thus, as shown in Fig. 1-5, 34.13% of the population will have values of z between 0 and 1 (or between 0 and −1); 13.59% of the population, between 1 and 2 (or between −1 and −2); 2.14% between 2 and 3 (or between −2 and −3); and .14% beyond 3 (or beyond −3). Figure 1-5 shows these percentages of the population in various intervals of z.

For example, suppose we know that the chamber pressures of a lot of ammunition may be represented by a normal distribution, with the average chamber pressure $m =$ 50,000 psi and standard deviation $\sigma =$ 5,000 psi. Then $z = \frac{X - 50{,}000}{5{,}000}$ and we know (Fig. 1-5) that if we fired the lot of ammunition in the prescribed manner we would expect 50% of the rounds to have a chamber pressure above 50,000 psi, 15.9% to have pressures above 55,000 psi, and 2.3% to have pressures above 60,000 psi, etc.

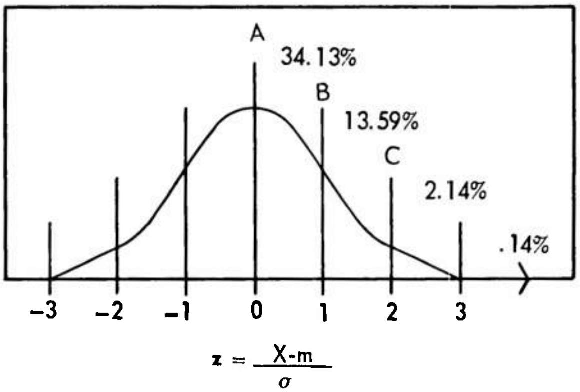

Figure 1-5. Percentage of the population in various intervals of a normal distribution.

1-6 ESTIMATION OF *m* and σ

In areas where a lot of experimental work has been done, it often happens that we know m or σ, or both, fairly accurately. However, in the majority of cases it will be our task to estimate them by means of a sample. Suppose we have n observations, $X_1, X_2, \ldots, X_n$ taken at random from a normal population. From a sample, what are the best estimates of m and σ? Actually, it is usual to compute the best unbiased estimates of m and σ^2, and then take the square root of the estimate of σ^2 as the estimate of σ. These recommended estimates of m and σ^2 are:*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

$\bar{X}$ and s^2 are the *sample mean* and *sample estimate of variance*, respectively. (s is often called "the sample standard deviation," but this is not strictly correct and we shall avoid the expression and simply refer to s.) For computational purposes, the following formula for s^2 is more convenient:

$$s^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n\,(n - 1)}$$

* The Greek symbol Σ is often used as shorthand for "the sum of." For example,

$$\sum_{i=1}^{4} X_i = X_1 + X_2 + X_3 + X_4$$

$$\sum_{i=1}^{3} (X_i + Y_i) = (X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3)$$

$$\sum_{i=1}^{3} X_i Y_i = X_1Y_1 + X_2Y_2 + X_3Y_3$$

$$\sum_{i=1}^{3} c = c + c + c = 3c$$

Nearly every sample will contain different individuals, and thus the estimates $\bar{X}$ and s^2 of m and σ^2 will differ from sample to sample. However, these estimates are such that "on the average" they tend to be equal to m and σ^2, respectively, and in this sense are *unbiased.* If, for example, we have a large number of random samples of size n, the average of their respective estimates of σ^2 will tend to be near σ^2. Furthermore, the amount of fluctuation of the respective s^2's about σ^2 (or of the $\bar{X}$'s about m, if we are estimating m) will be smaller in a certain well-defined sense than the fluctuation would be for any estimates other than the recommended ones. For these reasons, $\bar{X}$ and s^2 are called the "best unbiased" estimates of m and σ^2, respectively.*

As might be expected, the larger the sample size n, the more faith we can put in the estimates $\bar{X}$ and s^2. This is illustrated in Figures 1-6 and 1-7. Figure 1-6 shows the distribution of $\bar{X}$ (sample mean) for samples of various sizes from the same normal distribution. The curve for $n = 1$ is the distribution for individuals in the population. All of the curves are centered at m, the popula-

* On the other hand, s is not an unbiased estimator of σ. Thus, in samples of size n from a normal distribution, the situation is:

Sample size, n	s is an unbiased estimator of:
2	0.797 σ
3	0.886
4	0.921
5	0.940
6	0.952
7	0.959
8	0.965
9	0.969
10	0.973
20	0.987
30	0.991
40	0.994
60	0.996
120	0.998
∞	1.000

tion mean, but the scatter becomes less as n gets larger. Figure 1-7 shows the distribution of s^2 (sample variance) for samples of various sizes from the same normal distribution.

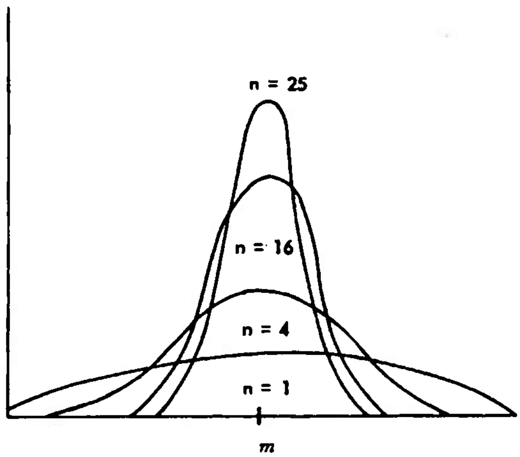

Figure 1-6. Sampling distribution of $\bar{X}$ for random samples of size n from a normal population with mean m.

Reproduced by permission from *The Methods of Statistics*, 4th ed., by L. H. C. Tippett, Copyright, 1952, John Wiley & Sons, Inc.

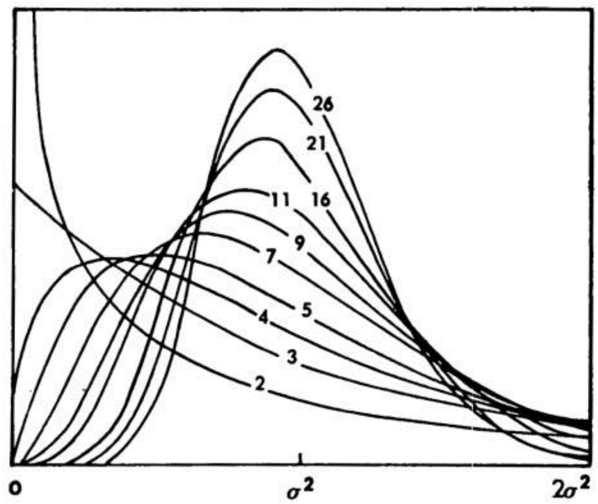

Figure 1-7. Sampling distribution of s^2 for sample size n from a normal population with $\sigma = 1$.

Adapted with permission from *Some Theory of Sampling*, by W. Edwards Deming, Copyright, 1950, John Wiley & Sons, Inc.

1-7 CONFIDENCE INTERVALS

Inasmuch as estimates of m and σ vary from sample to sample, interval estimates of m and σ may sometimes be preferred to "single-value" estimates. Provided we have a random sample from a normal population, we can make interval estimates of m or σ with a chosen degree of confidence. The level of confidence is not associated with a particular interval, but is associated with the method of calculating the interval. The interval obtained from a particular sample either brackets the true parameter value (m or σ, whichever we are estimating) or does not. The confidence coefficient γ is simply the proportion of samples of size n for which intervals computed by the prescribed method may be expected to bracket m (or σ). Such intervals are known as *confidence intervals*, and always are associated with a prescribed confidence coefficient. As we would expect, larger samples tend to give narrower confidence intervals for the same level of confidence.

Suppose we are given the lot of ammunition mentioned earlier (Par. 1-5) and wish to make a confidence interval estimate of the average chamber pressure of the rounds in the lot. The true average is 50,000 psi, although this value is unknown to us. Let us take a random sample of four rounds and from this sample, using the given procedure, calculate the upper and lower limits for our confidence interval. Consider all the possible samples of size 4 that could have been taken, and the resulting confidence intervals computed from each. If we compute 50% (90%) confidence intervals, then we expect 50% (90%) of the computed intervals to cover the true value, 50,000 psi. See Figure 1-8.

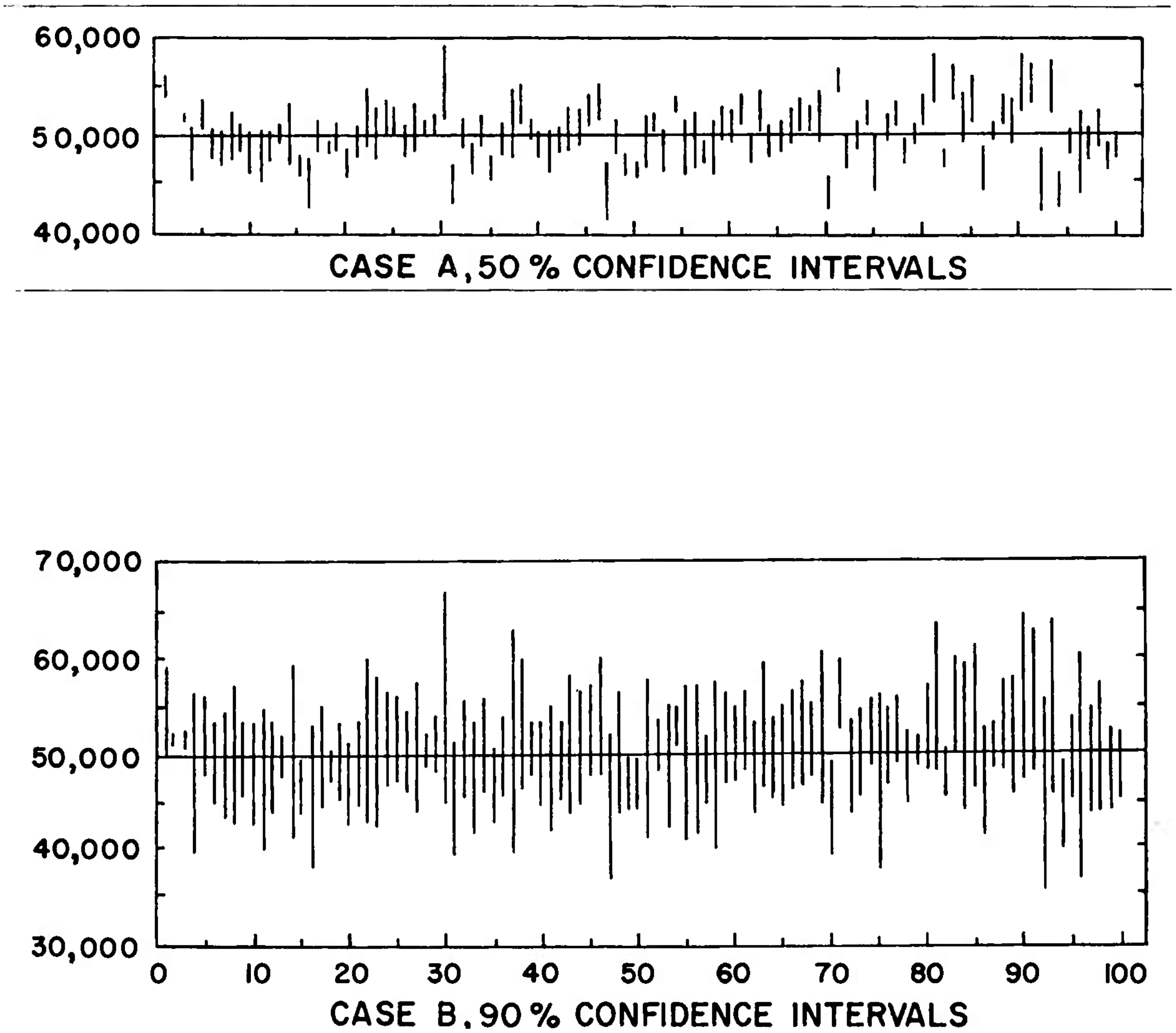

Figure 1-8. Computed confidence intervals for 100 samples of size 4 drawn at random from a normal population with m = 50,000 psi, σ = 5,000 psi. Case A shows 50% confidence intervals; Case B shows 90% confidence intervals.

Adapted with permission from *ASTM Manual on Quality Control of Materials*, Copyright, 1951, American Society for Testing Materials.

In Case A of Figure 1-8, 51 of the 100 intervals actually include the true mean. For 50% confidence interval estimates, we would expect in the long run that 50% of the intervals would include the true mean. Fifty-one out of 100 is a reasonable deviation from the expected 50%. In Case B, 90 out of 100 of the intervals contain the true mean. This is precisely the expected number for 90% intervals.

Note also (Fig. 1-8) that the successive confidence intervals vary both in position and width. This is because they were computed (see Par. 2-1.4) from the sample statistics $\bar{X}$ and s, both of which vary from sample to sample. If, on the other hand, the standard deviation of the population distribution σ were known, and the confidence intervals were computed from the successive $\bar{X}$'s and σ (procedure given in Par. 2-1.5), then the resulting confidence intervals would all be the same width, and would vary in position only.

Finally, as the sample size increases, confidence intervals tend not only to vary less in both position and width, but also to "pinch in" ever closer to the true value of the population parameter concerned, as illustrated in Figure 1-9.

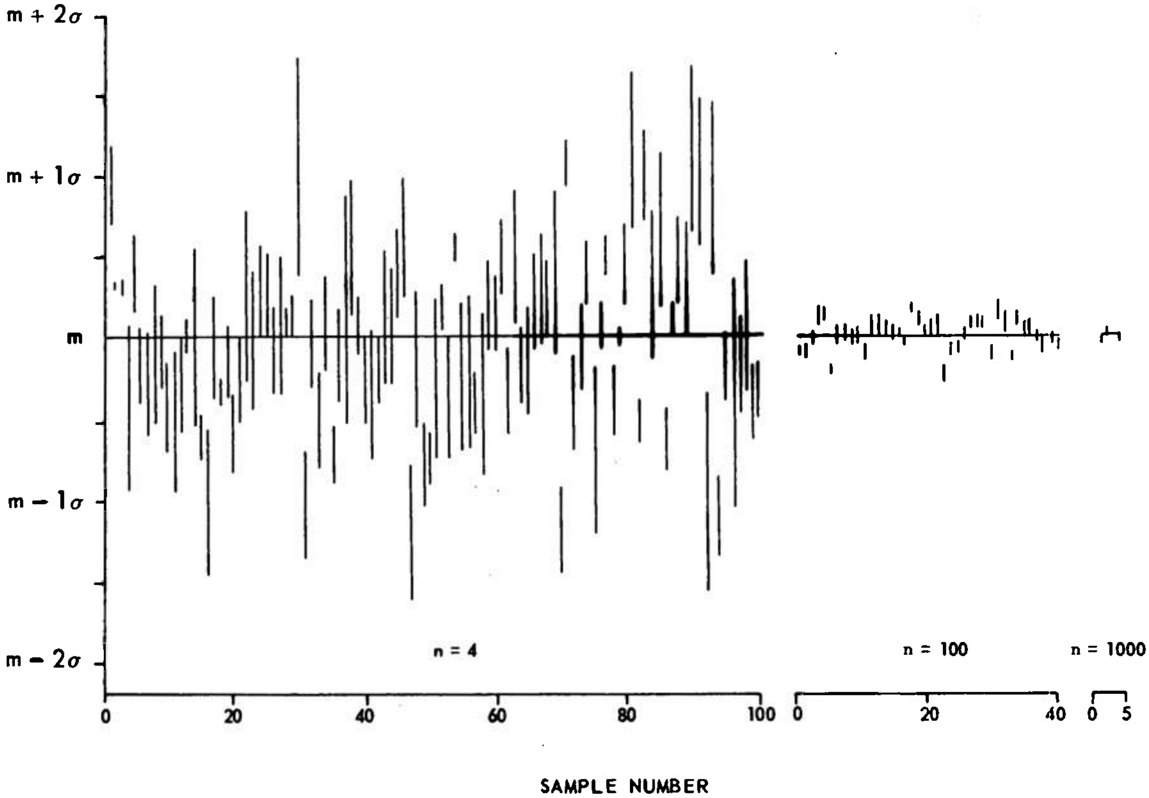

Figure 1-9. Computed 50% confidence intervals for the population mean m from 100 samples of 4, 40 samples of 100, and 4 samples of 1000.

Adapted with permission from *Statistical Method from the Viewpoint of Quality Control* by W. A. Shewhart (edited by W. Edwards Deming), Copyright, 1939, Graduate School, U.S. Department of Agriculture, Washington, D. C.

1-8 STATISTICAL TOLERANCE LIMITS

Sometimes what is wanted is not an estimate of the mean and variance of the population distribution but, instead, two outer values or limits which contain nearly all of the population values. For example, if extremely low chamber pressures or extremely high chamber pressures might cause serious problems, we may wish to know approximate limits to the range of chamber pressures in a lot of ammunition. More specifically, we may wish to know within what limits 99%, for example, of the chamber pressures lie. If we knew the mean m and standard deviation σ of chamber pressures in the lot, and if we knew the distribution of chamber pressures to be normal (or very nearly normal), then we could take $m - 3\sigma$ and $m + 3\sigma$ as our limits, and conclude that approximately 99.7% of the chamber pressures lie within these limits (see Fig. 1-5). If we do not know m and σ, then we may endeavor to approximate the limits with *statistical tolerance limits* of the form $\bar{X} - Ks$ and $\bar{X} + Ks$, based on the sample statistics $\bar{X}$ and s, with K chosen so that we may expect these limits to include *at least P* percent of the chamber pressures in the lot, at some prescribed level of confidence α.

Three sets of such limits for $P = 99.7\%$, corresponding to sample sizes $n = 4$, 100, and 1,000, are shown by the bars in Figure 1-10. It should be noted that for samples of size 4, the bars are very variable both in location and width, but that for $n = 100$ and $n = 1{,}000$, they are of nearly constant width

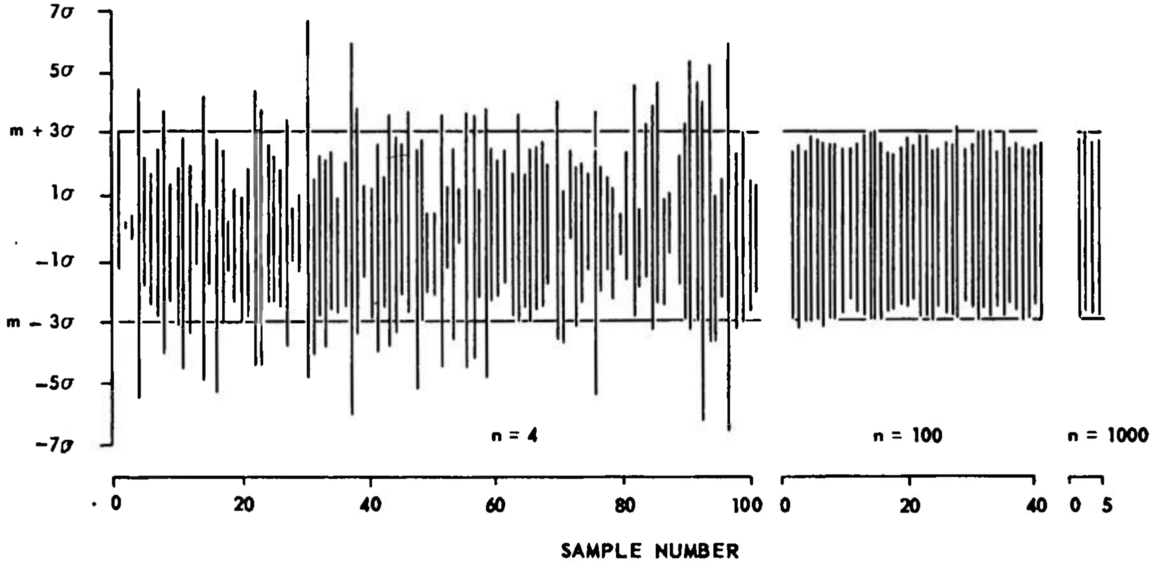

Figure 1-10. Computed statistical tolerance limits for 99.7% of the population from 100 samples of size 4, 40 samples of size 100, and 4 samples of size 1000.

Adapted with permission from *Statistical Method from the Viewpoint of Quality Control* by W. A. Shewhart (edited by W. Edwards Deming), Copyright, 1939, Graduate School, U.S. Department of Agriculture, Washington, D. C.

and position—and their end points approximate very closely to $m - 3\sigma$ and $m + 3\sigma$. In other words, statistical tolerance intervals tend to a fixed size (which depends upon P) as the sample size increases, whereas confidence intervals shrink down towards zero width with increasing sample size, as illustrated in Figure 1-9.

The difference in the meanings of the terms *confidence intervals, statistical tolerance limits*, and *engineering tolerance limits* should be noted. A *confidence interval* is an interval within which we estimate a given population parameter to lie (e.g., the population mean m with respect to some characteristic). *Statistical tolerance limits* for a given population are limits within which we expect a stated proportion of the population to lie with respect to some measurable characteristic. *Engineering tolerance limits* are specified outer limits of acceptability with respect to some characteristic usually prescribed by a design engineer.

1-9 USING STATISTICS TO MAKE DECISIONS

1-9.1 APPROACH TO A DECISION PROBLEM

Consider the following more-or-less typical practical situation: Ten rounds of a new type of shell are fired into a target, and the depth of penetration is measured for each round. The depths of penetration are 10.0, 11.1, 10.5, 10.5, 11.2, 10.8, 9.8, 12.2, 11.0, and 9.9 cm. The average penetration depth of the comparable standard shell is 10.0 cm. We wish to know whether the new type shells penetrate farther on the average than the standard type shells.

If we compute the arithmetic mean of the ten shells, we find it is 10.70 cm. Our first impulse might be to state that on the average the new shell will penetrate 0.7 cm. farther than the standard shell. This, indeed, is our best single guess, but how sure can we be that this actually is close to the truth? One thing that might catch our notice is the variability in the individual penetration depths of the new shells. They range from 9.8 cm. to 12.2 cm. The standard deviation as measured by s calculated from the sample is 0.73 cm. Might not our sample of ten shells have contained some atypical ones of the new type which have unusually high penetrating power? Could it be that the new shell is, on the average, no better than the standard one? If we were obliged to decide, on the basis of the results obtained from these ten shells alone, whether to keep on making the standard shells or to convert our equipment to making the new shell, how can we make a valid choice?

A very worthwhile step toward a solution in such situations is to compute, from the data in hand, a confidence interval for the unknown value of the population parameter of interest. The procedure (given in Par. 2-1.4) applied to the foregoing depth-of-penetration data for the new type of shell yields the interval from 10.18 to 11.22 cm. as a 95% confidence interval for the population mean depth of penetration of shells of the new type. Inasmuch as this interval lies entirely to the right of the mean for the standard shell, 10.00 cm., we are justified in concluding that the new shell *is*, on the average, better than the standard, with only a 5% risk of being in error. Nevertheless, taking other considerations into account (e.g., cost of the new type, cost of changing over, etc.), we may conclude finally that the improvement—which may be as little as 0.18 cm., and probably not more than 1.22 cm.—is not sufficient to warrant conversion to the new type. On the other hand, the evidence that the new type is almost certainly better plus the prospect that the improvement may be as great as 1.22 cm. may serve to recommend further developmental activity in the direction "pioneered" by the new type.

A somewhat different approach, which provides a direct answer to our question "Could it be that the new shell is on the average no better than the standard?" but not to the question of whether to convert to the new type, is to carry out a so-called *test of significance* (or test of a statistical hypothesis). In the case of the foregoing example, the formal procedure for the corresponding test of significance (Par. 3-2.2.1) turns out to be equivalent (as explained in AMCP 706-113, Chapter 21) to noting whether or not the confidence interval computed does or does not include the population mean for the standard shell (10.0 cm.). If, as in the present instance, the population mean for the standard shell is *not* included, this is taken to be a *negative* answer to our question. In other words, this is taken to be conclusive evidence (at the *5% level of significance*) *against* the *null hypothesis* that "the new shell is on the average *no better* than the standard." Rejection of the null hypothesis in this case is equivalent to accepting the indefinite *alternative hypothesis* that "the new shell *is better* on the average than the standard." If, on the other hand, the population mean for the standard shell *is* included in the confidence interval, this is taken as an *affirmative* answer to our question—not in the positive sense of definitely confirming the null hypothesis ("is no better"), but in the more-or-less neutral sense of the absence of conclusive evidence to the contrary.

As the foregoing example illustrates, an advantage of the confidence-interval approach to a decision problem is that the confidence interval gives an indication of how large the difference, if any, is likely to be, and thus provides some of the additional information usually needed to reach a final decision on the action to be taken next. For many purposes, this is a real advantage of confidence intervals over tests of significance.

However, all statistical decision problems are not amenable to solution via confidence intervals. For instance, the question at issue may be whether or not two particular characteristics of shell performance are mutually independent. In such a situation, any one of a variety of tests of significance can be used to test the null hypothesis of "no dependence." Some of these may have a reasonably good chance of rejecting the null hypothesis, and thus "discovering" the existence of a dependence when a dependence really exists—even though the exact nature of the dependence, if any, is not understood and a definitive measure of the extent of the dependence in the population is lacking.

A precise test of significance will be possible if: (a) the sampling distribution of some sample statistic is known (at least to a good approximation) for the case of "no dependence"; and (b) the effect of dependence on this statistic is known (e.g., tends to make it larger). For a confidence-interval approach to be possible, two conditions are necessary: (a) there must be agreement on what constitutes the proper measure (parameter) of dependence of the two characteristics in the population; and, (b) there must be a sample estimate of this dependence parameter whose sampling distribution is known, to a good approximation at least, for all values of the parameter. Confidence intervals tend to provide a more complete answer to statistical decision problems when they are available, but tests of significance are of wider applicability.

1-9.2 CHOICE OF NULL AND ALTERNATIVE HYPOTHESES

A statistical test always involves a *null hypothesis,* which is considered to be the hypothesis under test, as against a class of *alternative hypotheses.* The null hypothesis acts as a kind of "origin" or "base" (in the sense of "base line"), from which the alternative hypotheses deviate in one way or another to greater and lesser degrees. Thus, in the case of the classical problem of the tossing of a coin, the null or base hypothesis specifies that the probability of "heads" on any single trial equals 1/2. If, in a particular situation, the occurrence of "heads" were an *advantage*, then we might be particularly interested in the *one-sided* class of alternative hypotheses that the probability of "heads" on any single trial equals P, where P is some (unknown) fraction exceeding 1/2. If neither "heads" nor 'tails" were intrinsically advantageous, but a bias in favor of either could be employed to advantage, then we could probably be interested in the more general *two-sided* class of alternative hypotheses specifying that the probability of "heads" on any single toss equals P, where P is some fraction (less than, or greater than, but) *not* equal to 1/2.

The important point is that the null hypothesis serves as an origin or base. In the coin-tossing instance, it also happens to be a favored, or traditional, hypothesis. This is merely a characteristic of the example selected. Indeed, the null hypothesis is often the very antithesis of what we would really like to be the case.

1-9.3 TWO KINDS OF ERRORS

In basing decisions on the outcomes of statistical tests, we always run the risks of making either one or the other of two types of error. If we reject the null hypothesis when it is true, e.g., announce a difference which really does not exist, then we make an *Error of the First Kind.* If we fail to reject a null hypothesis when it is false, e.g., fail to find an improvement in the new shell over the old when an improvement exists, then we make what is called an *Error of the Second Kind.* Although we do not know in a given instance whether we have made an error of either kind, we can know the *probability* of making either type of error.

1-9.4 SIGNIFICANCE LEVEL AND OPERATING CHARACTERISTIC (OC) CURVE OF A STATISTICAL TEST

The risk of making an error of the first kind, α, equals what is by tradition called the *level of significance* of the test. The risk of making an error of the second kind, β, varies, as one would expect, with the magnitude of the real difference, and is summarized by the *Operating Characteristic (OC) Curve* of the test. See, for example, Figure 3-5. Also, the risk β of making an error of the second kind increases as the risk α of making an error of the first kind decreases. Compare Figure 3-5 with Figure 3-6. Only with "large" samples can we "have our cake and eat it too"—and then there is the cost of the test to worry about.

1-9.5 CHOICE OF THE SIGNIFICANCE LEVEL

The significance level of a statistical test is essentially an expression of our reluctance to give up or "reject" the null hypothesis. If we adopt a "stiff" significance level, 0.01 or even 0.001, say, this implies that we are very unwilling to reject the null hypothesis unjustly. A consequence of our ultraconservatism in this respect will usually be that the probability of not rejecting the null hypothesis when it is really false will be large unless the actual deviation from the null hypothesis is large. This is clearly an entirely satisfactory state of affairs if we are quite satisfied with the status quo and are only interested in making a change if the change represents a very substantial improvement. For example, we may be quite satisfied with the performance of the standard type of shell in all respects, and not be willing to consider changing to the new type unless the mean depth of penetration of the new type were at least, say, 20% better (12.0 cm.).

On the other hand, the standard shell may be unsatisfactory in a number of respects and the question at issue may be whether the new type shows promise of being able to replace it, either "as is" or with further development. Here "rejection" of the null hypothesis would not imply necessary abandonment of the standard type and shifting over to the new type, but merely that the new type shows "promise" and warrants further investigation. In such a situation, one could afford a somewhat higher risk of rejecting the null hypothesis falsely, and would take $\alpha = 0.05$ or 0.10 (or even 0.20, perhaps), in the interest of increasing the chances of detecting a small but promising improvement with a small-scale experiment. In such exploratory work, it is often more important to have a good chance of detecting a small but promising improvement than to protect oneself against crying "wolf, wolf" occasionally—because the "wolf, wolf" will be found out in due course, but a promising approach to improvement could be lost forever.

In summary, the significance level α of a statistical test should be chosen in the light of the attending circumstances, including costs. We are sometimes limited in the choice of significance level by the availability of necessary tables for some statistical tests. Two values of α, $\alpha = .05$ and $\alpha = .01$, have been most frequently used in research and development work; and are given in tabulations of test statistics. We have adopted these "standard" levels of significance for the purposes of this handbook.

1-9.6 A WORD OF CAUTION

Many persons who regularly employ statistical tests in the interpretation of research and development data do not seem to realize that all probabilities associated with such tests are calculated on the supposition that some definite set of conditions prevails. Thus, α, the level of significance (or probability of an error of the first kind), is computed on the assumption that the null hypothesis is strictly true in all respects; and β, the risk of an error of the second kind, is computed on the assumption that a particular specific alternative to the null hypothesis is true *and* that the statistical test concerned is carried out at the α-level of significance. Consequently, whatever may be the actual outcome of a statistical test, it is mathematically impossible to infer from the outcome anything whatsoever about the odds for or against some particular set of conditions being the truth.

Indeed, it is astonishing how often erroneous statements of the type "since r exceeds the 1% level of significance, the odds are 99 to 1 that there *is* a correlation between the variables" occur in research literature. How ridiculous this type of reasoning can be is brought out by the following simple example[5]: The *American Experience Mortality Table* gives .01008 as the probability of an individual aged 41 dying within the year. If we accept this table as being applicable to living persons today (which is analogous to accepting the published tables of the significance levels of tests which we apply to our data), and *if* a man's age really is 41, then the odds *are* 99 to 1 that he will live out the year. On the other hand, if we accept the table and happen to hear that some prominent individual has just died, then we *cannot* (and *would not*) conclude that the odds are 99 to 1 that his age was different from 41.

Suppose, on the other hand, that in some official capacity it is our practice to check the accuracy of age statements of all persons who say they are 41 and *then* die within the year. This practice (assuming the applicability of the American Experience Mortality Table) will lead us in the long run to suspect unjustly the word of one person in 100 whose age *was* 41, who told us so, and who then was unfortunate enough to die within the year. The *level of significance* of the test is in fact 0.01008 (1 in 100). On the other hand, this practice will also lead us to discover mis-statements of age of *all* persons professing to be 41 who are really some other age *and* who happen to die within the year. The probabilities of our discovering such mis-statements will depend on the actual ages of the persons making them. We shall, however, let slip by as correct all statements "age 41" corresponding to individuals who *are not* 41 but who do not happen to die within the year.

The moral of this is that all statistical tests can and should be viewed in terms of the consequences which may be expected to ensue from their repeated use in suitable circumstances. When viewed in this light, the great risks involved in drawing conclusions from exceedingly small samples becomes manifest to anyone who takes the time to study the OC curves for the statistical tests in common use.

REFERENCES

1. W. G. Cochran, F. Mosteller, and J. W. Tukey, "Principles of Sampling," *Journal of the American Statistical Association*, Vol. 49, pp. 13-35, 1954. (Copies of this article can be obtained from the American Statistical Association, 1757 K St., N.W., Washington 6, D. C. Price: 50 cents.)
2. W. G. Cochran, *Sampling Techniques*, John Wiley & Sons, Inc., New York, N. Y., 1953.
3. L. H. C. Tippett, *Random Sampling Numbers*, Tracts for Computers, No. 15, Cambridge University Press, 1927.
4. The Rand Corporation, *A Million Random Digits*, The Free Press, Glencoe, Ill., 1955.
5. C. Eisenhart, "The Interpretation of Tests of Significance," *Bulletin of the American Statistical Association*, Vol. 2, No. 3, pp. 79-80, April, 1941.

SOME RECOMMENDED ELEMENTARY TEXTBOOKS

A. H. Bowker and G. J. Lieberman, *Engineering Statistics*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1959.

W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis* (2d edition), McGraw-Hill Book Co., Inc., New York, N. Y., 1957.

M. J. Moroney, *Facts from Figures*, Penguin Books, Inc., Baltimore, Md., 1951.

L. H. C. Tippett, *The Methods of Statistics*, 4th edition), John Wiley & Sons, Inc., New York, N. Y., 1952.

W. A. Wallis and H. V. Roberts, *Statistics, A New Approach*, The Free Press, Glencoe, Ill., 1956.

DISCUSSION OF TECHNIQUES IN CHAPTERS 2 THROUGH 6

The techniques described in Chapters 2 through 6 apply to the analysis of results of experiments expressed as measurements in some conventional units on a continuous scale. They do not apply to the analysis of data in the form of proportions, percentages, or counts.

It is assumed that the underlying population distributions are normal or nearly normal. Where this assumption is not very important, or where the actual population distribution would show only slight departure from normality, an indication is given of the effect upon the conclusions derived from the use of the techniques. Where the normality assumption is critical, or where the actual population distribution shows substantial departure from normality, or both, suitable warnings are given.

Table A-37 is a table of three-decimal-place random normal deviates that exemplify sampling from a normal distribution with zero mean ($m = 0$) and unit standard deviation ($\sigma = 1$). To construct numbers that will simulate measurements that are normally distributed about a true value of, say, 0.12, with a standard deviation of, say, 0.02, multiply the table entries by 0.02 and then add 0.12. The reader who wishes to get a feel for the statistical behavior of sample data, and to try out and judge the usefulness of particular statistical techniques, is urged to carry out a few "dry runs" with such simulated measurements of known characteristics.

All A-Tables referenced in these Chapters are contained in AMCP 706-114, Section 5.

CHAPTER 2

CHARACTERIZING THE MEASURED PERFORMANCE OF A MATERIAL, PRODUCT, OR PROCESS

2-1 ESTIMATING AVERAGE PERFORMANCE FROM A SAMPLE

2-1.1 GENERAL

In this Chapter we present two important kinds of estimates of the average performance of a material, product, or process from a sample. These include the best single estimate, and *confidence interval* estimates.*

Specific procedures are given for obtaining confidence interval estimates when:

(a) we have a sample from a normal population whose variability is unknown; and,

(b) we have a sample from a normal population whose variability is known.

When the departures from normality are not great, or when the sample sizes are moderately large, interval estimates made as described in Paragraphs 2-1.4 and 2-1.5 will have confidence levels very little different from the chosen or nominal level.

The following data will serve to illustrate the application of the procedures.

Data Sample 2-1—Thickness of Mica Washers

Form: Measurements $X_1, X_2, \ldots, X_n$ of n items selected independently at random from a much larger group.

Example: Ten mica washers are taken at random from a large group, and their thicknesses measured in inches:

.123	.132
.124	.123
.126	.126
.129	.129
.120	.128

In general, what can we say about the larger group on the basis of our sample? We show how to answer two questions:

(a) What is our *best* guess as to the average thickness in the whole lot?

(b) Can we give an interval which we expect, with certain confidence, to bracket the true average—i.e., a *confidence interval?*

These two questions are answered in the paragraphs which follow, using the data shown above. Another question, which is sometimes confused with (b) above, is treated in Paragraph 2-5. This is the question of setting *statistical tolerance limits*, or estimating an interval which will include, with prescribed confidence, a specified proportion of the individual items in the population.

2-1.2 BEST SINGLE ESTIMATE

The most common and ordinarily the *best* single estimate of the population mean m is simply the arithmetic mean of the measurements.

* The reader who is not familiar with the meaning and interpretation of confidence intervals should refer to Chapter 1, and to Paragraph 2-1.3 of this Chapter.

Procedure

Compute the arithmetic mean $\bar{X}$ of the n measurements $X_1, X_2, \ldots, X_n$.

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)$$

Example

Compute the arithmetic mean $\bar{X}$ of 10 measurements (Data Sample 2-1):

$$\bar{X} = \frac{.123 + .124 + .126 + \ldots + .128}{10}$$

$$= \frac{1.260}{10}$$

$$= .1260 \text{ inch}$$

2-1.3 SOME REMARKS ON CONFIDENCE INTERVAL ESTIMATES

When we take a sample from a lot or a population, the sample average will seldom be exactly the same as the lot or population average. We do hope that it is fairly close, and we would like to state an interval which we are confident will bracket the lot mean. If we made such interval estimates in a particular fashion a large number of times, and found that these intervals actually did contain the true mean in 99% of the cases, we might say that we were operating at a 99% confidence level. Our particular kind of interval estimates might likewise be called "99% confidence intervals." Similarly, if our intervals included the true average "95% of the time"—strictly, in 95% of the times or instances involved—we would be operating at a 95% confidence level, and our intervals would be called 95% confidence intervals. In general, if in the long run we expect $100(1-\alpha)$ % of our intervals to contain the true value, we are operating at the $100(1-\alpha)$ % confidence level.

We may choose whatever confidence level we wish. Confidence levels γ commonly used are 99% and 95%, which correspond to $\alpha = .01$ and $\alpha = .05$. If we wish to estimate the mean of some characteristic of a large group (population) using the results of a random sample from that group, the procedures of Paragraphs 2-1.4 and 2-1.5 will allow us to make interval estimates at any chosen confidence level. It is assumed that the characteristic of interest has a normal distribution in the population. We may elect to make a two-sided interval estimate, expected to bracket the mean from both above and below; or we may make a one-sided interval estimate, limited on either the upper or the lower side, which is expected to contain the mean and to furnish either an upper or a lower bound to its magnitude.

2-1.4 CONFIDENCE INTERVALS FOR THE POPULATION MEAN WHEN KNOWLEDGE OF THE VARIABILITY CANNOT BE ASSUMED

2-1.4.1 Two-Sided Confidence Interval

This procedure gives an interval which is expected to bracket m, the true mean, 100 $(1-\alpha)$ % of the time.

Procedure

Problem: What is a two-sided $100\,(1-\alpha)$ % confidence interval for the true mean m?

(1) Choose the desired confidence level, $1-\alpha$

(2) Compute:
$\bar{X}$, the arithmetic mean (see Paragraph 2-1.2), and

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n\,(n-1)}}$$

Example

Problem: What is a two-sided 95% confidence interval for the mean thickness in the lot? (Data Sample 2-1)

(1) Let $1 - \alpha = .95$
$\alpha = .05$

(2)
$\bar{X} = .1260$ inch

$s = 0.00359$ inch

Procedure	Example
(3) Look up $t = t_{1-\alpha/2}$ for $n - 1$ degrees of freedom* in Table A-4.	(3) $t = t_{.975}$ for 9 degrees of freedom $= 2.262$
(4) Compute: $X_U = \bar{X} + t\dfrac{s}{\sqrt{n}}$; $X_L = \bar{X} - t\dfrac{s}{\sqrt{n}}$	(4) $X_U = .1260 + \dfrac{2.262\ (.00359)}{\sqrt{10}} = .1286$ inch ; $X_L = .1260 - \dfrac{2.262\ (.00359)}{\sqrt{10}} = .1234$ inch
Conclude: The interval from X_L to X_U is a $100\,(1-\alpha)\,\%$ confidence interval for the population mean; i.e., we may assert with $100\,(1-\alpha)\,\%$ confidence that $X_L < m < X_U$.	*Conclude:* The interval from .1234 to .1286 inch is a 95% confidence interval for the lot mean; i.e., we may assert with 95% confidence that .1234 inch < lot mean < .1286 inch.

2-1.4.2 One-Sided Confidence Interval

The preceding computations can be used to make another kind of confidence interval statement. We can say that $100(\alpha/2)\,\%$ of the time the entire interval in Paragraph 2-1.4.1 will lie above the true mean (i.e., X_L, the lower limit of the interval will be larger than the true mean). The rest of the time—namely $100(1-\alpha/2)\,\%$ of the time—X_L will be less than the true mean. Hence the interval from X_L to $+\infty$ is a $100(1-\alpha/2)\,\%$ one-sided confidence interval for the true mean. In the example, Paragraph 2-1.4.1, $100(1-\alpha/2)\,\%$ equals 97.5%. Thus, either of two open-ended intervals, "larger than .1234 inch," or "less than .1286 inch" can be called a 97.5% one-sided confidence interval for the population mean.

We now give the step-by-step procedure for determining a one-sided confidence interval for the population mean corresponding to a different choice of confidence level.

* In *A Dictionary of Statistical Terms*,(1) we find the following, under the phrase "degrees of freedom":

"This term is used in statistics in slightly different senses. It was introduced by Fisher on the analogy of the idea of degrees of freedom of a dynamical system, that is to say the number of independent coordinate values which are necessary to determine it. In this sense the degrees of freedom of a set of observations (which *ex hypothesi* are subject to sampling variation) is the number of values which could be assigned arbitrarily within the specification of the system; for example, in a sample of constant size n grouped into k intervals there are $k - 1$ degrees of freedom because, if $k - 1$ frequencies are specified, the other is determined by the total size n; . . .

A sample of n variate values is said to have n degrees of freedom, whether the variates are dependent or not, and a statistic calculated from it is, by a natural extension, also said to have n degrees of freedom. But, if k functions of the sample values are held constant, the number of degrees of freedom is reduced by k. For example, the statistic $\sum_{i=1}^{n} (x_i - \bar{x})^2$ where $\bar{x}$ is the sample mean, is said to have $n - 1$ degrees of freedom. . . ."

In this example, $s^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$ and has "$n - 1$" degrees of freedom.

Procedure

Problem: What is a one-sided 100 $(1 - \alpha)$ % confidence interval for the true mean?

(1) Choose the desired confidence level, $1 - \alpha$.

(2) Compute:

$\bar{X}$

s

(3) Look up $t = t_{1-\alpha}$ for $n - 1$ degrees of freedom in Table A-4.

(4) Compute:

$$X'_L = \bar{X} - t\frac{s}{\sqrt{n}}$$

(or compute

$$X'_U = \bar{X} + t\frac{s}{\sqrt{n}}\Big)$$

Conclude: We are 100 $(1 - \alpha)$ % confident that the lot mean m is greater than X'_L (or, we are 100 $(1 - \alpha)$ % confident that the lot mean m is less than X'_U), i.e., we may assert with 100 $(1 - \alpha)$ % confidence that $m > X_L$ (or, that $m < X_U$).

Example

Problem: What is a value which we expect, with 99% confidence, to be exceeded by the lot mean? (Alternatively, what is a value which we expect, with 99% confidence, to exceed the lot mean?) (Data Sample 2-1)

(1) Let $1 - \alpha = .99$

$\alpha = 0.01.$

(2)

$\bar{X} = .1260$ inch

$s = 0.00359$ inch

(3) $t = t_{.99}$ for 9 degrees of freedom $= 2.821$

(4)

$$X'_L = .1260 - \frac{(2.821)\ (.00359)}{\sqrt{10}}$$

$$= .1228$$

(or $X'_U = .1292$)

Conclude: We are 99% confident that the lot mean is greater than .1228 inch (or, we are 99% confident that the lot mean is less than .1292 inch), i.e., we may assert with 99% confidence that mean thickness in lot > .1228 inch (or, that mean thickness in lot < .1292 inch).

2-1.5 CONFIDENCE INTERVAL ESTIMATES WHEN WE HAVE PREVIOUS KNOWLEDGE OF THE VARIABILITY

We have assumed in the previous paragraph (2-1.4) that we had no previous information about the variability of performance of items and were limited to using an estimate of variability obtained from the sample at hand. Suppose that in the case of the mica washers we had taken samples many times previously from the same process and found that, although each lot had a different average, there was always essentially the same amount of variation within a lot. We would then be able to take σ, the standard deviation of the lot, as known and equal to the value indicated by this previous experience. This assumption should not be made casually, but only when warranted after real investigation of the stability of the variation among samples, using control chart techniques.

The procedure for computing these confidence intervals is simple. In the procedures of Paragraph 2-1.4, merely replace s by σ and t by z and the formulas remain the same. Values of z are given in Table A-2. Note that t_P for an infinite number of degrees of freedom (Table A-4) is exactly equal to z_P. The following procedure is for the two-sided confidence interval.

Procedure

Problem: Find a two-sided 100 $(1 - \alpha)$ % confidence interval for the lot mean, using known σ.

(1) Choose the desired confidence level, $1 - \alpha$.

(2) Compute:
$\bar{X}$

(3) Look up $z = z_{1-\alpha/2}$ in Table A-2.

(4) Compute:

$$X_U = \bar{X} + z\frac{\sigma}{\sqrt{n}}$$

$$X_L = \bar{X} - z\frac{\sigma}{\sqrt{n}}$$

Conclude: The interval from X_L to X_U is a 100 $(1 - \alpha)$ % confidence interval for the lot mean.

Example

Problem: What is a two-sided 95% confidence interval for the lot mean? (Data Sample 2-1; and σ is known to equal .0040 inch.)

(1) Let $1 - \alpha = .95$
$\alpha = .05$

(2)
$\bar{X} = .1260$ inch

(3) $z = z_{.975}$
$= 1.960$

(4)

$$X_U = .1260 + 1.960\frac{(.004)}{\sqrt{10}}$$
$$= .1285$$
$$X_L = .1235$$

Conclude: The interval from .1235 to .1285 inch is a 95% confidence interval for the lot mean.

Discussion: When the value of σ, the standard deviation in the population, is known, Procedure 2-1.5 should always be used in preference to Procedure 2-1.4, which is independent of our knowledge of σ. When available, Procedure 2-1.5 (σ known) will usually lead to a confidence interval for the population mean that is narrower than the confidence interval that would have been obtained by Procedure 2-1.4 (σ unknown). This is the case for our illustrative examples based on Data Sample 2-1, but the difference is very slight because σ and s were both very small—only 0.03% of the mean.

Whatever level of confidence is chosen, the t value required for the application of Procedure 2-1.4 (σ unknown) will always be larger than the corresponding z value required for Procedure 2-1.5 (σ known). This is evident from Table A-4. For very small samples, the difference can be considerable. Nevertheless, it can happen, as a result of unusual sampling fluctuations, that the value of s obtained in a particular sample is so small in comparison to σ that, if Procedure 2-1.4 (σ unknown) were used, the resulting confidence interval would be narrower than the confidence interval given by Procedure 2-1.5 (σ known). This would have been the case, for instance, if Data Sample 2-1 had yielded an s less than $1.960(0.0040)/2.262 = 0.00347$. With samples of size 10 (i.e., 9 degrees of freedom for s), the probability of such an occurrence is about one in three. In such a case, however, one must NOT adopt the confidence interval corresponding to Procedure 2-1.4 (σ unknown) because it is narrower. To choose between Procedure 2-1.4 (σ unknown) and Procedure 2-1.5 (σ known), when the value of σ IS known, by selecting the one which yields the narrower confidence interval in each instance, would result in a level of confidence somewhat lower than claimed.

2-2 ESTIMATING VARIABILITY OF PERFORMANCE FROM A SAMPLE

2-2.1 GENERAL

We take the standard deviation of performance in the population as our measure of the characteristic variability of performance. Presented here are various ways of estimating the population standard deviation, including:

(a) single-value estimates;

(b) confidence-interval estimates based on random samples from the population; and,

(c) techniques for estimating the population standard deviation when no appropriate random samples are available.

The first two procedures are illustrated by application to the following data.

Data Sample 2-2—Burning Time of Rocket Powder

Form: n independent measurements $X_1, X_2, \ldots, X_n$ selected at random from a much larger group.

Example: Ten unit amounts of rocket powder selected at random from a large lot were tested in a chamber and their burning times recorded as follows (seconds):

50.7	69.8
54.9	53.4
54.3	66.1
44.8	48.1
42.2	35.5

2-2.2 SINGLE ESTIMATES

2-2.2.1 s^2 and s

The best estimate of σ^2, the variance of a normal population, is:

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

For computational purposes, we usually find it more convenient to use the following formula:

$$s^2 = \frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

The formulas are algebraically identical. With any formula, it is important to carry a sufficient number of decimal places. If too few places are carried, the subtractions involved may result in a loss of significant figures in s^2. Excessive rounding may even lead to a negative value for s^2. The formula recommended for computational purposes is to be preferred on this account because only one subtraction is involved; and with a desk calculator one usually can retain all places in the computation of ΣX_i^2 and $(\Sigma X_i)^2$.

We take

$$s = \sqrt{s^2} = \sqrt{\frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}}$$

as our estimate of σ, the population standard deviation.

Example: Using Data Sample 2-2, $\Sigma X_i^2 = 27987.54$, $\Sigma X_i = 519.8$, and thus $s^2 = 107.593$; and $s = 10.37$ seconds.*

2-2.2.2 The Sample Range as an Estimate of the Standard Deviation

The *range* of n observations is defined as the difference between the highest and the

TABLE 2-1. TABLE OF FACTORS FOR CONVERTING THE RANGE OF A SAMPLE OF n TO AN ESTIMATE OF σ, THE POPULATION STANDARD DEVIATION. ESTIMATE OF σ = RANGE/d_n

Size of Sample n	d_n	$\frac{1}{d_n}$	$\sqrt{n}$ [See Note]
2	1.128	.8865	1.414
3	1.693	.5907	1.732
4	2.059	.4857	2.000
5	2.326	.4299	2.236
6	2.534	.3946	2.449
7	2.704	.3698	2.646
8	2.847	.3512	2.828
9	2.970	.3367	3.000
10	3.078	.3249	3.162
12	3.258	.3069	3.464
16	3.532	.2831	4.000

Note: d_n is approximately equal to $\sqrt{n}$ for $3 \leq n \leq 10$. Thus, for small n a quick estimate of σ can be obtained by dividing the range by $\sqrt{n}$.

* In a final report, values of s should be rounded to two significant figures, but as a basis for further calculations it is advisable to retain one or two additional figures. For fuller explanation, see Chapters 22 and 23, Section 4, AMCP 706-113.

lowest of the n values. For small samples, the sample range is a reasonably efficient substitute for s as an estimator of the standard deviation of a normal population—not as efficient as s, but easier to calculate. Using the range is particularly valuable for a "quick look" at data from small samples. As the sample size gets larger, the range is not only troublesome to calculate, but is a very inefficient estimator of σ. Table 2-1 gives the factors which convert from observed range in a sample of n observations to an estimate of population standard deviation σ.

2-2.3 CONFIDENCE INTERVAL ESTIMATES*

2-2.3.1 Two-Sided Confidence Interval Estimates

We are interested in determining an interval which we may confidently expect to bracket the true value of the standard deviation of a normal population.

Procedure

Problem: What is a two-sided $100(1-\alpha)\%$ confidence interval for σ?

(1) Choose the desired confidence level, $1-\alpha$.

(2) Compute:

$$s = \sqrt{\frac{n\Sigma X_i^2 - (\Sigma X_i)^2}{n(n-1)}}$$

(3) Look up B_U and B_L for $n-1$ degrees of freedom in Table A-20.

(4) Compute:

$$s_L = B_L s$$

$$s_U = B_U s$$

Conclude: The interval from s_L to s_U is a two-sided $100(1-\alpha)\%$ confidence interval estimate for σ; i.e., we may assert with $100(1-\alpha)\%$ confidence that $s_L < \sigma < s_U$.

Example

Problem: What is a 95% confidence interval for σ, the standard deviation of burning time in the lot of powder? (Data Sample 2-2)

(1) Let $1-\alpha = .95$

$\alpha = .05$

(2)

$s = 10.37$ seconds

(3) For 9 degrees of freedom,

$B_L = .6657$

$B_U = 1.746$

(4)

$s_L = (10.37)(.6657) = 6.90$ seconds

$s_U = (10.37)(1.746) = 18.11$ seconds

Conclude: The interval from 6.90 to 18.11 is a two-sided 95% confidence interval for σ; i.e., we may assert with 95% confidence that 6.90 seconds $< \sigma <$ 18.11 seconds.

2-2.3.2 One-Sided Confidence Interval Estimates

In some applications we are not particularly interested in placing both an upper and a lower bound on σ, but only in knowing whether the variability is excessively *large* (or, exceptionally *small*). We would like to make a statement such as the following: We can state

* The reader who is not familiar with the meaning and interpretation of confidence intervals should refer to Chapter 1, and to Paragraph 2-1.3 of this chapter. The remarks of Paragraph 2-1.3 concerning confidence intervals for the average carry over to confidence intervals for a measure of variability.

with $100(1-\alpha)\%$ confidence that the variability as measured by σ is less than some value s'_U computed from the sample. Similarly, but not simultaneously, we may wish to state with $100(1-\alpha)\%$ confidence that σ is greater than some value s'_L. Either statement is a one-sided confidence interval estimate.

Procedure	Example
Problem: What is a value s'_U such that we may have $100(1-\alpha)\%$ confidence that σ is less than s'_U?	*Problem:* What is a value s'_U such that we have 95% confidence that σ is less than s'_U? (Data Sample 2-2)
(1) Choose the desired confidence level, $1-\alpha$.	(1) Let $1-\alpha = .95$ $\alpha = .05$
(2) Compute: s	(2) $s = 10.37$ seconds
(3) Look up $A_{1-\alpha}$ for $n-1$ degrees of freedom in Table A-21.	(3) For 9 degrees of freedom, $A_{.95} = 1.645$
(4) Compute: $s'_U = A_{1-\alpha}\, s$	(4) $s'_U = (1.645)(10.37)$ $= 17.06$ seconds
(5) With $100(1-\alpha)\%$ confidence we can assert that σ is less than s'_U.	(5) We are 95% confident that the variability as measured by σ is *less than* $s'_U = 17.06$ seconds.

Should a lower bound to σ be desired, follow Procedure 2-2.3.2 with s'_U and $A_{1-\alpha}$ replaced by s'_L and A_α, respectively. Then it can be asserted with $100(1-\alpha)\%$ confidence that $\sigma > s'_L$.

2-2.4 ESTIMATING THE STANDARD DEVIATION WHEN NO SAMPLE DATA ARE AVAILABLE

It is often necessary to have some idea of the magnitude of the variation of some characteristic as measured by σ, its standard deviation in the population. In planning experiments, for example, the sample size required to meet certain requirements is a function of σ. In almost any situation, one can get at least a very rough estimate of σ. The minimum necessary information involves the form of the distribution and the spread of values. For example, if the values of the individual items can be assumed to follow a normal distribution, then either of the following rules can be used to get an estimate of σ:

(a) Estimate two values a_1 and b_1 between which you expect 99.7% (almost all) of the individuals to be. Then, estimate

$$\sigma \text{ as } \frac{|a_1 - b_1|}{6}$$

(b) Estimate two values a_2 and b_2 between which you expect 95% of the individuals to be. Then, estimate σ as $\dfrac{|a_2 - b_2|}{4}$

If the distribution concerned cannot be assumed to be normal but can be assumed to follow one of the top four forms in Figure 2-1, then the standard deviation may be estimated as indicated in the figure. This figure also illustrates the distribution and rules for (a) and (b) above.

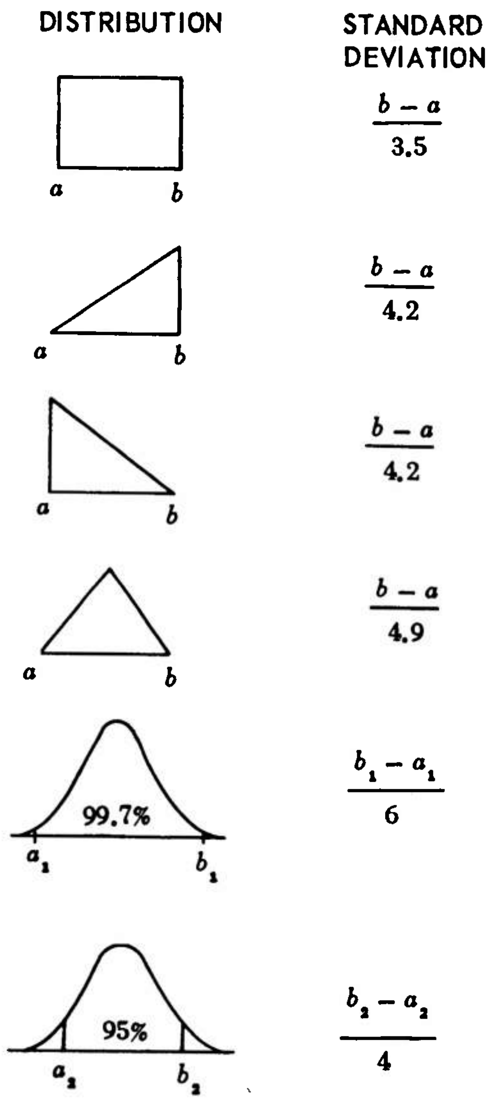

Figure 2-1. The standard deviation of some simple distributions.

Adapted with permission from *Some Theory of Sampling*, by W. Edwards Deming, Copyright, 1950, John Wiley & Sons, Inc.

2-3 NUMBER OF MEASUREMENTS REQUIRED TO ESTABLISH THE MEAN WITH PRESCRIBED ACCURACY

2-3.1 GENERAL

In planning experiments, we may need to know how many measurements or how large a sample to take in order to determine the mean of some distribution with prescribed accuracy. Suppose we are willing to allow a margin of error d, and a risk α that our estimate of m will be off by an amount d or greater. Since the sampling distribution of

$\bar{X}$ is "normal" to a good approximation for samples of four or more measurements from almost every population distribution likely to be met in practice, we can ascertain the required sample size n if we have an available estimate s of σ, or if we are willing to assume that we know σ. If we have not made an estimate or are unwilling to assume a value for σ, then we must use a two-stage sample. The two-stage method will usually result in a smaller total sample size. In the two-stage method, we must start by guessing a value of σ, but the end results do not depend upon how good or bad is the guess.

Sometimes we may have available to us one or more samples from the population of interest, from which we can derive an estimate s of σ based on ν degrees of freedom. Other times we may have one or more samples from some other population that has the same standard deviation as the population of interest, but possibly a different mean. Again, we can derive an estimate s of σ based on ν degrees of freedom. In either case, we can utilize this preliminary estimate of σ to determine the sample size n required to estimate the mean of the population of interest with prescribed accuracy.

2-3.2 ESTIMATION OF THE MEAN OF A POPULATION USING A SINGLE SAMPLE

Procedure

Problem: We wish to know the sample size required to ascertain the mean m of a population. We are willing to take a risk α that our estimate is off by d or more. There is available an estimate s of the population standard deviation σ, based on ν degrees of freedom.

(1) Choose d, the allowable margin of error, and α, the risk that our estimate of m will be off by d or more.

(2) Look up $t_{1-\alpha/2}$ for ν degrees of freedom in Table A-4.

(3) Compute:

$$n = \frac{t^2 s^2}{d^2}$$

Conclude: If we now compute the mean $\bar{X}$ of a random sample of size n from the population we may have $100(1-\alpha)\%$ confidence that the interval $\bar{X} - d$ to $\bar{X} + d$ will include the population mean m.

Example

Problem: We wish to know the average thickness of the washers in a given lot. We are willing to take a risk that 5 times in 100 the error in our estimate will be 0.002 inch or more. From a sample from another lot we have an estimate of the population standard deviation of $s = .00359$ with 9 degrees of freedom.

(1) Let $d = 0.002$ inch
$\alpha = .05$

(2) $t = t_{.975}$ for 9 degrees of freedom
$= 2.262.$

(3)

$$n = \frac{(2.262)^2 (.00359)^2}{(.002)^2} = 16.5$$

$= 17$ (conventionally rounded up to the next integer.)

Conclude: We may conclude that if we now compute the mean $\bar{X}$ of a random sample of size $n = 17$ from the lot of washers, we may have 95% confidence that the interval $\bar{X} - .002$ to $\bar{X} + .002$ will include the lot mean.

If we know σ, or assume some value for σ, replace s by σ and $t_{1-\alpha/2}$ by $z_{1-\alpha/2}$ in the above procedure. Values of $z_{1-\alpha/2}$ are given in Table A-2.

2-3.3 ESTIMATION USING A SAMPLE WHICH IS TAKEN IN TWO STAGES

It is possible that we do not have a good estimate of σ, the standard deviation of the population. When the cost of sampling is high, rather than take a larger sample than is really necessary, we might prefer to take the sample in two stages. The method (sometimes called

Stein's method) goes roughly as follows: Make a guess for the value of σ. From this determine n_1 the size of the first sample. The first sample will provide an estimate s of the population standard deviation. Use this value of s to determine how large the second sample should be.

Procedure

Problem: We wish to know the sample size required to ascertain the mean m of a population. We are willing to take a risk α that our estimate is off by d or more units.

(1) Choose d, the allowable margin of error, and α, the risk that our estimate of m will be off by d or more.

(2) Let σ' be the best possible guess for the value of σ, the standard deviation of the population (see Paragraph 2-2.4).

(3) Look up $z_{1-\alpha/2}$ in Table A-2.

(4) Compute:

$$n' = \left(\frac{z_{1-\alpha/2}\,\sigma}{d}\right)^2$$

n' is the first estimate of the total sample size required.

(5) Choose n_1 the size of the first sample. n_1 should be considerably less than n'. (If the guessed value of σ is too large, this will protect us against a first sample which is already larger than we need.) A rough rule might be to make $n_1 \geq 30$ unless $n' < 60$, in which case let n_1 be somewhere between $.5n'$ and $.7n'$.

(6) Make the necessary observations on the sample of n_1. Compute s_1, the standard deviation.

(7) Look up $t_{1-\alpha/2}$ for $n_1 - 1$ degrees of freedom in Table A-4.

(8) Compute

$$n = \frac{t^2 s_1^2}{d^2}$$

n is the total required sample size for the first and second samples combined. We then require a second sample size of $n_2 = n - n_1$.

Example

Problem: We have a large lot of devices, and wish to determine the average of some property. We are willing to take a risk of .05 of the estimate being in error by 30 units.

(1) Let $d = 30$

$\alpha = .05$

(2) From our knowledge of similar devices our best estimate of σ is 200 units.

(3) $z_{.975} = 1.960$

(4)

$$n' = \frac{(1.960)^2\,(200)^2}{(30)^2} = 170.7$$

(5) Let $n_1 = 50$

(6) From tests on 50 devices chosen at random, $s_1 = 160$ units.

(7) $t = t_{.975}$ for 49 degrees of freedom $= 2.01$.

(8)

$$n = \frac{(2.01)^2\,(160)^2}{(30)^2} = 114.9 = 115$$

$$n_2 = 115 - 50 = 65$$

We will require an additional 65 devices to be tested.

If now we obtain the second sample of size n_2 and compute the mean $\bar{X}$ of the total sample of size $n = n_1 + n_2$, we may have 100 $(1 - \alpha)$ % confidence that the interval $\bar{X} - d$ to $\bar{X} + d$ will include the population mean m.

2-4 NUMBER OF MEASUREMENTS REQUIRED TO ESTABLISH THE VARIABILITY WITH STATED PRECISION

We may wish to know the size of sample required to estimate the standard deviation with certain precision. If we can express this precision as a percentage of the true (unknown) standard deviation, we can use the curves in Figure 2-2.

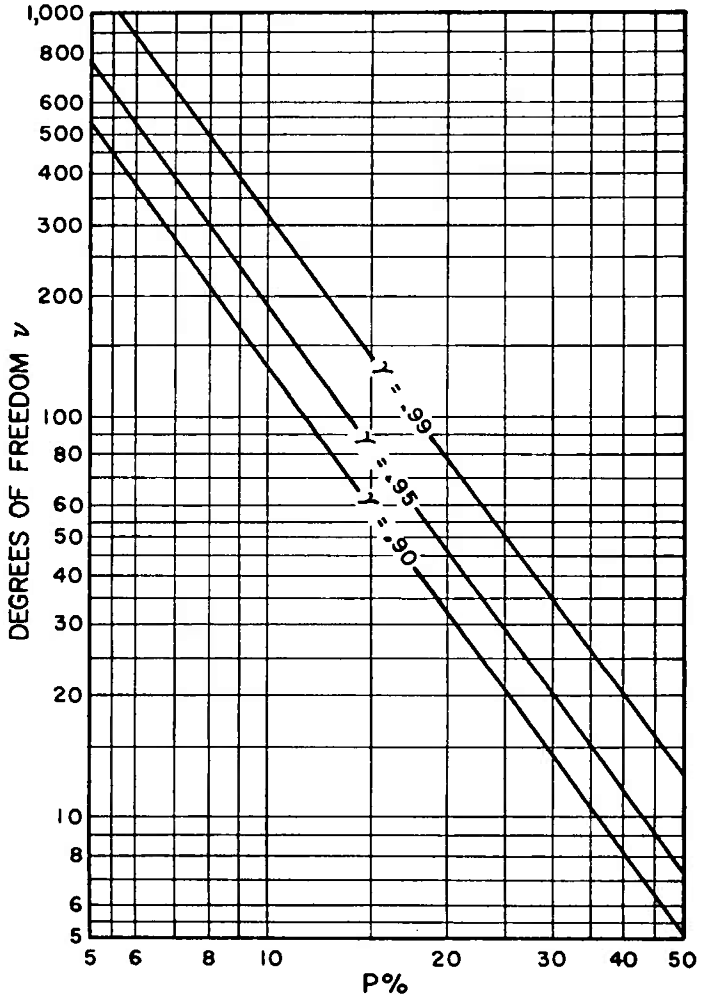

Figure 2-2. Number of degrees of freedom required to estimate the standard deviation within P% of its true value with confidence coefficient γ.

Adapted with permission from *Journal of the American Statistical Association*, Vol. 45 (1950), p. 258, from article entitled "Sample Size Required for Estimating the Standard Deviation as a Percent of its True Value" by J. A. Greenwood and M. M. Sandomire. The manner of graphing is adapted with permission from *Statistics Manual* by E. L. Crow, F. A. Davis, and M. W. Maxfield. NAVORD Report 3369, NOTS 948, U. S. Naval Ordnance Test Station, China Lake, Calif., 1955. (Reprinted by Dover Publications, Inc., New York, N.Y., 1960.)

Procedure

Problem: If we are to make a simple series of measurements, how many measurements are required to estimate the standard deviation within P percent of its true value, with prescribed confidence?

(1) Specify P, the allowable percentage deviation of the estimated standard deviation from its true value.

(2) Choose γ, the confidence coefficient.

(3) In Figure 2-2, find P on the horizontal scale, and use the curve for the appropriate γ. Read on the vertical scale the required degrees of freedom.

(4) For a simple series of measurements, the required number is equal to one plus the degrees of freedom.

Example

Problem: How large a sample would be required to estimate the standard deviation within 20% of its true value, with confidence coefficient equal to 0.95?

(1) Let $P = 20\%$

(2) Let $\gamma = .95$

(3) For $\gamma = .95$, $P = 20\%$, the required degrees of freedom equals 46.

(4) $n = 46 + 1$
$= 47$

2-5 STATISTICAL TOLERANCE LIMITS

2-5.1 GENERAL

Sometimes we are more interested in the approximate *range* of values in a lot or population than we are in its average value. Statistical tolerance limits furnish limits between, above, or below which we confidently expect to find a prescribed proportion of individual items of the population. Thus, we might like to be able to give two values A and B between which we can be fairly certain that at least a proportion P of the population will lie, (two-sided limits), or a value A above which at least a proportion P will lie, (one-sided limit).

Thus for the data on thickness of mica washers (Data Sample 2-1), we could give two thickness values, stating with chosen confidence that a proportion P (at least) of the washers in the lot have thicknesses between these two limits. We call the confidence coefficient γ, and it refers to the proportion of the time that our method will result in correct statements. If a normal distribution can be assumed, use the procedures of Paragraphs 2-5.2 and 2-5.3; otherwise use the procedures of Paragraph 2-5.4.

2-5.2 TWO-SIDED TOLERANCE LIMITS FOR A NORMAL DISTRIBUTION

When the mean m and standard deviation σ of a normally distributed quantity are known, symmetrical limits that include a prescribed proportion P of the distribution are readily obtained by adding and subtracting $z_\alpha \sigma$ from the known mean m, where z_α is read from Table A-2 with $\alpha = \frac{1}{2}(P+1)$. When m and σ are not known, we can use an interval of the form $\bar{X} \pm Ks$. Since both $\bar{X}$ and s will vary from sample to sample it is impossible to determine K so that the limits $\bar{X} \pm Ks$ will always include a specified proportion P of the underlying normal distribution. It is, however, possible to determine K so that in a long series of samples from the same or different normal distributions a definite proportion γ of the intervals $\bar{X} \pm Ks$ will include P or more of the underlying distribution(s).

Procedure

Problem: We would like to state two limits between which we are 100 γ percent confident that 100 P percent of the values lie.

(1) Choose P, the proportion, and γ, the confidence coefficient.

(2) Compute from the sample:

$\bar{X}$

s

(3) Look up K for chosen P and γ in Table A-6.

(4) Compute:

$$X_U = \bar{X} + Ks$$

$$X_L = \bar{X} - Ks$$

Conclude: With 100 γ % confidence we may predict that a proportion P of the individuals of the population have values between X_L and X_U.

Example

Problem: We would like to state thickness limits between which we are 95% confident that 90% of the values lie (Data Sample 2-1).

(1) Let $P = .90$

$\gamma = .95$

(2)

$\bar{X} = .1260$ inch

$s = 0.00359$ inch

(3) $K = 2.839$

(4)

$X_U = .1260 + 2.839\,(.00359)$

$= 0.136$ inch

$X_L = .1260 - 2.839\,(.00359)$

$= 0.116$ inch

Conclude: With 95% confidence, we may say that 90% of the washers have thicknesses between 0.116 and 0.136 inch.

2-5.3 ONE-SIDED TOLERANCE LIMITS FOR A NORMAL DISTRIBUTION

Sometimes we are interested only in estimating a value above which, or below which, a proportion P (at least) will lie. In this case the one-sided upper tolerance limit will be $X_U = \bar{X} + Ks$; and $X_L = \bar{X} - Ks$ will be the one-sided lower limit. The appropriate values for K are given in Table A-7 and are not the same as those of Paragraph 2-5.2.

Procedure

Problem: To find a single value above which we may predict with confidence γ that a proportion P of the population will lie.

(1) Choose P the proportion and γ, the confidence coefficient.

(2) Compute:

$\bar{X}$

s

(3) Look up K in Table A-7 for the appropriate n, γ, and P.

(4) $X_L = \bar{X} - Ks$

Example

Problem: To find a single value above which we may predict with 90% confidence that 99% of the population will lie. (Data Sample 2-1).

(1) Let $P = .99$

$\gamma = .90$

(2)

$\bar{X} = .1260$ inch

$s = 0.00359$ inch

(3) $K\,(10, .90, .99) = 3.532$

(4) $X_L = .1260 - 3.532\,(.00359)$

$= .1133$ inch

Thus we are 90% confident that 99% of the mica washers will have thicknesses above .113 inch.

Note: Factors for some values of n, γ, and P not covered in Table A-7 may be found in Sandia Corporation Monograph SCR-13[2]. Alternatively, one may compute K using the following formulas:

$$a = 1 - \frac{z_\gamma^2}{2(n-1)} \qquad \text{(where } z \text{ can be found in Table A-2)}$$

$$b = z_P^2 - \frac{z_\gamma^2}{n}$$

$$K = \frac{z_P + \sqrt{z_P^2 - ab}}{a}$$

2-5.4 TOLERANCE LIMITS WHICH ARE INDEPENDENT OF THE FORM OF THE DISTRIBUTION

The methods given in Paragraphs 2-5.2 and 2-5.3 are based on the assumption that the observations come from a normal distribution. If the distribution is not in fact normal, then the effect will be that the true proportion P of the population between the tolerance limits will vary from the intended P by an amount depending on the amount of departure from normality. If the departure from normality is more than slight we can use a procedure which assumes only that the distribution has no discontinuities. The tolerance limits so obtained will be substantially wider than those assuming normality.

2-5.4.1 Two-Sided Tolerance Limits (Distribution-Free)

Table A-30 gives values (r, s) such that we may assert with confidence at least γ that $100P\%$ of a population lies between the r^{th} smallest and the s^{th} largest of a random sample of n from that population. For example, from Table A-30 with $\gamma = .95$, $P = .75$, and $n = 60$, we may say that if we have a sample of $n = 60$, then we may have a confidence of at least $\gamma = .95$ that $100P\% = 75\%$ of the population will lie between the fifth largest $(s = 5)$ and the fifth smallest $(r = 5)$ of the sample values. That is, if we were to take many random samples of 60, and take the fifth largest and fifth smallest of each, we should expect to find that at least 95% of the resulting intervals would contain 75% of the population.

Table A-32 may be useful for sample sizes of $n \leq 100$. This table gives the confidence γ with which we may assert that $100P\%$ of the population lies between the largest and smallest values of the sample.

2-5.4.2 One-Sided Tolerance Limits (Distribution-Free)

Table A-31 gives the largest value of m such that we may assert with confidence at least γ that $100P\%$ of a population lies below the m^{th} largest (or above the m^{th} smallest) of a random sample of n from that population. For example, from Table A-31 with $\gamma = .95$, $P = .90$, and $n = 90$, we may say that we are 95% confident that 90% of a population will lie below the fifth largest value of a sample of size $n = 90$.

REFERENCES

1. M. G. Kendall and W. R. Buckland, *A Dictionary of Statistical Terms,* p. 79, Oliver and Boyd, London, 1957.
2. D. B. Owen, *Table of Factors for One-Sided Tolerance Limits for a Normal Distribution,* Sandia Corporation Monograph SCR-13, April 1958.

CHAPTER 3

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO AVERAGE PERFORMANCE

3-1 GENERAL REMARKS ON STATISTICAL TESTS

One of the most frequent uses of statistics is in testing for differences. If we wish to know whether a treatment applied to a standard round affects its muzzle velocity, we may conduct an experiment and apply a statistical test to the experimental results to see whether we would be justified in concluding that there is a difference between the performance of treated and untreated rounds. In another case, two manufacturing processes may be available—process A is cheaper and therefore preferable unless process B is demonstrated to be superior in some respect. Again, we apply a statistical test to the experimental results to see whether process B has demonstrated superiority.

Ordinarily, the statistical test applied to the results observed on a sample will point the way to decision between a pair of alternatives. For some tests, the two alternative decisions will be formally stated as follows:

(a) There *is* a difference between the (population) averages of two materials, products, processes, etc.

(b) No difference has been demonstrated.

In other cases, the formal statement of the two alternative decisions will be:

(a) The (population) average of product A *is* greater than that of product B.

(b) We have no reason to believe that the (population) average of product A is greater than that of product B.

In this Chapter and others, we shall consider a number of statistical tests of differences. The application of each statistical test will result in making one of two decisions, as in the pairs given. *In each case the pair of alternative decisions is chosen before the data are observed—this is important!*

Since we ordinarily obtain information on one or both of the products by means of a sample, we may sometimes make an erroneous decision. However, the chance of making the wrong decision can be reduced by increasing the number of observations. There are two ways in which we can make a wrong decision:

(a) When we conclude that there is a difference where in fact there is none, we say that we make an *Error of the First Kind;*

(b) When we fail to find a difference that really exists, then we say that we make an *Error of the Second Kind.*

In any particular case, we never can be absolutely sure that the correct decision has been made, but we can know the probability of making either type of error.

The probability of making an Error of the First Kind is usually denoted by α; and the probability of making an Error of the Second Kind is denoted by β. The ability of a given statistical test to detect a difference (e.g., between averages) will in general depend on the size of the difference δ; thus, β has no meaning unless associated with a particular difference δ. The value of β, $\beta(\delta)$, associated with a particular difference δ will decrease as δ increases. For a particular statistical test, the ability to detect a difference will be determined by three quantities:

α, $\beta(\delta)$, and n the sample size. The complementary quantity $1-\beta(\delta)$ is termed the *power* of the test to detect a difference δ with a sample of size n, when the test is carried out at the α-level of significance.

The decision procedure is a very logical one. Suppose we wish to test whether two types of vacuum tubes have the same resistance in ohms, on the average. We take samples of each type and measure their resistances. If the sample mean of one type of tube differs sufficiently from the sample mean of the other, we shall say that the two kinds of tubes differ in their average resistance. Otherwise, we shall say we failed to find a difference. How large must the difference be in order that we may conclude that the two types differ, or that the observed difference is "significant"?* This will depend on several factors: the amount of variability in the tubes of each type; the number of tubes of each type; and the risk we are willing to take of stating that a difference exists when there really is none, i.e., the risk of making an Error of the First Kind. We might proceed as follows: we would be willing to state that the true averages differ, if a difference larger than the observed difference could arise by chance less than five times in a hundred when the true averages are in fact equal. The probability of an Error of the First Kind is then $\alpha = .05$, or, as we commonly say, we have adopted a .05 *significance level.* The use of a *significance level* of .05 or .01 is common, and these levels are tabulated extensively for many tests. There is nothing unique about these levels, however, and a test user may choose any value for α that he feels is appropriate.

As we have mentioned, the ability to detect a difference will in general depend on the size of the difference δ. Let us denote by $\beta(\delta)$ the probability of failing to detect a specified difference δ. If we plot $\beta(\delta)$ versus the difference δ, we have what we call an Operating Characteristic (OC) curve. Actually, we usually plot $\beta(\delta)$ versus some convenient function of δ. Figures 3-1 through 3-8 show OC curves for a number of statistical tests when conducted at the $\alpha = .05$ or $\alpha = .01$ significance levels.

An OC curve depicts the discriminatory power of a particular statistical test. For specified values of n and α, there is a unique OC curve. The curve is useful in two ways. If we have specified n and α, we can use the OC curve to read $\beta(\delta)$ for various values of δ. If we are still at liberty to set the sample size for our experiment, and have a particular value of δ in mind, we can see what value of n is required by looking at the OC curves for specified α. If, for the α chosen, the sample size required to achieve a reasonably small $\beta(\delta)$ is too large, and if it really is important to detect a difference of δ when it exists, then a less conservative (i.e., larger) value of α must be used. Various uses of the OC curves shown in Figures 3-1 through 3-8 are described in detail in the appropriate paragraphs of this Chapter.

It is evident that for any $\beta(\delta)$, n will increase as δ decreases. It requires larger samples to recognize smaller differences. In some cases, the experiment as originally thought of will be seen to require prohibitively large sample sizes. We then must compromise between the sharp discriminatory power we think we need, the cost of the amount of testing required to achieve that power, and the risk of claiming a difference when none exists. If the experiment has already been run, and the sample size was fixed from other considerations, the OC curve will show what chance the experiment had of detecting a particular difference δ.

To use the OC curves in this Chapter, we must know the population standard deviation σ, or at least be willing to choose some range for σ. It is quite often possible to assign some upper bound to the variability, even without the use of past data (see Paragraph 2-2.4). After the experiment has been run, a possibly better estimate of σ will be available, and a hindsight look at the OC curve using this value will help to evaluate the experiment.

* Or more accurately, *statistically significant.* A difference may be *statistically significant* and yet be *practically unimportant.*

We outline a number of different tests in this Chapter. For each test, we give the procedure to be followed for a specified significance level α and sample size n. For most of the tests, we also give the OC curve which enables us to obtain the (approximate) value of β for any given difference. Tables are provided for determining n, the sample size required when α, δ, and $\beta(\delta)$ have been specified. The tests given are exact when:

(a) the observations for each item are taken randomly from a single population of possible observations; and,

(b) the quality characteristic measured is normally distributed within this population. Ordinarily, the assumption of normality is not crucial, particularly if the sample size is not very small.

Alternate procedures for most of the tests in this Chapter are given in AMCP 706-113, Chapters 15 and 16. Chapter 16 gives tests which require neither normality assumptions nor knowledge of the variability of the populations; but this greater generality is achieved at the price of somewhat reduced discriminating power when normality can be assumed and the knowledge about the variability of the populations, needed for the tests of this Chapter, is in hand. Chapter 15 gives shortcut tests for small samples from normal populations which involve less computation than the tests of this Chapter with negligible loss of efficiency.

3-2 COMPARING THE AVERAGE OF A NEW PRODUCT WITH THAT OF A STANDARD

The average performance of a standard product is known to be m_0. We shall consider three different problems:

(a) To determine whether the average of a new product *differs* from the standard, Paragraph 3-2.1.

(b) To determine whether the average of a new product *exceeds* the standard, Paragraph 3-2.2.

(c) To determine whether the average of a new product is *less* than the standard, Paragraph 3-2.3.

For summary of the procedures appropriate for each of these three problems, see Table 3-1.

It is necessary to decide which of the three problems is appropriate before taking the observations. If this is not done and the choice of the problem is influenced by the observations, (for example, Paragraph 3-2.1 vs. 3-2.2), the significance level of the test, i.e., the probability of an *Error of the First Kind*, and the operating characteristics of the test may differ considerably from their nominal values.

Ordinarily the variability of a new product is not known. At other times previous experience may enable us to state a value of σ. We shall outline the solutions of the three problems (Paragraphs 3-2.1, 3-2.2, and 3-2.3) for both cases, i.e., where the variability is estimated from the sample, and where σ is known from previous experience.

Symbols to be used:

m = average of new material, product or process (unknown).

m_0 = average of standard material, product or process (known).

$\bar{X}$ = average of sample of n measurements on new product.

s = standard deviation estimate computed from n measurements on the new product (used where σ is unknown).

σ = the known standard deviation of the new product.

Data Sample 3-2—Weight of Powder

For a certain type of shell, specifications state that the amount of powder should average 0.735 pound. In order to determine whether the average for a new stock meets the specification, 20 shells are taken at random, and the amount of powder contained in each is weighed.

The sample average $\bar{X}$ = .710 pound.

The sample standard deviation estimate s = .0504 pound. In illustrating the known-σ case, we assume σ known to be equal to 0.06 pound.

TABLE 3-1. SUMMARY OF TECHNIQUES FOR COMPARING THE AVERAGE OF A NEW PRODUCT WITH THAT OF A STANDARD (FOR DETAILS AND WORKED EXAMPLES SEE PARAGRAPHS 3-2.1, 3-2.2, AND 3-2.3)

We Wish to Test Whether	Paragraph Reference	Knowledge of Variation of New Item	Test to be Made	Operating Characteristics of the Test (for $\alpha = .05$ and $\alpha = .01$)	Sample Size Required n	Notes
m differs from m_0	3-2.1.1	σ unknown; s = estimate of σ from sample.	$\lvert \bar{X} - m_0 \rvert > u$	See Figs. 3-1 and 3-2*	Use Table A-8. For $\alpha = .05$, add 2 to tabular value. For $\alpha = .01$, add 4 to tabular value.	$u = t_{1-\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ (t for $n-1$ degrees of freedom)
	3-2.1.2	σ known	$\lvert \bar{X} - m_0 \rvert > u$	See Figs. 3-3 and 3-4	Use Table A-8.	$u = z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$
m is larger than m_0	3-2.2.1	σ unknown; s = estimate of σ from sample.	$(\bar{X} - m_0) > u$	See Figs. 3-5 and 3-6*	Use Table A-9. For $\alpha = .05$, add 2 to tabular value. For $\alpha = .01$, add 3 to tabular value.	$u = t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$ (t for $n-1$ degrees of freedom)
	3-2.2.2	σ known	$(\bar{X} - m_0) > u$	See Figs. 3-7 and 3-8	Use Table A-9.	$u = z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$
m is smaller than m_0	3-2.3.1	σ unknown; s = estimate of σ from sample.	$(m_0 - \bar{X}) > u$	See Figs. 3-5 and 3-6*	Use Table A-9. For $\alpha = .05$, add 2 to the tabular values. For $\alpha = .01$, add 3 to the tabular values.	$u = t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$ (t for $n-1$ degrees of freedom)
	3-2.3.2	σ known	$(m_0 - \bar{X}) > u$	See Figs. 3-7 and 3-8	Use Table A-9.	$u = z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$

* It is necessary to have some value for σ (or two bounding values) in order to use the Operating Characteristic curve. Although σ is unknown, in many situations it is possible to have some notion, however loose, about the magnitude of σ and thereby to get helpful information from the OC curve. Paragraph 2-2.4 gives assistance in estimating σ from general knowledge of the process.

3-2.1 TO DETERMINE WHETHER THE AVERAGE OF A NEW PRODUCT DIFFERS FROM THE STANDARD

3-2.1.1 Does the Average of the New Product Differ from the Standard (σ Unknown)?

[Two-sided t-test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $t_{1-\alpha/2}$ for $n-1$ degrees of freedom in Table A-4.

(3) Compute $\bar{X}$ and s from the n measurements.

(4) Compute

$$u = t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

(5) If $\lvert \bar{X} - m_0 \rvert > u$, decide that the average of the new type differs from that of the standard; otherwise, that there is no reason to believe that they differ.

(6) *Note:* The interval $\bar{X} \pm u$ is a $100(1-\alpha)$ % confidence interval estimate of the true average of the new type.

Example

(1) Let $\alpha = .05$

(2) $t_{.975}$ for 19 degrees of freedom = 2.093

(3)

$\bar{X} = .710$ pound

$s = .0504$ pound

(Data Sample 3-2)

(4)

$$u = \frac{(2.093)(.0504)}{\sqrt{20}} = .0236$$

(5) $\lvert \bar{X} - m_0 \rvert = \lvert .710 - .735 \rvert = .025$. We conclude that the average amount of powder in the new stock differs from 0.735, the specified standard amount.

(6) *Note:* $.710 \pm .0236$ is a 95% confidence interval estimate of the true average of the new stock.

Operating Characteristics of the Test. Figures 3-1 and 3-2 give the operating characteristic (OC) curves of the preceding test for $\alpha = .05$ and $\alpha = .01$, respectively, and various values of n.

Choose:

$\delta = |m - m_0|$,
the true absolute difference between the averages (unknown, of course)

Some value of σ.
(One may use an estimate from previous data; lacking such an estimate, see Paragraph 2-2.4. If the OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d = \frac{\delta}{\sigma} .$$

We then can read from the OC curve for a given significance level α and sample size n, a value of $\beta(\delta)$. The $\beta(\delta)$ read from the curve is $\beta(\delta \mid \sigma, \alpha, n)$, i.e., $\beta(\delta,$ *given* $\sigma, \alpha, n)$—the probability of failing to detect this difference when the given test is carried out with a sample of size n, at the α-level of significance, and the population standard deviation actually is σ.

If we use too large a value for σ, the effect is to underestimate d, and consequently to overestimate $\beta(\delta)$, the probability of not detecting a difference of δ when it exists. Conversely, if we choose too small a value of σ, then we shall overestimate d and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample size n and the significance level α employed, *and* the true value of σ.

Selection of Sample Size n. If we choose

$\delta = |m - m_0|$, the absolute value of the average difference that we desire to detect
α, the significance level of the test
β, the probability of failing to detect a difference δ

and compute

$$d = \frac{|m - m_0|}{\sigma}$$

then we may use Table A-8 to obtain a good approximation to the required sample size. If we take $\alpha = .01$, then we must add 4 to the value obtained from the table. If we take $\alpha = .05$, then we must add 2 to the table value. (In order to compute d, we must choose a value for σ. See Paragraph 2-2.4 if no other information is available.)

As an example, suppose that we plan to take $\alpha = .05$, and want to have $\beta = .50$ for a difference of .024 pound; that is, we wish to conduct a test at a significance level of .05 that will have a 50-50 chance of detecting a difference of 0.024 pound. What sample size should we require? Suppose previous experience suggests that σ lies between .04 and .06 pound.

Taking $\sigma = .04$, with $\delta = |m - m_0| = .024$, gives $d = 0.6$. Using Table A-8, with $\alpha = .05$, $1 - \beta = .50$, we find the required sample size as $n = 11 + 2 = 13$. Taking $\sigma = .06$, yields $d = .4$. From the same table, we find that the required sample size is $25 + 2 = 27$. To be safe, we would use $n = 27$. For $\sigma \leq .06$, with a significance level of .05, this would give the two-sided t test at least a 50% chance of detecting a difference of 0.024 pound.

If, when planning an investigation leading to a two-sided t-test, we overestimate σ, the consequences are two-fold: first, we overestimate the sample size required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if

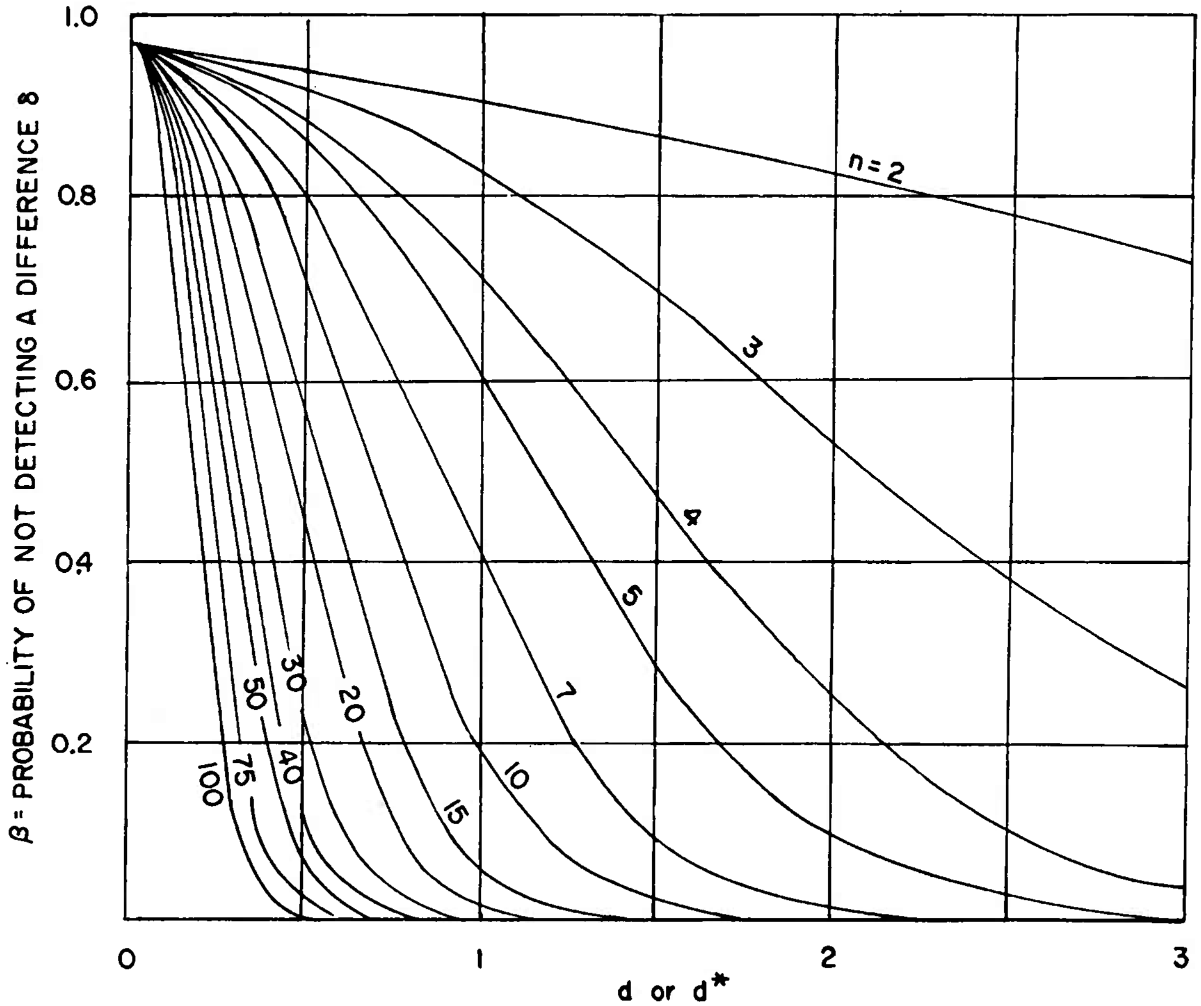

Figure 3-1. OC curves for the two-sided t-test ($\alpha = .05$).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, pp. 178-197, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Note: These curves apply to the following tests:

(a) Does the average m of a new product differ from a standard m_0?

$$\delta = |m - m_0|$$

$$d = \frac{|m - m_0|}{\sigma}$$ See Paragraph 3-2.1.1.

(b) Do the averages of two products differ?

$$\delta = |m_A - m_B|$$

$$d^* = \frac{|m_A - m_B|}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A\, n_B}{n_A + n_B}},$$

where $\sigma_A = \sigma_B = \sigma$ by assumption, and n_A and n_B are the respective sample sizes from products A and B. See Paragraph 3-3.1.1.

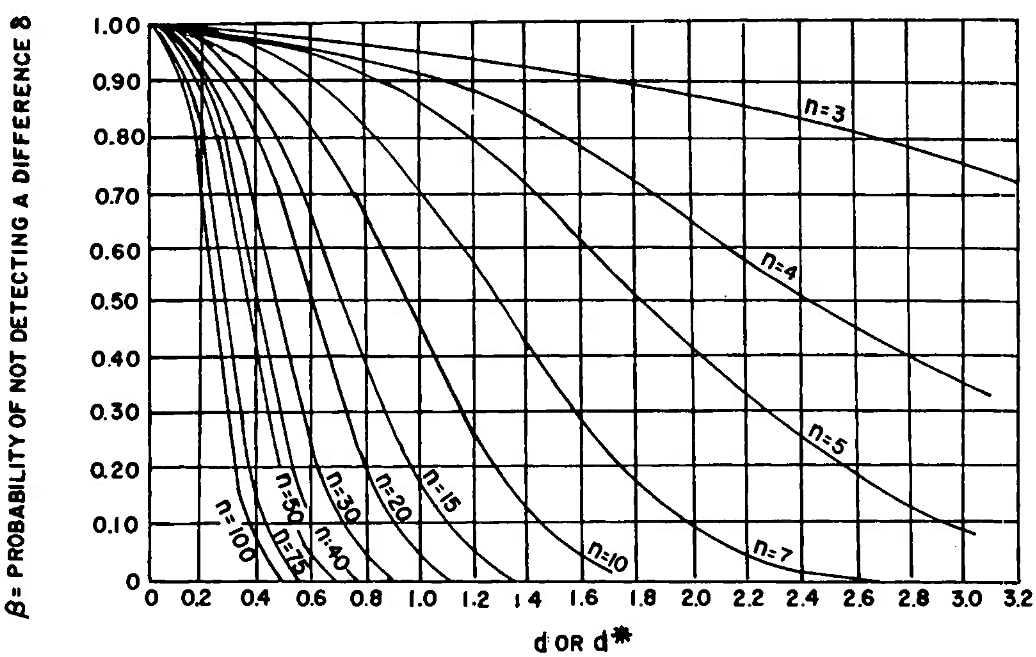

Figure 3-2. OC curves for the two-sided t-test (α = .01).

Adapted with permission from *Engineering Statistics* by A. H. Bowker and G. J. Lieberman, Copyright, 1959, Prentice-Hall, Inc.

Note: These curves apply to the following tests:

(a) Does the average m of a new product differ from a standard m_0?

$$\delta = |m - m_0|$$

$$d = \frac{|m - m_0|}{\sigma}$$ See Paragraph 3-2.1.1.

(b) Do the averages of two products differ?

$$\delta = |m_A - m_B|$$

$$d^* = \frac{|m_A - m_B|}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A n_B}{n_A + n_B}},$$

where $\sigma_A = \sigma_B = \sigma$ by assumption, and n_A and n_B are the respective sample sizes from products A and B. See Paragraph 3-3.1.1.

we underestimate σ, we shall underestimate the sample size actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened.

The following brief table, built around the preceding example, serves to illustrate these points numerically for a situation where $\alpha = .05$, and it is desired to have $\beta(\delta) = .50$ for $\delta = |m - m_0| = .024$, and σ in fact is equal to .04 though this is unknown.

Value of σ Assumed	Resulting Sample Size	Corresponding β (.024)
.08	45	.02
.06	27	.15
.04 (true value)	13	.50
.03	9	.64
.02	5	.80

Thus, if σ actually is .04, playing safe by taking $\sigma = .06$ has more than doubled the sample size actually needed, but we have gained a reduction in β from .50 to .15.

Finally, it should be noted that, inasmuch as the test criterion $u = t_{1-\alpha/2} \frac{s}{\sqrt{n}}$ does not depend on σ, an error in estimating σ when planning a two-sided t-test will not alter the level of significance of the test, which will be precisely equal to the value of α desired, *provided* that $t_{1-\alpha/2}$ is taken equal to the 100 $(1 - \alpha/2)$ percentile of the t distribution for $n - 1$ degrees of freedom, where n is the *sample size actually employed.*

3-2.1.2 Does the Average of the New Product Differ from the Standard (σ Known)? [Two-sided Normal Test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $z_{1-\alpha/2}$ in Table A-2.

(3) Compute $\bar{X}$, the mean of the n measurements.

(4) Compute

$$u = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(5) If $|\bar{X} - m_0| > u$ decide that the average of the new type differs from that of the standard; otherwise, that there is no reason to believe that they differ.

(6) Note that the interval $\bar{X} \pm u$ is a 100 $(1 - \alpha)$ % confidence interval estimate of the true average m of the new type.

Example

(1) Let $\alpha = .05$

(2) $z_{.975} = 1.960$

(3) $\bar{X} = .710$ pound (Data Sample 3-2)

(4) σ is known to be equal to .06 pound.

$$u = \frac{1.96\ (.06)}{\sqrt{20}} = .0263$$

(5) $|\bar{X} - m_0| = |.710 - .735| = .025$. We conclude that there is no reason to believe that the average amount of powder in the new stock differs from 0.735 (the specified standard amount).

(6) Note that $(.710 \pm .0263)$ is a 95% confidence interval estimate for the true average m of the new stock.

Operating Characteristics of the Test. Figures 3-3 and 3-4 give the operating characteristics of the preceding test for $\alpha = .05$ and $\alpha = .01$, respectively. For any given n and $d = \frac{|m - m_0|}{\sigma}$, the value of $\beta(\delta) = \beta(\delta | \sigma, \alpha, n)$, the probability of failing to detect a difference of absolute size $\delta = |m - m_0|$, can be read off directly.

Selection of Sample Size n. If we specify α, the significance level, and β, the probability or risk we are willing to take of not detecting a difference of absolute size $\delta = |m - m_0|$, then we can use Table A-8 to obtain n, the required sample size. As an example, if σ is known to be 0.04 pound, and we wish to have a 50-50 chance of detecting a difference of 0.024 pound, then $d = 0.6$. From Table A-8, we find that the required sample size is 11.

When we know the correct value of σ, we can achieve a desired value of $\beta(\delta)$ with fewer observations by using the normal test at the desired level of significance α than by using the corresponding t-test. The saving is 2 or 4 observations according as $\alpha = .05$ or $.01$, respectively.

Overestimating or underestimating σ when planning a two-sided normal test has somewhat different consequences than when planning a two-sided t-test. If we *overestimate* σ and choose $\sigma' > \sigma$, we also *overestimate* the sample size required as in the case of the t-test. In addition, we *overestimate* the correct test criterion $u = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ for the sample size n actually adopted, with the result that the effective significance level of the normal test is *reduced* to α', which is related to α by the equation

$$z_{1-\alpha'/2} = \left(\frac{\sigma'}{\sigma}\right) z_{1-\alpha/2}.$$

The actual probability of not detecting a difference of δ, $\beta'(\delta)$, is related to the intended risk $\beta(\delta)$ by the equation

$$z_{1-\beta'} = \left(\frac{\sigma'}{\sigma}\right) z_{1-\beta}.$$

$\beta'(\delta)$ will be less than $\beta(\delta)$ when $\sigma' > \sigma$ for all (large) δ for which $\beta(\delta) < 0.50$; $\beta'(\delta)$ will be larger than $\beta(\delta)$ for all (small) δ for which $\beta'(\delta) > 0.50$. For the particular δ for which $\beta(\delta) = 0.50$, $\beta'(\delta)$ also will equal 0.50. Conversely, if we *underestimate* σ, then we not only *underestimate* the sample size required but also the test criterion for the sample size actually used, so that the actual risk of an Error of the First Kind α' will be *larger* than α, and the risk of an Error of the Second Kind $\beta'(\delta)$ will be *increased* for large δ, and *decreased* for small δ.

The following calculations serve to illustrate these points numerically for situations bordering on the conditions assumed in the preceding sample-size calculation:

Intended significance level $\alpha = 0.05$.
Intended risk of *Error of the Second Kind* $\beta(\delta) = 0.50$ for $\delta = 0.024$.

TWO-SIDED NORMAL TEST

Value of σ Assumed	Sample Size Indicated	Actual Significance Level, α'	Actual Risk of Error of Second Kind, β' (0.024)
.08	43 (45)*	.00009 (.05)*	0.50 (.02)*
.06	25 (27)	.003 (.05)	0.50 (.15)
.04 (true value)	11 (13)	.05 (.05)	0.50 (.50)
.03	7 (9)	.14 (.05)	0.50 (.64)
.02	3 (5)	.33 (.05)	0.50 (.80)

* Values in parentheses are for corresponding two-sided *t*-test.

To obtain a numerical illustration of the more general case where $\beta(\delta) \neq 0.50$, let us modify the foregoing example by taking $\beta(\delta) = 0.20$, say, as the intended risk of an Error of the Second Kind for $\delta = 0.024$:

Intended significance level $\alpha = 0.05$.
Intended risk of Error of the Second Kind $\beta(\delta) = 0.20$ for $\delta = 0.024$.

TWO-SIDED NORMAL TEST

Value of σ Assumed	Sample Size Indicated	Actual Significance Level, α'	Actual Risk of Error of Second Kind, β' (0.024)
.08	88 (90)*	.00009 (.05)*	.046 (.0004)*
.06	50 (52)	.003 (.05)	.103 (.01)
.04 (true value)	22 (24)	.05 (.05)	.20 (.20)
.03	13 (15)	.14 (.05)	.26 (.43)
.02	6 (8)	.33 (.05)	.34 (.70)

* Values in parentheses are for corresponding two-sided *t*-test.

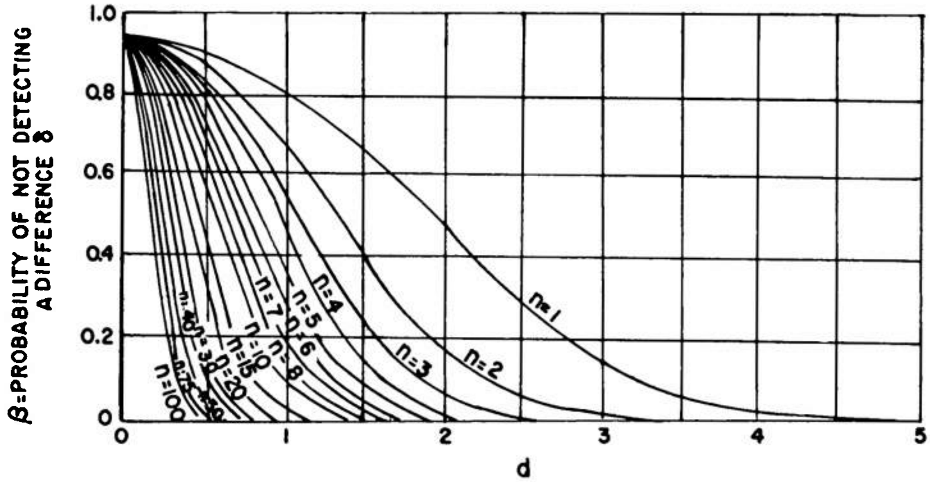

Figure 3-3. OC curves for the two-sided normal test (α = .05).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, pp. 178-197, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Note: These curves apply to the following tests:

(a) Does the average m of a new product differ from a standard m_0?

$$\delta = |m - m_0|$$

$$d = \frac{|m - m_0|}{\sigma}$$ See Paragraph 3-2.1.2.

(b) Do the averages of two products differ?

$$\delta = |m_A - m_B|$$

$$d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}};$$ σ_A and σ_B are known. See Paragraph 3-3.1.3.

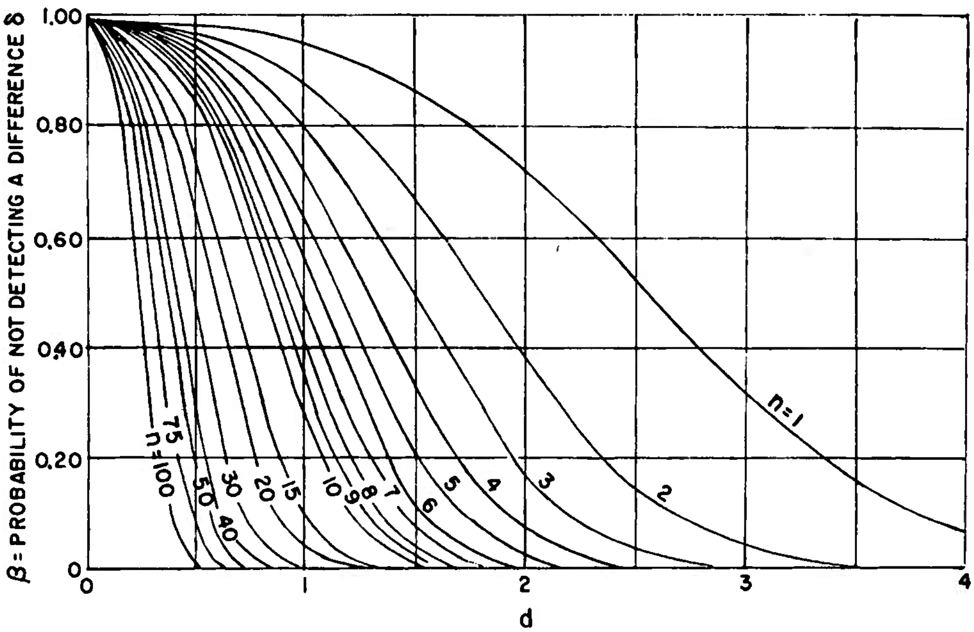

Figure 3-4. OC curves for the two-sided normal test (α = .01).

Adapted with permission from *Engineering Statistics* by A. H. Bowker and G. J. Lieberman, Copyright, 1959, Prentice-Hall, Inc.

Note: These curves apply to the following tests:

(a) Does the average m of a new product differ from a standard m_0?

$$\delta = |m - m_0|$$

$$d = \frac{|m - m_0|}{\sigma}.$$ See Paragraph 3-2.1.2.

(b) Do the averages of two products differ?

$$\delta = |m_A - m_B|$$

$$d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}};$$ σ_A and σ_B are known. See Paragraph 3-3.1.3.

3-2.2 TO DETERMINE WHETHER THE AVERAGE OF A NEW PRODUCT EXCEEDS THE STANDARD

3-2.2.1 Does the Average of the New Product Exceed the Standard (σ Unknown)?

[One-sided *t*-test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $t_{1-\alpha}$ for $n - 1$ degrees of freedom in Table A-4.

(3) Compute $\bar{X}$ and s

(4) Compute

$$u = t_{1-\alpha} \frac{s}{\sqrt{n}}$$

(5) If $(\bar{X} - m_0) > +u$, decide that the average of the new type exceeds that of the standard; otherwise, that there is no reason to believe that the average of the new type exceeds that of the standard.

(6) Note that the open interval from $(\bar{X} - u)$ to $+\infty$ is a one-sided $100(1 - \alpha)\%$ confidence interval for the true mean of the new product.

Example

(1) Let $\alpha = .05$

(2) $t_{.95}$ for 19 degrees of freedom $= 1.729$

(3)

$\bar{X} = .710$ pound

$s = .0504$ pound

(Data Sample 3.2)

(4)

$$u = \frac{1.729\ (.0504)}{\sqrt{20}} = 0.019$$

(5) $(\bar{X} - m_0) = (.710 - .735) = -.025$. We conclude that there is no reason to believe that the average of the new product exceeds the specified standard.

(6) Note that the open interval from .691 to $+\infty$ is a one-sided 95% confidence interval for true average of the new product.

Operating Characteristics of the Test. Figures 3-5 and 3-6 give the operating characteristic (OC) curves of the above test for $\alpha = .05$, and $\alpha = .01$, respectively, and various values of n.

Choose:

$\delta = (m - m_0)$, the true difference between averages, (unknown, of course)

Some value of σ. (We may use an estimate from previous data; lacking such an estimate, see Paragraph 2-2.4. If OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d = \frac{\delta}{\sigma} \ .$$

We then can read from the OC curve for a given significance level α and sample size n, a value of $\beta(\delta)$. The $\beta(\delta)$ read from the curve is $\beta(\delta \mid \sigma, \alpha, n)$, i.e., $\beta(\delta$ *given* $\sigma, \alpha, n)$—the probability of failing to detect this difference when the given test is carried out with a sample of size n, at the α-level of significance, and the population standard deviation actually is σ.

If we use too large a value for σ, the effect is to underestimate d and consequently to overestimate $\beta(\delta)$, the probability of not detecting a difference of δ when it exists. Conversely, if we choose too small a value of σ, then we shall overestimate d and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample size n and significance level α employed, *and* the true value of σ.

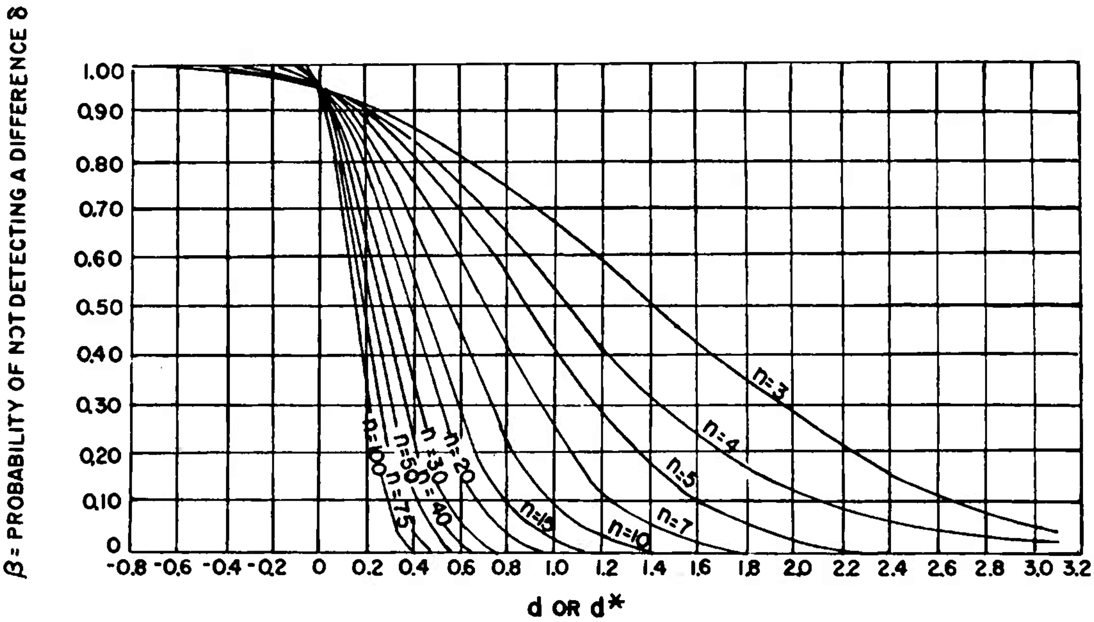

Figure 3-5. OC curves for the one-sided t-test (α = .05).

Adapted with permission from *Engineering Statistics* by A. H. Bowker and G. J. Lieberman, Copyright, 1959, Prentice-Hall, Inc.

Note: These curves apply to the following tests:

(a) Does the average m of a new product exceed a standard m_0?

$$\delta = m - m_0$$

$$d = \frac{m - m_0}{\sigma}$$ See Paragraph 3-2.2.1.

(b) Is the average m of a new product less than a standard m_0?

$$\delta = m_0 - m$$

$$d = \frac{m_0 - m}{\sigma}$$ See Paragraph 3-2.3.1.

(c) Does the average of product A exceed that of product B?

$$\delta = m_A - m_B$$

$$d^* = \frac{m_A - m_B}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A n_B}{n_A + n_B}},$$

where $\sigma_A = \sigma_B = \sigma$ by assumption, and n_A and n_B are the respective sample sizes from products A and B. See Paragraph 3-3.2.1.

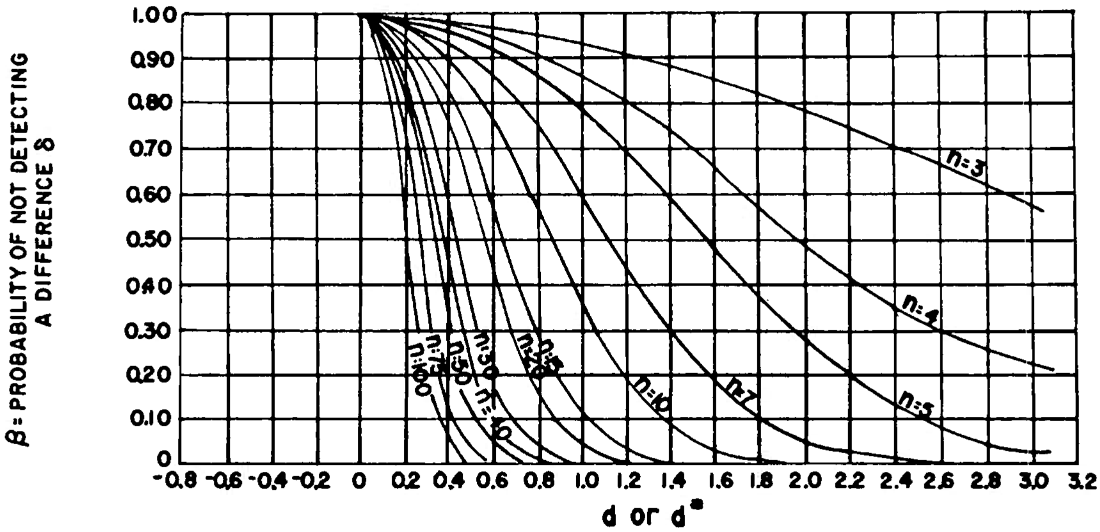

Figure 3-6. OC curves for the one-sided t-test ($\alpha = .01$).

Adapted with permission from *Engineering Statistics* by A. H. Bowker and G. J. Lieberman, Copyright, 1959, Prentice-Hall, Inc.

Note: These curves apply to the following tests:

(a) Does the average m of a new product exceed a standard m_0?

$$\delta = m - m_0$$

$$d = \frac{m - m_0}{\sigma}$$ See Paragraph 3-2.2.1.

(b) Is the average m of a new product less than a standard m_0?

$$\delta = m_0 - m$$

$$d = \frac{m_0 - m}{\sigma}$$ See Paragraph 3-2.3.1.

(c) Does the average of product A exceed that of product B?

$$\delta = m_A - m_B$$

$$d^* = \frac{m_A - m_B}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A\, n_B}{n_A + n_B}},$$

where $\sigma_A = \sigma_B = \sigma$ by assumption, and n_A and n_B are the respective sample sizes from products A and B. See Paragraph 3-3.2.1.

Selection of Sample Size n. If we choose

$\delta = (m - m_0)$,

α, the significance level of the test

β, the probability of failing to detect a positive difference of size $(m - m_0)$

and compute

$$d = \frac{m - m_0}{\sigma}$$

then we may use Table A-9 to obtain a good approximation to the required sample size. If we are using $\alpha = .01$, then we must add 3 to the table value. If we are using $\alpha = .05$, then we must add 2 to the table value. (In order to compute d, we must choose a value for σ; see Paragraph 2-2.4 when no other information is available.)

If, when planning an investigation leading to a one-sided t-test, we overestimate σ, the consequences are two-fold: first, we overestimate the sample size required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if we underestimate σ, we shall underestimate the sample size actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened. (A numerical example for the *two*-sided t-test is given in Paragraph 3-2.1.1. The one-sided case is similar).

Finally, it should be noted, that inasmuch as the test criterion $u = t_{1-\alpha} \frac{s}{\sqrt{n}}$ does not depend on σ, an error in estimating σ when planning a one-sided t-test does not alter the level of significance of the test, which will be precisely equal to the value of α desired, *provided* that $t_{1-\alpha}$ is taken equal to the 100 $(1 - \alpha)$ percentile of the t distribution for $n - 1$ degrees of freedom, where n is the *sample size actually employed.*

3-2.2.2 Does the Average of the New Product Exceed the Standard (σ Known)?

[One-sided Normal Test]

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Look up $z_{1-\alpha}$ in Table A-2.	(2) $z_{.95} = 1.645$
(3) Compute $\bar{X}$, the sample mean	(3) $\bar{X} = 0.710$ pound (Data Sample 3-2)
(4) Compute $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$	(4) σ is known to be equal to .06 pound. $u = \frac{1.645\ (.06)}{\sqrt{20}} = .022$
(5) If $(\bar{X} - m_0) > u$, decide that the average performance of the new type exceeds that of the standard; otherwise, that there is no reason to believe that the average of the new type exceeds that of the standard.	(5) $(\bar{X} - m_0) = .710 - .735 = -.025$, which is not larger than u. We conclude that there is no reason to believe that the average of the new product exceeds that of the standard.
(6) Note that the open interval from $(\bar{X} - u)$ to $+\infty$ is a one-sided 100 $(1 - \alpha)$ % confidence interval for the true mean of the new product.	(6) Note that the open interval from .688 to $+\infty$ is a 95% one-sided confidence interval for the true mean of the new product.

Operating Characteristics of the Test. Figures 3-7 and 3-8 give the operating characteristics of the above test for $\alpha = .05$ and $\alpha = .01$, respectively. For any given n and $d = \frac{m - m_0}{\sigma}$, the value of $\beta(\delta) = \beta(\delta | \sigma, \alpha, n)$, the probability of failing to detect a positive difference $\delta = (m - m_0)$, can be read off directly.

Selection of Sample Size n. If we specify

$\delta = (m - m_0)$, the magnitude of a positive difference of interest to us
α, the significance level of the test
β, the probability of failing to detect a positive difference of size δ

and compute

$$d = \frac{m - m_0}{\sigma}$$

then we may use Table A-9 to obtain the required sample size.

When we know the correct value of σ, we can achieve a desired value of $\beta(\delta)$ with fewer observations by using the normal test at the desired level of significance α than by using the corresponding t-test. The saving is 2 or 3 observations according as $\alpha = .05$ or $.01$, respectively.

Overestimating or underestimating σ when planning a one-sided normal test has somewhat different consequences than when planning a one-sided t-test. If we *overestimate* σ and choose $\sigma' > \sigma$, we also *overestimate* the sample size required as in the case of the t-test. In addition, we *overestimate* the correct test criterion $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ for the sample size n actually adopted, with the result that the effective significance level of the normal test is *reduced* to α', which is related to α by the equation

$$z_{1-\alpha'} = \left(\frac{\sigma'}{\sigma}\right) z_{1-\alpha}.$$

The actual probability of not detecting a difference of δ, $\beta'(\delta)$, is related to the intended risk $\beta(\delta)$ by the equation

$$z_{1-\beta'} = \left(\frac{\sigma'}{\sigma}\right) z_{1-\beta}.$$

$\beta'(\delta)$ will be less than $\beta(\delta)$ when $\sigma' > \sigma$ for all (large) δ for which $\beta(\delta) < 0.50$; $\beta'(\delta)$ will be larger than $\beta(\delta)$ for all (small) δ for which $\beta'(\delta) > 0.50$. For the particular δ for which $\beta(\delta) = 0.50$, $\beta'(\delta)$ also will equal 0.50. Conversely, if we *underestimate* σ, then we not only *underestimate* the sample size required but also the test criterion for the sample size actually used, so that the actual risk of an Error of the First Kind α' will be *larger* than α, and the risk of an Error of the Second Kind $\beta'(\delta)$ will be *increased* for large δ, and *decreased* for small δ. (Numerical examples for the two-sided normal test are given in Paragraph 3-2.1.2. The one-sided case is similar.)

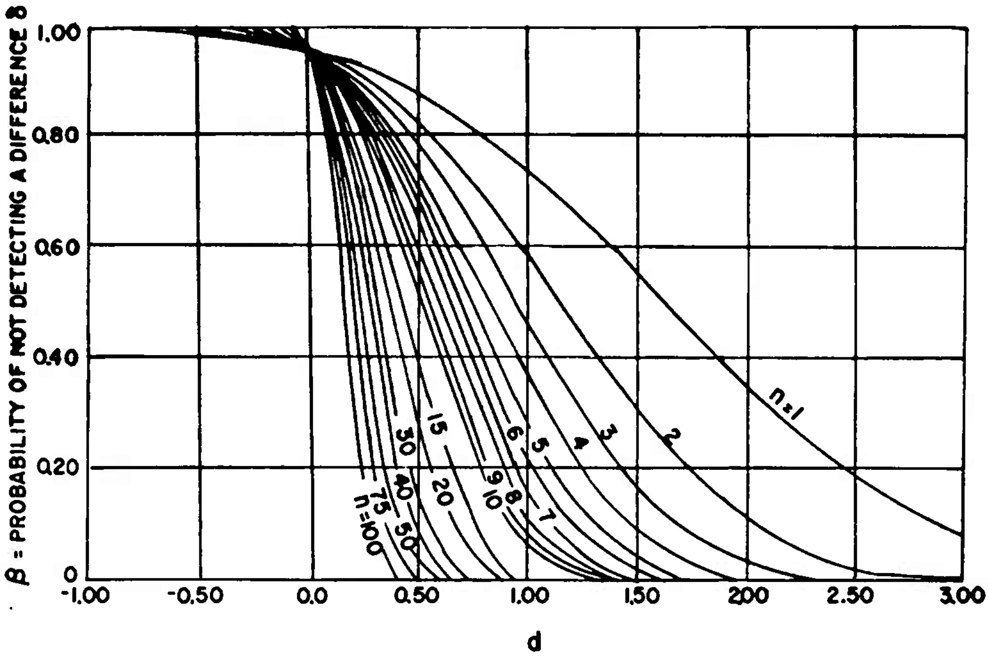

Figure 3-7. OC curves for the one-sided normal test ($\alpha = .05$).

Note: These curves apply to the following tests:

(a) Does the average m of a new product exceed a standard m_0?

$$\delta = m - m_0$$

$$d = \frac{m - m_0}{\sigma}$$ See Paragraph 3-2.2.2.

(b) Is the average m of a new product less than a standard m_0?

$$\delta = m_0 - m$$

$$d = \frac{m_0 - m}{\sigma}$$ See Paragraph 3-2.3.2.

(c) Does the average of product A exceed that of product B?

$$\delta = m_A - m_B$$

$$d = \frac{m_A - m_B}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$ σ_A and σ_B are known. See Paragraph 3-3.2.3.

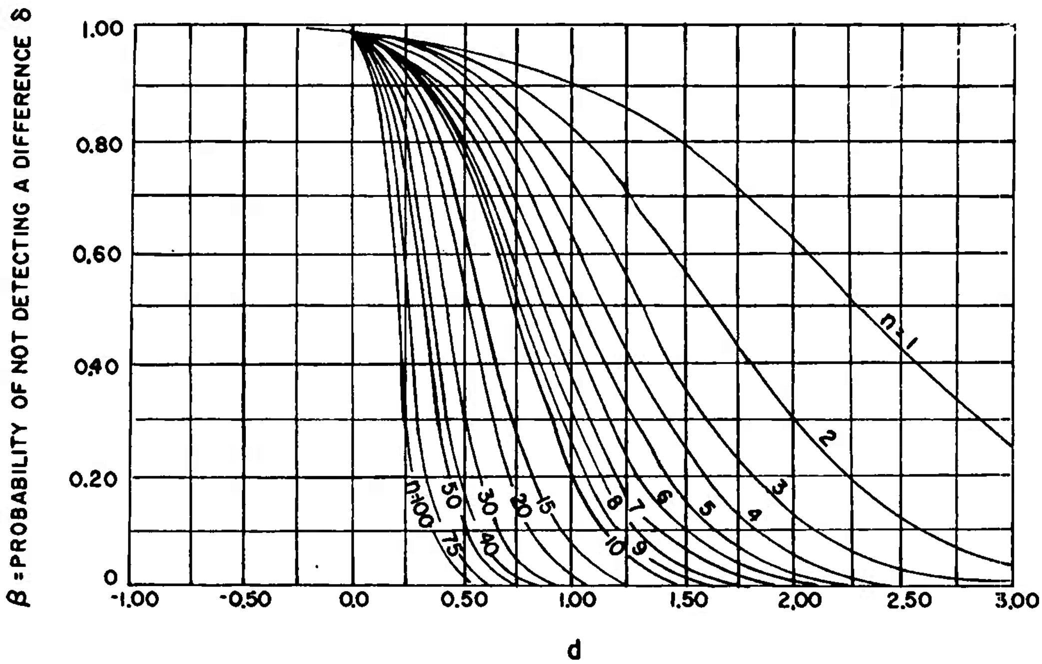

Figure 3-8. OC curves for the one-sided normal test (α = *.01*).

Adapted with permission from *Engineering Statistics* by A. H. Bowker and G. J. Lieberman, Copyright, 1959, Prentice-Hall, Inc.

Note: These curves apply to the following tests:

(a) Does the average m of a new product exceed a standard m_0?

$$\delta = m - m_0$$

$$d = \frac{m - m_0}{\sigma} \qquad \text{See Paragraph 3-2.2.2.}$$

(b) Is the average m of a new product less than a standard m_0?

$$\delta = m_0 - m$$

$$d = \frac{m_0 - m}{\sigma} \qquad \text{See Paragraph 3-2.3.2.}$$

(c) Does the average of product A exceed that of product B?

$$\delta = m_A - m_B$$

$$d = \frac{m_A - m_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \quad \sigma_A \text{ and } \sigma_B \text{ are known. See Paragraph 3-3.2.3.}$$

3-2.3 TO DETERMINE WHETHER THE AVERAGE OF A NEW PRODUCT IS LESS THAN THE STANDARD

3-2.3.1 Is the Average of the New Product Less than the Standard (σ Unknown)?

[One-sided t-test]

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Look up $t_{1-\alpha}$ for $n - 1$ degrees of freedom in Table A-4.	(2) $t_{.95}$ for 19 degrees of freedom $= 1.729$
(3) Compute $\bar{X}$ and s	(3) $\bar{X} = .710$ pound $s = .0504$ pound (Data Sample 3-2)
(4) Compute $u = t_{1-\alpha} \frac{s}{\sqrt{n}}$	(4) $u = \frac{1.729\ (.0504)}{\sqrt{20}}$ $= 0.019$
(5) If $(m_0 - \bar{X}) > u$, decide that the average of the new type is less than that of the standard; otherwise, that there is no reason to believe that the average of the new type is less than the standard.	(5) $.735 - .710 = .025$. We conclude that the average of the new type is less than that of the standard.
(6) Note that the open interval from $-\infty$ to $(\bar{X} + u)$ is a one-sided $100\ (1 - \alpha)\ \%$ confidence interval for the true mean of the new type.	(6) Note that the open interval from $-\infty$ to .729 is a one-sided 95% confidence interval for the true mean of the new type.

Operating Characteristics of the Test. Figures 3-5 and 3-6 give the operating characteristic (OC) curves of the above test for $\alpha = .05$, and $\alpha = .01$, respectively, for various values of n.

Choose:

$\delta = (m_0 - m)$, the true difference between averages (unknown, of course)

Some value of σ. (We may use an estimate from previous data; lacking such an estimate, see Paragraph 2-2.4. If OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d = \frac{\delta}{\sigma} .$$

We then can read from the OC curve for a given significance level α and sample size n, a value of $\beta(\delta)$. The $\beta(\delta)$ read from the curve is $\beta(\delta \mid \sigma, \alpha, n)$, i.e., $\beta(\delta$ *given* $\sigma, \alpha, n)$—the probability of failing to detect this difference when the given test is carried out with a sample of size n, at the α-level of significance, and the population standard deviation actually is σ.

If we use too large a value for σ, the effect is to underestimate d and consequently to overestimate $\beta(\delta)$, the probability of not detecting a difference of δ when it exists. Conversely, if we choose too small a value of σ, then we shall overestimate d and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample size n and significance level α employed, *and* the true value of σ.

Selection of Sample Size n. If we choose

$\delta = (m_0 - m)$,

α, the significance level of the test

β, the probability of failing to find a negative difference of size $(m_0 - m)$;

and compute

$$d = \frac{m_0 - m}{\sigma}$$

then we may use Table A-9 to obtain a good approximation to the required sample size. If we are using $\alpha = .01$, then we must add 3 to the table value. If we are using $\alpha = .05$, then we must add 2 to the table value. (In order to use the table, we must have a value for σ. See Paragraph 2-2.4 if no other information is available.)

The effect of overestimating or underestimating σ is the same as when a one-sided t-test is to be used to detect a *positive* difference of magnitude $\delta = m - m_0$. See Paragraph 3-2.2.1.

3-2.3.2 Is the Average of the New Product Less Than That of the Standard (σ Known)? [One-sided Normal Test]

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Look up $z_{1-\alpha}$ in Table A-2.	(2) $z_{.95} = 1.645$
(3) Compute $\bar{X}$, the sample mean	(3) $\bar{X} = 0.710$ pound (Data Sample 3-2)
(4) Compute $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$	(4) σ is known to be equal to .06 pound. $u = \frac{1.645\ (.06)}{\sqrt{20}} = 0.022$
(5) If $(m_0 - \bar{X}) > u$, decide that the average of the new type is less than that of the standard; otherwise, that there is no reason to believe that the average of the new type is less than that of the standard.	(5) $(m_0 - \bar{X}) = (.735 - .710) = .025$, which is larger than u. We conclude that the average of the new type is less than the standard.
(6) Note that the open interval from $-\infty$ to $(\bar{X} + u)$ is a one-sided $100\ (1 - \alpha)\ \%$ confidence interval for the true mean of the new type.	(6) Note that the open interval from $-\infty$ to .732 is a one-sided 95% confidence interval for the true mean of the new type.

Operating Characteristics of the Test. Figures 3-7 and 3-8 give the operating characteristics of the test for $\alpha = .05$ and $\alpha = .01$, respectively. For any given n and $d = \frac{m_0 - m}{\sigma}$ the value of $\beta(\delta) = \beta(\delta | \sigma, \alpha, n)$, the probability of failing to detect a negative difference of size $(m_0 - m)$, can be read off directly.

Selection of Sample Size n. If we specify

$\delta = (m_0 - m)$, the magnitude of a negative difference of interest to us
α, the significance level of the test
β, the probability of failing to detect a negative difference of size δ,

and compute

$$d = \frac{m_0 - m}{\sigma}$$

then we may use Table A-9 to obtain the required sample size.

The effect of overestimating or underestimating σ is the same as when the one-sided normal test is to be used to detect a *positive* difference of magnitude $\delta = m - m_0$. See Paragraph 3-2.2.2.

3-3 COMPARING THE AVERAGES OF TWO MATERIALS, PRODUCTS, OR PROCESSES

We consider two problems:

(a) We wish to test whether the averages of two materials, products, or processes differ, and we are not particularly concerned which is larger, Paragraph 3-3.1.

(b) We wish to test whether the average of material, product, or process A exceeds that of material, product, or process B, Paragraph 3-3.2.

TABLE 3-2. SUMMARY OF TECHNIQUES FOR COMPARING THE AVERAGE PERFORMANCE OF TWO PRODUCTS (FOR DETAILS AND WORKED EXAMPLES, SEE PARAGRAPHS 3-3.1 AND 3-3.2)

We Wish to Test Whether	Paragraph Reference	Knowledge of Variation	Test to be Made	Operating Characteristics of Test	Determination of Sample Size n	Notes
m_A differs from m_B	3-3.1.1	$\sigma_A \simeq \sigma_B$; both unknown	$\lvert\bar{X}_A - \bar{X}_B\rvert > u$, where $u = t_{1-\alpha/2}\, s_p \sqrt{\frac{n_A + n_B}{n_A n_B}}$	For $\alpha = .05$ and $\alpha = .01$ see Figs. 3-1 and 3-2* and Par. 3-3.1.1.	Use Table A-8. For $\alpha = .05$, add 1 to the tabular value. For $\alpha = .01$, add 2 to the tabular value.	$s_p = \sqrt{\frac{(n_A - 1)\, s_A^2 + (n_B - 1)\, s_B^2}{n_A + n_B - 2}}$
	3-3.1.2	$\sigma_A \neq \sigma_B$; both unknown	$\lvert\bar{X}_A - \bar{X}_B\rvert > u$, where $u = t' \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$ See Notes.			t' is the value of $t_{1-\alpha/2}$ for the effective number of degrees of freedom $f = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{(s_A^2/n_A)^2}{n_A + 1} + \frac{(s_B^2/n_B)^2}{n_B + 1}} - 2$
	3-3.1.3	σ_A, σ_B; both known	$\lvert\bar{X}_A - \bar{X}_B\rvert > u$, where $u = z_{1-\alpha/2} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$	For $\alpha = .05$ and $\alpha = .01$, see Figs. 3-3 and 3-4.	Use Table A-8.	
m_A is greater than m_B	3-3.2.1	$\sigma_A \simeq \sigma_B$; both unknown	$(\bar{X}_A - \bar{X}_B) > u$, where $u = t_{1-\alpha}\, s_p \sqrt{\frac{n_A + n_B}{n_A n_B}}$	For $\alpha = .05$ and $\alpha = .01$ see Figs. 3-5 and 3-6* and Par. 3-3.2.1.	Use Table A-9. For $\alpha = .05$, add 1 to the tabular value. For $\alpha = .01$, add 2 to the tabular value.	$s_p = \sqrt{\frac{(n_A - 1)\, s_A^2 + (n_B - 1)\, s_B^2}{n_A + n_B - 2}}$
	3-3.2.2	$\sigma_A \neq \sigma_B$; both unknown	$(\bar{X}_A - \bar{X}_B) > u$, where $u = t' \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$			t' is the value of $t_{1-\alpha}$ for the effective number of degrees of freedom $f = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{(s_A^2/n_A)^2}{n_A + 1} + \frac{(s_B^2/n_B)^2}{n_B + 1}} - 2$
	3-3.2.3	σ_A, σ_B; both known	$(\bar{X}_A - \bar{X}_B) > u$, where $u = z_{1-\alpha} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$	For $\alpha = .05$ and $\alpha = .01$ see Figs. 3-7 and 3-8.	Use Table A-9.	

* Although the common σ is unknown, useful information may be obtained from the OC curve if a value (or 2 bounding values) of σ can be assumed.

It again is important to decide which problem is appropriate before making the observations. If this is not done and the choice of the problem is influenced by the observations, the significance level of the test, i.e., the probability of an Error of the First Kind, and the operating characteristics of the test may differ considerably from their nominal values. It is assumed that the appropriate problem has been selected and that n_A and n_B observations are taken from products A and B, respectively.

Ordinarily, we will not know σ_A or σ_B. In some cases, it may be safe to assume that σ_A is approximately equal to σ_B.* We give the solutions for the two problems (Paragraphs 3-3.1 and 3-3.2) for three situations with regard to knowledge of the variability, and for the special case where the observations are paired.

Case 1—The variability in performance of each of A and B is unknown but can be assumed to be about the same.

Case 2—The variability in performance of each of A and B is unknown, and it is not reasonable to assume that they both have the same variability.

Case 3—The variability in performance of each of A and B is known from previous experience. The standard deviations are σ_A and σ_B, respectively.

Case 4—The observations are paired.

3-3.1 DO THE PRODUCTS A AND B DIFFER IN AVERAGE PERFORMANCE?

3-3.1.1 (Case 1)—Variability of A and B is Unknown, But Can Be Assumed to be Equal.

Data Sample 3-3.1.1—Latent Heat of Fusion of Ice

Two methods were used in a study of the latent heat of fusion of ice. Both Method A (an electrical method) and Method B (a method of mixtures) were conducted with the specimens cooled to −0.72°C. The data represent the change in total heat from −0.72°C to water at 0°C, in calories per gram of mass.

Method A	Method B
79.98	80.02
80.04	79.94
80.02	79.98
80.04	79.97
80.03	79.97
80.03	80.03
80.04	79.95
79.97	79.97
80.05	
80.03	
80.02	
80.00	
80.02	

* For a procedure to test whether σ_A and σ_B differ, see Chapter 4.

[Two-sided t-test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $t_{1-\alpha/2}$ for $\nu = (n_A + n_B - 2)$ degrees of freedom in Table A-4.

(3) Compute: $\bar{X}_A$ and s_A^2, $\bar{X}_B$ and s_B^2, for the n_A and n_B measurements from A and B.

(4) Compute

$$s_P = \sqrt{\frac{(n_A - 1)\, s_A^2 + (n_B - 1)\, s_B^2}{n_A + n_B - 2}}$$

(5) Compute

$$u = t_{1-\alpha/2}\, s_P \sqrt{\frac{n_A + n_B}{n_A\, n_B}}$$

(6) If $|\bar{X}_A - \bar{X}_B| > u$, decide that A and B differ with regard to their average performance; otherwise, that there is no reason to believe A and B differ with regard to their average performance.

(7) Let m_A, m_B be the true average performances of A and B (unknown of course). It is worth noting that the interval $(\bar{X}_A - \bar{X}_B) \pm u$ is a $100\,(1 - \alpha)\,\%$ confidence interval estimate of $(m_A - m_B)$.

Example

(1) Let $\alpha = .05$

(2)
$n_A = 13$
$n_B = 8$
$\nu = 19$ degrees of freedom
$t_{.975}$ for 19 d.f. $= 2.093$

(3)
$\bar{X}_A = 80.02$
$s_A^2 = .000574$
$\bar{X}_B = 79.98$
$s_B^2 = .000984$

(4)

$$s_P = \sqrt{\frac{12\,(.000574) + 7\,(.000984)}{19}}$$
$$= \sqrt{.000725}$$
$$= .0269$$

(5)

$$u = 2.093\,(.0269)\sqrt{\frac{21}{104}}$$
$$= (.05630)\,(.4493)$$
$$= .025$$

(6) $|\bar{X}_A - \bar{X}_B| = .04$, which is larger than u. Conclude that A and B differ with regard to average performance.

(7) The interval $.04 \pm .025$, i.e., the interval from .015 to .065, is a 95% confidence interval for the true difference between the averages of the methods.

Operating Characteristics of the Test. Figures 3-1 and 3-2 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of $n = n_A + n_B - 1$.

Choose:

$\delta = |m_A - m_B|$, the true absolute difference between the averages

Some value of $\sigma\ (= \sigma_A = \sigma_B)$, the common standard deviation.

(We may use an estimate from previous data; lacking such an estimate, see Paragraph 2-2.4. If OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d^* = \frac{|m_A - m_B|}{\sigma}\,\frac{1}{\sqrt{n_A + n_B - 1}}\sqrt{\frac{n_A\, n_B}{n_A + n_B}}$$

We then can read a value of $\beta(\delta)$ from the OC curve for a given significance level and effective sample size $n = n_A + n_B - 1$. The $\beta(\delta)$ read from the curve is $\beta(\delta \mid \sigma, \alpha, n_A, n_B)$ i.e., $\beta(\delta$, given σ, α, n_A and $n_B)$ the probability of failing to detect a real difference between the two population means of magnitude $\delta = \pm(m_A - m_B)$ when the test is carried out with samples of sizes n_A and n_B, respectively, at the α-level of significance, *and* the two population standard deviations actually *are* both equal to σ.

If we use too large a value for σ, the effect is to make us underestimate d^* and consequently to overestimate $\beta(\delta)$. Conversely, if we choose too small a value of σ, then we shall overestimate d^* and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample sizes n_A and n_B and significance level α actually employed, and the true value of σ $(= \sigma_A = \sigma_B)$.

Since the test criterion u does not depend on the value of σ $(= \sigma_A = \sigma_B)$, an error in estimating σ will not alter the significance level of the test, which will be precisely equal to the value of α desired, *provided that* the value of $t_{1-\alpha/2}$ is taken equal to the 100 $(1 - \alpha/2)$ percentile of the t-distribution for $n_A + n_B - 2$ degrees of freedom, where n_A and n_B are the sample sizes actually employed, *and* it actually *is* true that $\sigma_A = \sigma_B$.

If $\sigma_A \neq \sigma_B$, then, whatever may be the ratio σ_A/σ_B, the effective significance level α' will not differ seriously from the intended value α, provided that $n_A = n_B$, except possibly when both are as small as two. If, on the other hand, unequal sample sizes are used, and $\sigma_A \neq \sigma_B$, then the effective level of significance α' can differ considerably from the intended value α, as shown in Figure 3-9 where $\alpha = .05$.

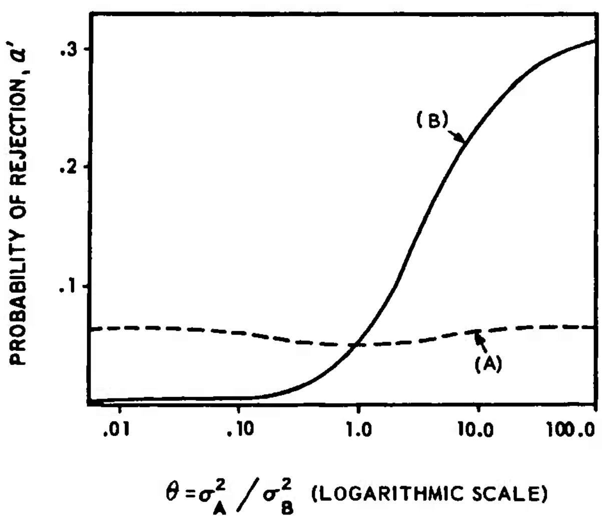

Figure 3-9. Probability of rejection of hypothesis $m_A = m_B$ when true, plotted against θ.

(A) $n_A = n_B = 10$, $P(|u|) > 2.101$;
(B) $n_A = 5$, $n_B = 15$, $P(|u|) > 2.101$.

Adapted with permission from *Biometrika*, Vol. XXIX, Parts III and IV, February 1938, from article entitled "The significance of the difference between two means when the population variances are unequal" by B. L. Welch.

Selection of Sample Size n. If we choose

$\delta = |m_A - m_B|$, the absolute value of the average difference that we desire to detect

α, the significance level of the test

β, the probability of failing to detect a difference of absolute size δ,

and compute

$$d = \frac{|m_A - m_B|}{\sqrt{2\sigma^2}}, \text{ where } \sigma = \sigma_A = \sigma_B ,$$

then we may use Table A-8 to obtain a good approximation to the required sample size $n\ (= n_A = n_B)$. If we take $\alpha = .01$, then we must add 2 to the value obtained from the table. If we take $\alpha = .05$, then we must add 1 to the table value.

In order to compute d, we must choose a value for $\sigma\ (= \sigma_A = \sigma_B)$. (See Paragraph 2-2.4 if no other information is available.) If we overestimate σ, the consequences are two-fold: first, we overestimate the sample size $n\ (= n_A = n_B)$ required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if we underestimate σ, we shall underestimate the sample size actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened. These effects of overestimating or underestimating $\sigma\ (= \sigma_A = \sigma_B)$ will be similar in magnitude to those considered and illustrated in Paragraph 3-2.1.1 for the case of comparing the mean m of a new material, product, or process, with a standard value m_0.

As explained in the preceding discussion of the operating characteristics of the test, an error in estimating $\sigma\ (= \sigma_A = \sigma_B)$ will have no effect on the significance level of the test, provided that the value of $t_{1-\alpha/2}$ is taken equal to the 100 $(1 - \alpha/2)$ percentile of the t-distribution for $n_A + n_B - 2$ degrees of freedom, where n_A and n_B are the sample sizes actually employed; and if $\sigma_A \neq \sigma_B$, the effect will not be serious provided that the sample sizes are taken equal.

3-3.1.2 (Case 2)—Variability of A and B is Unknown, Cannot Be Assumed Equal.

Data Sample 3-3.1.2—Compressive Strength of Concrete

Two investigators using somewhat different techniques obtained specimen cores to determine the compressive strength of the concrete in a poured slab. The following results in psi were reported:

A	B
3128	1939
3219	1697
3244	3030
3073	2424
	2020
	2909
	1815
	2020
	2310

Procedure*

(1) Choose α, the significance level of the test. (Actually, the procedure outlined will give a significance level of only approximately α).

(2) Compute: $\bar{X}_A$ and s_A^2, $\bar{X}_B$ and s_B^2, for the n_A and n_B measurements from A and B.

(3) Compute:

$$V_A = \frac{s_A^2}{n_A}$$

and

$$V_B = \frac{s_B^2}{n_B},$$

the estimated variances of $\bar{X}_A$ and $\bar{X}_B$, respectively.

(4) Compute the "effective number of degrees of freedom"

$$f = \frac{(V_A + V_B)^2}{\frac{V_A^2}{n_A + 1} + \frac{V_B^2}{n_B + 1}} - 2$$

(5) Look up $t_{1-\alpha/2}$ for f' degrees of freedom in Table A-4, where f' is the integer nearest to f; denote this value by $t'_{1-\alpha/2}$.

(6) Compute

$$u = t'_{1-\alpha/2} \sqrt{V_A + V_B}$$

(7) If $|\bar{X}_A - \bar{X}_B| > u$, decide that A and B differ with regard to their average performance; otherwise, decide that there is no reason to believe A and B differ in average performance.

Example

(1) Let $\alpha = .05$

(2)

$$\bar{X}_A = 3166.0$$
$$s_A^2 = 6328.67$$
$$n_A = 4$$
$$\bar{X}_B = 2240.4$$
$$s_B^2 = 221{,}661.3$$
$$n_B = 9$$

(3)

$$V_A = \frac{6328.67}{4} = 1582.17$$

$$V_B = \frac{221{,}661.3}{9} = 24629.03$$

(4)

$$f = \frac{(26211.20)^2}{500652.4 + 60658911.9} - 2$$
$$= \frac{687027005}{61159564} - 2$$
$$= 11.233 - 2$$
$$= 9.233$$

(5)

$$f' = 9$$
$$t'_{.975} = 2.262$$

(6)

$$u = 2.262 \sqrt{26211.20}$$
$$= 2.262\,(161.9)$$
$$= 366.2$$

(7) $|\bar{X}_A - \bar{X}_B| = 925.6$, which is larger than u. Conclude that A and B differ with regard to average performance.

---

* See footnote on page 3-28.

Procedure*	**Example**
(8) If m_A, m_B are the true average performances of A and B (unknown of course), then it is worth noting that the interval $(\bar{X}_A - \bar{X}_B) \pm u$ is approximately a $100\,(1-\alpha)$ % confidence interval estimate of $m_A - m_B$.	(8) The interval 925.6 ± 366.2, i.e., the interval from 559.4 psi to 1291.8 psi is a 95% confidence interval for the true difference between the averages of the two methods.

Discussion.

To gain some understanding of the nature, properties, and limitations of this approximate procedure, note first that V_A and V_B are unbiased estimates of the true variances σ_A^2/n_A and σ_B^2/n_B of the means $\bar{X}_A$ and $\bar{X}_B$, respectively. Consequently, $V_A + V_B$ is an unbiased estimate of the true variance of the difference $\bar{X}_A - \bar{X}_B$, provided only that $\bar{X}_A$ and $\bar{X}_B$ are the means of *independent* random samples of n_A and n_B observations from populations A and B, respectively. Note next that the effective number of degrees of freedom f, defined by the expression in step (4), also can be expressed in the form

$$\frac{1}{f+2} = \frac{c^2}{f_A+2} + \frac{(1-c)^2}{f_B+2}$$

where

$$f_A = n_A - 1 \quad \text{and} \quad f_B = n_B - 1$$

are the degrees of freedom associated with the variance estimates V_A and V_B, respectively, and,

$$c = \frac{V_A}{V_A + V_B} \quad \text{and} \quad 1 - c = \frac{V_B}{V_A + V_B}$$

are the fractions of the estimated variance of the difference $\bar{X}_A - \bar{X}_B$ that are associated with $\bar{X}_A$ and $\bar{X}_B$, respectively. From this expression for f, it is evident that f can never be less than the smaller of f_A $(= n_A - 1)$ and f_B $(= n_B - 1)$, and f cannot be larger than

$$(f_A + 2) + (f_B + 2) - 2 = n_A + n_B.$$

When V_A is so large in comparison to V_B that V_B is negligible, then $c \simeq 1$ and $f \simeq f_A$, which is intuitively reasonable—the f_B degrees of freedom upon which V_B is based are not making a useful contribution to the estimate of the variance of the difference $\bar{X}_A - \bar{X}_B$. Similarly, when V_B dominates the situation, then $c \simeq 0$ and $f \simeq f_B$. In intermediate situations where neither V_A nor V_B can be neglected, both the f_A and the f_B degrees of freedom make useful contributions, and the effective number of degrees of freedom f expresses the *sum* of their joint contributions. Thus, in our illustrative example, $f_A = 3$ and $f_B = 8$, but $f = 9^+$. Both samples make their maximum contributions, that is, f achieves its maximum of $n_A + n_B$, only when $V_A/V_B = (n_A + 1)/(n_B + 1)$, i.e., when $s_A^2/s_B^2 = n_A\,(n_A + 1)/n_B\,(n_B + 1)$.

* The test procedure given here is an approximation, i.e., the stated significance level is only approximately achieved. The approximation is good provided n_A and n_B are not too small. A more accurate procedure is given in *Biometrika Tables for Statisticians*,[1] which (in the notation of the present procedure) provides 10% and 2% significance levels of $|v| = |(\bar{X}_A - \bar{X}_B) - (m_A - m_B)|/\sqrt{V_A + V_B}$ for $n_A \geq 6$, $n_B \geq 6$, and $0 \leq V_A/(V_A + V_B) \leq 1$. 5% and 1% significance levels of $|v|$ for $n_A \geq 8$ and $n_B \geq 8$ and the same range of $V_A/(V_A + V_B)$ are given by Trickett, Welch, and James.[2] (When using either of the tables (1) or (2), it should be noticed that our "α" corresponds to their "2α".)
The appropriate modification when the value of the ratio of the variances $\theta = \sigma_A^2/\sigma_B^2$ is known, but not their respective values, is indicated at the end of the Discussion that follows this procedure.

When samples of equal size ($n_A = n_B = n$) are involved, the present approximate procedure for Case 2 (σ_A and σ_B both unknown and presumably *unequal*) in Paragraph 3-3.1.2, and the exact procedure for Case 1 (σ_A and σ_B presumably *equal*, but their common value unknown) given in Paragraph 3-3.1.1 are the same in all respects *except* for the value of $t_{1-\alpha/2}$ to be used. In the exact procedure for Case 1, the value of $t_{1-\alpha/2}$ to be used when $n_A = n_B = n$ is the 100 ($1 - \alpha/2$) percentile of the t-distribution for $\nu = 2(n - 1)$ degrees of freedom, and is completely determined by the choice of significance level α and the common sample size n. In contrast, the value of $t_{1-\alpha/2}$ to be used in the approximate procedure for Case 2 when $n_A = n_B = n$ is the 100 ($1 - \alpha/2$) percentile for the integral number of degrees of freedom f' nearest to the effective degrees of freedom

$$f = (n + 1)\frac{(s_A^2 + s_B^2)^2}{s_A^4 + s_B^4} - 2,$$

and thus depends not only on the choice of significance level α and common sample size n, but also on the ratio s_A^2/s_B^2 of the sample estimates of σ_A^2 and σ_B^2. Furthermore, since f can vary from $(n - 1)$ to $2n$, and equals $2n$ only when $s_A^2 = s_B^2$, it is clear that the two procedures may lead to different results when $\sigma_A \simeq \sigma_B$. Consequently, *when samples of equal size ($n_A = n_B = n$) are involved, the procedure for Case 1 of Paragraph 3-3.1.1 should be used even when it cannot be assumed that $\sigma_A \simeq \sigma_B$.* If in fact $\sigma_A = \sigma_B$, then the effective significance level α' will be identically equal to the intended significance level α, and the test will have maximum sensitivity with respect to any real difference between the population means m_A and m_B. If, on the other hand, $\sigma_A \neq \sigma_B$, then the effective significance level α' will differ from the significance level α intended, but only slightly, as shown by curve (A) in Figure 3-9; and the test will tend to have greater sensitivity with respect to any real difference between m_A and m_B than would be the case if the procedure of the present section were used.

In contrast, *when the samples are of unequal size ($n_A \neq n_B$), the procedure of the present section should always be used unless it is known for certain that $\sigma_A = \sigma_B$.* Otherwise, the effective significance level α' may differ considerably from the significance level α intended, even when $\sigma_A \simeq \sigma_B$ as shown by curve (B) in Figure 3-9.

When the smaller sample comes from the more variable population, the effective number of degrees of freedom f to be used with the procedure of the present section is likely to be much smaller than $n_A + n_B - 2$, the degrees of freedom to be used with the procedure of Paragraph 3-3.1.1. Nevertheless, the small advantage of greater sensitivity to real differences between m_A and m_B that the procedure of Paragraph 3-3.1.1 provides when $\sigma_A = \sigma_B$ is rapidly offset, as the inequality of σ_A and σ_B increases, by the much firmer control of the effective significance level by the procedure of the present section, except when f is very small (say < 6).

Finally, it should be remarked that the effective number of degrees of freedom appropriate to the procedure of the present section is given more accurately by

$$f^* = \frac{(v_A + v_B)^2}{\dfrac{v_A^2}{n_A - 1} + \dfrac{v_B^2}{n_B - 1}} = \frac{(n_B\theta + n_A)^2}{\dfrac{n_B^2\theta^2}{n_A - 1} + \dfrac{n_A^2}{n_B^2}}$$

where

$$v_A = \frac{\sigma_A^2}{n_A} \quad \text{and} \quad v_B = \frac{\sigma_B^2}{n_B}$$

are the true variances of $\bar{X}_A$ and $\bar{X}_B$, respectively, and $\theta = \sigma_A^2/\sigma_B^2$.

It easily is shown that f^* never is less than the smaller of $n_A - 1$ and $n_B - 1$, and never exceeds $n_A + n_B - 2$. If we know the values of σ_A^2 and σ_B^2, then we could evaluate f^*; but under these circumstances we should use the procedure of Paragraph 3-3.1.3, not the present approximate procedure. If we do not know the values of σ_A^2 and σ_B^2, but do know their ratio θ, then the exact procedure (Case 3) of Paragraph 3-3.1.3 cannot be applied, but f^* can be evaluated. Under these circumstances, the approximate procedure of the present section should be followed, with f replaced by f^*. When we do not know the values of σ_A^2 and σ_B^2, nor even their ratio θ, then we must rely on the best available sample estimate of f^*; namely, f defined in Step (4) of the present procedure.

3-3.1.3 (Case 3)—Variability in Performance of Each of A and B is Known from Previous Experience, and the Standard Deviations are σ_A and σ_B, respectively.

Data Sample 3-3.1.3—Latent Heat of Fusion of Ice

The observational data are those of Data Sample 3-3.1.1 and, in addition, it now is assumed to be known that $\sigma_A = 0.024$ and $\sigma_B = 0.033$.

[Two-sided Normal Test]

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up $z_{1-\alpha/2}$ in Table A-2.	(2) $z_{.975} = 1.960$
(3) Compute: $\bar{X}_A$ and $\bar{X}_B$, the means of the n_A and n_B measurements from A and B.	(3) $\bar{X}_A = 80.02$ $\sigma_A^2 = 0.000576$ $n_A = 13$ $\bar{X}_B = 79.98$ $\sigma_B^2 = 0.001089$ $n_B = 8$
(4) Compute $u = z_{1-\alpha/2}\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}$	(4) $u = 1.960\sqrt{\dfrac{0.000576}{13} + \dfrac{0.001089}{8}}$ $= 1.960\ (.0134)$ $= 0.026$
(5) If $\|\bar{X}_A - \bar{X}_B\| > u$, decide that A and B differ with regard to their average performance; otherwise, decide that there is no reason to believe that A and B differ in average performance.	(5) $\|\bar{X}_A - \bar{X}_B\| = .04$, which is larger than u. Conclude that methods A and B differ with regard to their averages.
(6) Let m_A, m_B be the true average performances of A and B (unknown of course). It is worth noting that the interval $(\bar{X}_A - \bar{X}_B) \pm u$ is a $100(1-\alpha)\ \%$ confidence interval estimate of $(m_A - m_B)$.	(6) The interval $.04 \pm .026$ i.e., the interval from .014 to .066 is a 95% confidence interval for the true difference between the averages of the methods.

Operating Characteristics of the Test. Figures 3-3 and 3-4 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of n.

If $n_A = n_B = n$, and $(m_A - m_B)$ is the true difference between the two averages, then putting

$$d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$

we can read β, the probability of failing to detect a difference of size $\pm (m_A - m_B)$.

If $n_A = cn_B$, we can put $d = \dfrac{|m_A - m_B|}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$ and, using $n = n_A$, we can read β, the probability of failing to detect a difference of size $\pm (m_A - m_B)$.

Selection of Sample Size. We choose

α, the significance level of the test
β, the probability of failing to detect a difference of size $(m_A - m_B)$.

If we wish $n_A = n_B = n$, we compute

$$d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}}$$

and we may use Table A-8 directly to obtain the required sample size n.

If we wish to have n_A and n_B such that $n_A = cn_B$, then we may compute

$$d = \frac{|m_A - m_B|}{\sigma_A^2 + c\sigma_B^2}$$

and use Table A-8 to obtain $n = n_A$.

3-3.1.4 (Case 4)—The Observations are Paired.

Often, an experiment is, or can be, designed so that the observations are taken in pairs. The two units of a pair are chosen in advance so as to be as nearly alike as possible in all respects other than the characteristic to be measured, and then one member of each pair is assigned at random to treatment A, and the other to treatment B. For instance, the experimenter may wish to compare the effects of two different treatments on a particular type of device, material, or process. The word "treatments" here is to be understood in a broad sense: the two "treatments" may be different operators; different environmental conditions to which a material may be exposed, or merely two different methods of measuring one of its properties; two different laboratories in an interlaboratory test of a particular process of measurement or manufacture. Since the comparison of the two treatments is made *within* pairs, two advantages result from such pairing. First, the effect of extraneous variation is reduced and there is consequent increase in the precision of the comparison, and in its sensitivity to real differences between the treatments with respect to the measured characteristic. Second, the test may be carried out under a wide range of conditions representative of actual use without sacrifice of sensitivity and precision, thereby assuring wider applicability of any conclusions reached.

Data Sample 3-3.1.4—Capacity of Batteries

The data below are measurements of the capacity (in ampere hours) of paired batteries, one from each of two different manufacturers:

A	B	$X_d = X_A - X_B$
146	141	5
141	143	−2
135	139	−4
142	139	3
140	140	0
143	141	2
138	138	0
137	140	−3
142	142	0
136	138	−2

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Compute: $\bar{X}_d$ and s_d for the n differences, X_d. (Each X_d represents an observation on A minus the paired observation on B).	(2) $\bar{X}_d = -0.1$ $s_d = 2.807$
(3) Look up $t_{1-\alpha/2}$ for $n - 1$ degrees of freedom in Table A-4.	(3) $t_{.975}$ (9 d.f.) $= 2.262$
(4) Compute $u = t_{1-\alpha/2} \frac{s_d}{\sqrt{n}}$	(4) $u = 2.262 \left(\frac{2.807}{3.162}\right) = 2.008$
(5) If $\lvert\bar{X}_d\rvert > u$, decide that the averages differ; otherwise, that there is no reason to believe they differ.	(5) $\lvert\bar{X}_d\rvert = 0.1$, which is less than u. Conclude that batteries of the two manufacturers do not differ in average capacity.
(6) *Note:* The interval $\bar{X}_d \pm u$ is a $100\,(1 - \alpha)$ % confidence interval estimate of the average difference (A minus B).	(6) The interval -0.1 ± 2.0, i.e., the interval -2.1 to $+1.9$ is a 95% confidence interval estimate of the average difference in capacity between the batteries of the two manufacturers.

Operating Characteristics of the Test. Figures 3-1 and 3-2 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of n, the number of pairs involved.

Choose:

$\delta = |m_A - m_B|$, the true absolute difference between the averages (unknown, of course)

Some value of σ ($= \sigma_d$), the true standard deviation of a signed difference X_d.
(We may use an estimate from previous data. If OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d = \frac{\delta}{\sigma} .$$

We then can read from the OC curve for a given significance level α and sample size n, a value of of $\beta(\delta)$. The $\beta(\delta)$ read from the curve is $\beta(\delta | \sigma, \alpha, n)$, i.e., $\beta(\delta$, *given* $\sigma, \alpha, n)$—the probability of failing to detect a difference of $\pm (m_A - m_B)$ when it exists, if the given test is carried out with n pairs, at the α-level of significance, *and* the standard deviation of signed differences X_d actually *is* σ.

If we use too large a value for σ, the effect is to underestimate d, and consequently to overestimate $\beta(\delta)$, the probability of not detecting a difference of δ when it exists. Conversely, if we choose too small a value of σ, then we shall overestimate d and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample size n and the significance level α employed, *and* the true value of σ ($= \sigma_d$).

Selection of Number of Pairs n required. If we choose

$\delta = |m_A - m_B|$, the absolute value of the average difference that we desire to detect
α, the significance level of the test
β, the probability of failing to detect a difference of δ

and compute

$$d = \frac{|m_A - m_B|}{\sigma}$$

where σ is the standard deviation of the population of signed differences X_d for the type of pairs concerned, then we may use Table A-8 to obtain a good approximation to the required number of pairs n. If we take $\alpha = .01$, then we must add 4 to the value obtained from the table. If we take $\alpha = .05$, then we must add 2 to the table value. In order to compute d, we must choose a value for σ.

If, when planning the test, we overestimate σ, the consequences are two-fold: first, we overestimate the number of pairs required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if we underestimate σ, we shall underestimate the number of pairs actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened.

Finally, it should be noted, that inasmuch as the test criterion $u = t_{1-\alpha/2} \frac{s_d}{\sqrt{n}}$ does not depend on σ, an error in estimating σ when planning the test will not alter the level of significance, which will be precisely equal to the value of α desired, *provided* that $t_{1-\alpha/2}$ is taken equal to the 100 $(1 - \alpha/2)$ percentile of the t-distribution for $n - 1$ degrees of freedom, where n is the *number of pairs actually employed.*

3-3.2 DOES THE AVERAGE OF PRODUCT A EXCEED THE AVERAGE OF PRODUCT B?

3-3.2.1 (Case 1)—Variability of A and B is Unknown, but can be Assumed to be Equal.

Data Sample 3-3.2.1—Surface Hardness of Steel Plates

A study was made of the effect of two grinding conditions on the surface hardness of steel plates used for intaglio printing. Condition A represents surfaces "as ground" and Condition B represents surfaces after light polishing with emery paper. The observations are hardness indentation numbers.

Condition A	Condition B
187	157
157	152
152	148
164	158
159	161
164	
172	

[One-sided t-test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $t_{1-\alpha}$ for $\nu = n_A + n_B - 2$ degrees of freedom in Table A-4.

(3) Compute: $\bar{X}_A$ and s_A^2, $\bar{X}_B$ and s_B^2, from the n_A and n_B measurements from products A and B, respectively.

(4) Compute

$$s_p = \sqrt{\frac{(n_A - 1)\, s_A^2 + (n_B - 1)\, s_B^2}{n_A + n_B - 2}}$$

(5) Compute

$$u = t_{1-\alpha}\, s_p \sqrt{\frac{n_A + n_B}{n_A\, n_B}}$$

(6) If $(\bar{X}_A - \bar{X}_B) > u$, decide that the average of A exceeds the average of B; otherwise, decide there is no reason to believe that the average of A exceeds the average of B.

(7) Let m_A and m_B be the true averages of A and B. Note that the interval from $\{(\bar{X}_A - \bar{X}_B) - u\}$ to ∞ is a $1 - \alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$.

Example

(1) Let $\alpha = .05$

(2)
$n_A = 7$
$n_B = 5$
$\nu = 10$
$t_{.95}$ for 10 d.f. $= 1.812$

(3)
$\bar{X}_A = 165$
$s_A^2 = 134$
$\bar{X}_B = 155.2$
$s_B^2 = 26.7$

(4)

$$s_p = \sqrt{\frac{6\,(134) + 4\,(26.7)}{10}} = \sqrt{91.08} = 9.544$$

(5)

$$u = (1.812)\,(9.544)\sqrt{\frac{12}{35}} = 17.294\,(.5855) = 10.1$$

(6) $(\bar{X}_A - \bar{X}_B) = 9.8$, which is not larger than u. There is no reason to believe that the average hardness for Condition A exceeds the average hardness for Condition B.

(7) $(\bar{X}_A - \bar{X}_B) - u = 9.8 - 10.1 = -0.3$. The interval from -0.3 to ∞ is a 95% one-sided confidence interval estimate of the true difference between averages.

Operating Characteristics of the Test. Figures 3-5 and 3-6 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of $n = n_A + n_B - 1$.

Choose:

$\delta = (m_A - m_B)$, the true difference between the averages

Some value of σ $(= \sigma_A = \sigma_B)$, the common standard deviation (We may use an estimate from previous data; lacking such an estimate, see Paragraph 2-2.4. If OC curve is consulted after the experiment, we may use the estimate from the experiment).

Compute

$$d^* = \frac{(m_A - m_B)}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A\, n_B}{n_A + n_B}} .$$

We then can read a value of $\beta(\delta)$ from the OC curve for a given significance level and effective sample size n. The $\beta(\delta)$ read from the curve is $\beta(\delta | \sigma, \alpha, n_A, n_B)$ i.e., $\beta(\delta$, given σ, α, n_A, and $n_B)$ the probability of failing to detect a real difference between the two population means of magnitude $\delta = +(m_A - m_B)$ when the test is carried out with samples of sizes n_A and n_B, respectively, at the α-level of significance, *and* the two population standard deviations actually *are* both equal to σ.

If we use too large a value for σ, the effect is to make us underestimate d^*, and consequently to overestimate $\beta(\delta)$. Conversely, if we choose too small a value of σ, then we shall overestimate d^* and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the sample sizes (n_A and n_B) and significance level α actually employed, *and* the true value of σ $(= \sigma_A = \sigma_B)$.

Since the test criterion u does not depend on the value of σ $(= \sigma_A = \sigma_B)$, an error in estimating σ will not alter the significance level of the test, which will be precisely equal to the value of α desired, *provided that* the value of $t_{1-\alpha}$ is taken equal to the 100 $(1 - \alpha)$ percentile of the t-distribution for $n_A + n_B - 2$ degrees of freedom, where n_A and n_B are the sample sizes actually employed, *and* it actually *is* true that $\sigma_A = \sigma_B$.

If $\sigma_A \neq \sigma_B$, then, whatever may be the ratio σ_A/σ_B, the effective significance level α' will not differ seriously from the intended value α, *provided that* $n_A = n_B$, except possibly when both are as small as two. If, on the other hand, unequal sample sizes are used, and $\sigma_A \neq \sigma_B$, then the effective level of significance α' can differ considerably from the intended value α, as shown in Figure 3-9.

Selection of Sample Size n. If we choose

$\delta = (m_A - m_B)$, the value of the average difference that we desire to detect

α, the significance level of the test

β, the probability of failing to detect a difference of size δ

and compute

$$d = \frac{(m_A - m_B)}{\sqrt{2\sigma^2}}, \text{ where } \sigma = \sigma_A = \sigma_B ,$$

then we may use Table A-9 to obtain a good approximation to the required sample size n $(= n_A = n_B)$. If we take $\alpha = .01$, then we must add 2 to the table value. If we take $\alpha = .05$, then we must add 1 to the table value.

In order to compute d, we must choose a value for σ $(= \sigma_A = \sigma_B)$. (See Paragraph 2-2.4 if no other information is available.) If we overestimate σ, the consequences are two-fold: first, we overestimate the sample size n $(= n_A = n_B)$ required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if we underestimate σ, we shall underestimate the sample size actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened. These effects of overestimating or underestimating σ ($= \sigma_A = \sigma_B$) will be similar in magnitude to those considered and illustrated in Paragraph 3-2.2.1 for the case of comparing the mean m of a new material, product, or process, with a standard value m_0.

As explained in the preceding discussion of the Operating Characteristics of the Test, an error in estimating σ ($= \sigma_A = \sigma_B$) will have no effect on the significance level of the test, provided that the value of $t_{1-\alpha}$ is taken equal to the 100 $(1 - \alpha)$ percentile of the t-distribution for $n_A + n_B - 2$ degrees of freedom, where n_A and n_B are the sample sizes actually employed; and if $\sigma_A \neq \sigma_B$, the effect will not be serious provided that the sample sizes *are* taken equal.

3-3.2.2 (Case 2)—Variability of A and B is Unknown, Cannot Be Assumed Equal.

Consider the data of Data Sample 3-3.1.2. Suppose that (from a consideration of the methods, and *not* after looking at the results) the question to be asked was whether the average for Method A *exceeded* the average for Method B.

Procedure*	**Example**
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Compute: $\bar{X}_A$ and s_A^2, $\bar{X}_B$ and s_B^2, from the n_A and n_B measurements from A and B.	(2) $\bar{X}_A = 3166.0$ $s_A^2 = 6328.67$ $n_A = 4$ $\bar{X}_B = 2240.4$ $s_B^2 = 221{,}661.3$ $n_B = 9$
(3) Compute: $V_A = \dfrac{s_A^2}{n_A}$ and $V_B = \dfrac{s_B^2}{n_B}$, the estimated variances of $\bar{X}_A$ and $\bar{X}_B$, respectively.	(3) $V_A = \dfrac{6328.67}{4} = 1582.17$ $V_B = \dfrac{221{,}661.3}{9} = 24629.03$
(4) Compute the "effective number of degrees of freedom" $f = \dfrac{(V_A + V_B)^2}{\dfrac{V_A^2}{n_A + 1} + \dfrac{V_B^2}{n_B + 1}} - 2$	(4) $f = \dfrac{(26211.20)^2}{500652.4 + 60658911.9} - 2$ $= 11.233 - 2$ $= 9.233$
(5) Look up $t_{1-\alpha}$ for f' degrees of freedom in Table A-4, where f' is the integer nearest to f; denote this value by $t'_{1-\alpha}$.	(5) $f' = 9$ $t'_{.95} = 1.833$
(6) Compute $u = t'_{1-\alpha}\sqrt{V_A + V_B}$	(6) $u = 1.833\sqrt{26211.20}$ $= 1.833\ (161.90)$ $= 296.76$

* See footnote on page 3-37.

Procedure*

(7) If $(\bar{X}_A - \bar{X}_B) > u$, decide that the average of A exceeds the average of B; otherwise, decide that there is no reason to believe that the average of A exceeds the average of B.

(8) Let m_A and m_B be the true averages of A and B. Note that the interval from $\{(\bar{X}_A - \bar{X}_B) - u\}$ to ∞ is approximately a one-sided 100 $(1 - \alpha)$ % confidence interval estimate of the true difference $(m_A - m_B)$.

Example

(7) $\bar{X}_A - \bar{X}_B = 925.6$, which is larger than u. Conclude that the average for Method A exceeds the average for Method B.

(8) $(\bar{X}_A - \bar{X}_B) - u = 925.6 - 296.76 = 628.8$. The interval from 628.8 to ∞ is approximately a one-sided 95% confidence interval estimate of the true difference between the averages for the methods.

3-3.2.3 (Case 3)—Variability in Performance of Each of A and B is Known from Previous Experience and the Standard Deviations are σ_A and σ_B, Respectively.

Data Sample 3-3.2.3

The observational data are those of Data Sample 3-3.2.1 on surface hardness of steel plates. In addition, it now is assumed that the variability for the two conditions was known from previous experience to be $\sigma_A = 10.25$ and $\sigma_B = 5.00$.

[One-sided Normal Test]

Procedure

(1) Choose α, the significance level of the test.

(2) Look up $z_{1-\alpha}$ in Table A-2.

(3) Compute: $\bar{X}_A$ and $\bar{X}_B$, the means of the n_A and n_B measurements from A and B.

(4) Compute

$$u = z_{1-\alpha}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

(5) If $(\bar{X}_A - \bar{X}_B) > u$, decide that the average of A exceeds the average of B; otherwise, decide that there is no reason to believe that the average of A exceeds the average of B.

(6) Let m_A and m_B be the true averages of A and B. Note that the interval from $\{(\bar{X}_A - \bar{X}_B) - u\}$ to ∞ is a $1 - \alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$.

Example

(1) Let $\alpha = .05$

(2) $z_{1-\alpha} = 1.645$

(3)
$$\begin{aligned} \bar{X}_A &= 165 \\ \sigma_A^2 &= 105 \\ n_A &= 7 \\ \bar{X}_B &= 155.2 \\ \sigma_B^2 &= 25 \\ n_B &= 5 \end{aligned}$$

(4)
$$\begin{aligned} u &= 1.645\sqrt{15 + 5} \\ &= 1.645\,(4.472) \\ &= 7.4 \end{aligned}$$

(5) $(\bar{X}_A - \bar{X}_B) = 9.8$, which is larger than u. Conclude that the average hardness for Condition A exceeds the average hardness for Condition B.

(6) The interval from 2.4 to ∞ is a 95% one-sided confidence interval estimate of the true difference between averages.

* See footnotes, and also the discussion of the properties and limitations of this type of procedure, in Paragraph 3-3.1.2.

Operating Characteristics of the Test. Figures 3-7 and 3-8 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of n.

If $n_A = n_B = n$ and $(m_A - m_B)$ is the true positive difference between the averages, then putting

$$d = \frac{(m_A - m_B)}{\sqrt{\sigma_A^2 + \sigma_B^2}}$$

we can read β, the probability of failing to detect a difference of size $(m_A - m_B)$.

If $n_A = cn_B$, we can put

$$d = \frac{(m_A - m_B)}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$$

and again read β, the probability of failing to detect a difference of size $(m_A - m_B)$.

Selection of Sample Size. We choose

α, the significance level of the test
β, the probability of failing to detect a difference of size $(m_A - m_B)$.

If we wish $n_A = n_B = n$, we compute

$$d = \frac{(m_A - m_B)}{\sqrt{\sigma_A^2 + \sigma_B^2}}$$

and we may use Table A-9 directly to obtain the required sample size n.

If we wish to have n_A and n_B such that $n_A = cn_B$, then we may compute

$$d = \frac{(m_A - m_B)}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$$

and use Table A-9 to obtain $n = n_A$.

3-3.2.4 (Case 4)—The Observations are Paired.

Often, an experiment is, or can be, designed so that the observations are taken in pairs. The two units of a pair are chosen in advance so as to be as nearly alike as possible in all respects other than the characteristic to be measured, and then one member of each pair is assigned at random to Treatment A, and the other to Treatment B. For a discussion of the advantage of this approach, see Paragraph 3-3.1.4.

Data Sample 3-3.2.4—Molecular Weight of Dextrons

During World War II bacterial polysaccharides (dextrons) were considered and investigated for use as blood plasma extenders. Sixteen samples of hydrolyzed dextrons were supplied by various manufacturers in order to assess two chemical methods for determining the average molecular weight of dextrons.

Method A	Method B	$X_d = X_A - X_B$
62,700	56,400	6,300
29,100	27,500	1,600
44,400	42,200	2,200
47,800	46,800	1,000
36,300	33,300	3,000
40,000	37,100	2,900
43,400	37,300	6,100
35,800	36,200	− 400

Method A	Method B	$X_d = X_A - X_B$
33,900	35,200	−1,300
44,200	38,000	6,200
34,300	32,200	2,100
31,300	27,300	4,000
38,400	36,100	2,300
47,100	43,100	4,000
42,100	38,400	3,700
42,200	39,900	2,300

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$
(2) Compute the $\bar{X}_d$ and s_d for the n differences, X_d. Each X_d represents an observation on A minus the paired observation on B.	(2) $\bar{X}_d = 2875$ $s_d = 2182.2$ $n = 16$
(3) Look up $t_{1-\alpha}$ for $n - 1$ degrees of freedom in Table A-4.	(3) $t_{.95}$ for 15 d.f. $= 1.753$
(4) Compute $u = t_{1-\alpha} \frac{s_d}{\sqrt{n}}$	(4) $u = 1.753 \left(\frac{2182.2}{4}\right)$ $= 1.753\ (545.6)$ $= 956.4$
(5) If $\bar{X}_d > u$, decide that the average of A exceeds that of B; otherwise, there is no reason to believe the average of A exceeds that of B.	(5) $\bar{X}_d = 2875$, which is larger than u. Conclude that the average for Method A exceeds the average for Method B.
(6) Note that the open interval from $\bar{X}_d - u$ to $+\infty$ is a one-sided $100\,(1-\alpha)\,\%$ confidence interval for the true difference $(m_A - m_B)$.	(6) $\bar{X}_d - u = (2875 - 956) = 1919$. The interval from 1919 to $+\infty$ is a one-sided 95% confidence interval for the true difference between the averages of the two methods.

Operating Characteristics of the Test. Figures 3-5 and 3-6 give the operating characteristic (OC) curves of the test for $\alpha = .05$ and $\alpha = .01$, respectively, for various values of n, the number of pairs involved.

Choose:

$\delta = (m_A - m_B)$, the true difference between the averages (unknown, of course)

Some value of $\sigma\ (= \sigma_d)$, the true standard deviation of a signed difference X_d. (We may use an estimate from previous data. If OC curve is consulted after the experiment, we may use the estimate from the experiment.)

Compute

$$d = \frac{\delta}{\sigma}\ .$$

We can then read from the OC curve for a given significance level α and number of pairs n, a value of $\beta(\delta)$. The $\beta(\delta)$ read from the curve is $\beta(\delta | \sigma, \alpha, n)$, i.e., $\beta(\delta$, *given* $\sigma, \alpha, n)$—the probability of failing to detect a difference $(m_A - m_B)$ of magnitude $+\delta$ when the given test is carried out with n pairs, at the α-level of significance, and the population standard deviation of the differences X_d actually *is* σ.

If we use too large a value for σ, the effect is to underestimate d, and consequently to overestimate $\beta(\delta)$, the probability of not detecting a difference $(m_A - m_B)$ of size $+\delta$ when it exists. Conversely, if we choose too small a value of σ, then we shall overestimate d and underestimate $\beta(\delta)$. The true value of $\beta(\delta)$ is determined, of course, by the actual number of pairs n, the significance level α employed, *and* the true value of σ $(= \sigma_d)$.

Selection of Number of Pairs (n). If we choose

$\delta = (m_A - m_B)$, the value of the (positive) average difference that we desire to detect
α, the significance level of the test
β, the probability of failing to detect a difference of $+\delta$

and compute

$$d = \frac{(m_A - m_B)}{\sigma}$$

where σ $(= \sigma_d)$ is the standard deviation of the population of signed differences X_d of the type concerned, then we may use Table A-9 to obtain a good approximation to the required number of pairs n. If we take $\alpha = .01$, then we must add 3 to the table value. If we take $\alpha = .05$, then we must add 2 to the table value. (In order to compute d, we must choose a value for σ.)

If, when planning the test, we overestimate σ, the consequences are two-fold: first, we overestimate the number of pairs required, and thus unnecessarily increase the cost of the test; but, by employing a sample size that is larger than necessary, the actual value of $\beta(\delta)$ will be somewhat less than we intended, which will be all to the good. On the other hand, if we underestimate σ, we shall underestimate the number of pairs actually required, and by using too small a sample size, $\beta(\delta)$ will be somewhat larger than we intended, and our chances of detecting real differences when they exist will be correspondingly lessened.

Finally, it should be noted, that inasmuch as the test criterion $u = t_{1-\alpha} \frac{s_d}{\sqrt{n}}$ does not depend on σ, an error in estimating σ when planning the test will not alter the level of significance, which will be precisely equal to the value of α desired, *provided* that $t_{1-\alpha}$ is taken equal to the 100 $(1 - \alpha)$ percentile of the t-distribution for $n - 1$ degrees of freedom, where n is the *number of pairs actually employed.*

3-4 COMPARING THE AVERAGES OF SEVERAL PRODUCTS

Do the averages of t products 1, 2, . . . , t differ? We shall assume that $n_1 = n_2 = \ldots = n_t = n$. If the n's are in fact not all equal, but differ only slightly, then in the following procedure we may replace n by the harmonic mean of the n's,

$$n_H = t/(1/n_1 + 1/n_2 + \ldots + 1/n_t)$$

and obtain a satisfactory approximation.

Data Sample 3-4—Breaking-strength of Cement Briquettes

The following data relate to breaking-strength of cement briquettes (in pounds per square inch). The question to be answered is: Does the average breaking-strength differ for the different groups?

	Group				
	1	2	3	4	5
	518	508	554	555	536
	560	574	598	567	492
	538	528	579	550	528
	510	534	538	535	572
	544	538	544	540	506
ΣX_i	2670	2682	2813	2747	2634
n_i	5	5	5	5	5
$\bar{X}_i$	534.0	536.4	562.6	549.4	526.8
ΣX^2	1427404	1440924	1585141	1509839	1391364
$\frac{(\Sigma X)^2}{n}$	1425780	1438624.8	1582593.8	1509201.8	1387591.2
$\Sigma X^2 - \frac{(\Sigma X)^2}{n}$	1624	2299.2	2547.2	637.2	3772.8
s^2	406	574.8	636.8	159.3	943.2

Excerpted with permission from *Statistical Exercises, Part II, Analysis of Variance and Associated Techniques*, by N. L. Johnson, Copyright, 1957 Department of Statistics, University College, London.

Procedure	Example
(1) Choose α, the significance level (the risk of concluding that the averages differ, when in fact all averages are the same).	(1) Let $\alpha = .01$
(2) Compute: $s_1^2, s_2^2, \ldots, s_t^2$.	(2) $s_1^2 = 406.0$ $s_2^2 = 574.8$ $s_3^2 = 636.8$ $s_4^2 = 159.3$ $s_5^2 = 943.2$
(3) Compute $s_e^2 = \frac{1}{t}(s_1^2 + s_2^2 + \ldots + s_t^2)$ If the n_i are not all equal, the following formula usually is to be preferred: $s_e^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2 + \ldots + (n_t - 1)\, s_t^2}{(n_1 + n_2 + \ldots + n_t) - t}$	(3) $s_e^2 = \frac{2720.1}{5}$ $= 544.0$ $s_e = 23.32$

Procedure	Example
(4) Look up $q_{1-\alpha}(t, \nu)$ in Table A-10 where $\nu = (n_1 + n_2 + \ldots + n_t) - t$.	(4) $\nu = 25 - 5$ $= 20$ $t = 5$ $q_{.99}(5,20) = 5.29$
(5) Compute $w = \frac{q_{1-\alpha}\, s_e}{\sqrt{n}}$	(5) $w = \frac{5.29\,(23.32)}{\sqrt{5}} = \frac{123.36}{2.236} = 55.2$
(6) If the absolute difference between any two sample means exceeds w, decide that the averages differ; otherwise, decide that there is no reason to believe the averages differ.	(6) The greatest difference between sample means is 562.6 − 526.8 = 35.8, which is less than w. We, therefore, have no reason to believe that the group averages differ.

Note: It is worth noting that we simultaneously can make confidence interval estimates for each of the $\frac{t(t-1)}{2}$ pairs of differences between product averages, with a confidence of $1 - \alpha$ that *all* of the estimates are correct. The confidence intervals are $(\bar{X}_i - \bar{X}_j) \pm w$, where $\bar{X}_i$, $\bar{X}_j$, are sample means of the i^{th} and j^{th} products.

REFERENCES

1. E. S. Pearson and H. O. Hartley, *Biometrika Tables For Statisticians*, Vol. I, (2d edition), pp. 27, 136-7, Cambridge University Press, 1958.
2. W. H. Trickett, B. L. Welch, and G. S. James, "Further Critical Values for the Two-Means Problem," *Biometrika*, Vol. 43, pp. 203-205, 1956.

CHAPTER 4

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO VARIABILITY OF PERFORMANCE

4-1 COMPARING A NEW MATERIAL OR PRODUCT WITH A STANDARD WITH RESPECT TO VARIABILITY OF PERFORMANCE

The variability of a standard material, product, or process, as measured by its standard deviation, is known to be σ_0. We consider the following three problems:

(a) Does the variability of the new product *differ* from that of the standard? See Paragraph 4-1.1.

(b) Does the variability of the new product *exceed* that of the standard? See Paragraph 4-1.2.

(c) Is the variability of the new product *less* than that of the standard? See Paragraph 4-1.3.

It is important to decide which of the three problems is appropriate before taking the observations. If this is not done, and the choice of problem is influenced by the observations, both the significance level of the test (i.e., the probability of an Error of the First Kind) and the operating characteristics of the test may differ considerably from their nominal values.

The tests given are exact when:

(a) the observations for an item, product, or process are taken randomly from a single population of possible observations; and,

(b) within the population, the quality characteristic measured is normally distributed.

4-1.1 DOES THE VARIABILITY OF THE NEW PRODUCT DIFFER FROM THAT OF THE STANDARD?

The variability in the performance of a standard material, product, or process, as measured by its standard deviation, is known to be σ_0. We wish to determine whether a given item *differs* in variability of performance from the standard. We wish, from analysis of the data, to make one of the following decisions:

(a) The variability in performance of the new product *differs* from that of the standard.

(b) There is no reason to believe the variability of the new product is different from that of the standard.

Data Sample 4-1.1—Capacity of Batteries

The standard deviation σ_0 of capacity for batteries of a standard type is known to be 1.66 ampere hours. The following capacities (ampere hours) were recorded for 10 batteries of a new type: 146, 141, 135, 142, 140, 143, 138, 137, 142, 136.

We wish to compare the new type of battery with the standard type with regard to variability of capacity. The question to be answered is: Does the new type *differ* from the standard type with respect to variability of capacity (either a decrease or an increase is of interest)?

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up B_U and B_L both for $n - 1$ degrees of freedom in Table A-20.	(2) $n - 1 = 9$ B_U for 9 d.f. $= 1.746$ B_L for 9 d.f. $= .6657$
(3) Compute s, from the n observations $s = \sqrt{\dfrac{\Sigma X^2 - (\Sigma X)^2/n}{n-1}}$	(3) $s = \sqrt{\dfrac{196108 - 196000}{9}}$ $= \sqrt{\dfrac{108}{9}}$ $= \sqrt{12}$ $= 3.464$
(4) Compute: $s_L = B_L\, s$ $s_U = B_U\, s$	(4) $s_L = (.6657)\,(3.464) = 2.31$ $s_U = (1.746)\,(3.464) = 6.05$
(5) If σ_0 does not lie between s_L and s_U, decide that the variability in performance of the new product *differs* from that of the standard; otherwise, that there is no reason to believe the new product differs from the standard with regard to variability.	(5) Since $\sigma_0 = 1.66$ does not lie between the limits 2.31 to 6.05, conclude that the variability for the new type *does* differ from the variability for the standard type.
(6) It is worth noting that the interval from s_L to s_U is a 100 $(1 - \alpha)$ % confidence interval estimate of σ, the standard deviation of the new product. (See Par. 2-2.3.1).	(6) The interval from 2.31 to 6.05 ampere hours is a 95% confidence interval estimate for the standard deviation of the new type.

Operating Characteristics of the Test. Operating-characteristic (OC) curves for this Neyman-Pearson "unbiased Type A" test of the null hypothesis that $\sigma = \sigma_0$ relative to the alternative that $\sigma \neq \sigma_0$ are not currently available except for two special cases considered in the original Neyman-Pearson memoir.(1) These special cases and more general considerations indicate that the OC curves for this test will not differ greatly, except for the smallest sample sizes, from the OC curves for the corresponding traditional "equal-tail" test (see Figures 6.15 and 6.16 of Bowker and Lieberman(2)). The OC curve for the present test for a given significance level and sample size n will lie above the OC curve of the corresponding "equal-tail" test for $\sigma > \sigma_0$ and below the OC curve for the "equal-tail" test for $\sigma < \sigma_0$. In other words, the chances of failing to detect that σ exceeds σ_0 are somewhat greater with the present test than with the "equal-tail" test, and somewhat less of failing to detect that σ is less than σ_0. The reader is reminded, however, that if there is special interest in determining whether $\sigma > \sigma_0$, or special interest in determining whether $\sigma < \sigma_0$, the problem and procedure of this Paragraph is not at all appropriate, and Paragraph 4-1.2 or 4-1.3 should be consulted.

4-1.2 DOES THE VARIABILITY OF THE NEW PRODUCT EXCEED THAT OF THE STANDARD?

The variability in performance of a standard material, product, or process, as measured by its standard deviation, is known to be σ_0. We wish to determine whether the variability in performance of a new product *exceeds* that of the standard. We wish, from analysis of the data, to make one of the following decisions:

(a) The variability in performance of the new product *exceeds* that of the standard.

(b) There is no reason to believe the variability of the new product exceeds that of the standard.

In terms of Data Sample 4-1.1, let us suppose that—in advance of looking at the data!—the important question is: Does the variability of the new type *exceed* that of the standard?

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up A_α for $n - 1$ degrees of freedom in Table A-21.	(2) $n - 1 = 9$ $A_{.05}$ for 9 d.f. $= .7293$
(3) Compute s, from the n observations.	(3) $s = 3.464$
(4) Compute $s_L = A_\alpha s$	(4) $s_L = .7293\ (3.464)$ $= 2.53$
(5) If s_L exceeds σ_0, decide that the variability of the new product exceeds that of the standard; otherwise, that there is no reason to believe that the new product exceeds the standard with regard to variability.	(5) Since 2.53 exceeds 1.66, conclude that the variability of the new type exceeds that of the standard type.
(6) It is worth noting that the interval above s_L is a $100\ (1 - \alpha)\ \%$ confidence interval estimate of σ, the standard deviation of the new product. (See Par. 2-2.3.2).	(6) The interval from 2.53 to $+\infty$ is a 95% confidence interval estimate of the standard deviation of the new type.

Operating Characteristics of the Test. Figure 4-1 provides operating-characteristic (OC) curves of the test for $\alpha = 0.05$ and various values of n. Let σ_1 denote the true standard deviation of the new product. Then the OC curves of Figure 4-1 show the probability $\beta = \beta\ (\lambda \mid .05, n)$ of failing to conclude that σ_1 exceeds σ_0 when $\sigma_1 = \lambda\sigma_0$ and the test is carried out at the $\alpha = 0.05$ level of significance using a value of s derived from a sample of size n. Similar OC curves for the case of $\alpha = 0.01$ are given in Figure 6.18 of Bowker and Lieberman.(2) OC curves are easily constructed for other values of n — and, if desired, other values of α — by utilizing the fact that if the test is conducted at the α level of significance using a value of s based on a sample of size n, then the probability of failing to conclude that σ_1 exceeds σ_0 when $\sigma_1 = \lambda\sigma_0$ is exactly β for

$$\lambda = \lambda\ (\alpha, \beta, n) = \sqrt{\chi^2_{1-\alpha}\ (n-1)/\chi^2_\beta\ (n-1)}\,,$$

where $\chi^2_P\ (\nu)$ is the P-probability level of χ^2 for ν degrees of freedom, as given in Table A-3. Values of $\rho\ (\alpha, \beta, n_1) = \lambda^2\ (\alpha, \beta, n)$ corresponding to $\alpha = 0.05$ and $\alpha = 0.01$, for $\beta = 0.005$, 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975, 0.99, and 0.995 are given in Tables 8.1 and 8.2 of Eisenhart(3) for $n_1 = n - 1 = 1(1)30(10)100, 120, \infty$.

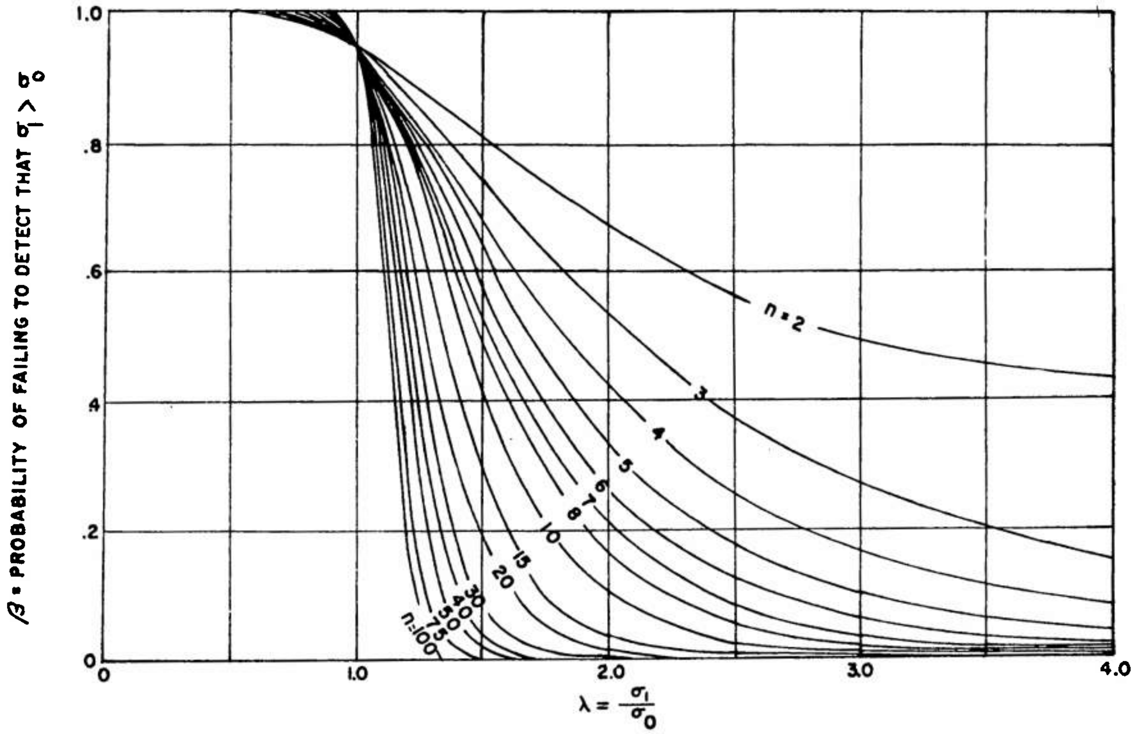

Figure 4-1. Operating characteristics of the one-sided χ^2-test to determine whether the standard deviation σ_1 of a new product exceeds the standard deviation σ_0 of a standard (α = .05).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Selection of Sample Size. If we choose

$$\lambda = \frac{\sigma_1}{\sigma_0}$$

α, the significance level of the test

and, β, the probability of failing to detect that σ_1 exceeds σ_0 when $\sigma_1 = \lambda\sigma_0$

then for $\alpha = 0.05$ we may use the OC curves of Figure 4-1 to determine the necessary sample size n.

Example: Choose

$$\lambda = \frac{\sigma_1}{\sigma_0} = 1.5$$

$$\alpha = 0.05$$

$$\beta = 0.05$$

then from Figure 4-1 it is seen that $n = 30$ is not quite sufficient, and $n = 40$ is more than sufficient. Visual interpolation suggests $n = 35$.

Alternatively, one may compute the necessary sample size from the approximate formula

$$n = n\,(\alpha, \beta, \lambda) = 1 + \frac{1}{2}\left(\frac{z_{1-\alpha} + \lambda \cdot z_{1-\beta}}{\lambda - 1}\right)^2$$

where z_P is the P-probability point of the standard normal variable z, values of which are given in Table A-2 for various values of P. Thus, in the foregoing example we find

$$n = 1 + \frac{1}{2}\left(\frac{1.645 + (1.5)\,(1.645)}{1.5 - 1}\right)^2$$

$$= 1 + \frac{1}{2}\left(\frac{4.1125}{0.5}\right)^2 = 1 + \frac{1}{2}\,(8.225)^2 = 1 + \frac{1}{2}\,(67.65)$$

$$= 34.8$$

which rounds to $n = 35$. Chand[4] has found this formula generally quite satisfactory, and that "even for such a small value as $n = 5$" it "errs on the safe side in the sense that it gives (at least for $\alpha = \beta$) a sample size which will always be sufficient."

Check: For $n = 35$,

$$\lambda\,(.05, .05, 35) = \sqrt{\frac{\chi^2_{.95}\,(34)}{\chi^2_{.05}\,(34)}} = \sqrt{\frac{48.3}{21.4}} = \sqrt{2.26}$$

$$= 1.50$$

Hence $\beta = 0.05$ for $\lambda = 1.50$.

4-1.3 IS THE VARIABILITY OF THE NEW PRODUCT LESS THAN THAT OF THE STANDARD?

The variability in performance of a standard material, product, or process, as measured by its standard deviation, is known to be σ_0. We wish to determine whether the variability in performance of the new product is less than that of the standard. We wish, from analysis of the data, to make one of the following decisions:

(a) The variability in performance of the new product is *less than* that of the standard.

(b) There is no reason to believe the variability in performance of the new product is less than that of the standard.

Data Sample 4-1.3—Cutoff Bias of Tubes

A manufacturer has recorded the cutoff bias of a sample of ten tubes, as follows (volts):

12.1, 12.3, 11.8, 12.0, 12.4, 12.0, 12.1, 11.9, 12.2, 12.2.

The variability of cutoff bias for tubes of a standard type as measured by the standard deviation is $\sigma_0 = 0.208$ volt.

Let us assume with respect to Data Sample 4-1.3 that the important question is: Is the variability of the new type with respect to cutoff bias *less than* that of the standard type?

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up $A_{1-\alpha}$ for $n - 1$ degrees of freedom in Table A-21.	(2) $n - 1 = 9$ $A_{.95}$ for 9 d.f. $= 1.645$
(3) Compute s, from the n observations $s = \sqrt{\dfrac{\Sigma X^2 - (\Sigma X)^2/n}{n - 1}}$	(3) $s = \sqrt{\dfrac{.30}{9}}$ $= \sqrt{.0333}$ $= .1826$

Procedure	Example
(4) Compute $s_U = A_{1-\alpha}\, s$	(4) $s_U = 1.645\ (.1826)$ $= 0.300$
(5) If s_U is less than σ_0, decide that the variability in performance of the new product is less than that of the standard; otherwise, that there is no reason to believe the new product is less variable than the standard.	(5) Since .300 is *not* less than .208, conclude that there is no reason to believe that the new type is less variable than the standard.
(6) It is worth noting that the interval below s_U is a $100\ (1 - \alpha)\ \%$ confidence interval estimate of σ, the standard deviation of the new product. (See Par. 2-2.3.2.)	(6) The interval below 0.300 is a 95% confidence interval estimate of the standard deviation of the new type.

Operating Characteristics of the Test. Figure 4-2 provides operating-characteristic (OC) curves of the test for $\alpha = 0.05$ and various values of n. Let σ_1 denote the true standard deviation of the new product. Then the OC curves of Figure 4-2 show the probability $\beta = \beta\ (\lambda \mid .05, n)$ of failing to conclude that σ_1 is less than σ_0 when $\sigma_1 = \lambda\sigma_0$ and the test is carried out at the $\alpha = 0.05$ level of significance using a value of s derived from a sample of size n. Similar OC curves for the case of $\alpha = 0.01$ are given in Figure 6.20 of Bowker and Lieberman.[2] OC curves are easily constructed

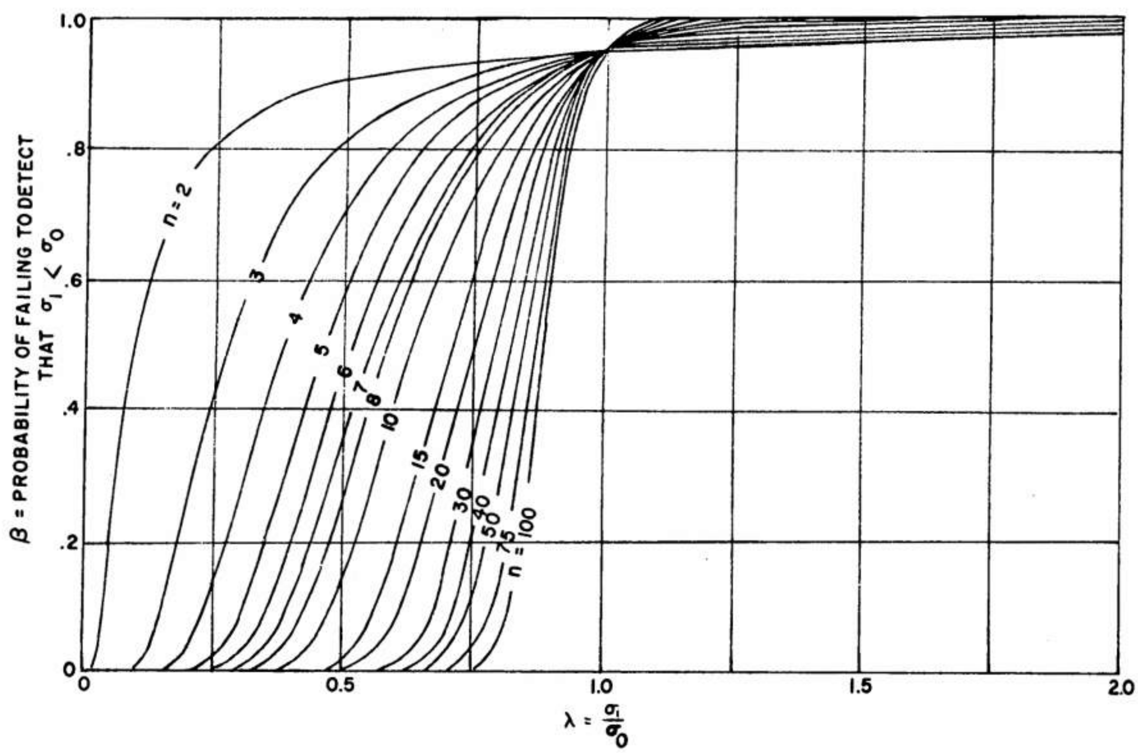

Figure 4-2. Operating characteristics of the one-sided χ^2-test to determine whether the standard deviation σ_1 of a new product is less than the standard deviation σ_0 of a standard ($\alpha = .05$).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

for other values of n — and, if desired, other values of α — by utilizing the fact that if the test is conducted at the α level of significance using a value of s based on a sample of size n, then the probability of failing to conclude that σ_1 is less than σ_0 when $\sigma_1 = \lambda\sigma_0$ is exactly β for

$$\lambda = \lambda\ (\alpha,\ \beta,\ n) = \sqrt{\chi^2_{\alpha}\ (n-1)/\chi^2_{1-\beta}\ (n-1)}\,,$$

where $\chi^2_P\ (\nu)$ is the P-probability level of χ^2 for ν degrees of freedom, as given in Table A-3.

Selection of Sample Size. If we choose

$$\lambda = \frac{\sigma_1}{\sigma_0}$$

α, the significance level of the test

and, β, the probability of failing to detect that σ_1 is less than σ_0 when $\sigma_1 = \lambda\sigma_0$

then for $\alpha = 0.05$ we may use the OC curves of Figure 4-2 to determine the necessary sample size n.

Example: Choose

$$\lambda = \frac{\sigma_1}{\sigma_0} = 0.5$$

$$\alpha = 0.05$$

$$\beta = 0.05$$

then from Figure 4-2 it is seen that $n = 10$ is not quite sufficient, and $n = 15$ is more than sufficient. Visual interpolation suggests $n = 14$.

Alternatively, one may compute the necessary sample size from the approximate formula

$$n = n\ (\alpha,\ \beta,\ \lambda) = 1 + \frac{1}{2}\left(\frac{z_{1-\alpha} + \lambda \cdot z_{1-\beta}}{1-\lambda}\right)^2$$

where z_P is the P-probability point of the standard normal variable z, values of which are given in Table A-2 for various values of P. Thus, in the foregoing example we find

$$n = 1 + \frac{1}{2}\left(\frac{1.645 + (0.5)\ (1.645)}{1 - 0.5}\right)^2$$

$$= 1 + \frac{1}{2}\left(\frac{2.4675}{0.5}\right)^2 = 1 + \frac{1}{2}\ (4.935)^2 = 1 + \frac{1}{2}\ (24.35)$$

$$= 13.18$$

which rounds to $n = 13$.

Check: For $n = 13$,

$$\lambda\ (.05,\ .05,\ 13) = \sqrt{\frac{\chi^2_{.05}\ (12)}{\chi^2_{.95}\ (12)}} = \sqrt{\frac{5.23}{21.03}} = \sqrt{0.2487}$$

$$= 0.499 < 0.50$$

Hence, $\beta = 0.05$ for $\lambda = 0.50$.

4-2 COMPARING TWO MATERIALS OR PRODUCTS WITH RESPECT TO VARIABILITY OF PERFORMANCE

We consider two problems:

(a) Does the variability of product A *differ* from that of product B? (We are not concerned which is larger). See Paragraph 4-2.1.

(b) Does the variability of product A *exceed* that of product B? See Paragraph 4-2.2.

It is important to decide which of these two problems is appropriate before taking the observations. If this is not done, and the choice of problem is influenced by the observations, both the significance level of the test (i.e., the probability of an Error of the First Kind) and the operating characteristics of the test may differ considerably from their nominal values. The tests given are exact when:

(a) the observations for an item, product, or process are taken randomly from a single population of possible observations; and,

(b) within the population, the quality characteristic measured is normally distributed.

In the following, it is assumed the appropriate problem is selected and then n_A, n_B observations are taken from items, processes, or products A and B, respectively.

4-2.1 DOES THE VARIABILITY OF PRODUCT A DIFFER FROM THAT OF PRODUCT B?

We wish to test whether the variability of performance of two materials, products, or processes differ, and we are not particularly concerned which is larger. We wish, from analysis of the data, to make one of the following decisions:

(a) The two products differ with regard to their variability.

(b) There is no reason to believe the two products differ with regard to their variability.

Data Sample 4-2.1—Dive-bombing Methods

The performance of each of two different dive-bombing methods is measured a dozen times with the following results:

Method A	Method B
526	414
406	430
499	419
627	453
585	504
459	459
415	337
460	598
506	425
450	438
624	456
506	385

Let us suppose that, in the case of Data Sample 4-2.1, the question to be answered is: Do the two methods *differ* in variability (it being of interest if either is more variable than the other)?

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up $F_{1-\alpha/2}$ for $(n_A - 1, n_B - 1)$ degrees of freedom, and $F_{1-\alpha/2}$ for $(n_B - 1, n_A - 1)$ degrees of freedom, in Table A-5.	(2) $n_A - 1 = 11$ $n_B - 1 = 11$ $F_{.975}(11, 11) = 3.48$
(3) Compute s_A^2 and s_B^2 from the observations from A and B, respectively.	(3) $s_A^2 = 5545$ $s_B^2 = 4073$
(4) Compute $F = s_A^2/s_B^2$	(4) $F = 5545/4073$ $= 1.36$
(5) If $F > F_{1-\alpha/2}(n_A - 1, n_B - 1)$ or $F < \dfrac{1}{F_{1-\alpha/2}(n_B - 1, n_A - 1)}$ decide that the two products differ with regard to their variability; otherwise, there is no reason to believe that they differ.	(5) $F_{.975}(11, 11) = 3.48$ $\dfrac{1}{F_{.975}(11, 11)} = 0.29$ Since F is not larger than 3.48, and is not smaller than 0.29, there is no reason to believe that the two bombing methods differ in variability.
(6) It is worth noting that the interval between $\dfrac{1}{F_{1-\alpha/2}(n_A - 1, n_B - 1)}\left(\dfrac{s_A^2}{s_B^2}\right)$ and $F_{1-\alpha/2}(n_B - 1, n_A - 1)\left(\dfrac{s_A^2}{s_B^2}\right)$ is a 100 $(1 - \alpha)$ % confidence interval estimate of the ratio σ_A^2/σ_B^2.	(6) The interval between 0.39 (i.e., 0.29 × 1.36) and 4.73 (i.e., 3.48 × 1.36) is a 95% confidence interval estimate of the ratio of the true variances, σ_A^2/σ_B^2.

Operating Characteristics of the Test. Operating-characteristic (OC) curves for this traditional "equal-tail" test of the null hypothesis that $\sigma_A = \sigma_B$ relative to the alternative $\sigma_A \neq \sigma_B$ are given in Figures 7.1 and 7.2 of Bowker and Lieberman[2] for the case of equal sample sizes $n_A = n_B = n$, and significance levels $\alpha = 0.05$ and $\alpha = 0.01$, respectively. These curves may be used to determine the common sample size $n_A = n_B = n$ needed to achieve a preassigned risk β of failing to detect that $\sigma_A/\sigma_B = \lambda$ when the test is carried out at the $\alpha = 0.05$ or $\alpha = 0.01$ level of significance. The reader is reminded, however, that if there is special interest in determining whether $\sigma_A > \sigma_B$, the problem and procedure of this Paragraph is not at all appropriate, and Paragraph 4-2.2 should be consulted.

4-2.2 DOES THE VARIABILITY OF PRODUCT A EXCEED THAT OF PRODUCT B?

We wish to test whether the variability in performance of product A exceeds that of product B. We wish, as a result of analysis of the data, to make one of the following decisions:

(a) The variability of product A exceeds that of product B.

(b) There is no reason to believe that the variability of product A exceeds the variability of product B.

In terms of Data Sample 4-2.1, let us suppose that—in advance of looking at the data!—the important question is: Does the variability of Method A *exceed* that of Method B?

Procedure	Example
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .05$
(2) Look up $F_{1-\alpha}$ for $n_A - 1$, $n_B - 1$ degrees of freedom, in Table A-5.	(2) $n_A - 1 = 11$ $n_B - 1 = 11$ $F_{.95}\,(11, 11) = 2.82$
(3) Compute s_A^2, s_B^2, the sample variances of the observations from A and B, respectively.	(3) $s_A^2 = 5545$ $s_B^2 = 4073$
(4) Compute $F = s_A^2/s_B^2$.	(4) $F = 1.36$
(5) If $F > F_{1-\alpha}$, decide that the variability of product A exceeds that of B; otherwise, there is no reason to believe that the variability of A is greater than that of B.	(5) Since 1.36 is not larger than 2.82, there is no reason to believe that the variability of Method A is greater than the variability of Method B.
(6) Note that the interval above $\frac{1}{F_{1-\alpha}\ (n_A - 1, n_B - 1)} \left(\frac{s_A^2}{s_B^2}\right)$ is a 100 $(1 - \alpha)$ % confidence interval estimate of σ_A^2/σ_B^2.	(6) $\frac{1}{F_{.95}\,(11, 11)} = 0.35$ The interval above 0.48 (i.e., 0.35×1.36) is a 95% confidence interval estimate of the ratio of the true variances, σ_A^2/σ_B^2.

Operating Characteristics of the Test. Figures 4-3, 4-4, and 4-5 provide operating-characteristic (OC) curves of the test for $\alpha = 0.05$ and various combinations of n_A and n_B. Let σ_A and σ_B denote the true standard deviations of the products A and B, respectively. These OC curves show the probability $\beta = \beta\,(\lambda\,|\,.05, n)$ of failing to conclude that σ_A exceeds σ_B when $\sigma_A = \lambda\sigma_B$ with $\lambda > 1$ and the test is carried out at the $\alpha = 0.05$ level of significance using the values of s_A and s_B derived from samples of size n_A and n_B, respectively. Similar OC curves for the case of $\alpha = 0.01$ and $n_A = n_B$ are given in Figure 7.4 of Bowker and Lieberman.[2] OC curves are easily constructed for other values of n_A and n_B — and, if desired, other values of α — by utilizing the fact that if the test is conducted at the α level of significance using values of s_A and s_B based on samples of size n_A and n_B, respectively, then the probability of failing to conclude that σ_A exceeds σ_B when $\sigma_A = \lambda\sigma_B$ is exactly β for

$$\lambda = \lambda\,(\alpha, \beta, n_A, n_B) = \sqrt{\frac{F_{1-\alpha}\,(n_A - 1, n_B - 1)}{F_\beta\,(n_A - 1, n_B - 1)}}$$

$$= \sqrt{F_{1-\alpha}\,(n_A - 1, n_B - 1) \cdot F_{1-\beta}\,(n_B - 1, n_A - 1)}$$

where $F_P\,(n_1, n_2)$ is the P-probability level of F for n_1 and n_2 degrees of freedom, as given in Table A-5. Values of $\phi\,(\alpha, \beta, n_1, n_2) = \lambda^2\,(\alpha, \beta, n_A, n_B)$ corresponding to $\alpha = 0.05$ and $\alpha = 0.01$, for $\beta = 0.005, 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975, 0.99$, and 0.995 are given in Tables 8.3 and 8.4 of Eisenhart[3] for all combinations of values of $n_1 = n_A - 1$ and $n_2 = n_B - 1$ derivable from the sequence 1(1)30(10)100, 120, ∞.

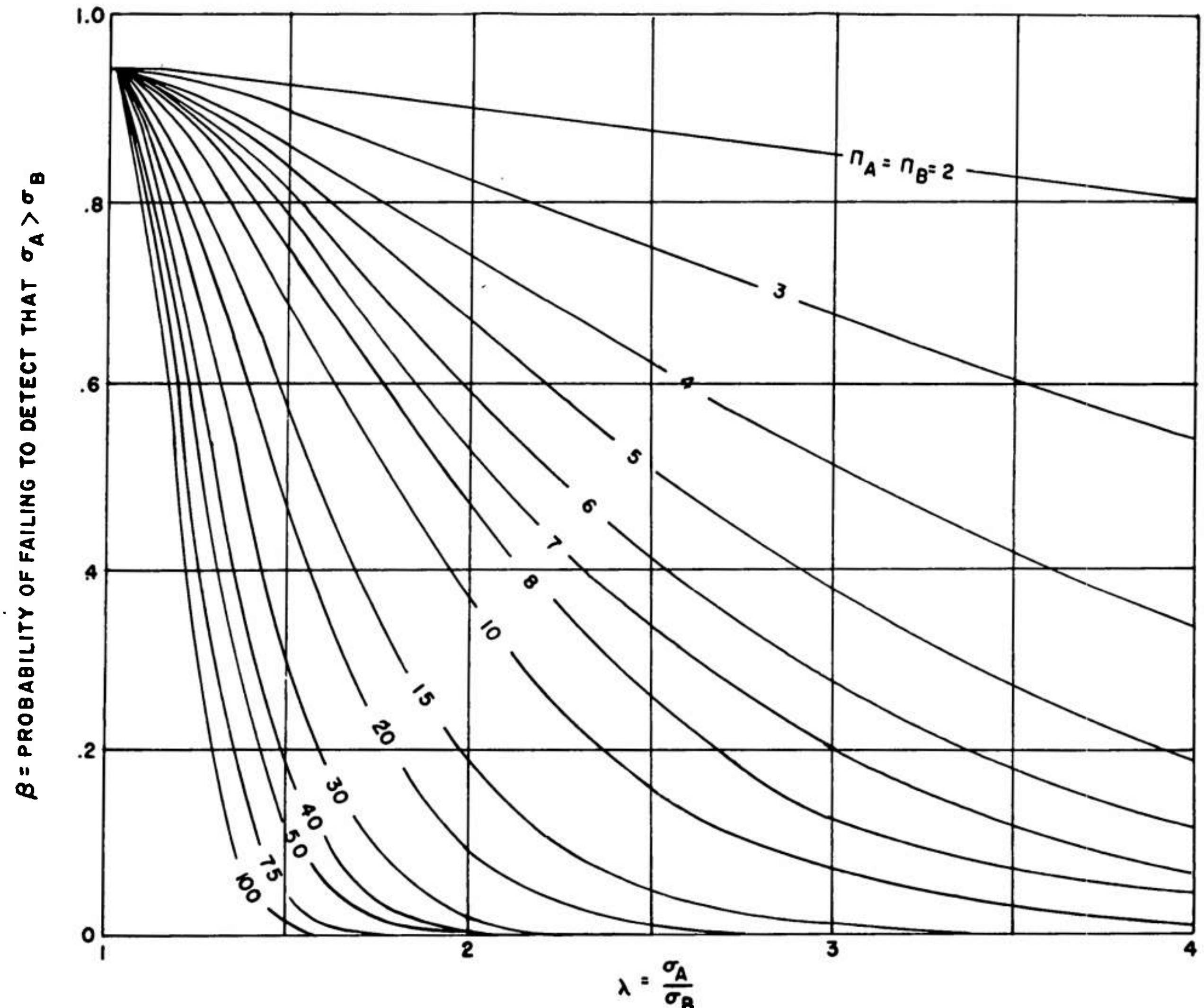

Figure 4-3. Operating characteristics of the one-sided F-test to determine whether the standard deviation σ_A of product A exceeds the standard deviation σ_B of product B ($\alpha = .05$; $n_A = n_B$).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Selection of Sample Size. If we choose

$n_A = n_B = n$

$\lambda = \frac{\sigma_A}{\sigma_B}$

α, the significance level of the test

β, the probability of failing to detect that σ_A exceeds σ_B when $\sigma_A = \lambda\sigma_B$

then for $\alpha = 0.05$, we may use the OC curve of Figure 4-3 to determine the necessary common sample size n.

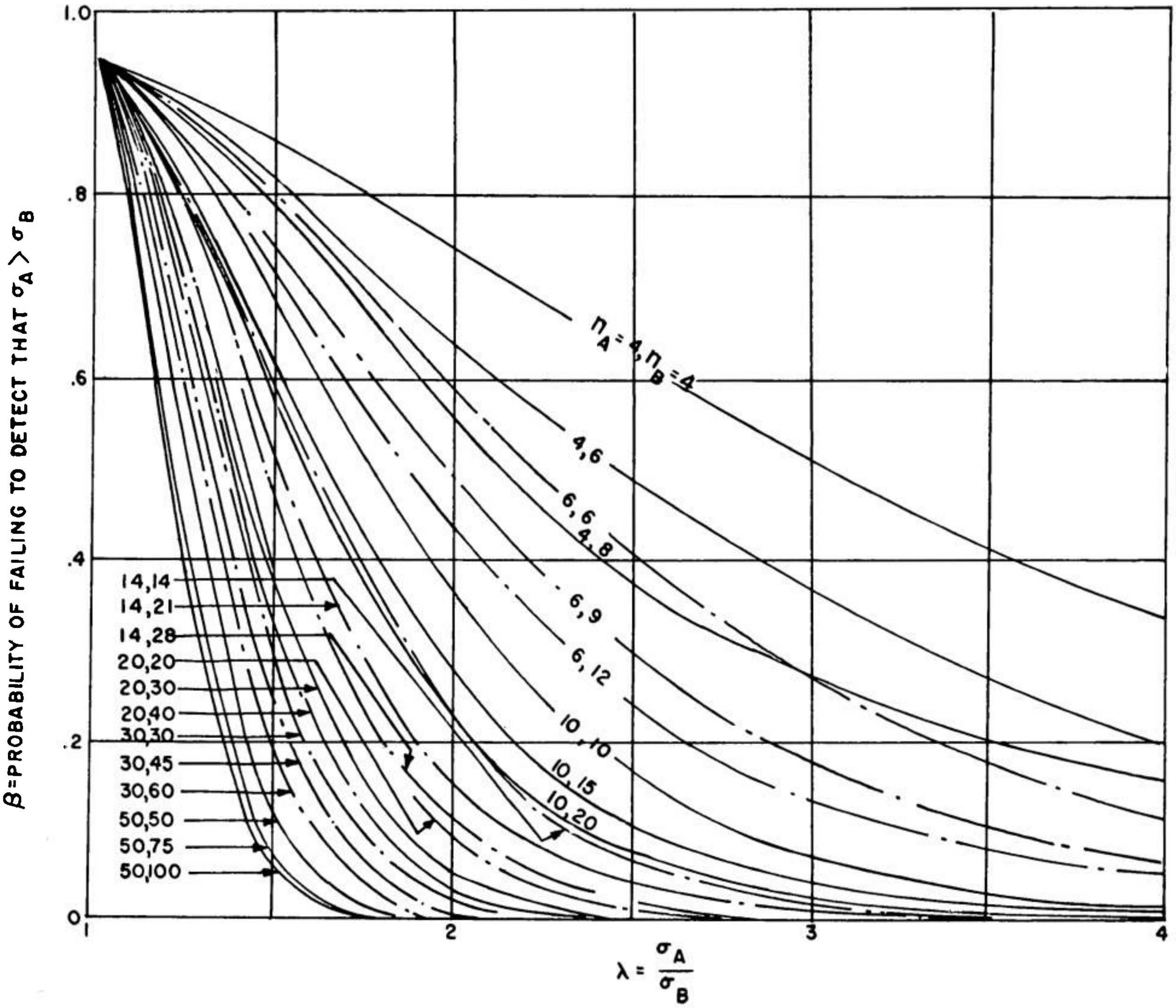

Figure 4-4. Operating characteristics of the one-sided F-test to determine whether the standard deviation σ_A of product A exceeds the standard deviation σ_B of product B ($\alpha = .05$; $n_A = n_B$, $3n_A = 2n_B$, $2n_A = n_B$).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Example: Choose

$\lambda = \frac{\sigma_A}{\sigma_B} = 1.5$

$\alpha = 0.05$

$\beta = 0.05$

then from Figure 4-3 it is seen that $n = 50$ is too small and $n = 75$ a bit too large. Visual interpolation suggests $n = 70$.

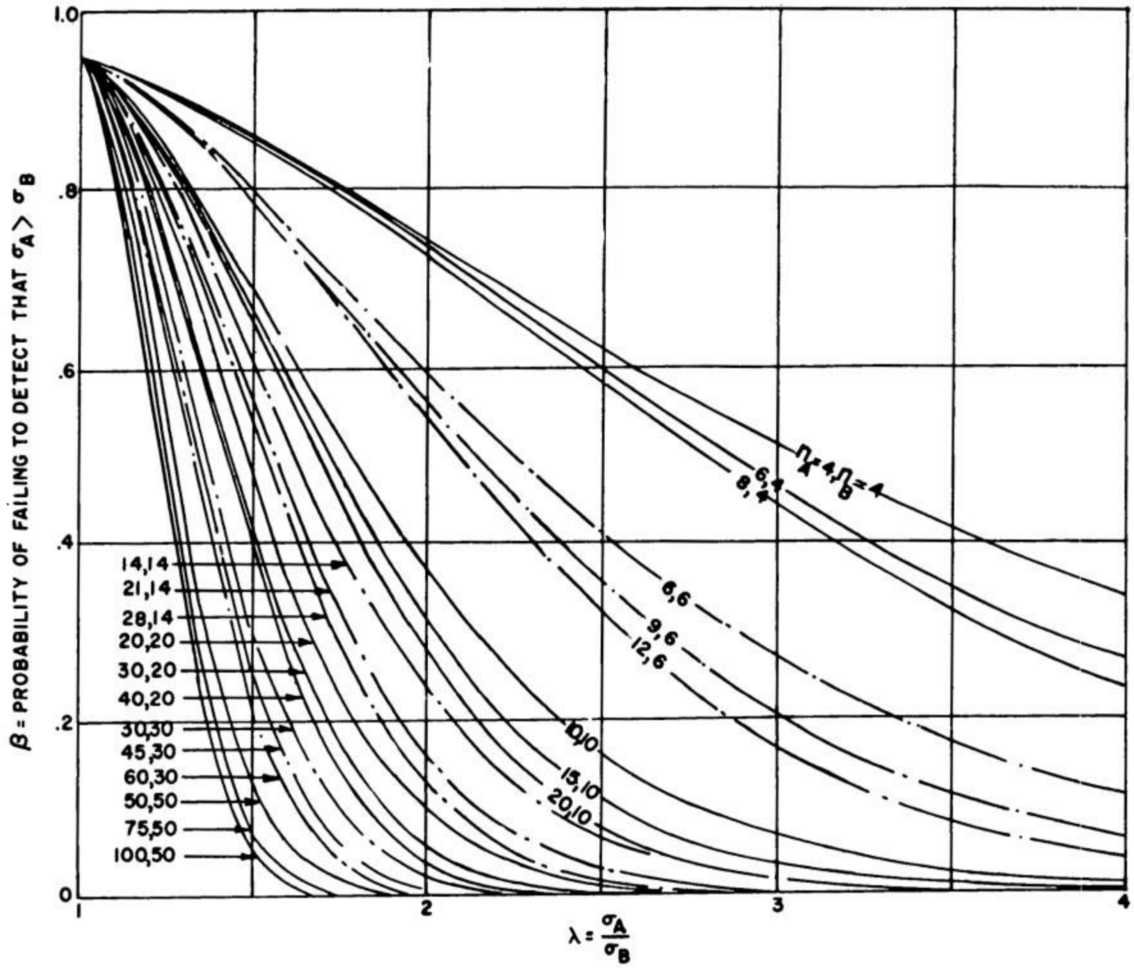

Figure 4-5. Operating characteristics of the one-sided F-test to determine whether the standard deviation σ_A of product A exceeds the standard deviation σ_B of product B ($\alpha = .05$; $n_A = n_B$, $2n_A = 3n_B$, $n_A = 2n_B$).

Adapted with permission from *Annals of Mathematical Statistics*, Vol. 17, No. 2, June 1946, from article entitled "Operating Characteristics for the Common Statistical Tests of Significance" by C. D. Ferris, F. E. Grubbs, and C. L. Weaver.

Alternatively, for $n_A = n_B = n$ one may compute the necessary sample size from the approximate formula

$$n = n(\alpha, \beta, \lambda) = 2 + \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log_e \lambda}\right)^2$$

where z_P is the P-probability point of the standard normal variable z, values of which are given in

Table A-2 for various values of P. Thus in the foregoing example we find

$$n = 2 + \left(\frac{1.645 + 1.645}{.4055}\right)^2$$
$$= 2 + \left(\frac{3.290}{.4055}\right)^2 = 2 + (8.11)^2 = 2 + 65.8$$
$$= 68.$$

If instead of choosing $n_A = n_B$ we choose $3n_A = 2n_B$ or $2n_A = n_B$, then for $\alpha = 0.05$ we may use the OC curves of Figure 4-4 to determine the necessary combination of sample sizes n_A and n_B. Similarly, Figure 4-5 may be used if it is desired to have $2n_A = 3n_B$ or $n_A = 2n_B$. Alternatively, one may evaluate the harmonic mean h of $n_A - 2$ and $n_B - 2$ from the approximate formula

$$h = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log_e \lambda}\right)^2$$

and then determine the integer values of n_A and n_B (satisfying any additional requirements, e.g., $n_A = 2n_B$) that most closely satisfy the equation

$$\frac{1}{h} = \frac{1}{2}\left\{\frac{1}{n_A - 2} + \frac{1}{n_B - 2}\right\}.$$

REFERENCES

1. J. Neyman and E. S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses, I. Unbiased Critical Regions of Type A and Type A_1," *Statistical Research Memoirs*, Vol. 1, Department of Statistics, University College, University of London, 1936.
2. A. H. Bowker and G. J. Lieberman, *Engineering Statistics*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1959.
3. Churchill Eisenhart, "Planning and Interpreting Experiments for Comparing Two Standard Deviations," Chapter 8 of *Selected Techniques of Statistical Analysis* (Edited by Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis), McGraw-Hill Book Co., Inc., New York, N. Y., 1947.
4. Uttam Chand, "On the Derivation and Accuracy of Certain Formulas for Sample Sizes and Operating Characteristics of Nonsequential Sampling Procedures," *Journal of Research of the National Bureau of Standards*, Vol. 47, No. 6, pp. 491-501, December 1951.

CHAPTER 5

CHARACTERIZING LINEAR RELATIONSHIPS BETWEEN TWO VARIABLES

5-1 INTRODUCTION

In many situations it is desirable to know something about the relationships between two characteristics of a material, product, or process. In some cases, it may be known from theoretical considerations that two properties are functionally related, and the problem is to find out more about the structure of this relationship. In other cases, there is interest in investigating whether there exists a degree of association between two properties which could be used to advantage. For example, in specifying methods of test for a material, there may be two tests available, both of which reflect performance, but one of which is cheaper, simpler, or quicker to run. If a high degree of association exists between the two tests, we might wish to run regularly only the simpler test.

In this chapter, we deal only with linear relationships. Curvilinear relationships are discussed in Chapter 6 (see Paragraph 6-5). It is worth noting that many nonlinear relationships may be expressed in linear form by a suitable transformation (change of variable). For example, if the relationship is of the form $Y = aX^b$, then $\log Y = \log a + b \log X$. Putting $Y_T = \log Y$, $b_0 = \log a$, $b_1 = b$, $X_T = \log X$, we have the linear expression $Y_T = b_0 + b_1 X_T$ in terms of the new (transformed) variables X_T and Y_T. A number of common linearizing transformations are summarized in Table 5-4 and are discussed in Paragraph 5-4.4.

5-2 PLOTTING THE DATA

Where only two characteristics are involved, the natural first step in handling the experimental results is to plot the points on graph paper. Conventionally, the *independent variable X* is plotted on the horizontal scale, and the *dependent variable Y* is plotted on the vertical scale.

There is no substitute for a plot of the data to give some idea of the general spread and shape of the results. A pictorial indication of the probable form and sharpness of the relationship, if any, is indispensable and sometimes may save needless computing. When investigating a structural relationship, the plotted data will show whether a hypothetical linear relationship is borne out; if not, we must consider whether there is any theoretical basis for fitting a curve of higher degree. When looking for an empirical association of two characteristics, a glance at the plot will reveal whether such association is likely or whether there is only a patternless scatter of points.

In some cases, a plot will reveal unsuspected difficulties in the experimental setup which must be ironed out before fitting any kind of relationship. An example of this occurred in measuring the time required for a drop of dye to travel between marked distances along a water channel. The channel was marked with distance markers spaced at equal distances, and an observer recorded the time at which the dye passed each marker. The device used for recording time consisted of two clocks hooked up so that when one was stopped, the other started: Clock 1 recorded the times for Distance Markers 1, 3, 5, etc.; and Clock 2 recorded times for the even-numbered distance markers. When the elapsed times were plotted, they looked somewhat as shown in Figure 5-1. It is obvious that there was a systematic time difference between odd and even markers (presumably a lag in the circuit connecting the two clocks). One could easily have fitted a straight line to the odd-numbered distances and a different line to the even-numbered distances, with approximately constant difference between the two lines. The effect was so consistent, however, that the experimenter quite properly decided to find a better means of recording travel times before fitting any line at all.

If no obvious difficulties are revealed by the plot, and the relationship appears to be linear, then a line $Y = b_0 + b_1X$ ordinarily should be fitted to the data, according to the procedures given in this Chapter. Fitting by eye usually is inadequate for the following reasons:

(a) No two people would fit exactly the same line, and, therefore, the procedure is not objective;

(b) We always need some measure of how well the line does fit the data, and of the uncertainties inherent in the fitted line as a representation of the true underlying relationship—and these can be obtained only when a formal, well-defined mathematical procedure of fitting is employed.

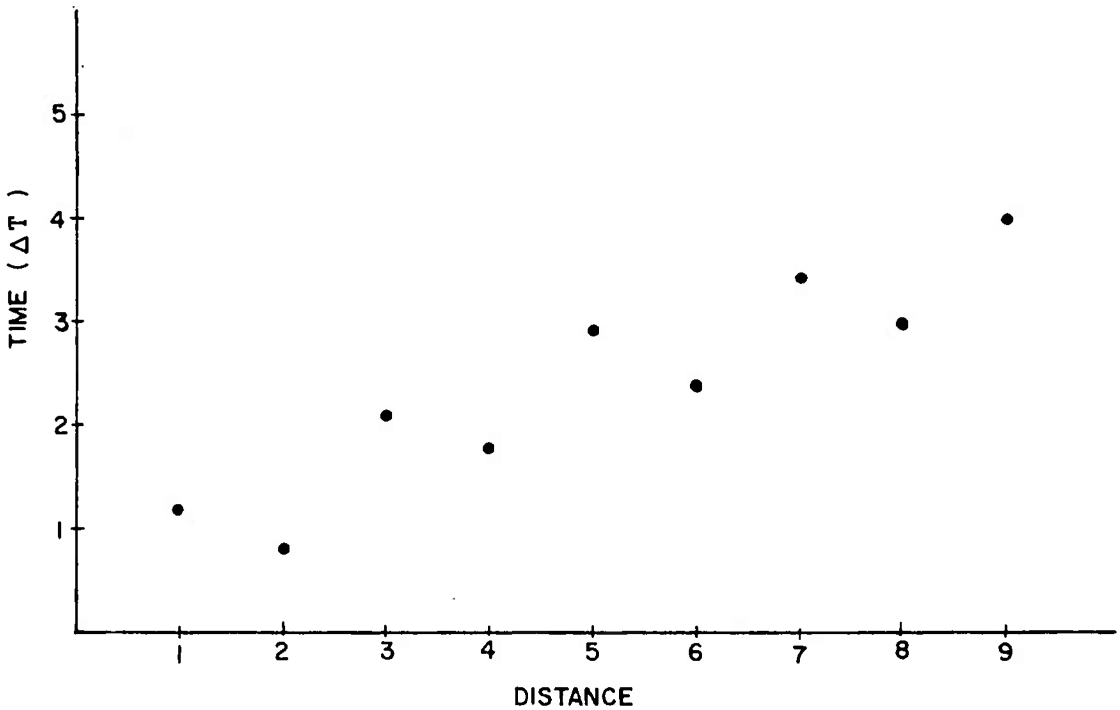

Figure 5-1. Time required for a drop of dye to travel between distance markers.

5-3 TWO IMPORTANT SYSTEMS OF LINEAR RELATIONSHIPS

Before giving the detailed procedure for fitting a straight line, we discuss different physical situations which can be described by a linear relationship between two variables. The methods of description and prediction may be different, depending upon the underlying system. In general, we recognize two different and important systems which we call *Statistical* and *Functional*. It is not possible to decide which is the appropriate system from looking at the data. The distinction must be made before fitting the line—indeed, before taking the measurements.

5-3.1 FUNCTIONAL RELATIONSHIPS

In the case of a Functional Relationship, there exists an exact mathematical formula (y as a function of x) relating the two variables, and the only reason that the observations do not fit this equation exactly is because of disturbances or errors of measurement in the observed values of one or both variables. We discuss two cases of this type:

FI—Errors of measurement affect only one variable (Y). (See Fig. 5-2).

FII—Both variables (X and Y) are subject to errors of measurement. (See Fig. 5-3).

Common situations that may be described by Functional Relationships include calibration lines, comparisons of analytical procedures, and relationships in which time is the X variable.

For instance, we may regard Figure 5-2 as portraying the calibration of a straight-faced spring balance in terms of a series of weights whose masses are accurately known. By Hooke's Law, the extension of the spring, and hence the position y of the scale pointer, should be determined exactly by the mass x upon the pan through a linear functional relationship* $y = \beta_0 + \beta_1 x$. In practice, however, if a weight of mass x_1 is placed upon the pan repeatedly and the position of the pointer is read in each instance, it usually is found that the readings Y_1 are not identical, due to variations in the performance of the spring and to reading errors. Thus, corresponding to the mass x_1 there is a distribution of pointer readings Y_1; corresponding to mass x_2, a distribution of pointer readings Y_2; and so forth—as indicated in Figure 5-2. It is customary to assume that these distributions are normal (or, at least symmetrical and all of the same form) and that the mean of the distribution of Y_i's coincides with the *true value* $y_i = \beta_0 + \beta_1 x_i$.

If, instead of calibrating the spring balance in terms of a series of accurately known weights, we were to calibrate it in terms of another spring balance by recording the corresponding pointer positions when a series of weights are placed first on the pan of one balance and then on the pan of the other, the resulting readings (X and Y) would be related by a linear structural relationship FII, as shown in Figure 5-3, inasmuch as both X and Y are affected by errors of measurement. In this case, corresponding to the repeated weighings of a single weight w_1 (whose true mass need not be known), there is a joint distribution of the pointer readings (X_1 and Y_1) on the two balances, represented by the little transparent *mountain* centered over the *true point* (x_1, y_1) in Figure 5-3; similarly at points (x_2, y_2) and (x_3, y_3), corresponding to repeated weighings of other weights w_2 and w_3, respectively. Finally, it should be noticed that this FII model is more general than the FI model in that it does *not* require linearity of response of each instrument to the independent variable w, but merely that the response curves

* *Note on Notation for Functional Relationships:*
We have used x and y to denote the true or accurately known values of the variables, and X and Y to denote their values measured with error. In the FI Relationship, the independent variable is always without error, and therefore in our *discussions* of the FI case and in the paragraph headings we always use x. In the Worksheet, and Procedures and Examples for the FI case, however, we use X and Y because of the computational similarity to other cases discussed in this Chapter (i.e., the computations for the Statistical Relationships).
In the FII case, both variables are subject to error, and clearly we use X and Y everywhere for the observed values.

of the two instruments be linearly related, that is, that $X = a + b \cdot f(w)$ and $Y = c + d \cdot f(w)$, where $f(w)$ may be linear, quadratic, exponential, logarithmic, or whatever.

Table 5-1 provides a concise characterization of FI and FII relationships. Detailed problems and procedures with numerical examples for FI relationships are given in Paragraphs 5-4.1 and 5-4.2, and for FII relationships in Paragraph 5-4.3.

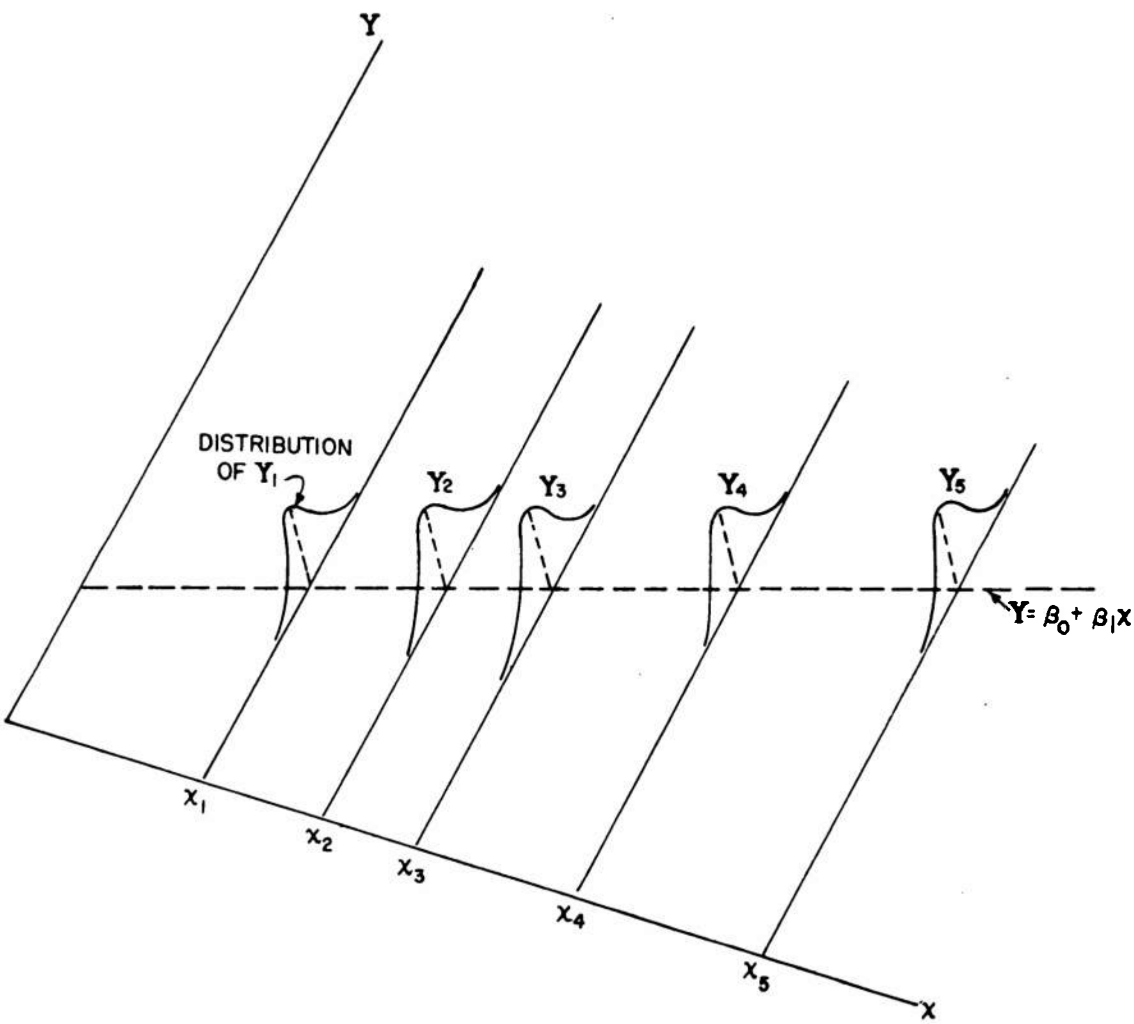

Figure 5-2. Linear functional relationship of Type FI (only Y affected by measurement errors).

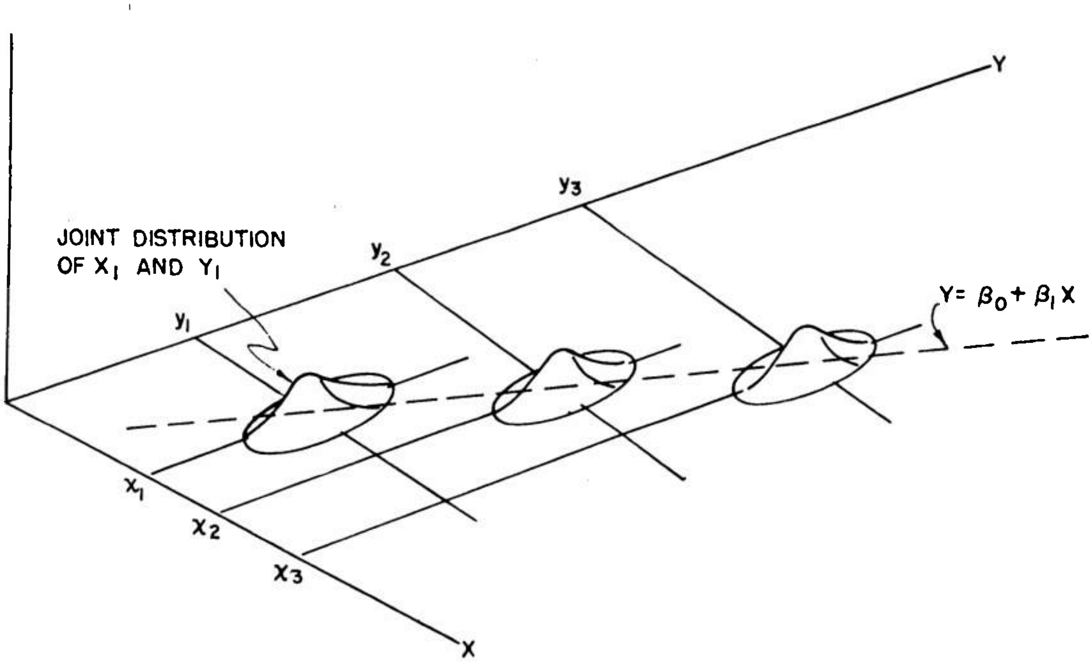

Figure 5-3. Linear functional relationship of Type FII (both X and Y affected by measurement errors).

5-3.2 STATISTICAL RELATIONSHIPS

In the case of a Statistical Relationship, there is no exact mathematical relationship between X and Y; there is only a statistical association between the two variables as characteristics of individual items from some particular population. If this statistical association is of bivariate normal type as shown in Figure 5-4, then the *average* value of the Y's associated with a particular value of X, say $\bar{Y}_X$, is found to depend linearly on X, i.e., $\bar{Y}_X = \beta_0 + \beta_1 X$; similarly, the *average* value of the X's associated with a particular value of Y, say $\bar{X}_Y$, depends linearly on Y (Fig. 5-4) i.e., $\bar{X}_Y = \beta_0' + \beta_1' Y$; but—and this is important!—the two lines are *not* the same, i.e., $\beta_1' \neq \frac{1}{\beta_1}$ and $\beta_0' \neq -\frac{\beta_0}{\beta_1}$.*

* Strictly, we should write

$$m_{Y \cdot X} = \beta_0 + \beta_1 X ,$$

and

$$m_{X \cdot Y} = \beta_0' + \beta_1' Y$$

to conform to our notation of using m to signify a population mean. But this more exact notation tends to conceal the parallelism of the curve-fitting processes in the FI and SI situations. Consequently, to preserve appearances here and in the sequel, we use $\bar{Y}_X$ in place of $m_{Y \cdot X}$ and $\bar{X}_Y$ in place of $m_{X \cdot Y}$—and it should be remembered that these signify *population means*.

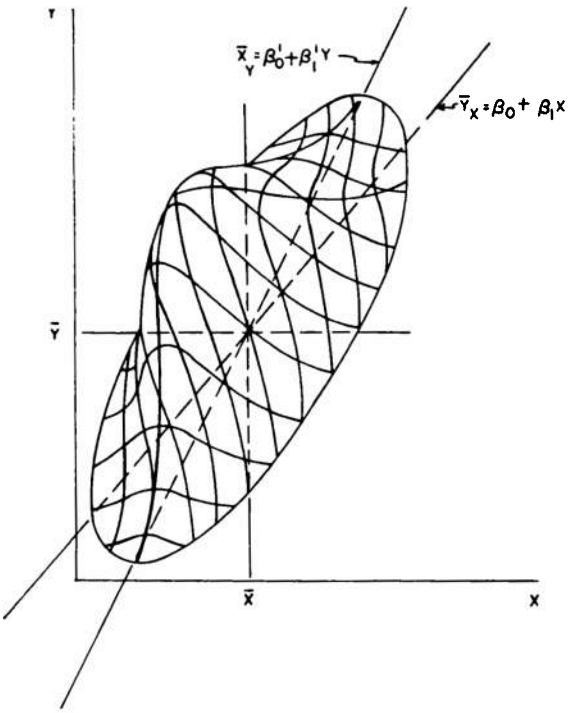

Figure 5-4. A normal bivariate frequency surface.

If a random sample of items is drawn from the population, and the two characteristics X and Y are measured on each item, then typically it is found that errors of measurement are negligible in comparison with the variation of each characteristic over the individual items. This general case is designated SI. A special case (involving preselection or restriction of the range of one of the variables) is denoted by SII.

SI Relationships. In this case, a random sample of items is drawn from some definite population (material, product, process, or people), and two characteristics are measured on each item.

A classic example of this type is the relationship between height and weight of men. Any observant person knows that weight tends to vary with height, but also that individuals of the same height may vary widely in weight. It is obvious that the errors made in measuring height or weight are very small compared to this inherent variation between individuals. We surely would not expect to predict the exact weight of one individual from his height, but we might expect to be able to estimate the average weight of all individuals of a given height.

The height-weight example is given as one which is universally familiar. Such examples also exist in the physical and engineering sciences, particularly in cases involving the interrelation of two test methods. In many cases there may be two tests that, strictly speaking, measure two basically different properties of a material, product, or process, but these properties are statistically related to each other in some complicated way and both are related to some performance characteristic of particular interest, one usually more directly than the other. Their interrelationship may be obscured by inherent variations among sample units (due to varying density, for example). We would be very interested in knowing whether the relationship between the two is sufficient to enable us to predict with reasonable accuracy, from a value given by one test, the average value to be expected for the other—particularly if one test is considerably simpler or cheaper than the other.

The choice of which variable to call X and which variable to call Y is arbitrary—actually there are two regression lines. If a statistical association is found, ordinarily the variable which is easier to measure is called X. Note well that this is the only case of linear relationship in which it may be appropriate to fit two different lines, one for predicting Y from X and a different one for predicting X from Y, *and* the only case in which the sample correlation coefficient r is meaningful as an estimate of the degree of association of X and Y *in the population* as measured by the population coefficient of correlation $\rho = \sqrt{\beta_1 \beta_1'}$. The six sets of contour ellipses shown in Figure 5-5 indicate the manner in which the location, shape, and orientation of the normal bivariate distribution varies with changes of the population means (m_X and m_Y) and standard deviations (σ_X and σ_Y) of X and Y and their coefficient of correlation in the population (ρ_{XY}).

If $\rho = \pm 1$, all the points lie on a line and $Y = \beta_0 + \beta_1 X$ and $X = \beta_0' + \beta_1' Y$ coincide. If $\rho = +1$, the slope is positive, and if $\rho = -1$, the slope is negative. If $\rho = 0$, then X and Y are said to be uncorrelated.

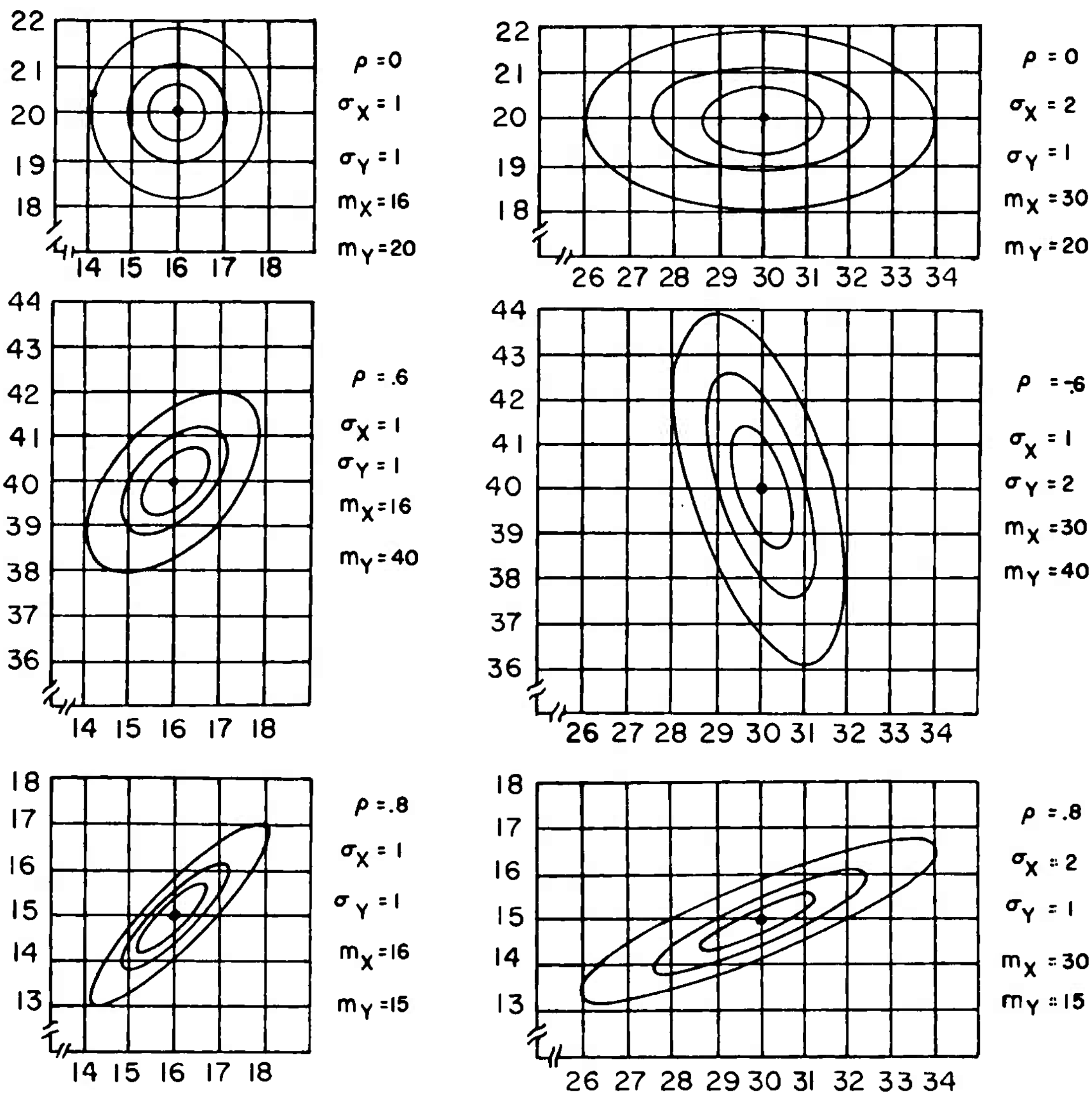

Figure 5-5. Contour ellipses for normal bivariate distributions having different values of the five parameters m_X, m_Y, σ_X, σ_Y, ρ_{XY}.

Adapted with permission from *Statistical Inference* by Helen M. Walker and Joseph Lev, copyright, 1953, Holt, Rinehart and Winston, Inc., New York, N. Y.

SII Relationships. The general case described above (SI) is the most familiar example of a statistical relationship, but we also need to consider a common case of Statistical Relationship (SII) that must be treated a bit differently. In SII, one of the two variables, although a random variable in the population, is sampled only within a limited range (or at selected preassigned values). In the height-weight example, suppose that the group of men included only those whose heights were between 5′4″ and 5′8″. We now are able to fit a line predicting weight from height, but are unable to determine the correct line for predicting height from weight. A correlation coefficient computed from such data is not a measure of the true correlation among height and weight in the (unrestricted) population.

The restriction of the range of X, when it is considered as the independent variable, does not spoil the estimates of $\bar{Y}_X$ when we fit the line $\bar{Y}_X = b_0 + b_1X$. The restriction of the range of the dependent variable (i.e., of Y in fitting the foregoing line, or of X in fitting the line $\bar{X}_Y = b_0' + b_1'Y$), however, gives a seriously distorted estimate of the true relationship. This is evident from Figure 5-6, in which the contour ellipses of the top diagram serve to represent the bivariate distribution of X and Y in the unrestricted population, and the "true" regression lines of $\bar{Y}_X$ on X and $\bar{X}_Y$ on Y are indicated. The central diagram portrays the situation when consideration is restricted to items in the population for which $a < X < b$. It is clear that for any particular X in this interval, the distribution and hence the mean $\bar{Y}_X$ of the corresponding Y's is the same as in the unrestricted case (top diagram). Consequently, a line of the form $\bar{Y}_X = b_0 + b_1X$ fitted to data involving either a random or selected set of values of X between $X = a$ and $X = b$, but with *no* selection or restrictions on the corresponding Y's, will furnish an unbiased estimate of the *true* regression line $\bar{Y}_X = \beta_0 + \beta_1X$ in the population at large. In contrast, if consideration is restricted to items for which $c < Y < d$, as indicated in the bottom diagram, then it is clear that the mean value, say $\bar{Y}'_X$, of the (restricted) Y's associated with any particular value of $X > m_X$ will be less than the corresponding mean value $\bar{Y}_X$ in the population as a whole. Likewise, if $X < m_X$, then the mean $\bar{Y}'_X$ of the corresponding (restricted) Y's will be greater than $\bar{Y}_X$ in the population as a whole. Consequently, a line of the form $\bar{Y}'_X = b_0 + b_1X$ fitted to data involving selection or restriction of Y's will *not* furnish an unbiased estimate of the true regression line $\bar{Y}_X = \beta_0 + \beta_1X$ in the population as a whole, and the distortion may be serious. In other words, introducing a restriction with regard to X does not bias inferences with regard to Y, when Y is considered as the dependent variable, but restricting Y will distort the dependence of $\bar{Y}_X$ on X so that the relationship observed will not be representative of the true underlying relationship in the population as a whole. Obviously, there is an equivalent statement in which the roles of X and Y are reversed. For further discussion and illustration of this point, and of the corresponding distortion of the sample correlation coefficient r as a measure of the true coefficient of correlation ρ in the populations, when *either* X or Y is restricted, see Eisenhart[1] and Ezekiel.[2]

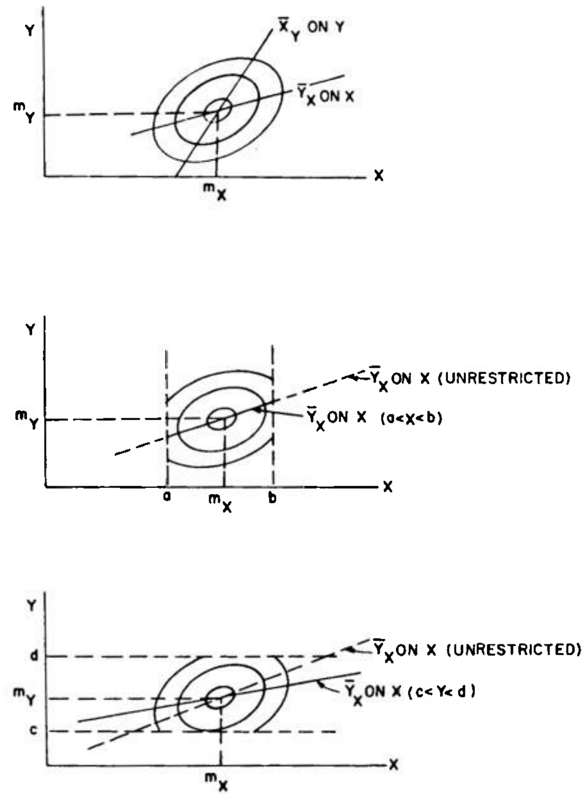

Figure 5-6. Diagram showing effect of restrictions of X or Y on the regression of Y on X.

As an engineering example of SII, consider a study of watches to investigate whether there was a relationship between the cost of a stop watch and its temperature coefficient. It was suggested that a correlation coefficient be computed. This was not possible because the watches had not been selected at random from the total watch production, but a deliberate effort had been made to obtain a fixed number of low-priced, medium-priced, and high-priced stop watches.

In any given case, consider carefully whether one is measuring samples as they come (and thereby accepting the values of both properties that come with the sample) which is an SI Relationship, or whether one selects samples which

TABLE 5-1. SUMMARY OF FOUR CASES OF LINEAR RELATIONSHIPS

	Functional (F)		Statistical (S)	
	FI	FII	SI	SII
Distinctive Features and Example	x and y are linearly related by a mathematical formula, $y = \beta_0 + \beta_1 x$, or $x = \beta_0' + \beta_1' y$, which is not observed exactly because of disturbances or errors in one or both variables. Example: Determination of elastic constant of a spring which obeys Hooke's law. x = accurately-known weight applied, Y = measured value of corresponding elongation y.		X = Height Y = Weight Both measured on a random sample of individuals. X is *not* selected but "comes with" sample unit.	X = Height (preselected values) Y = Weight of individuals of preselected height X is measured beforehand; only *selected* values of X are used at which to measure Y.
Errors of Measurement	Measurement error affects Y only.	X and Y both subject to error.	Ordinarily negligible compared to variation among individuals.	Same as in SI.
Form of Line Fitted	$Y = b_0 + b_1 x$	See Paragraph 5-4.3.	$\bar{Y}_X = b_0 + b_1 X$ $\bar{X}_Y = b_0' + b_1' Y$	$\bar{Y}_X = b_0 + b_1 X$ only.
Procedure for Fitting	See Paragraphs 5-4.1, 5-4.2, and basic worksheet.	Procedure depends on what assumptions can be made. See Paragraph 5-4.3.	See Paragraph 5-5.1 and basic worksheet.	See Paragraph 5-5.2 and basic worksheet.
Correlation Coefficient	Not applicable	Not applicable	Sample estimate is $r = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$ See Paragraph 5-5.1.5.	Correlation may exist in the population, but r computed from *such* an experiment would provide a distorted estimate of the correlation.

are known to have a limited range of values of X (which is an SII Relationship).

Table 5-1 gives a brief summary characterization of SI and SII Relationships. Detailed problems and procedures with numerical examples are given for SI relationships in Paragraph 5-5.1 and for SII relationships in Paragraph 5-5.2.

BASIC WORKSHEET FOR ALL TYPES OF LINEAR RELATIONSHIPS

X denotes ______	Y denotes ______
ΣX = ______	ΣY = ______
$\bar{X}$ = ______	$\bar{Y}$ = ______

Number of points: n = ______

Step (1) ΣXY = ______

(2) $(\Sigma X)(\Sigma Y)/n$ = ______

(3) S_{xy} = Step (1) − Step (2)

(4) ΣX^2 = ______	(7) ΣY^2 = ______
(5) $(\Sigma X)^2/n$ = ______	(8) $(\Sigma Y)^2/n$ = ______
(6) S_{xx} = Step (4) − Step (5)	(9) S_{yy} = Step (7) − Step (8)
(10) $b_1 = \frac{S_{xy}}{S_{xx}}$ = Step (3) ÷ Step (6)	(14) $\frac{(S_{xy})^2}{S_{xx}}$ = ______
(11) $\bar{Y}$ = ______	(15) $(n-2)\, s_Y^2$ = Step (9) − Step (14)
(12) $b_1\bar{X}$ = ______	(16) s_Y^2 = Step (15) ÷ $(n-2)$
(13) $b_0 = \bar{Y} - b_1\bar{X}$ = Step (11) − Step (12)	s_Y = ______

Equation of the line:

$Y = b_0 + b_1 X$

______

s_{b_1} = ______

s_{b_0} = ______

Estimated variance of the slope:

$s_{b_1}^2 = \frac{s_Y^2}{S_{xx}}$ = Step (16) ÷ Step (6)

Estimated variance of intercept:

$s_{b_0}^2 = s_Y^2 \left\{\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right\}$ = ______

Note: The following are algebraically identical:

$S_{xx} = \Sigma(X - \bar{X})^2$; $S_{yy} = \Sigma(Y - \bar{Y})^2$; $S_{xy} = \Sigma(X - \bar{X})(Y - \bar{Y})$.

Ordinarily, in hand computation, it is preferable to compute as shown in the steps above. Carry all decimal places obtainable—i.e., if data are recorded to two decimal places, carry four places in Steps (1) through (9) in order to avoid losing significant figures in subtraction.

5-4 PROBLEMS AND PROCEDURES FOR FUNCTIONAL RELATIONSHIPS

5-4.1 FI RELATIONSHIPS (General Case)

There is an underlying mathematical (functional) relationship between the two variables, of the form $y = \beta_0 + \beta_1 x$. The variable x can be measured relatively accurately. Measurements Y of the value of y corresponding to a given x follow a normal distribution with mean $\beta_0 + \beta_1 x$ and variance $\sigma^2_{Y \cdot x}$ which is independent of the value of x. Furthermore, we shall assume that the deviations or *errors* of a series of observed Y's, corresponding to the same or different x's, all are mutually independent. See Paragraph 5-3.1 and Table 5-1.

The general case is discussed here, and the special case where it is known that $\beta_0 = 0$ (i.e., a line known to pass through the origin) is discussed in Paragraph 5-4.2. The procedure discussed here also will be valid if in fact $\beta_0 = 0$ even though this fact is not known beforehand. However, when it is known that $\beta_0 = 0$, the procedures of Paragraph 5-4.2 should be followed because they are simpler and somewhat more efficient.

It will be noted that SII, Paragraph 5-5.2, is handled computationally in exactly the same manner as FI, but both the underlying assumptions and the interpretation of the end results are different.

Data Sample 5-4.1—Young's Modulus vs. Temperature for Sapphire Rods

Observed values (Y) of Young's modulus (y) for sapphire rods measured at different temperatures (x) are given in the following table. There is assumed to be a linear functional relationship between the two variables x and y. (For the purpose of computation, the observed Y values were coded by subtracting 4000 from each. To express the line in terms of the original units, add 4000 to the computed intercept; the slope will not be affected.) The observed data are plotted in Figure 5-7.

x = Temperature °C	Y = Young's Modulus	Coded Y = Young's Modulus minus 4000
30	4642	642
100	4612	612
200	4565	565
300	4513	513
400	4476	476
500	4433	433
600	4389	389
700	4347	347
800	4303	303
900	4251	251
1000	4201	201
1100	4140	140
1200	4100	100
1300	4073	73
1400	4024	24
1500	3999	−1

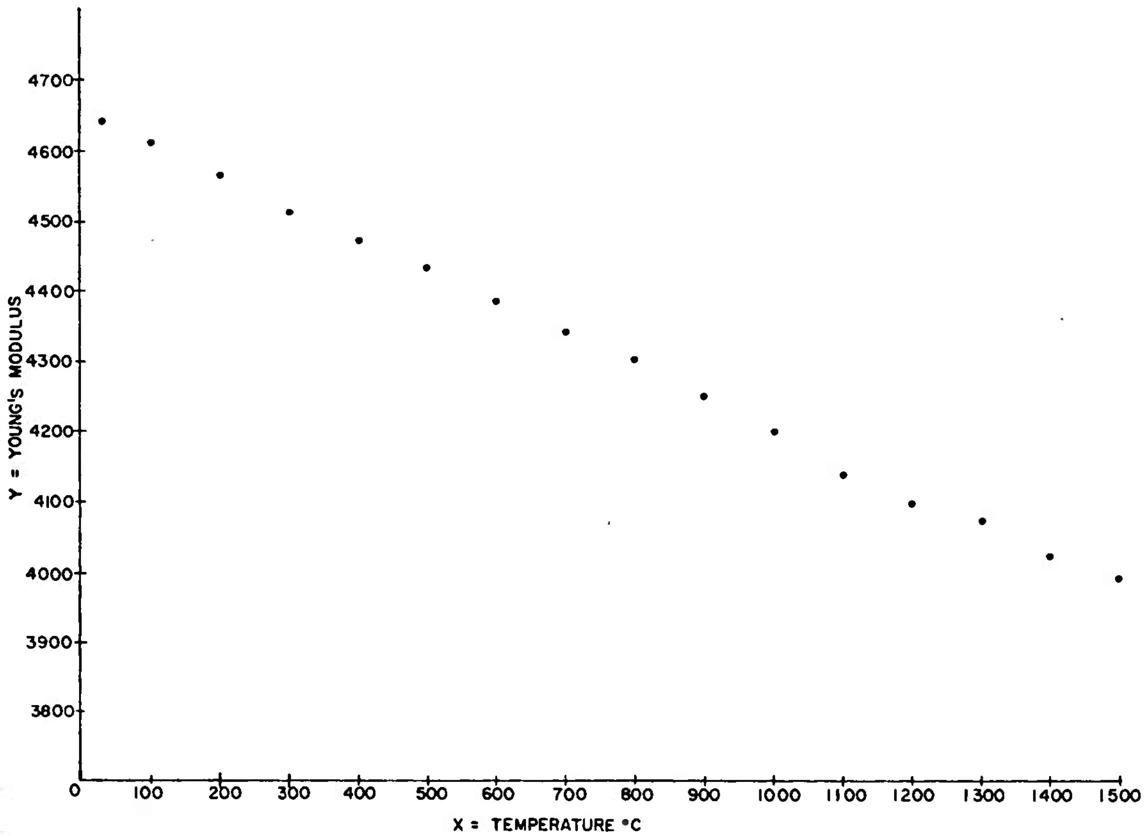

Figure 5-7. Young's modulus of sapphire rods as a function of temperature—an FI relationship.

5-4.1.1 What is the Best Line to be Used for Estimating y From Given Values of x?

CAUTION: Extrapolation, i.e., use of the line for prediction outside the range of data from which the line was computed, may lead to highly erroneous conclusions.

Procedure

Using Worksheet (See Worksheet 5-4.1), compute the line $Y = b_0 + b_1 x$. This is an estimate of the true equation $y = \beta_0 + \beta_1 x$. The method of fitting a line given here is a particular application of the general method of least squares. From Data Sample 5-4.1, the equation of the fitted line (in original units) is:

$$Y = 4654.9846 - 0.44985482\,x.$$

The equation in original units is obtained by adding 4000 to the computed intercept b_0. Since the Y's were coded by subtracting a constant, the computed slope b_1 was not affected. In Figure 5-8, the line is drawn and confidence limits for the line (computed as described in Paragraph 5-4.1.2.1) also are shown.

WORKSHEET 5-4.1
EXAMPLE OF FI RELATIONSHIP
YOUNG'S MODULUS AS FUNCTION OF TEMPERATURE

X denotes Temperature, °C — Y denotes Young's Modulus − 4000

$\Sigma X =$ 12030 — $\Sigma Y =$ 5068

$\bar{X} =$ 751.875 — $\bar{Y} =$ 316.75

Number of points: $n =$ 16

(1) ΣXY = 2,300,860

(2) $(\Sigma X)(\Sigma Y)/n$ = 3,810,502.5

(3) S_{xy} = −1,509,642.5

(4)	ΣX^2 = 12,400,900		(7)	ΣY^2 = 2,285,614	
(5)	$(\Sigma X)^2/n$ = 9,045,056.25		(8)	$(\Sigma Y)^2/n$ = 1,605,289.	
(6)	S_{xx} = 3,355,843.75		(9)	S_{yy} = 680,325.	

(10) $b_1 = \dfrac{S_{xy}}{S_{xx}}$ = −.449,854,82

(11) $\bar{Y}$ = 316.75

(12) $b_1\bar{X}$ = −338.2346

(13) $b_0 = \bar{Y} - b_1\bar{X}$ = 654.9846

b_0 (in original units) = 4654.9846

(14) $\dfrac{(S_{xy})^2}{S_{xx}}$ = 679,119.9614

(15) $(n-2)\, s_Y^2$ = 1,205.0386

(16) s_Y^2 = 86.074 1857

s_Y = 9.277617

Equation of the line:
(in original units)

$Y = b_0 + b_1 X$

4654.9846 − .449,854,82 x

s_{b_1} = .005 064

s_{b_0} = 4.458 638

Estimated variance of the slope:

$s_{b_1}^2 = \dfrac{s_Y^2}{S_{xx}}$ = .000 025 649 045

Estimated variance of intercept:

$s_{b_0}^2 = s_Y^2 \left\{\dfrac{1}{n} + \dfrac{\bar{X}^2}{S_{xx}}\right\}$ = 19.879 452

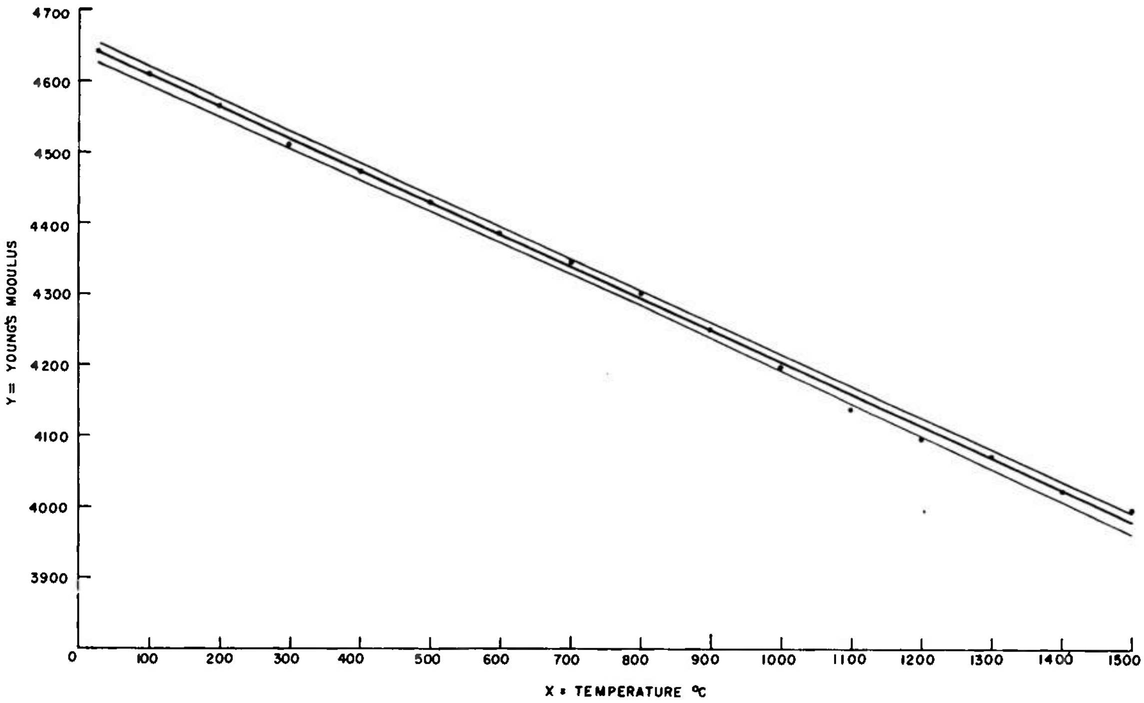

Figure 5-8. Young's modulus of sapphire rods as a function of temperature—showing computed regression line and confidence interval for the line.

Using the Regression Equation for Prediction. The fitted regression equation may be used for two kinds of predictions:

(a) To estimate the *true* value of y associated with a particular value of x, e.g., given $x = x'$ to estimate the value of $y' = \beta_0 + \beta_1 x'$; or,

(b) To predict a single new observed value Y corresponding to a particular value of x, e.g., given $x = x'$ to predict the value of a single measurement of y'.

Which prediction should be made? In some cases, it is sufficient to say that the *true* value of y (for given x) lies in a certain interval, and in other cases we may need to know how large (or how small) an individual observed Y value is likely to be associated with a particular value of x. The question of what to predict is similar to the question of what to specify (e.g., whether to specify average tensile strength or to specify minimum tensile strength) and can be answered only with respect to a particular situation. The difference is that here we are concerned with relationships between two variables and therefore must always talk about the value of y, or Y, for fixed x.

The predicted y' or Y' value is obtained by substituting the chosen value (x') of x in the fitted equation. For a particular value of x, either type of prediction ((a) or (b)) gives the same numerical answer for y' or Y'. The uncertainty associated with the prediction, however, does depend on whether we are estimating the *true* value of y', or predicting the value Y' of an individual measurement of y'. If the experiment could be repeated many times, each time obtaining n pairs of (x, Y) values, consider the range of Y values which would be obtained for a given x. Surely the individual Y values in all the sets will spread over a larger range than will the collection consisting of the average Y's (one from each set).

To estimate the *true* value of y associated with the value x', use the equation

$$y'_c = b_0 + b_1 x'.$$

The variance of y'_c as an estimate of the *true* value $y' = \beta_0 + \beta_1 x'$ is

$$\text{Var } y'_c = s^2_{Y \cdot x} \left[\frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}} \right]$$

This variance is the variance of estimate of a point on the fitted line.

For example, using the equation relating Young's modulus to temperature, we predict a value for y at $x = 1200$:

$$y'_c = 4654.9846 - .44985482\ (1200)$$
$$y'_c = 4115.16$$
$$\text{Var } y'_c = 86.074 \left[.0625 + \frac{(1200 - 751.875)^2}{3{,}355{,}843.75} \right]$$
$$= 86.074\ (.0625 + .0598)$$
$$= 86.074\ (.1223)$$
$$\text{Var } y'_c = 10.53\ .$$

To predict a single observed value of Y corresponding to a given value (x') of x, use the same equation

$$Y'_c = b_0 + b_1 x'.$$

The variance of Y'_c as an estimate of a single new (additional, future) measurement of y' is

$$\text{Var } Y'_c = s^2_{Y \cdot x} \left[1 + \frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}} \right]$$

The equation for our example is

$$Y = 4654.9846 - .44985482\, x.$$

To predict the value of a single determination of Young's modulus at $x = 750$, substitute in this equation and obtain:

$$Y'_c = 4654.9846 - .44985482\ (750)$$
$$= 4317.59$$

$$\text{Var } Y'_c = s^2_Y \left[1 + \frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}} \right]$$
$$= 86.074 \left[1 + .0625 + \frac{(750 - 751.875)^2}{3{,}355{,}843.75} \right]$$
$$= 86.074\ (1.0625)$$
$$= 91.45\ .$$

5-4.1.2 What are the Confidence Interval Estimates for: the Line as a Whole; a Point on the Line; a Future Value of Y Corresponding to a Given Value of x?

Once we have fitted the line, we want to make predictions from it, and we want to know how good our predictions are. Often, these predictions will be given in the form of an interval together with a confidence coefficient associated with the interval—i.e., confidence interval estimates. Several kinds of confidence interval estimates may be made:

(a) A confidence band for the line as a whole.

(b) A confidence interval for a point on the line—i.e., a confidence interval for y' (the *true* value of y and the *mean* value of Y) corresponding to a single value of $x = x'$.

If the fitted line is, say, a calibration line which will be used over and over again, we will want to make the interval estimate described in (a). In other cases, the line as such may not be so important. The line may have been fitted only to investigate or check the structure of the relationship, and the interest of the experimenter may be centered at one or two values of the variables.

Another kind of interval estimate sometimes is required:

(c) A single observed value (Y') of Y corresponding to a new value of $x = x'$.

These three kinds of confidence interval statements have somewhat different interpretations. The confidence interval for (b) is interpreted as follows:

Suppose that we repeated our experiment a large number of times. *Each time*, we obtain n pairs of values (x_i, Y_i), fit the line, and compute a confidence interval estimate for $y' = \beta_0 + \beta_1 x'$, the value of y corresponding to the particular value $x = x'$. Such interval estimates of y' are expected to be correct (i.e., include the *true* value of y') a proportion $(1 - \alpha)$ of the time. If we were to make an interval estimate of y'' corresponding to another value of $x = x''$, these interval estimates also would be expected to include y'' the same proportion $(1 - \alpha)$ of the time. However, taken together, these intervals do not constitute a joint confidence statement about y' and y'' which would be expected to be correct exactly a proportion $(1 - \alpha)$ of the time; nor is the effective level of confidence $(1 - \alpha)^2$, because the two statements are not independent but are correlated in a manner intimately dependent on the values x' and x'' for which the predictions are to be made.

The confidence band for the whole line (a) implies the same sort of repetition of the experiment except that our confidence statements are not now limited to one x at a time, but we can talk about any number of x values simultaneously—about the whole line. Our confidence statement applies to the line as a whole, and therefore the confidence intervals for y corresponding to all the chosen x values will simultaneously be correct a proportion $(1 - \alpha)$ of the time. It will be noted that the intervals in (a) are larger than the intervals in (b) by the ratio $\sqrt{2F}/t$. This wider interval is the "price" we pay for making joint statements about y for any number of or for all of the x values, rather than the y for a single x.

Another *caution* is in order. We cannot use the same computed line in (b) and (c) to make a large number of predictions, and claim that 100 $(1 - \alpha)$ % of the predictions will be correct. The *estimated* line may be very close to the *true line*, in which case nearly all of the interval predictions may be correct; or the line may be considerably different from the *true line*, in which case very few may be correct. In practice, provided our situation is *in control*, we should always revise our estimate of the line to include additional information in the way of new points.

5-4.1.2.1 What is the $(1 - \alpha)$ Confidence Band for the Line as a Whole?

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 9.277617$ from Worksheet 5-4.1
(3) Look up $F_{1-\alpha}$ for $(2, n - 2)$ degrees of freedom in Table A-5.	(3) $F_{.95}(2, 14) = 3.74$
(4) Choose a number of values of X (within the range of the data) at which to compute points for drawing the confidence band.	(4) Let: $X = 30$ $X = 400$ $X = 800$ $X = 1200$ $X = 1500$, for example.
(5) At each selected value of X, compute: $Y_c = \bar{Y} + b_1 (X - \bar{X})$ and $W_1 = \sqrt{2F}\, s_Y \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$	(5) See Table 5-2 for a convenient computational arrangement and the example calculations.
(6) A $(1 - \alpha)$ confidence band for the whole line is determined by $Y_c \pm W_1$.	(6) See Table 5-2.

Procedure

(7) To draw the line and its confidence band, plot Y_c at two of the extreme selected values of X. Connect the two points by a straight line. At each selected value of X, also plot $Y_c + W_1$ and $Y_c - W_1$. Connect the upper series of points, and the lower series of points, by smooth curves.

If more points are needed for drawing the curves for the band, note that, because of symmetry, the calculation of W_1 at n values of X actually gives W_1 at $2n$ values of X.

Example

(7) See Figure 5-8.

For example: W_1 (but not Y_c) has the same value at $X = 400$ (i.e., $\bar{X} - 351.875$) as at $X = 1103.75$ (i.e., $\bar{X} + 351.875$).

TABLE 5-2. COMPUTATIONAL ARRANGEMENT FOR PROCEDURE 5-4.1.2.1

X	$(X - \bar{X})$	Y_c	$\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}}$	$s^2_{Y_c}$	s_{Y_c}	W_1	$Y_c + W_1$	$Y_c - W_1$
30	−721.875	4641.49	.21778	18.7452	4.3296	11.84	4653.33	4629.65
400	−351.875	4475.04	.09940	8.5558	2.9250	8.00	4483.04	4467.04
800	48.125	4295.10	.06319	5.4390	2.3322	6.38	4301.48	4288.72
1200	448.125	4115.16	.12234	10.5303	3.2450	8.88	4124.04	4106.28
1500	748.125	3980.20	.22928	19.7351	4.4424	12.15	3992.35	3968.05

$\bar{X} = 751.875$

coded $\bar{Y} = 316.75$

$\bar{Y}$ (original units) $= 4316.75$

$s^2_Y = 86.0741857$

$\frac{1}{n} = .0625$

$b_1 = -.44985482$

$S_{xx} = 3,355,843.75$

$\sqrt{2F} = 2.735$

$Y_c = \bar{Y} + b_1 (X - \bar{X})$

$s^2_{Y_c} = s^2_Y \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{xx}}\right]$

$W_1 = 2.735\, s_{Y_c}$

5-4.1.2.2 Give a $(1 - \alpha)$ Confidence Interval Estimate for a Single Point on the Line (i.e., the Mean Value of Y Corresponding to a Chosen Value of $x = x'$)

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 9.277617$ from Worksheet 5-4.1
(3) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(3) $t_{.975}\ (14) = 2.145$
(4) Choose X', the value of X at which we want to make an interval estimate of the mean value of Y.	(4) Let $X' = 1200$
(5) Compute: $W_2 = t_{1-\alpha/2}\, s_Y \left[\frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$ and $Y_c = \bar{Y} + b_1 (X' - \bar{X})$	(5) $W_2 = 2.145\ (3.2451)$ $= 6.96$ $Y_c = 4115.16$
(6) A $(1 - \alpha)$ confidence interval estimate for the mean value of Y corresponding to $X = X'$ is given by $Y_c \pm W_2$.	(6) A 95% confidence interval estimate for the mean value of Y corresponding to $X = 1200$ is 4115.16 ± 6.96 $= 4108.20$ to 4122.12.

Note: An interval estimate of the intercept of the line (β_0) is obtained by setting $X' = 0$ in the above procedure.

5-4.1.2.3 Give a $(1 - \alpha)$ Confidence Interval Estimate for a Single (Future) Value (Y') of Y Corresponding to a Chosen Value (x') of x.

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 9.277617$ from Worksheet 5-4.1
(3) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(3) $t_{.975}\ (14) = 2.145$
(4) Choose X', the value of X at which we want to make an interval estimate of a single value of Y.	(4) Let $X' = 1200$
(5) Compute: $W_3 = t_{1-\alpha/2}\, s_Y \left[1 + \frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$ and $Y_c = \bar{Y} + b_1 (X' - \bar{X})$	(5) $W_3 = 2.145\ (9.8288)$ $= 21.08$ $Y_c = 4115.16$
(6) A $(1 - \alpha)$ confidence interval estimate for Y' (the single value of Y corresponding to X') is $Y_c \pm W_3$.	(6) A 95% confidence interval estimate for a single value of Y corresponding to $X' = 1200$ is 4115.16 ± 21.08 $= 4094.08$ to 4136.24 .

5-4.1.3 What is the Confidence Interval Estimate for β_1, the Slope of the True Line $y = \beta_0 + \beta_1 x$?

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(2) $t_{.975}\ (14) = 2.145$
(3) Obtain s_{b_1} from Worksheet.	(3) $s_{b_1} = .005064$ from Worksheet 5.4.1
(4) Compute $W_4 = t_{1-\alpha/2}\, s_{b_1}$	(4) $W_4 = 2.145\ (.005064)$ $= .010862$
(5) A $(1 - \alpha)$ confidence interval estimate for β_1 is $b_1 \pm W_4$.	(5) $b_1 = -.449855$ $W_4 = .010862$ A 95% confidence interval for β_1 is the interval $-.449855 \pm .010862$, i.e., the interval from $-.460717$ to $-.438993$.

5-4.1.4 If We Observe n' New Values of Y (with Average $\bar{Y}'$), How Can We Use the Fitted Regression Line to Obtain an Interval Estimate of the Value of x that Produced These Values of Y?

Example: Suppose that we obtain 10 new measurements of Young's modulus (with average, $\bar{Y}' = 4500$) and we wish to use the regression line to make an interval estimate of the temperature (x) at which the measurements were made.

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(2) $t_{.975}\,(14) = 2.145$
(3) Obtain b_1 and $s^2_{b_1}$ from Worksheet.	(3) From Worksheet 5-4.1, $b_1 = -.449855$ $s^2_{b_1} = .0000256490$
(4) Compute $C = b_1^2 - (t_{1-\alpha/2})^2\, s^2_{b_1}$	(4) $C = .202370 - .000118$ $= .202252$
(5) A $(1 - \alpha)$ confidence interval estimate for the X corresponding to $\bar{Y}'$ is computed from $X' = \bar{X} + \dfrac{b_1\,(\bar{Y}' - \bar{Y})}{C} \pm \dfrac{t_{1-\alpha/2}\, s_Y}{C} \sqrt{\dfrac{(\bar{Y}' - \bar{Y})^2}{S_{xx}} + \left(\dfrac{1}{n} + \dfrac{1}{n'}\right) C}$	(5) A 95% confidence interval would be computed as follows:

$$X' = 751.875 - \frac{.449855\,(4500 - 4316.75)}{.202252} \pm \frac{2.145\,(9.277617)}{.202252} \times \sqrt{\frac{(183.25)^2}{3{,}355{,}843.75} + (.1625)\,(.202252)}$$

$$= 751.875 - 407.590 \pm 98.39452\,\sqrt{.0100066 + .0328660}$$

$$= 344.285 \pm 98.39452\,\sqrt{.0428726}$$

$$= 344.285 \pm 98.39452\,(.20706)$$

$$= 344.285 \pm 20.374$$

The interval from $X = 323.911$ to $X = 364.659$ is a 95% confidence interval for the value of temperature which produced the 10 measurements whose mean Young's modulus was 4500.

5-4.1.5 Using the Fitted Regression Line, How Can We Choose a Value (x') of x Which We May Expect with Confidence ($1 - \alpha$) Will Produce a Value of Y Not Less Than Some Specified Value Q?

Example: What value (x') of temperature (x) can be expected to produce a value of Young's modulus not less than 4300?

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$; and choose Q	(1) Let: $1 - \alpha = .95$ $\alpha = .05$ $Q = 4300$
(2) Look up $t_{1-\alpha}$ for $n - 2$ degrees of freedom in Table A-4.	(2) $t_{.95}\ (14) = 1.761$
(3) Obtain b_1 and $s^2_{b_1}$ from Worksheet.	(3) From Worksheet 5-4.1, $b_1 = -.449855$ $s^2_{b_1} = .0000256490$
(4) Compute $C = b_1^2 - (t_{1-\alpha})^2\, s^2_{b_1}$	(4) $C = .202370 - .000080$ $= .202290$
(5) Compute $X' = \bar{X} + b_1\left(\frac{Q - \bar{Y}}{C}\right) \pm \frac{t_{1-\alpha}\, s_Y}{C}\sqrt{\frac{(Q - \bar{Y})^2}{S_{xx}} + \left(\frac{n+1}{n}\right) C}$ where the sign before the last term is + if b_1 is positive or − if b_1 is negative. We have confidence $(1 - \alpha)$ that a value of $X = X'$ will correspond to (produce) a value of Y not less than Q. (See discussion of "confidence" in straight-line prediction in Paragraph 5-4.1.2).	(5) The value of X' is computed as follows: $X' = 751.875 + \frac{-.449855\,(4300 - 4316.75)}{.202290} - \frac{1.761\ (9.277617)}{.202290} \times \sqrt{\frac{(4300 - 4316.75)^2}{3{,}355{,}843.75} + \left(\frac{17}{16}\right) C}$ $= 751.875 + 37.249 - 80.764662\sqrt{.000084 + .214933}$ $= 751.875 + 37.249 - 80.764662\sqrt{.215017}$ $= 751.875 + 37.249 - 37.450$ $= 751.674$

5-4.1.6 Is the Assumption of Linear Regression Justified?

This involves a test of the assumption that the mean Y values ($\bar{Y}_x$) for given x values do lie on a straight line (we assume that for any given value of x, the corresponding individual Y values are normally distributed with variance σ_Y^2, which is independent of the value of x). A simple test is available provided that we have more than one observation on Y at one or more values of x. Assume that there are n pairs of values (x_i , Y_i), and that among these pairs there occur only k values of x (where k is less than n).

For example, see the data recorded in Table 5-3 which shows measurements of Young's modulus (coded) of sapphire rods as a function of temperature.

Each x is recorded in Column 1, and the corresponding Y values (varying in number from 1 to 3 in the example) are recorded opposite the appropriate x. The remaining columns in the table are convenient for the required computations.

TABLE 5-3. COMPUTATIONAL ARRANGEMENT FOR TEST OF LINEARITY

X = Temperature	Y = Young's Modulus Minus 3000			ΣY	$(\Sigma Y)^2$	ΣY^2	n_i	$n_i X_i$	$n_i X_i^2$	ΣXY	$\frac{(\Sigma Y)^2}{n_i}$
500	328			328	107584	107584	1	500	250000	164000	107584
550	296			296	87616	87616	1	550	302500	162800	87616
600	266			266	70756	70756	1	600	360000	159600	70756
603	260	244		504	254016	127136	2	1206	727218	303912	127008
650	240	232	213	685	469225	156793	3	1950	1267500	445250	156408.3
700	204	203	184	591	349281	116681	3	2100	1470000	413700	116427
750	174	175	154	503	253009	84617	3	2250	1687500	377250	84336.3
800	152	146	124	422	178084	59796	3	2400	1920000	337600	59361.3
850	117	94		211	44521	22525	2	1700	1445000	179350	22260.5
900	97	61		158	24964	13130	2	1800	1620000	142200	12482
950	38			38	1444	1444	1	950	902500	36100	1444
1000	30	5		35	1225	925	2	2000	2000000	35000	612.5
			TOTAL	4037 $= T_1$		849003 $= T_2$	24 $= n$	18006 $= T_3$	13952218 $= T_4$	2756762 $= T_5$	846296 $= T_6$

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let: $\alpha = .05$ $1 - \alpha = .95$
(2) Compute: $\bar{Y} = \frac{T_1}{n}$ $\bar{X} = \frac{T_3}{n}$, the weighted average of X.	(2) $\bar{Y} = \frac{4037}{24} = 168.21$ $\bar{X} = \frac{18006}{24} = 750.25$
(3) Compute $S_1 = T_6 - \frac{(T_1)^2}{n}$	(3) $\frac{(T_1)^2}{n} = 679057.04$ $S_1 = 846296 - 679057.04 = 167238.96$
(4) Compute $b = \dfrac{T_5 - \frac{T_3 T_1}{n}}{T_4 - \frac{(T_3)^2}{n}}$	(4) $b = \frac{2756762 - 3028759.25}{13952218 - 13509001.5} = \frac{-271997.25}{443216.5} = -0.6136894$
(5) Compute $S_2 = b\left(T_5 - \frac{T_3 T_1}{n}\right)$	(5) $S_2 = -0.6136894\,(-271997.25) = 166921.83$
(6) Compute $S_3 = T_2 - \frac{(T_1)^2}{n}$	(6) $S_3 = 849003 - 679057.04 = 169945.96$
(7) Look up $F_{1-\alpha}$ for $(k - 2, n - k)$ degrees of freedom in Table A-5.	(7) $n = 24$ $k = 12$ $F_{.95}$ for (10, 12) degrees of freedom $= 2.75$
(8) Compute $F = \left(\frac{S_1 - S_2}{S_3 - S_1}\right)\left(\frac{n - k}{k - 2}\right)$	(8) $F = \left(\frac{317.13}{2707}\right)\left(\frac{24 - 12}{10}\right) = (.11715)\,(1.2) = 0.14$
(9) If $F > F_{1-\alpha}$, decide that the "array means" $\bar{Y}_x$ do not lie on a straight line. If $F < F_{1-\alpha}$, the hypothesis of linearity is not disproved.	(9) Since F is less than $F_{1-\alpha}$, the hypothesis of linearity is not disproved.

5-4.2 FI RELATIONSHIPS WHEN THE INTERCEPT IS KNOWN TO BE EQUAL TO ZERO (LINES THROUGH THE ORIGIN)

In Paragraph 5-4.1, we assumed:

(a) that there is an underlying linear functional relationship between x and y of the form $y = \beta_0 + \beta_1 x$, with intercept β_0 and slope β_1 both different from zero;

(b) that our data consist of observed values $Y_1, Y_2, \ldots, Y_n$ of y, corresponding to accurately-known values $x_1, x_2, \ldots, x_n$ of x; and,

(c) that the Y's can be regarded as being independently and normally distributed with means equal to their respective *true* values (i.e, mean of $Y_i = \beta_0 + \beta_1 x_i$, $i = 1, 2, \ldots, n$) and constant variance $\sigma^2_{Y \cdot x} = \sigma^2$ for all x.

Furthermore, we gave: a procedure (Paragraph 5-4.1.2.2 with $X' = 0$) for determining confidence limits for β_0, and hence for testing the hypothesis that $\beta_0 = 0$, in the absence of prior knowledge of the value of β_1; and a procedure that is independent of the value of β_0 (Paragraph 5-4.1.3) for determining confidence limits for β_1, and hence for testing the hypothesis that $\beta_1 = 0$.

We now consider the analysis of data corresponding to an FI structural relationship when it is known that $y = 0$ when $x = 0$, so that the line must pass through the origin, i.e., *when it is known that* $\beta_0 = 0$. To begin with, we assume as in (b) and (c) above, that our data consist of observed values $Y_1, Y_2, \ldots, Y_n$, of a *dependent* variable y corresponding to accurately-known values $x_1, x_2, \ldots, x_n$ of the *independent* variable x and that these Y's can be regarded as being *independently* and normally distributed with means $\beta_1 x_1, \beta_1 x_2, \ldots, \beta_1 x_n$, respectively, and variances $\sigma^2_{Y \cdot x}$ that may depend on x. We consider explicitly the cases of constant variance ($\sigma^2_{Y \cdot x} = \sigma^2$), variance proportional to x ($\sigma^2_{Y \cdot x} = x\sigma^2$), and standard deviation proportional to x ($\sigma_{Y \cdot x} = x\sigma$). Finally, we consider briefly the case of *cumulative data* where $x_1 < x_2 < \ldots < x_n$ and the error in Y_i is of the form $e_1 + e_2 + \ldots + e_{i-1} + e_i$, that is, is the sum of the errors of all preceding Y's plus a "private error" e_i of its own. Following Mandel,[3] we assume that the errors (e_i) are independently and normally distributed with zero means and with variances proportional to the length of their generation intervals, i.e., $\sigma^2_{e_i} = (x_i - x_{i-1})\sigma^2$. Under these circumstances, the Y's will be normally distributed with means $\beta_1 x_1, \beta_1 x_2, \ldots, \beta_1 x_n$, respectively, as before; and with variances $\sigma^2_{Y_i} = x_i \sigma^2$, respectively; but will not be independent owing to the overlap among their respective errors.

5-4.2.1 Line Through Origin, Variance of Y's Independent of x.

The slope of the best-fitting line of the form $Y = b_1 x$ is given by

$$b_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

and the estimated variance of b_1 is

$$s^2_{b_1} = \frac{s^2_Y}{\sum_{i=1}^{n} x_i^2}$$

where

$$s^2_Y = \frac{\sum_{i=1}^{n} (Y_i - b_1 x_i)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{1}^{n} x_i Y_i\right)^2}{\sum_{1}^{n} x_i^2}}{n-1} .$$

Consequently, we may effect a simplification of our Basic Worksheet—see Worksheet 5-4.2.1.

Using the values of b_1 and s_{b_1} so obtained, confidence limits for β_1, the slope of the true line through the origin, $y = \beta_1 x$, can be obtained by following the procedure of Paragraph 5-4.1.3 using $t_{1-\alpha/2}$ for $n - 1$ degrees of freedom. Confidence limits for the line as a whole then are obtained simply by plotting the lines $y = \beta_1^U x$ and $y = \beta_1^L x$, where β_1^U and β_1^L are the upper and lower confidence limits for β_1 obtained in the manner just described. The limiting lines, in this instance, also furnish confidence limits for the value y' of y corresponding to a particular point on the line, say for $x = x'$, so that an additional procedure is unnecessary. Confidence limits for a single future observed Y corresponding to $x = x'$ are given by

$$b_1 x' \pm t_{1-\alpha/2} \sqrt{s^2_Y + (x')^2 s^2_{b_1}} \;,$$

where s^2_Y and s_{b_1} are from our modified worksheet and $t_{1-\alpha/2}$ corresponds to $n - 1$ degrees of freedom.

WORKSHEET 5-4.2.1

WORKSHEET FOR FI RELATIONSHIPS WHEN THE INTERCEPT IS KNOWN TO BE ZERO AND THE VARIANCES OF THE Y's IS INDEPENDENT OF x

X denotes __________ Y denotes __________

ΣX = __________ ΣY = __________

$\bar{X}$ = __________ $\bar{Y}$ = __________

Number of points: n = __________

Step (1) ΣXY = __________

(2) ΣX^2 = __________

(3) ΣY^2 = __________

(4) $b_1 = \dfrac{\Sigma XY}{\Sigma X^2}$ = Step (1) ÷ Step (2)

(5) $\dfrac{(\Sigma XY)^2}{\Sigma X^2}$ = __________

(6) $(n-1)\,s_Y^2$ = Step (3) − Step (5)

(7) s_Y^2 = Step (6) ÷ $(n-1)$

s_Y = __________

Equation of the Line:

$Y = b_1 X$

____

Estimated variance of the slope:

$s_{b_1}^2 = \dfrac{s_Y^2}{\Sigma X^2}$ = Step (7) ÷ Step (2)

s_{b_1} = __________

5-4.2.2 Line Through Origin, Variance Proportional to x ($\sigma_{Y\cdot x}^2 = x\sigma^2$). The slope of the best-fitting line of form $Y = b_1 x$ is given by

$$b_1 = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{Y}}{\bar{x}} ,$$

the *ratio of the averages*, and the estimated variance of b_1 is

$$s_{b_1}^2 = \frac{s^2}{\sum_{i=1}^{n} x_i}$$

where

$$(n-1)\,s^2 = \sum_{i=1}^{n}\left(\frac{Y_i^2}{x_i}\right) - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{\sum_{i=1}^{n} x_i} .$$

Using the values of b_1 and s_{b_1} so obtained, confidence limits for β_1, the slope of the true line through the origin, $y = \beta_1 x$, can be obtained by following the procedure of Paragraph 5-4.1.3 using $t_{1-\alpha/2}$ for $n - 1$ degrees of freedom. Confidence limits for the line as a whole then are obtained simply by plotting the lines $y = \beta_1^U x$ and $y = \beta_1^L x$ where β_1^U and β_1^L are the upper and lower confidence limits for β_1 obtained in the manner just described. The limiting lines, in this instance, also furnish confidence limits for the value y' corresponding to a particular point on the line, say for $x = x'$. Confidence limits for a single future observed Y corresponding to $x = x'$, are given by

$$b_1 x' \pm t_{1-\alpha/2} \sqrt{x'\, s^2 + (x')^2\, s_{b_1}^2} ,$$

where s_{b_1} is computed as shown above and $t_{1-\alpha/2}$ corresponds to $n - 1$ degrees of freedom.

5-4.2.3 Line Through Origin, Standard Deviation Proportional to x ($\sigma_{Y \cdot x} = x\sigma$). The slope of the best-fitting line of form $Y = b_1 x$ is given by

$$b_1 = \sum_{i=1}^{n} \left(\frac{Y_i}{x_i}\right) / n,$$

the average of the ratios $\left(\frac{Y_i}{x_i}\right)$,

and the estimated variance of b_1 is

$$s_{b_1}^2 = \frac{s^2}{n}$$

where

$$(n-1)\, s^2 = \sum_{i=1}^{n} \left(\frac{Y_i}{x_i}\right)^2 - \frac{\left[\sum_{i=1}^{n} \left(\frac{Y_i}{x_i}\right)\right]^2}{n}$$

that is,

$$s^2 = \frac{\sum_{i=1}^{n} R_i - \frac{(\Sigma R_i)^2}{n}}{n\,(n-1)}$$

for $\quad R_i = \dfrac{Y_i}{x_i}$.

Using the values of b_1 and s_{b_1} so obtained, confidence limits for β_1, the slope of the true line through the origin, $y = \beta_1 x$, can be obtained by following the procedure of Paragraph 5-4.1.3 using $t_{1-\alpha/2}$ for $n - 1$ degrees of freedom. Confidence limits for the line as a whole are then obtained simply by plotting the lines $y = \beta_1^U x$ and $y = \beta_1^L x$ where β_1^U and β_1^L are the upper and lower confidence limits for β_1 obtained in the manner just described. The limiting lines, in this instance, also furnish confidence limits for the value y' of y corresponding to a particular point on the line, say for $x = x'$. Confidence limits for a single future observed Y corresponding to $x = x'$, are given by

$$b_1 x' \pm t_{1-\alpha/2}\, x' \sqrt{s^2 + s_{b_1}^2}\ ,$$

where s_{b_1} is computed as shown above and $t_{1-\alpha/2}$ corresponds to $n - 1$ degrees of freedom.

5-4.2.4 Line Through Origin, Errors of Y's Cumulative (Cumulative Data). In many engineering tests and laboratory experiments the observed values $Y_1, Y_2, \ldots, Y_i, \ldots,$ of a dependent variable y represent the cumulative magnitude of some effect at successive values $x_1 < x_2 < x_3 < \ldots$ of the independent variable x. Thus, $Y_1, Y_2, \ldots,$ may denote: the total weight loss of a tire under road test, measured at successive mileages $x_1, x_2, \ldots$; or the weight gain of some material due to water absorption at successive times $x_1, x_2, \ldots$; or the total deflection of a beam (or total compression of a spring) under continually increasing load, measured at loads $x_1, x_2, \ldots$; and so forth. In such cases, even though the underlying functional relationship takes the form of a line through the origin, $y = \beta x$, none of the procedures that we have presented thus far will be applicable, because of the cumulative effect of errors of technique on the successive Y's; the deviation of Y_i from its true or expected value y_i will include the deviation $(Y_{i-1} - y_{i-1})$ of Y_{i-1} from its true or expected value, plus an individual "private deviation or error" e_i of its own. Hence, the total error of Y_i will be the sum $(e_1 + e_2 + \ldots + e_{i-1} + e_i)$ of the individual error contributions of $Y_1, Y_2, \ldots, Y_{i-1}$, and its own additional deviation.

If the test or experiment starts at $x_0 = 0$, and the x's form an uninterrupted sequence $0 < x_1 < x_2 < \ldots < x_n$, and if we may regard the individual error contributions $e_1, e_2, \ldots,$ as independently and normally distributed with zero means and variances proportional to the lengths of the x-intervals over which they accrue, i.e., if $\sigma_{e_i}^2 = (x_i - x_{i-1})\,\sigma^2$, then the best estimate of the slope of the underlying linear functional relation $y = \beta_1 x$ is given by

$$b_1 = \frac{Y_n}{x_n}$$

and estimated variance of b_1

$$s_{b_1}^2 = \frac{1}{(n-1)\, x_n} \left\{ \sum_{i=1}^{n} \frac{(Y_i - Y_{i-1})^2}{x_i - x_{i-1}} - \frac{Y_n^2}{x_n} \right\}$$

in which $x_0 = 0$ and $Y_0 = 0$ by hypothesis.

Using the values of b_1 and s_{b_1} so obtained, confidence limits for β_1, the slope of the true line through the origin, $y = \beta_1 x$, can be obtained by following the procedure of Paragraph 5-4.1.3 using $t_{1-\alpha/2}$ for $n - 1$ degrees of freedom. Confidence limits for the line as a whole then are obtained simply by plotting the lines $y = \beta_1^U x$ and $y = \beta_1^L x$, where β_1^U and β_1^L are the upper and lower confidence limits for β_1 obtained in the manner just described. These limit lines also provide confidence limits for a particular point on the line, say the value y' corresponding to $x = x'$. For the fitting of lines of this sort to cumulative data under more general conditions, and for other related matters, see Mandel's article.[3]

5-4.3 FII RELATIONSHIPS

Distinguishing Features. There is an underlying mathematical (functional) relationship between the two variables, of the form

$$y = \beta_0 + \beta_1 x.$$

Both X and Y are subject to errors of measurement. Read Paragraph 5-3.1 and Table 5-1.

The full treatment of this case depends on the assumptions we are willing to make about error distributions. For complete discussion of the problem, see Acton.[4]

5-4.3.1 A Simple Method of Fitting the Line In the General Case.

There is a quick and simple method of fitting a line of the form $Y = b_0 + b_1 X$ which is generally applicable when both X and Y are subject to errors of measurement. This method is described in Bartlett,[5] and is illustrated in this paragraph. Similar methods had been used previously by other authors.

(a) For the location of the fitted straight line, use as the pivot point the center of gravity of all n observed points (X_i, Y_i), that is, the point with the mean coordinates $(\bar{X}, \bar{Y})$. In consequence, the fitted line will be of the form $Y = b_0 + b_1 X$ with $b_0 = \bar{Y} - b_1 \bar{X}$, just as in the least-squares method in Paragraph 5-4.1.

(b) For the slope, divide the n plotted points into three non-overlapping groups when considered in the X direction. There should be an equal number of points, k, in each of the two extreme groups, with k as close to $\frac{n}{3}$ as possible.

Take, as the slope of the line,

$$b_1 = \frac{\bar{Y}_3 - \bar{Y}_1}{\bar{X}_3 - \bar{X}_1} ,$$

where

$\bar{Y}_3$ = average Y for 3rd group
$\bar{Y}_1$ = average Y for 1st group
$\bar{X}_3$ = average X for 3rd group
$\bar{X}_1$ = average X for 1st group.

Data Sample 5-4.3.1—Relation of Two Colorimetric Methods

The following data are coded results of two colorimetric methods for the determination of a chemical constituent. (The data have been coded for a special purpose which has nothing to do with this illustration). The interest here, of course, is in the relationship between results given by the two methods, and it is presumed that there is a functional relationship with both methods subject to errors of measurement.

Sample	Method I X	Method II Y
1	3720	5363
2	4328	6195
3	4655	6428
4	4818	6662
5	5545	7562
6	7278	9184
7	7880	10070
8	10085	12519
9	11707	13980

(a) The fitted line must pass through the point $(\bar{X}, \bar{Y})$, where

$$\bar{X} = 6668.4$$
$$\bar{Y} = 8662.6$$

(b) To determine the slope, divide the points into 3 groups. Since there are 9 points, exactly 3 equal groups are obtained.

$$\bar{Y}_3 = 12190$$
$$\bar{Y}_1 = 5995$$
$$\bar{X}_3 = 9891$$
$$\bar{X}_1 = 4234$$
$$b_1 = \frac{\bar{Y}_3 - \bar{Y}_1}{\bar{X}_3 - \bar{X}_1} = \frac{12190 - 5995}{9891 - 4234} = \frac{6195}{5657} = 1.0951$$
$$b_0 = \bar{Y} - b_1 \bar{X} = 8662.6 - \frac{6195}{5657}(6668.4) = 1360.0$$

The fitted line

$$Y = 1360.0 + 1.0951\,X$$

is shown in Figure 5-9.

Procedures are given in Bartlett[5] for determining 100 $(1 - \alpha)$ % confidence limits for the true slope β_1; and for determining a 100 $(1 - \alpha)$ % confidence ellipse for β_0 and β_1 jointly, from which 100 $(1 - \alpha)$ % confidence limits for the line as a whole can be derived. For strict validity, they require that the measurement errors affecting the observed X_i be sufficiently small in comparison with the spacing of their true values x_i that the allocation of the observational points (X_i, Y_i) to the three groups is unaffected. These procedures are formally

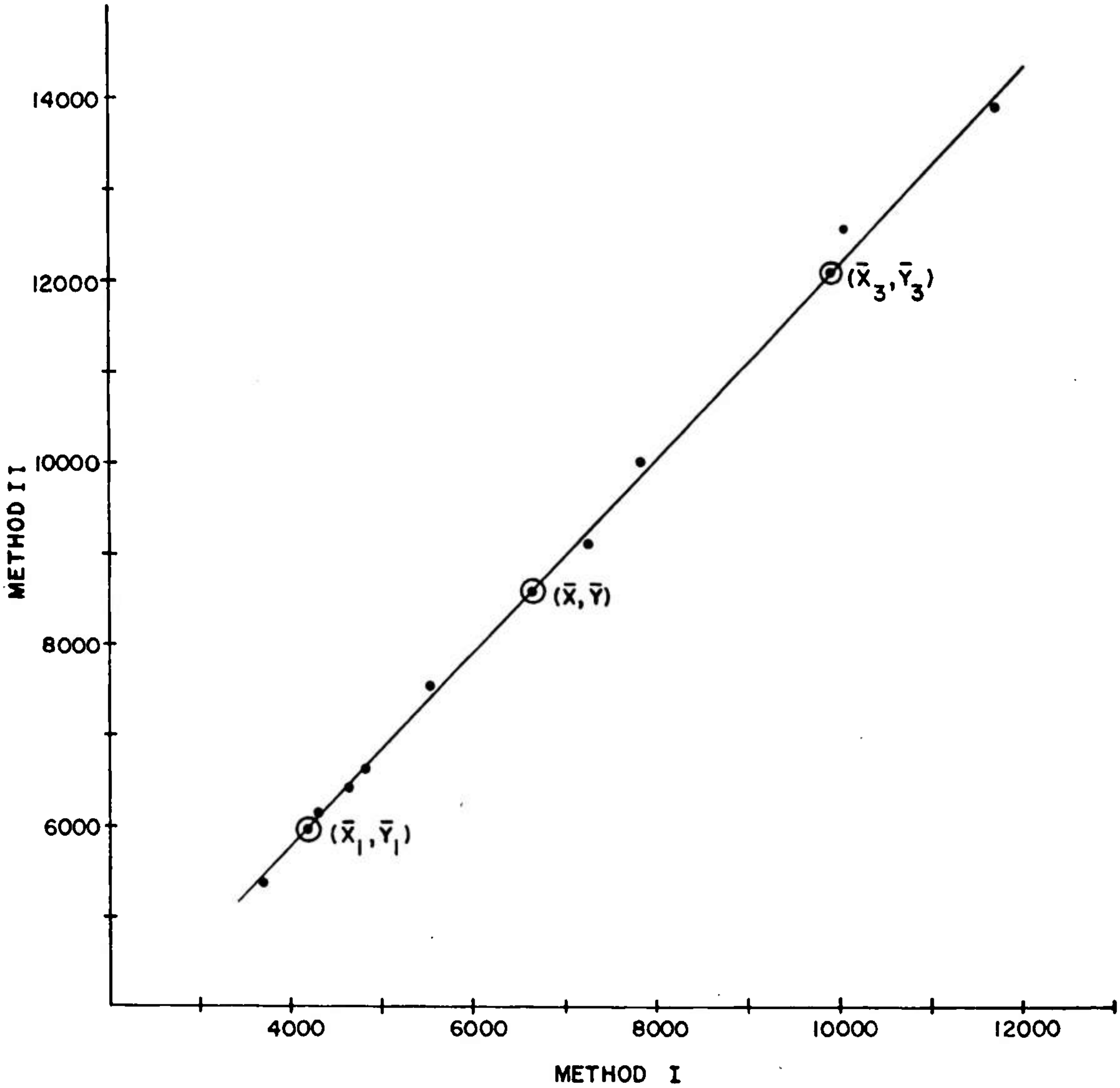

Figure 5-9. Relationship between two methods of determining a chemical constituent—an FII relationship.

similar to those appropriate to the least-squares method in FI situations, but involve more complex calculations. We do not consider them further here.

5-4.3.2 An Important Exceptional Case. Until comparatively recently it was not realized that there is a broad class of controlled experimental situations in which both X and Y are subject to errors of measurement, yet all of the techniques appropriate to the FI case (x's accurately known, measurement errors affect the Y's only) are strictly applicable without change.

As an example, let us consider the case of an analytical chemist who, in order to obtain an accurate determination of the concentration of a potassium sulphate solution, decides to proceed as follows: From a burette he will draw off 5, 10, 15, and 20 *ml* samples of the solution. Volume of solution is his independent variable x, and his target values are $x_1 = 5$, $x_2 = 10$, $x_3 = 15$, and $x_4 = 20$, respectively. The volumes of solution that he actually draws off X_1, X_2, X_3, and X_4 will, of course, differ from the nominal or target values as a result of errors of technique, and he will not attempt to measure their volumes accurately. These four samples of the potassium sulphate solution then will be treated with excess barium chloride, and the precipitated barium sulphate dried and weighed. Let Y_1, Y_2, Y_3, and Y_4 denote the corresponding yields of barium sulphate. These yields actually will correspond, of course, to the actual inputs X_1, X_2, X_3, and X_4, respectively; and will differ from the true yields associated with these inputs, say $y_1(X_1)$, $y_2(X_2)$, $y_3(X_3)$, and $y_4(X_4)$, respectively, as a result of errors of weighing and analytical technique. The sulphate concentration of the original potassium sulphate solution then will be determined by evaluating the slope b_1 of the best fitting straight line $Y = b_0 + b_1x$, relating the observed barium sulphate yields (Y_1, Y_2, Y_3, and Y_4) to the nominal or target volumes of solution (x_1, x_2, x_3, and x_4)—the intercept b_0 of the line making appropriate allowance for the possibility of bias of the analytical procedure resulting in a non-zero blank.

Without going into the merits of the foregoing as an analytical procedure, let us note a number of features that are common to *controlled experiments:* First, the experimental program involves a number of *preassigned* nominal or target values (x_1, x_2, . . .) of the independent variable x, to which the experimenter equates the independent variable in his experiment as best he can, and then observes the corresponding yields (Y_1, Y_2, . . .) of the dependent variable y; Second, the experimenter, in his notebook, records the *observed* yields (Y_1, Y_2, . . .) as corresponding to, and treats them as if they were produced by, the *nominal* or *target* values (x_1, x_2, . . .) of the independent variable—whereas, strictly they correspond to, and were produced by, the actual input values (X_1, X_2, . . .), which ordinarily will differ somewhat from the nominal or target values (x_1, x_2, . . .) as a result of errors of technique. Furthermore, the *effective values* (X_1, X_2, . . .) of the independent variable actually realized in the experiment are not recorded at all—nor even measured!

It is surprising but nevertheless true that an underlying linear structural relationship of the form $y = \beta_0 + \beta_1x$ can be estimated validly from the results of such experiments, by fitting a line of the form $Y = b_0 + b_1x$ in accordance with the procedures for FI situations (x's known accurately, Y's only subject to error). This fact was emphatically brought to the attention of the scientific world by Joseph Berkson in a paper[6] published in 1950, and for its validity requires only the usual assumptions regarding the randomness and independence of the errors of measurement and technique affecting both of the variables (i.e., causing the deviations of the actual *inputs* X_1, X_2, . . . , from their target values x_1, x_2, . . . , *and* the deviations of the observed *outputs* Y_1, Y_2, . . . , from their *true values* of $y_1(X_1)$, $y_2(X_2)$, . . .). The conclusion also extends to the many-variable case considered in Chapter 6, *provided that* the relationship is linear, i.e., that

$$y = \beta_0 + \beta_1x + \beta_2u + \beta_3v + \ldots .$$

If the underlying relationship is a polynomial in x (e.g., $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$), then Geary[7] has found that Berkson's conclusion carries over to the extent that the usual least-squares estimates (given in Chapter 6) of the coefficients of the two highest powers of x (i.e., of β_2 and β_3 here) retain their optimum properties of unbiasedness and minimum variance, but the confidence-interval and tests-of-significance procedures require modification.

5-4.4 SOME LINEARIZING TRANSFORMATIONS

If the form of a non-linear relationship between two variables is known, it is sometimes possible to make a transformation of one or both variables such that the relationship between the transformed variables can be expressed as a straight line. For example, we might know that the relationship is of the form $Y = ab^X$. If we take logs of both sides of this equation, we obtain

$$\log Y = \log a + X \log b,$$

which will be recognized to be a straight line whose intercept on the log Y scale is equal to log a, and whose slope is equal to log b. The procedure for fitting the relationship is given in the following steps.

(1) Make the transformation $Y_T = \log Y$ (i.e., take logs of all the observed Y values).

(2) Use the procedure of Paragraph 5-4.1.1 to fit the line $Y_T = b_0 + b_1 X$, substituting Y_T everywhere for Y.

(3) Obtain the constants of the original equation by substituting the calculated values of b_0 and b_1 in the following equations:

$$b_0 = \log a$$
$$b_1 = \log b,$$

and taking the required antilogs.

Some relationships between X and Y which can easily be transformed into straight-line form are shown in Table 5-4. This table gives the appropriate change of variable for each relationship, and gives the formulas to convert the constants of the resulting straight line to the constants of the relationship in its original form. In addition to the ones given in Table 5-4, some more-complicated relationships can be handled by using special tricks which are not described here, but can be found in Lipka,[8] Rietz,[9] and Scarborough.[10]

It should be noted that the use of these transformations is certain to accomplish one thing only—i.e., to yield a relationship in straight-line form. The transformed data will not necessarily satisfy certain assumptions which are theoretically necessary in order to apply the procedures of Paragraph 5-4.1.1, for example, the assumption that the variability of Y given X is the same for all X. However, for practical purposes and within the range of the data considered, the transformations often do help in this regard.

Thus far, our discussion has centered on the use of transformations to convert a *known* relationship to linear form. The existence of such linearizing transformations also makes it possible to determine the form of a relationship empirically. The following possibilities, adapted from Scarborough,[10] are suggested in this regard:

(1) Plot Y against $\frac{1}{X}$ on ordinary graph paper. If the points lie on a straight line, the relationship is

$$Y = a + \frac{b}{X}$$

(2) Plot $\frac{1}{Y}$ against X on ordinary graph paper. If the points lie on a straight line, the relationship is

$$Y = \frac{1}{a + bX}, \quad \text{or}$$
$$\frac{1}{Y} = a + bX \; .$$

(3) Plot X against Y on semilog paper (X on the arithmetic scale, Y on the logarithmic scale). If the points lie on a straight line, the variables are related in the form

$$Y = ae^{bX}, \text{ or}$$
$$Y = ab^X \; .$$

(4) Plot Y against X on log-log paper. If the points lie on a straight line, the variables are related in the form

$$Y = aX^b \; .$$

TABLE 5-4. SOME LINEARIZING TRANSFORMATIONS

If the Relationship Is of the Form:	Plot the Transformed Variables		Fit the Straight Line $Y_T = b_0 + b_1X_T$	Convert Straight Line Constants (b_0 and b_1) To Original Constants:	
	$Y_T =$	$X_T =$		$b_0 =$	$b_1 =$
$Y = a + \frac{b}{X}$	Y	$\frac{1}{X}$	Use the procedures of Paragraph 5-4.1.1. In all formulas given there, substitute values of Y_T for Y and values of X_T for X, as appropriate.	a	b
$Y = \frac{1}{a + bX}$, or $\frac{1}{Y} = a + bX$	$\frac{1}{Y}$	X		a	b
$Y = \frac{X}{a + bX}$	$\frac{X}{Y}$	X		a	b
$Y = ab^X$	$\log Y$	X		$\log a$	$\log b$
$Y = ae^{bX}$	$\log Y$	X		$\log a$	$b \log e$
$Y = aX^b$	$\log Y$	$\log X$		$\log a$	b
$Y = a + bX^n$, where n is known	Y	X^n		a	b

5-5 PROBLEMS AND PROCEDURES FOR STATISTICAL RELATIONSHIPS

5-5.1 SI RELATIONSHIPS

In this case, we are interested in an association between two variables. See Paragraph 5-3.2 and Table 5-1.

We usually make the assumption that for any fixed value of X, the corresponding values of Y form a normal distribution with means $\bar{Y}_X = \beta_0 + \beta_1 X$ and variance $\sigma^2_{Y \cdot X}$ (read as "variance of Y given X") which is constant for all values of X.* Similarly, we usually assume that for any fixed value of Y, the corresponding values of X form a normal distribution with mean $\bar{X}_Y = \beta'_0 + \beta'_1 Y$ and variance $\sigma^2_{X \cdot Y}$, (variance of X given Y) which is constant for all values of Y.* Taken together, these two sets of assumptions imply that X and Y are jointly distributed according to the bivariate normal distribution. In practical situations, we usually have only a sample from all the possible pairs of values X and Y, and therefore we cannot determine either of the *true* regression lines, $\bar{Y}_X = \beta_0 + \beta_1 X$ or $\bar{X}_Y = \beta'_0 + \beta'_1 Y$, exactly. If we have a random sample of n pairs of values (X_1, Y_1), (X_2, Y_2), . . . , (X_n, Y_n), we can estimate either line, or both. Our method of fitting the line gives us best predictions in the sense that, for a given $X = X'$ our estimate of the corresponding value of $Y = Y'$ will:

(a) on the average equal $\bar{Y}_{X'}$ the mean value of Y for $X = X'$ (i.e., it will be on the *true* line $\bar{Y}_X = \beta_0 + \beta_1 X$); and

(b) have a smaller variance than had we used any other method for fitting the line.

* Strictly, we should write

$m_{Y \cdot X} = \beta_0 + \beta_1 X$

and

$m_{X \cdot Y} = \beta'_0 + \beta'_1 Y$.

See Footnote in Paragraph 5-3.2.

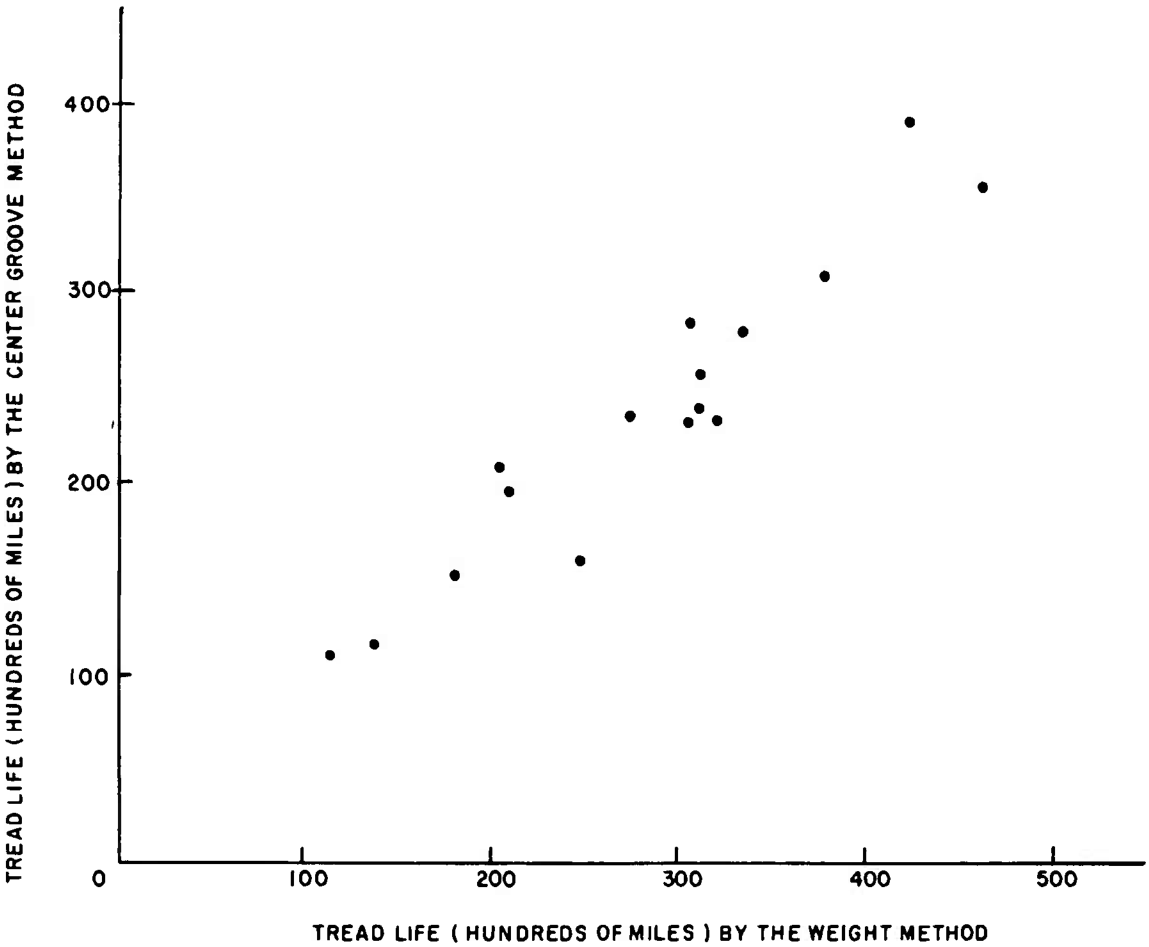

Figure 5-10. Relationship between the weight method and the center groove method of estimating tread life— an SI relationship.

Data Sample 5-5.1—Estimated Tread Wear of Tires

The data used for illustration are from a study of two methods of estimating tread wear of commercial tires (Stiehler and others(11)). The data are shown here and plotted in Figure 5-10. The variable which is taken as the independent variable X is the estimated tread life in hundreds of miles by the *weight-loss* method. The associated variable Y is the estimated tread life by the *groove-depth* method (center grooves). The plot seems to indicate a relationship between X and Y, but the relationship is statistical rather than functional or exact. The scatter of the points stems primarily from product variability and variation of tread wear under normal operating conditions, rather than from errors of measurement of weight loss or groove depth. Descriptions and predictions are applicable only "on the average."

X = Tread Life (Hundreds of Miles) Estimated By Weight Method	Y = Tread Life (Hundreds of Miles) Estimated By Center Groove Method
459	357
419	392
375	311
334	281
310	240
305	287
309	259
319	233
304	231
273	237
204	209
245	161
209	199
189	152
137	115
114	112

5-5.1.1 What is the Best Line To Be Used for Estimating $\bar{Y}_X$ for Given Values of X?

Procedure

The procedure is identical to that of Paragraph 5-4.1.1. Using Basic Worksheet (see Worksheet 5-5.1), compute the line

$$Y = b_0 + b_1X.$$

This is an estimate of the true regression line

$$\bar{Y}_X = \beta_0 + \beta_1X.$$

Using Data Sample 5-5.1, the equation of the fitted line is

$$Y = 13.506 + 0.790212\,X.$$

In Figure 5-11, the line is drawn, and confidence limits for the line (see Paragraph 5-5.1.2) are shown.

WORKSHEET 5-5.1

EXAMPLE OF SI RELATIONSHIP

X denotes Tread Life Estimated by Weight Method

ΣX = 4505

$\bar{X}$ = 281.5625

Y denotes Tread Life Estimated by Center Groove Method

ΣY = 3776

$\bar{Y}$ = 236

Number of points: n = 16

Step (1) ΣXY = 1,170,731

(2) $(\Sigma X)(\Sigma Y)/n$ = 1,063,180

(3) S_{xy} = 107551

(4) ΣX^2 = 1,404,543

(5) $(\Sigma X)^2/n$ = 1,268,439.0625

(6) S_{xx} = 136103.9375

(7) ΣY^2 = 985740

(8) $(\Sigma Y)^2/n$ = 891136

(9) S_{yy} = 94604

(10) $b_1 = \dfrac{S_{xy}}{S_{xx}}$ = .790212

(11) $\bar{Y}$ = 236

(12) $b_1\bar{X}$ = 222.494

(13) $b_0 = \bar{Y} - b_1\bar{X}$ = 13.506

(14) $\dfrac{(S_{xy})^2}{S_{xx}}$ = 84988.119

(15) $(n-2)\, s_Y^2$ = 9615.881

(16) s_Y^2 = 686.849

s_Y = 26.21

Equation of the line:

$Y = b_0 + b_1 X$

= 13.506 + .790212 X

s_{b_1} = 0.0710387

s_{b_0} = 21.048

Estimated variance of the slope:

$s_{b_1}^2 = \dfrac{s_Y^2}{S_{xx}}$ = .005046504

Estimated variance of intercept:

$s_{b_0}^2 = s_Y^2 \left\{\dfrac{1}{n} + \dfrac{\bar{X}^2}{S_{xx}}\right\}$ = 443.002

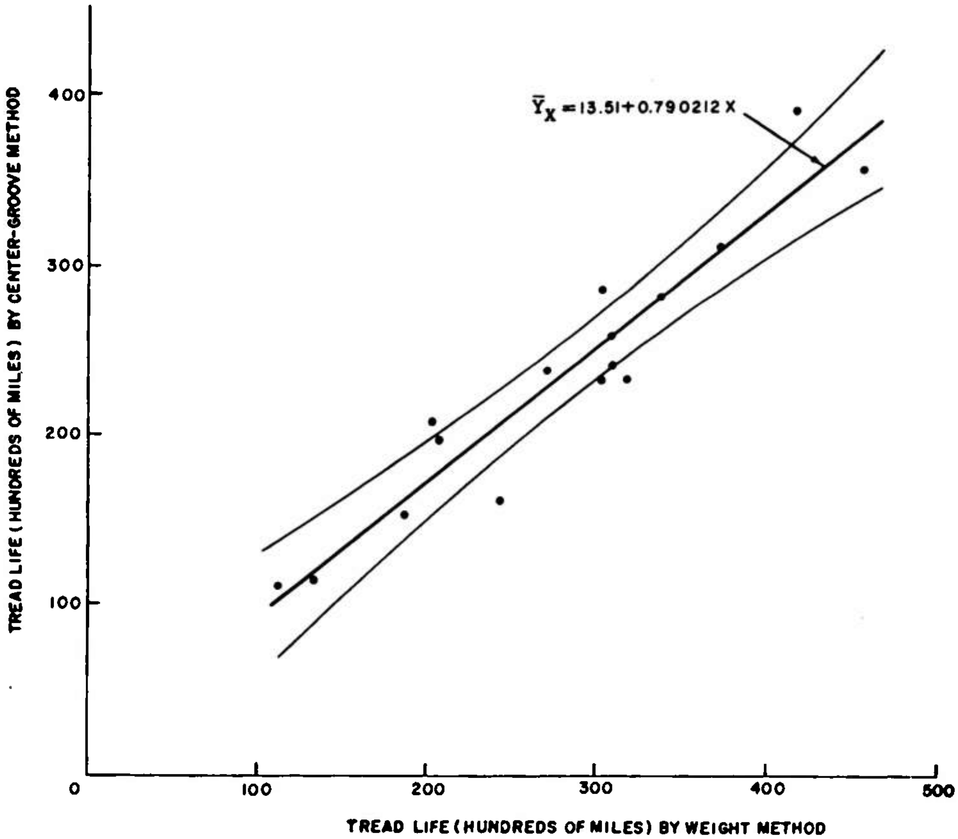

Figure 5-11. Relationship between weight method and center groove method—the line shown with its confidence band is for estimating tread life by center groove method from tread life by weight method.

Using the Regression Line for Prediction. The equation of the fitted line may be used to predict $\bar{Y}_X$, the average value of Y associated with a value of X. For example, using the fitted line, $Y = 13.506 + 0.790212\,X$, the following are some predicted values for $\bar{Y}_X$.

X	$\bar{Y}_X$
200	172
250	211
300	251
350	290
400	330
450	369

5-5.1.2 What are the Confidence Interval Estimates for: the Line as a Whole; a Point on the Line; a Single Y Corresponding to a New Value of X?

Read the discussion of the interpretation of three types of confidence intervals in Paragraph 5-4.1.2, in order to decide which is the appropriate kind of confidence interval.

The solutions are identical to those given in Paragraph 5-4.1.2, and are illustrated for the tread wear of commercial tires example (Data Sample 5-5.1).

5-5.1.2.1 What Is the $(1 - \alpha)$ Confidence Band for the Line as a Whole?

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 26.21$
(3) Look up $F_{1-\alpha}$ for $(2, n - 2)$ degrees of freedom in Table A-5.	(3) $n = 16$ $F_{.95}\ (2, 14) = 3.74$.
(4) Choose a number of values of X (within the range of the data) at which to compute points for drawing the confidence band.	(4) Let: $X = 200$ $X = 250$ $X = 300$ $X = 350$ $X = 400$, for example.
(5) At each selected value of X, compute: $Y_c = \bar{Y} + b_1 (X - \bar{X})$ and $W_1 = \sqrt{2F}\, s_Y \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$	(5) See Table 5-5 for a convenient computational arrangement, and the example calculations.
(6) A $(1 - \alpha)$ confidence band for the whole line is determined by $Y_c \pm W_1$.	(6) See Table 5-5.
(7) To draw the line and its confidence band, plot Y_c at two of the extreme selected values of X. Connect the two points by a straight line. At each selected value of X, plot also $Y_c + W_1$ and $Y_c - W_1$. Connect the upper series of points, and the lower series of points, by smooth curves.	(7) See Figure 5-11.
If more points are needed for drawing the curves, note that, because of symmetry, the calculation of W_1 at n values of X actually gives W_1 at $2n$ values of X.	For example: W_1 (but not Y_c) has the same value at $X = 250$ (i.e., $\bar{X} - 31.56$) as at $X = 313.12$ (i.e., $\bar{X} + 31.56$).

TABLE 5-5. COMPUTATIONAL ARRANGEMENT FOR PROCEDURE 5-5.1.2.1

X	$(X - \bar{X})$	Y_c	$\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{xx}}$	$s^2_{Y_c}$	s_{Y_c}	W_1	$Y_c + W_1$	$Y_c - W_1$
200	−81.56	171.6	0.111375	76.50	8.746	23.9	195.5	147.7
250	−31.56	211.1	0.069818	47.95	6.925	18.9	230.0	192.2
300	+18.44	250.6	0.064998	44.64	6.681	18.3	268.9	232.3
350	68.44	290.1	0.096915	66.57	8.159	22.3	312.4	267.8
400	118.44	329.6	0.165569	113.72	10.66	29.2	358.8	300.4

$\bar{X} = 281.5625$

$\bar{Y} = 236$

$s^2_Y = 686.849$

$\frac{1}{n} = .0625$

$b_1 = 0.790212$

$S_{xx} = 136103.9375$

$Y_c = \bar{Y} + b_1 (X - \bar{X})$

$s^2_{Y_c} = s^2_Y \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{xx}}\right]$

$\sqrt{2F} = \sqrt{7.48} = 2.735$

$W_1 = \sqrt{2F}\, s_{Y_c}$

5-5.1.2.2 Give a (1 − α) Confidence Interval Estimate For a Single Point On the Line, i.e., the Mean Value of Y Corresponding to $X = X'$.

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 26.21$
(3) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(3) $n = 16$ $t_{.975}$ for 14 d.f. $= 2.145$
(4) Choose X', the value of X at which we want to make an interval estimate of the mean value of Y.	(4) Let $X' = 250$, for example.
(5) Compute: $W_2 = t_{1-\alpha/2}\, s_Y \left[\frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$ and $Y_c = \bar{Y} + b_1 (X' - \bar{X})$	(5) $W_2 = (2.145)\,(26.21)\,(.2642) = 14.85$ $Y_c = 211.1$
(6) A $(1 - \alpha)$ confidence interval estimate for the mean value of Y corresponding to $X = X'$ is given by $Y_c \pm W_2$.	(6) A 95% confidence interval estimate for the mean value of Y corresponding to $X = 250$ is 211.1 ± 14.8 , the interval from 196.3 to 225.9 .

5-5.1.2.3 Give a $(1-\alpha)$ Confidence Interval Estimate For a Single (Future) Value of Y Corresponding to a Chosen Value of $X = X'$.

Procedure	Example
(1) Choose the desired confidence level, $1-\alpha$	(1) Let: $1-\alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) $s_Y = 26.21$
(3) Look up $t_{1-\alpha/2}$ for $n-2$ degrees of freedom in Table A-4.	(3) $n = 16$ $t_{.975}$ for 14 d.f. $= 2.145$
(4) Choose X', the value of X at which we want to make an interval estimate of a single value of Y.	(4) Let $X' = 250$, for example.
(5) Compute: $W_3 = t_{1-\alpha/2}\, s_Y \left[1 + \frac{1}{n} + \frac{(X'-\bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$ and $Y_c = \bar{Y} + b_1 (X' - \bar{X})$	(5) $W_3 = (2.145)\ (26.21)\ (1.0343)$ $= 58.1$ $Y_c = 211.1$.
(6) A $(1-\alpha)$ confidence interval estimate for Y' (the single value of Y corresponding to X') is $Y_c \pm W_3$.	(6) A 95% confidence interval estimate for a single value of Y corresponding to $X' = 250$ is 211.1 ± 58.1, the interval from 153.0 to 269.2 .

5-5.1.3 Give a Confidence Interval Estimate For β_1, the Slope of the True Regression Line, $\bar{Y}_X = \beta_0 + \beta_1 X$.

The solution is identical to that of Paragraph 5-4.1.3 and is illustrated here for Data Sample 5-5.1.

Procedure	Example
(1) Choose the desired confidence level, $1-\alpha$	(1) Let: $1-\alpha = .95$ $\alpha = .05$
(2) Look up $t_{1-\alpha/2}$ for $n-2$ degrees of freedom in Table A-4.	(2) $n = 16$ $t_{.975}$ for 14 d.f. $= 2.145$
(3) Obtain s_{b_1} from Worksheet.	(3) $s_{b_1} = 0.0710387$
(4) Compute $W_4 = t_{1-\alpha/2}\, s_{b_1}$	(4) $W_4 = (2.145)\ (.0710387)$ $= 0.152378$
(5) A $(1-\alpha)$ confidence interval estimate for β_1 is $b_1 \pm W_4$.	(5) $b_1 = 0.790212$ $W_4 = 0.152378$ A 95% confidence interval estimate for β_1 is the interval 0.790212 ± 0.152378, i.e., the interval from 0.637834 to 0.942590 .

5-5.1.4 What Is the Best Line For Predicting $\bar{X}_Y$ From Given Values of Y?

For this problem, we fit a line $X = b_0' + b_1' Y$ (an estimate of the true line $\bar{X}_Y = \beta_0' + \beta_1' Y$). To fit this line we need to interchange the roles of the X and Y variables in the computations outlined in Worksheet 5-5.1 and proceed as in Paragraph 5-5.1.1.

That is, the fitted line will be:

$$X = b_0' + b_1' Y \ ,$$

where

$$b_0' = \bar{X} - b_1' \bar{Y}$$

and

$$b_1' = \frac{S_{xy}}{S_{yy}} \ .$$

From Data Sample 5-5.1:

$$b_1' = \frac{107551}{94604} = 1.136855$$

$$b_0' = 281.5625 - (1.136855)\ (236) = 13.26$$

The equation of the fitted line is:

$$X = 13.26 + 1.136855\ Y \ ,$$

and this line is shown in Figure 5-12, along with the line for predicting Y from X.

In order to obtain confidence intervals, we need the following formulas:

$$s_X^2 = \frac{S_{xx} - \dfrac{(S_{xy})^2}{S_{yy}}}{n - 2}$$

$$s_{b'_1}^2 = \frac{s_X^2}{S_{yy}}$$

$$s_{b'_0}^2 = s_X^2 \left\{\frac{1}{n} + \frac{(\bar{Y})^2}{S_{yy}}\right\} \ .$$

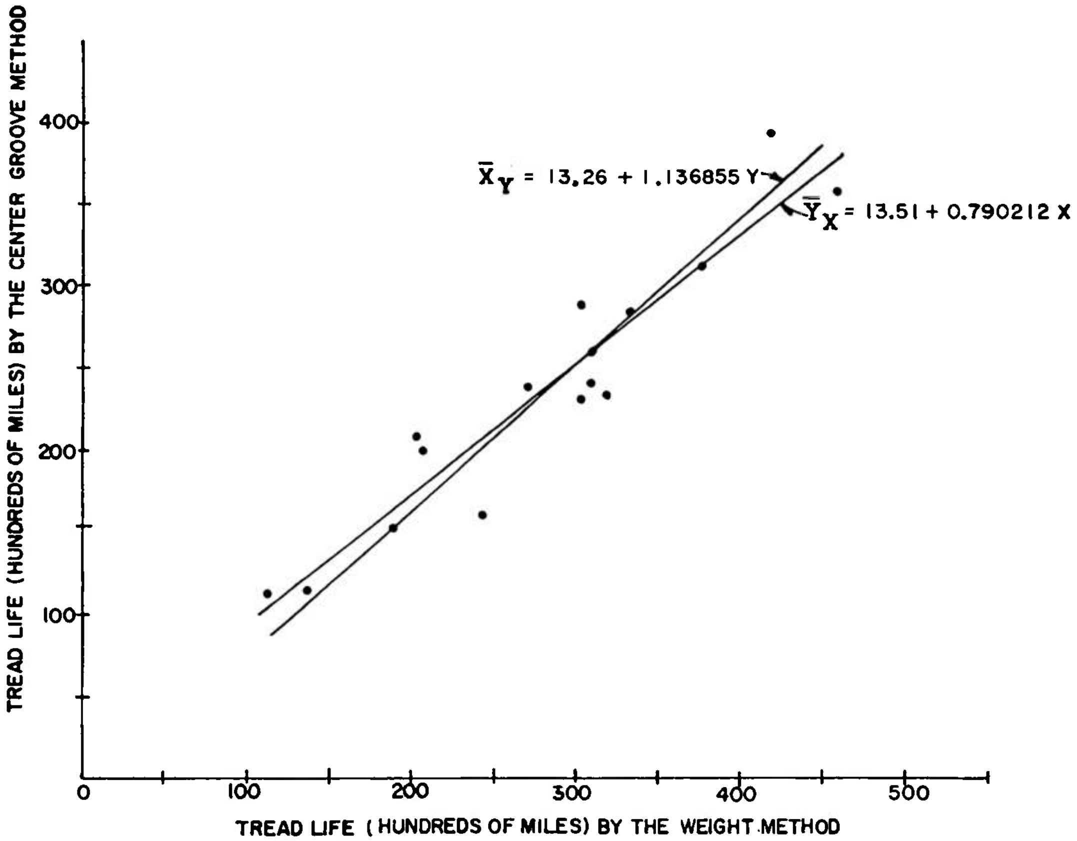

Figure 5-12. Relationship between weight method and center groove method—showing the two regression lines.

5-5.1.5 What is the Degree of Relationship of the Two Variables X and Y as Measured by ρ, the Correlation Coefficient?

Procedure

(1) Compute

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

(2) A 95% confidence interval for ρ can be obtained from Table A-17, using the appropriate n and r. If the confidence interval does not include $\rho = 0$, we may state that the data give reason to believe that there is a relationship (measured by $\rho \neq 0$) between the two variables; otherwise, we may state that the data are consistent with the possibility that the two variables are uncorrelated ($\rho = 0$).

Example

(1) Using Worksheet 5-5.1,

$$r = \frac{107551}{\sqrt{136103.94}\sqrt{94604}}$$

$$= \frac{107551}{(368.92)\ (307.58)}$$

$$= 0.95\ .$$

(2) $n = 16$

$r = 0.95$

From Table A-17, the 95% confidence interval estimate of ρ is the interval from 0.85 to 0.98. Since this interval does not include $\rho = 0$, we may state that the data give reason to believe that there is a relationship between the two methods of estimating tread wear of tires.

5-5.2 SII RELATIONSHIPS

In this case, we are interested in an association between two variables. This case differs from SI in that one variable has been measured at only preselected values of the other variable. (See Paragraph 5-3.2 and Table 5-1.)

For any given value of X, the corresponding values of Y have a normal distribution with mean $\bar{Y}_X = \beta_0 + \beta_1 X$, and variance $\sigma^2_{Y \cdot X}$ which is independent of the value of X. We have n pairs of values $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, in which X is the independent variable. (The X values are selected, and the Y values are thereby determined.) We wish to describe the line which will enable us to make the best estimate of values of Y corresponding to given values of X.

We have seen that for SI there are two lines, one for predicting Y from X and one for predicting X from Y. When we use only selected values of X, however, the only appropriate line to fit is $Y = b_0 + b_1 X$.

It should be noted that SII is handled computationally in the same manner as FI, but both the underlying assumptions and the interpretation of the end results are different.

Data Sample 5-5.2—Estimated Tread Wear of Tires

For our example, we use part of the data used in Data Sample 5-5.1 (the SI example). Suppose that, due to some limitation, we were only able to measure X values between $X = 200$ and $X = 400$, or that we had taken but had lost the data for $X < 200$ and $X > 400$. From Figure 5-10, we use only the 11 observations whose X values are between these limits. The "selected" data are recorded in the following table.

X = Tread Life (Hundreds of Miles) Estimated By Weight Method	Y = Tread Life (Hundreds of Miles) Estimated By Center Groove Method
375	311
334	281
310	240
305	287
309	259
319	233
304	231
273	237
204	209
245	161
209	199

5-5.2.1 What Is the Best Line To Be Used for Estimating $\bar{Y}_X$ From Given Values of X?

Procedure

Using Basic Worksheet (see Worksheet 5-5.2), compute the line $Y = b_0 + b_1X$. This is an estimate of the true line $\bar{Y}_X = \beta_0 + \beta_1 X$.

Using Data Sample 5-5.2, the fitted line is

$$Y = 48.965 + 0.661873\, X.$$

The fitted line is shown in Figure 5-13, and the confidence band for the line (see the procedure of Paragraph 5-5.2.2.1) also is shown.

WORKSHEET 5-5.2

EXAMPLE OF SII RELATIONSHIP

X denotes Tread Life Estimated by Weight Method

Y denotes Tread Life Estimated by Center Groove Method

$\Sigma X =$	3187	$\Sigma Y =$	2648
$\bar{X} =$	289.727	$\bar{Y} =$	240.727

Number of points: $n =$ 11

Step (1)	$\Sigma XY =$	785369
(2)	$(\Sigma X)(\Sigma Y)/n =$	767197.818
(3)	$S_{xy} =$	18171.182

(4)	$\Sigma X^2 =$	950815	(7)	$\Sigma Y^2 =$	655754
(5)	$(\Sigma X)^2/n =$	923360.818	(8)	$(\Sigma Y)^2/n =$	637445.818
(6)	$S_{xx} =$	27454.182	(9)	$S_{yy} =$	18308.182
(10)	$b_1 = \dfrac{S_{xy}}{S_{xx}} =$	0.661873	(14)	$\dfrac{(S_{xy})^2}{S_{xx}} =$	12027.015
(11)	$\bar{Y} =$	240.727	(15)	$(n-2)\, s_Y^2 =$	6281.167
(12)	$b_1\bar{X} =$	191.762	(16)	$s_Y^2 =$	697.9074
(13)	$b_0 = \bar{Y} - b_1\bar{X} =$	48.965		$s_Y =$	26.418

Equation of the line:

$Y = b_0 + b_1 X$

$=$ 48.965 + 0.661873 X

$s_{b_1} =$ 0.159439

$s_{b_0} =$ 46.88

Estimated variance of the slope:

$$s_{b_1}^2 = \frac{s_Y^2}{S_{xx}} = .0254208$$

Estimated variance of intercept:

$$s_{b_0}^2 = s_Y^2 \left\{\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right\} = 2197.313$$

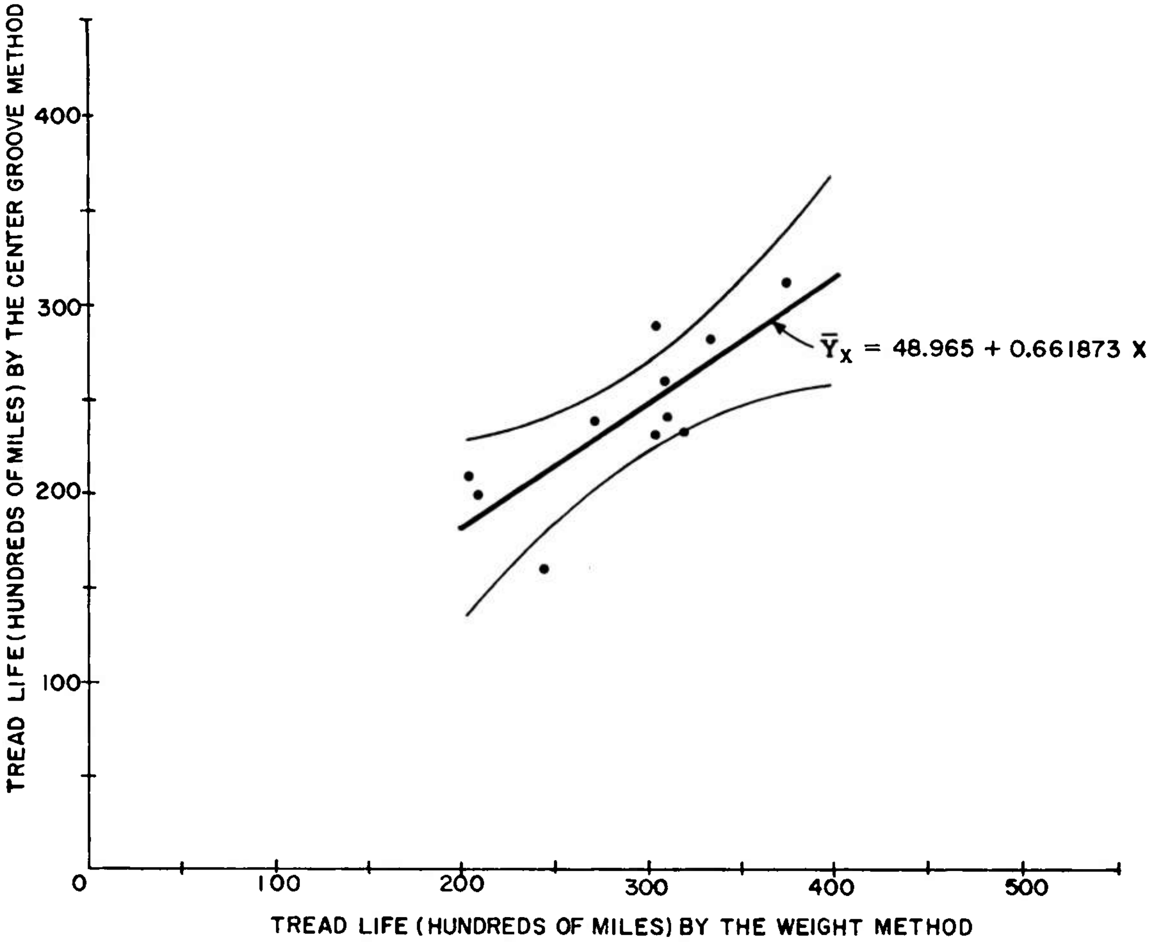

Figure 5-13. Relationship between weight method and center groove method when the range of the weight method has been restricted—an SII relationship.

5-5.2.2 What are the Confidence Interval Estimates for: the Line as a Whole; a Point on the Line; a Single Y Corresponding to a New Value of X?

Read the discussion of the interpretation of these three types of confidence intervals in Paragraph 5-4.1.2 in order to decide which is the appropriate kind of confidence interval.

5-5.2.2.1 What Is the $(1-\alpha)$ Confidence Band For the Line as a Whole?

The solution is identical to that of Procedure 5-4.1.2.1 and is illustrated here for Data Sample 5-5.2.

Procedure	Example
(1) Choose the desired confidence level, $1-\alpha$	(1) Let: $1-\alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Worksheet.	(2) From Worksheet 5-5.2 $s_Y = 26.418$
(3) Look up $F_{1-\alpha}$ for $(2, n-2)$ degrees of freedom in Table A-5.	(3) $n = 11$ $F_{.95}(2, 9) = 4.26$
(4) Choose a number of values of X (within the range of the data) at which to compute points for drawing the confidence band.	(4) Let: $X = 200$ $X = 250$ $X = 300$ $X = 350$ $X = 400$, for example.
(5) At each selected value of X, compute: $Y_c = \bar{Y} + b_1 (X - \bar{X})$ and $W_1 = \sqrt{2F}\, s_Y \left[\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$	(5) See Table 5-6 for a convenient computational arrangement and the example calculations.
(6) A $(1-\alpha)$ confidence band for the whole line is determined by $Y_c \pm W_1$.	(6) See Table 5-6.
(7) To draw the line and its confidence band, plot Y_c at two of the extreme selected values of X. Connect the two points by a straight line. At each selected value of X, also plot $Y_c + W_1$ and $Y_c - W_1$. Connect the upper series of points, and the lower series of points, by smooth curves.	(7) See Figure 5-13.

If more points are needed for drawing the curves for the band, note that, because of symmetry the calculation of W_1 at n values of X actually gives W_1 at $2n$ values of X.

For example: W_1 (but not Y_c) has the same value at $X = 250$ (i.e., $\bar{X} - 39.73$) as at $X = 329.5$ (i.e., $\bar{X} + 39.73$).

TABLE 5-6. COMPUTATIONAL ARRANGEMENT FOR PROCEDURE 5-5.2.2.1

X	$(X - \bar{X})$	Y_c	$\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}}$	$s^2_{Y_c}$	s_{Y_c}	W_1	$Y_c + W_1$	$Y_c - W_1$
200	−89.73	181.3	0.384179	268.12	16.37	47.8	229.1	133.5
250	−39.73	214.4	0.148404	103.57	10.18	29.7	244.1	184.7
300	+10.27	247.5	0.094751	66.127	8.132	23.7	271.2	223.8
350	60.27	280.6	0.223219	155.79	12.48	36.4	317.0	244.2
400	110.27	313.7	0.533810	372.55	19.30	56.3	370.0	257.4

$\bar{X} = 289.727$
$\bar{Y} = 240.727$

$s^2_Y = 697.9074$
$\frac{1}{n} = 0.0909091$
$b_1 = 0.661873$
$S_{xx} = 27454.182$

$Y_c = \bar{Y} + b_1(X - \bar{X})$
$s^2_{Y_c} = s^2_Y \left[\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}}\right]$
$\sqrt{2F} = \sqrt{8.52} = 2.919$
$W_1 = \sqrt{2F}\, s_{Y_c}$

5-5.2.2.2 Give a $(1 - \alpha)$ Confidence Interval For a Single Point On the Line, i.e., the Mean Value of Y Corresponding To a Chosen Value of X (X′).

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$	(1) Let: $1 - \alpha = .95$ $\alpha = .05$
(2) Obtain s_Y from Basic Worksheet.	(2) From Worksheet 5-5.2 $s_Y = 26.418$
(3) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.	(3) $n = 11$ $t_{.975}$ for 9 d.f. $= 2.262$
(4) Choose X', the value of X at which we want to make an interval estimate of the mean value of Y.	(4) Let $X' = 300$, for example.
(5) Compute: $W_2 = t_{1-\alpha/2}\, s_Y \left[\frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$ and $Y_c = \bar{Y} + b_1(X' - \bar{X})$	(5) $W_2 = (2.262)\,(26.418)\,(0.3078) = 18.4$ $Y_c = 247.5$
(6) A $(1 - \alpha)$ confidence interval estimate for the mean value of Y corresponding to $X = X'$ is given by $\bar{Y} + b_1(X - \bar{X}) \pm W_2 = Y_c \pm W_2$.	(6) A 95% confidence interval estimate for the mean value of Y at $X = 300$ is the interval 247.5 ± 18.4, i.e., the interval from 229.1 to 265.9 .

5-5.2.2.3 Give a $(1 - \alpha)$ Confidence Interval Estimate For a Single (Future) Value of Y Corresponding To a Chosen Value of $X = X'$.

Procedure

(1) Choose the desired confidence level, $1 - \alpha$

(2) Obtain s_Y from Worksheet.

(3) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.

(4) Choose X', the value of X at which we want to make an interval estimate of a single value of Y.

(5) Compute:

$$W_3 = t_{1-\alpha/2}\, s_Y \left[1 + \frac{1}{n} + \frac{(X' - \bar{X})^2}{S_{xx}}\right]^{\frac{1}{2}}$$

and

$$Y_c = \bar{Y} + b_1 (X' - \bar{X})$$

(6) A $(1 - \alpha)$ confidence interval estimate for Y' (the single value of Y corresponding to X') is given by

$$\bar{Y} + b_1 (X' - \bar{X}) \pm W_3 = Y_c \pm W_3 \ .$$

Example

(1) Let: $1 - \alpha = .95$
$\alpha = .05$

(2) From Worksheet 5-5.2
$s_Y = 26.418$

(3) $t_{.975}$ for 9 d.f. $= 2.262$

(4) Let $X' = 300$,
for example.

(5)

$$W_3 = (2.262)\ (26.418)\ (1.0463) = 62.5$$

$$Y_c = 247.5$$

(6) A 95% confidence interval estimate for Y at $X = 300$ is the interval 247.5 ± 62.5, i.e., the interval from 185.0 to 310.0 .

5-5.2.3 What Is the Confidence Interval Estimate for β_1, the Slope of the True Line, $\bar{Y}_X = \beta_0 + \beta_1 X$?

Procedure

(1) Choose the desired confidence level, $1 - \alpha$

(2) Look up $t_{1-\alpha/2}$ for $n - 2$ degrees of freedom in Table A-4.

(3) Obtain s_{b_1} from Worksheet.

(4) Compute

$$W_4 = t_{1-\alpha/2}\, s_{b_1}$$

(5) A $(1 - \alpha)$ confidence interval estimate for β_1 is

$$b_1 \pm W_4 \ .$$

Example

(1) Let: $1 - \alpha = .95$
$\alpha = .05$

(2) $n = 11$
$t_{.975}$ for 9 d.f. $= 2.262$

(3) From Worksheet 5-5.2
$s_{b_1} = 0.159439$

(4)

$$W_4 = 2.262\ (0.159439) = 0.360651$$

(5) $b_1 = 0.661873$
$W_4 = 0.360651$

A 95% confidence interval estimate for β_1 is the interval 0.661873 ± 0.360651, i.e., the interval from 0.301222 to 1.022524 .

REFERENCES

1. C. Eisenhart, "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," *Annals of Mathematical Statistics*, Vol. 10, No. 2, pp. 162-186, June, 1939.
2. M. Ezekiel, *Methods of Correlation Analysis* (2d edition), Chapter 20, John Wiley & Sons, Inc., New York, N.Y., 1941.
3. J. Mandel, "Fitting a Straight Line to Certain Types of Cumulative Data," *Journal of the American Statistical Association*, Vol. 52, pp. 552-566, 1957.
4. F. S. Acton, *Analysis of Straight-Line Data*, John Wiley & Sons, Inc., New York, N.Y., 1959.
5. M. S. Bartlett, "Fitting a Straight Line When Both Variables are Subject to Error," *Biometrics*, Vol. 5, No. 3, pp. 207-212, 1949.
6. J. Berkson, "Are There Two Regressions?", *Journal of the American Statistical Association*, Vol. 45, pp. 164-180, 1950.
7. R. C. Geary, "Non-linear Functional Relationships Between Two Variables When One is Controlled," *Journal of the American Statistical Association*, Vol. 48, pp. 94-103, 1953.
8. J. Lipka, *Graphical and Mechanical Computation*, John Wiley & Sons, Inc., New York, N.Y., 1918.
9. H. L. Rietz, (ed.), *Handbook of Mathematical Statistics*, Houghton Mifflin Company, Boston, Mass., 1924.
10. J. B. Scarborough, *Numerical Mathematical Analysis* (2d edition), The Johns Hopkins Press, Baltimore, Md., 1950.
11. R. D. Stiehler, G. G. Richey, and J. Mandel, "Measurement of Treadwear of Commercial Tires," *Rubber Age*, Vol. 73, No. 2, May, 1953.

CHAPTER 6

POLYNOMIAL AND MULTIVARIABLE RELATIONSHIPS ANALYSIS BY THE METHOD OF LEAST SQUARES

6-1 INTRODUCTION

In this Chapter, we give methods for estimating the coefficients of, and for answering various questions about, multivariable functional relationships of the form

$$y = \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1} \tag{6-1}$$

between a *dependent variable* y and a number of *independent variables* $x_0, x_1, \ldots, x_{k-1}$. We restrict our discussion, however, to the case in which the values of the independent variables $x_0, x_1, \ldots, x_{k-1}$, are known exactly, and errors of measurement affect only the *observed* values Y of y, that is, to many-variable analogs of the FI functional relationships considered in Paragraphs 5-3.1 and 5-4.1.

Methods for the analysis of many-variable relationships in which errors of measurement affect the values of the x's involved as well as the *observed* Y's, i.e., the multivariable analogs of the FII structural relationships considered in Paragraphs 5-3.1 and 5-4.3, are not discussed *per se* in this Chapter. If, however, the *errors* that affect the x's are not errors of measurement, but rather are errors of control in the sense of Paragraph 5-4.3.2, i.e., are errors made in attempting to set $X_0, X_1, \ldots, X_{k-1}$, equal to their respective *nominal values* $x_0', x_1', \ldots, x_{k-1}'$, then the methods of this Chapter are applicable, *provided that* the errors made in adjusting $X_0, X_1, \ldots, X_{k-1}$, to their respective nominal values are mutually independent (or, at least, are uncorrelated).

The techniques presented in this Chapter are general. They are applicable whenever we know the functional form of the relation between y and the x's, and are primarily concerned with estimating the unknown values of the coefficients of the respective terms of the relationship. Thus, taking $x_0 = 1$, $x_1 = x$, $x_2 = x^2, \ldots, x_m = x^m$, the methods of this Chapter enable us to estimate the coefficients of, and to answer various questions about, an mth degree polynomial relationship

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m \tag{6-2}$$

between a *dependent variable* y and a single *independent variable* x. Alternatively, taking $x_0 = 1$, $x_1 = x$, $x_2 = z$, $x_3 = x^2$, $x_4 = xz$, and $x_5 = z^2$, the techniques of this Chapter can be used to investigate the nature of a quadratic *surface* relationship

$$y = \beta_0 + (\beta_1 x + \beta_2 z) + (\beta_3 x^2 + \beta_4 xz + \beta_5 z^2) \tag{6-3}$$

between a *dependent variable* y and two *independent variables* x and z. For example, we may wish to test the hypothesis that the *surface* actually is a plane, i.e., that β_3, β_4, and β_5, in Equation (6-3) are equal to zero, and so forth.

Multivariate* statistical relationships analogous to the SI and SII situations considered in Paragraphs 5-3.2 and 5-5.1 are not considered *per se* in this Chapter. If, however, Y and X_1, $X_2, \ldots, X_{k-1}$, have a joint multivariate frequency (probability) distribution in some definite population, and if a sample of size n is drawn from this population, with or without selection or restrictions on the values of the X's but *without selection or restriction on the Y's*, then the methods of this Chapter, taking $X_0 \equiv 1$ throughout, are directly applicable to estimating the coefficients of, and to answering various questions about, the *multivariate regression* of Y on $X_1, X_2, \ldots,$ and X_{k-1}, namely,

$$\bar{Y}_{\{x\}} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-1} X_{k-1}, \qquad (6\text{-}4)$$

where $\bar{Y}_{\{x\}}$ is *shorthand* for $m_{Y \cdot X_1 X_2 \ldots X_{k-1}}$, the mean value of all of the Y's that are associated in the population with the particular indicated combination $X_1, X_2 \ldots X_{k-1}$, of values of the X's (see footnote of Par. 5-3.2)—and, where

$$\beta_0 = m_Y - \beta_1 m_{X_1} - \beta_2 m_{X_2} - \ldots - \beta_{k-1} m_{X_{k-1}}, \qquad (6\text{-}5)$$

$m_y, m_{x_1}, \ldots, m_{x_{k-1}}$, are the population means of $Y_1, X_1, \ldots, X_{k-1}$, respectively. The *fitted regression*, yielded by the application of the methods of this Chapter to observational data of this kind, will be of the form

$$\bar{Y}_{\{x\}} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_{k-1} X_{k-1} \qquad (6\text{-}6)$$

with $b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \ldots - b_{k-1} \bar{X}_{k-1}$ (6-7)

where $\bar{Y}, \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_{k-1}$, are the means of $Y, X_1, X_2, \ldots,$ and X_{k-1}, *in the sample;* and each b will be a *best* (i.e., minimum variance unbiased) estimate of the corresponding *true* β.

When, as in all of the previously mentioned situations, the relationship between y and the x's is *linear in the coefficients* whose values are to be determined from the data in hand, the Method of Least Squares is the most generally accepted procedure for estimating the unknown values of the coefficients, and for answering questions about the relationship as a whole. A widely applicable Least Squares Theorem is given in Paragraph 6-2; and its application to a general linear situation is presented in detail in Paragraph 6-3, with worked examples. Special applications to polynomial and other situations are discussed in subsequent paragraphs of this Chapter.

The numerical calculations required for least-squares analysis of multivariable relationships often are lengthy and tedious. Hence, this Chapter is directed toward arrangement of the work for automatic computation on modern electronic computers. Consequently, basic equations called for in the calculations are written both in traditional and in matrix forms. This Chapter concludes with a discussion of matrix operations that are useful both in formulating and in carrying out the requisite calculations, Paragraph 6-9.

In most instances, related Procedures and Examples appear on facing pages in this Chapter.

* The important distinction in statistical work between a *variable* and a *variate* is drawn in the Kendall-Buckland *Dictionary of Statistical Terms*[1] as follows:

Variable—Generally, any quantity which varies. More precisely, a variable in the mathematical sense, i.e., a quantity which may take any one of a specified set of values. It is convenient to apply the same word to denote non-measurable characteristics, e.g., "sex" is a variable in this sense, since any human individual may take one of two "values", male or female.

It is useful, but far from being the general practice, to distinguish between a variable as so defined and a random variable or variate (q.v.).

Variate—In contradistinction to a variable (q.v.) a variate is a quantity which may take any of the values of a specified set with a specified relative frequency or probability. The variate is therefore often known as a random variable. It is to be regarded as defined, not merely by a set of permissible values like an ordinary mathematical variable, but by an associated frequency (probability) function expressing how often those values appear in the situation under discussion.

6-2 LEAST SQUARES THEOREM

If the n measurements $Y_1, Y_2, \ldots, Y_n$ are statistically independent with common variance σ^2 and have expected values $E(Y_i)$,

$$\begin{aligned} E(Y_1) &= \beta_0 x_{01} + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_{k-1} x_{k-1,1} \\ E(Y_2) &= \beta_0 x_{02} + \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_{k-1} x_{k-1,2} \\ &\ldots \\ E(Y_n) &= \beta_0 x_{0n} + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots + \beta_{k-1} x_{k-1,n} \end{aligned} \tag{6-8}$$

then the best linear unbiased estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{k-1}$, of the unknown coefficients are given by the solution of k simultaneous equations, called the *normal equations*,

$$\begin{aligned} \hat{\beta}_0 \Sigma x_0^2 + \hat{\beta}_1 \Sigma x_0 x_1 + \ldots + \hat{\beta}_{k-1} \Sigma x_0 x_{k-1} &= \Sigma x_0 Y \\ \hat{\beta}_0 \Sigma x_1 x_0 + \hat{\beta}_1 \Sigma x_1^2 + \ldots + \hat{\beta}_{k-1} \Sigma x_1 x_{k-1} &= \Sigma x_1 Y \\ &\ldots \\ \hat{\beta}_0 \Sigma x_{k-1} x_0 + \hat{\beta}_1 \Sigma x_{k-1} x_1 + \ldots + \hat{\beta}_{k-1} \Sigma x_{k-1}^2 &= \Sigma x_{k-1} Y \end{aligned} \tag{6-9}$$

where the summation is over all of the n values of the variables involved; e.g.,

$$\Sigma x_1 x_2 = \sum_{i=1}^{n} x_{1i} x_{2i},$$

and the estimate of σ^2 is given by

$$\begin{aligned} s^2 &= \frac{1}{n-k} \sum_1^n [Y_i - (\hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_{k-1} x_{k-1,i})]^2 \\ &= \frac{1}{n-k} \left\{ \sum_1^n Y_i^2 - \sum_1^k \hat{\beta}_j (\Sigma x_j Y) \right\}. \end{aligned} \tag{6-10}$$

If no unique solution to Equation (6-9) exists (which will occur when one or more of the x's are linearly dependent, for example, if $x_1 = ax_2 + bx_3$), then not all k coefficients can be estimated from the data. Variables may be deleted or several variables may be replaced by a linear function of those variables so that a solvable system involving fewer equations results.

In situations where the variance of the Y's is not the same for all Y's and/or there is correlation among the Y's, a transformation of variables is required. The methods for these cases are discussed later in this Chapter.

This theorem can be restated using matrix notation as follows:

Let, $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_n \end{bmatrix}$, $X = \begin{bmatrix} x_{01} & x_{11} & \ldots & x_{k-1,1} \\ x_{02} & x_{12} & \ldots & x_{k-1,2} \\ \cdot & \cdot & & \cdot \\ x_{0n} & x_{1n} & \ldots & x_{k-1,n} \end{bmatrix}$, and $\beta = \begin{bmatrix} \beta_0 \\ \beta_2 \\ \cdot \\ \beta_{k-1} \end{bmatrix}$

The expected values of the Y's then is expressed as

$$E(Y) = X\beta, \tag{6-8M}$$

and the condition of independence and common variance is expressed by

$$\text{Var}(Y) = V = \sigma^2 I.$$

Under these conditions, the minimum variance unbiased estimates $\hat{\beta}$ of β are given by the solution of the normal equations

$$X'X\hat{\beta} = X'Y. \tag{6-9M}$$

The estimate of σ^2 is given by

$$s^2 = \frac{1}{n-k}\{(Y - X\hat{\beta})'(Y - X\hat{\beta})\}$$
$$= \frac{1}{n-k}(Y'Y - \hat{\beta}'X'Y). \tag{6-10M}$$

Equations (6-8), (6-9), and (6-10) are given in the usual algebraic notation, and the corresponding equations in matrix notation are (6-8M), (6-9M), and (6-10M).

6-3 MULTIVARIABLE FUNCTIONAL RELATIONSHIPS

6-3.1 USE AND ASSUMPTIONS

Least-squares methods for estimating the coefficients of a functional relation of the form

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k-1} x_{k-1} \tag{6-1}$$

are used in a number of situations:

(a) when it is known from theoretical considerations in the subject matter field that the relationship *is* of this form;

(b) when the exact expression relating y and the x's either is unknown or is too complicated to be used directly and it is assumed that an approximation of this type will be satisfactory.

In the latter case, the approximation often can be justified on the grounds that, for the limited range of the x's considered, the surface representing y as a function of the x's is very nearly the hyperplane given by Equation (6-1). The method is strictly valid in (a), but in (b) there is danger of obtaining misleading results, analogous to the bias arising in the straight-line case from the assumption that the functional relation involved is *linear* when in fact it is not linear.

In addition to the validity of Equation (6-1), the following assumptions must be satisfied:*

(a) the random errors in the Y's have mean zero and a common variance σ^2;

(b) the random errors in the Y's are mutually independent in the statistical sense.

For strict validity of the usual tests of significance, and confidence interval estimation procedures in Paragraph 6-3.3 (Steps 8 and 9), an additional assumption must be satisfied:

(c) the random errors affecting the Y's are normally distributed.

The x variables may be powers or other functions of some basic variables, and several different functions of the same x variable may be used. (See, for example, Equation (6-2) or (6-3)).

The data for analysis consist of the n *points* $(x_{01}, x_{11}, \ldots, x_{k-1,1}, Y_1)$ $(x_{02}, x_{12}, \ldots, x_{k-1,2}\ Y_2) \ldots,$ $(x_{0n}, x_{1n}, \ldots, x_{k-1,n}, Y_n)$, and usually are represented in tabular form as:

X_0	X_1	X_2	. . .	X_{k-1}	Y
x_{01}	x_{11}	x_{21}		$x_{k-1,1}$	Y_1
x_{02}	x_{12}	x_{22}		$x_{k-1,2}$	Y_2
.				.	
.				.	
.				.	
x_{0n}	x_{1n}	x_{2n}	. . .	$x_{k-1,n}$	Y_n

* When these assumptions are not satisfied, see Paragraph 6-6 for the case of inequality of variance, and Paragraph 6-7 for the case of correlation among the Y's.

Alternatively, the data may be expressed in the form of *observational equations*,

$$\begin{aligned}
\beta_0 x_{01} + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_{k-1} x_{k-1,1} &= Y_1 = y_1 + e_1 \\
\beta_0 x_{02} + \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_{k-1} x_{k-1,2} &= Y_2 = y_2 + e_2 \\
&\;\;\vdots \\
\beta_0 x_{0n} + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots + \beta_{k-1} x_{k-1,n} &= Y_n = y_n + e_n
\end{aligned} \qquad (6\text{-}11)$$

where $e_1, e_2, \ldots, e_n$ denote the *errors* of the Y's as measured values of the corresponding true y's. When the number of observational equations exceeds the number of unknown coefficients, i.e., when $n > k$, the observational equations ordinarily are mutually contradictory; that is, the values of $\beta_0, \beta_1, \ldots,$ and β_{k-1} implied by any chosen solvable selection of k of the equations do not satisfy one or more of the remaining $n - k$ equations. Hence, there is a need for *best* estimates of the β's based on the data as a whole.

For a unique least-squares solution, n must not be less than k, and the normal equations (6-9) must be uniquely solvable. If not, some variables must be deleted or suitably combined with other variables.

6-3.2 DISCUSSION OF PROCEDURES AND EXAMPLES

In setting forth the steps in the solution, the formulas are given in the usual algebraic notation and also in matrix notation where appropriate.

Data Sample 6-3.2, selected for arithmetical simplicity, serves to illustrate the worked examples of numerical procedures involved in estimating the coefficients of, and in answering various questions about, multivariable functional relationships.

Data Sample 6-3.2

x_1	x_2	x_3	Y
1	8	1	2
2	8	7	4
2	6	0	4
3	1	2	4
4	2	7	3
4	5	1	3

We assume that these data correspond to a situation in which the functional dependence of y on x_1, x_2, and x_3, is of the form

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \qquad (6\text{-}12)$$

which is a special case of Equation (6-1) with the term $\beta_0 x_0$ omitted; i.e., with β_0 taken equal to zero. Equation (6-12) implies that the functional dependence of y on x_1, x_2, and x_3, takes the form of a hyperplane* that passes through the origin (0, 0, 0, 0) of the four-dimensional Euclidean *space*

* A flat surface in four or more dimensions is termed a *hyperplane* when it is the locus of points that vary in more than two dimensions.

whose coordinates are x_1, x_2, x_3, and y. If we wished to allow for the possibility that the dependence of y on x_1, x_2, and x_3, may take the form of a hyperplane that intersects the y-axis at some point $(0, 0, 0, \beta_0)$, not necessarily the origin, then we would substitute

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

for Equation (6-12), and take $x_0 \equiv 1$; i.e., amend Data Sample 6-3.2 by adding an x_0 column of 1's.

By analogy with Equations (6-11), Data Sample 6-3.2 and the assumed functional relationship Equation (6-12) can be summarized *symbolically* by *observational equations* of the form

$$\begin{aligned} \beta_1 x_{11} + \beta_2 x_{21} + \beta_3 x_{31} &= Y_1 \\ \beta_1 x_{12} + \beta_2 x_{22} + \beta_3 x_{32} &= Y_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ \beta_1 x_{16} + \beta_2 x_{26} + \beta_3 x_{36} &= Y_6 \end{aligned} \tag{6-13}$$

Substitution of the values of the x's and Y's of Data Sample 6-3.2 in Equation (6-13) gives

$$\begin{aligned} \beta_1 \cdot 1 + \beta_2 \cdot 8 + \beta_3 \cdot 1 &= 2 \\ \beta_1 \cdot 2 + \beta_2 \cdot 8 + \beta_3 \cdot 7 &= 4 \\ \beta_1 \cdot 2 + \beta_2 \cdot 6 + \beta_3 \cdot 0 &= 4 \\ \beta_1 \cdot 3 + \beta_2 \cdot 1 + \beta_3 \cdot 2 &= 4 \\ \beta_1 \cdot 4 + \beta_2 \cdot 2 + \beta_3 \cdot 7 &= 3 \\ \beta_1 \cdot 4 + \beta_2 \cdot 5 + \beta_3 \cdot 1 &= 3 \end{aligned} \tag{6-14}$$

as the observational equations corresponding to Data Sample 6-3.2.

6-3.3 PROCEDURES AND EXAMPLES

Step 1 Procedure—Formation of Normal Equations. The *normal equations* are formed from the sums of squares and cross products as follows:

$$\begin{aligned} \beta_0 \Sigma x_0^2 + \beta_1 \Sigma x_0 x_1 + \ldots + \beta_{k-1} \Sigma x_0 x_{k-1} &= \Sigma x_0 Y \\ \beta_0 \Sigma x_1 x_0 + \beta_1 \Sigma x_1^2 + \ldots + \beta_{k-1} \Sigma x_1 x_{k-1} &= \Sigma x_1 Y \\ \ldots & \\ \beta_0 \Sigma x_{k-1} x_0 + \beta_1 \Sigma x_{k-1} x_1 + \ldots + \beta_{k-1} \Sigma x_{k-1}^2 &= \Sigma x_{k-1} Y \end{aligned} \tag{6-9}$$

or in matrix form

$$X'X\hat{\beta} = X'Y = Q \tag{6-9M}$$

where $Q' = (q_1, q_2, \ldots, q_r)$,

and $q_j = \sum_{i=1}^{n} x_{ji} Y_i, \ (j = 0, 1, \ldots, k-1).$

Step 1 Example—Formation of Normal Equations. The *normal equations* (See Equations (6-9)) corresponding to the observational equations (6-13) are

$$\begin{aligned} \beta_1\, \Sigma x_1^2 + \beta_2\, \Sigma x_1 x_2 + \beta_3\, \Sigma x_1 x_3 &= \Sigma x_1 Y \\ \beta_1\, \Sigma x_1 x_2 + \beta_2\, \Sigma x_2^2 + \beta_3\, \Sigma x_2 x_3 &= \Sigma x_2 Y \\ \beta_1\, \Sigma x_1 x_3 + \beta_2\, \Sigma x_2 x_3 + \beta_3\, \Sigma x_3^2 &= \Sigma x_3 Y \end{aligned} \tag{6-15}$$

or in matrix form

$$X'\, X\, \beta = X' Y \tag{6-15M}$$

where $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$.

Numerical evaluation of the requisite sums of squares and sums of cross products for Data Sample 6-3.2 and substitution in Equation (6-15), yields

$$\begin{aligned} \beta_1 \cdot 50 + \beta_2 \cdot 67 + \beta_3 \cdot 53 &= 54 \\ \beta_1 \cdot 67 + \beta_2 \cdot 194 + \beta_3 \cdot 85 &= 97 \\ \beta_1 \cdot 53 + \beta_2 \cdot 85 + \beta_3 \cdot 104 &= 62 \end{aligned} \tag{6-16}$$

and the matrices involved in Equation (6-15M) become

$$(X'X) = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 & \Sigma x_1 x_3 \\ \Sigma x_1 x_2 & \Sigma x_2^2 & \Sigma x_2 x_3 \\ \Sigma x_1 x_3 & \Sigma x_2 x_3 & \Sigma x_3^2 \end{bmatrix} = \begin{bmatrix} 50 & 67 & 53 \\ 67 & 194 & 85 \\ 53 & 85 & 104 \end{bmatrix}$$

$$(X'Y) = \begin{bmatrix} \Sigma x_1 Y \\ \Sigma x_2 Y \\ \Sigma x_3 Y \end{bmatrix} = \begin{bmatrix} 54 \\ 97 \\ 62 \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = Q \tag{6-16M}$$

Step 2 Procedure—Solution of Normal Equations. Equations (6-9) can be solved by a number of methods giving values for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots$, which can be expressed as

$$\begin{aligned} \hat{\beta}_0 &= c_{00}q_0 + c_{01}q_1 + \ldots + c_{0,k-1}q_{k-1} \\ \hat{\beta}_1 &= c_{10}q_0 + c_{11}q_1 + \ldots + c_{1,k-1}q_{k-1} \\ &\ldots \\ \hat{\beta}_{k-1} &= c_{k-1,0}q_0 + c_{k-1,1}q_1 + \ldots + c_{k-1,k-1}q_{k-1}. \end{aligned} \tag{6-17}$$

A solution for the $\hat{\beta}_i$'s can be arrived at without explicitly computing the c_{ij}'s, of course, but in the following computations the c_{ij}'s are needed. The values of the c_{ij}'s depend only on the sums of squares and cross products of the *independent* variables $x_0, x_1, \ldots, x_k$, so that the estimates of the β_i's can be expressed as a linear function of the Y's.

In matrix notation, this step is given by computing the inverse of the matrix of normal equations, i.e.,

$$(X'X)^{-1} = \begin{bmatrix} c_{00} & c_{01} & \ldots & c_{0,k-1} \\ c_{10} & c_{11} & \ldots & c_{1,k-1} \\ \ldots & & & \\ c_{k-1,0} & c_{k-1,1} & \ldots & c_{k-1,k-1} \end{bmatrix}$$

and Equations (6-17) become

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}Q. \end{aligned} \tag{6-17M}$$

Step 2 Example—Solution of Normal Equations. The values $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, that constitute the solutions of the normal equations can be expressed (See Equations (6-17)) in the form

$$\begin{aligned}
\hat{\beta}_1 &= c_{11}\,q_1 + c_{12}\,q_2 + c_{13}\,q_3 \\
&= c_{11}\cdot 54 + c_{12}\cdot 97 + c_{13}\cdot 62 \\
\hat{\beta}_2 &= c_{21}\,q_1 + c_{22}\,q_2 + c_{23}\,q_3 \\
&= c_{21}\cdot 54 + c_{22}\cdot 97 + c_{23}\cdot 62 \\
\hat{\beta}_3 &= c_{31}\,q_1 + c_{32}\,q_2 + c_{33}\,q_3 \\
&= c_{31}\cdot 54 + c_{32}\cdot 97 + c_{33}\cdot 62
\end{aligned} \tag{6-18}$$

where the c's are the elements of the inverse matrix

$$(X'X)^{-1} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

$(X'X)^{-1}$ may be computed in many ways.* The exact inverse of the matrix $(X'X)$ determined by the first equation of Equations (6-16M) is

$$(X'X)^{-1} = \frac{1}{239418}\begin{bmatrix} 12951 & -2463 & -4587 \\ -2463 & 2391 & -699 \\ -4587 & -699 & 5211 \end{bmatrix} \tag{6-18M}$$

where the factor in front of the matrix is to be applied to the individual terms in the matrix.

Using the first equation of Equations (6-18), we get

$$\begin{aligned}
\hat{\beta}_1 &= \frac{1}{239418}\{(12951)\,(54) + (-2463)\,(97) + (-4587)\,(62)\} \\
&= \frac{1}{239418}\{699354 - 238911 - 284394\} \\
&= \frac{176049}{239418} \\
&= 0.735\ 320\ 652.
\end{aligned}$$

The other coefficients are obtained similarly:

$$\begin{aligned}
\hat{\beta}_2 &= 0.232\ 175\ 526 \\
\hat{\beta}_3 &= 0.031\ 664\ 286.
\end{aligned}$$

The prediction equation, therefore, is

$$\hat{Y} = 0.735\ 320\ 652\ x_1 + 0.232\ 175\ 526\ x_2 + 0.031\ 664\ 286\ x_3.$$

* The advent of automatic electronic digital computers has reduced the inversion of matrices of even moderate size to a matter of seconds. Routines for matrix inversion are standard tools of automatic computation. In contrast, matrix inversion by desk calculators is a time-consuming and tedious affair. Detailed illustration at this juncture of any one of the common methods of matrix inversion by desk calculator would not only constitute a distractive interruption to the orderly presentation of the essential features of this Chapter, but would lengthen it considerably. The two most common methods of matrix inversion by desk calculator—the *Doolittle method*, and the *abbreviated Doolittle method* (also called the *Gauss-Doolittle method*)—are described and illustrated by numerical examples in various statistical textbooks, e.g., in Chapter 15 of Anderson and Bancroft.[2] Details of the square-root method, favored by some computers, are given, with a numerical illustration, in Appendix 11A of O. L. Davies' book.[3] All of the common methods of matrix inversion by desk calculators are described in considerable detail, illustrated by numerical examples, and compared with respect to advantages and disadvantages in a paper by L. Fox, *Practical Solution of Linear Equations and Inversion of Matrices*, included in Taussky.[4] Reference also may be made to the book of Dwyer.[6] The reader of this Handbook who is faced with matrix inversion by desk calculator is referred to these standard sources for guidance and details.

Step 3 Procedure—Calculation of Deviation Between Predicted and Observed Value of the Y's. The predicted value $\hat{Y}_i$ at a given point $(x_{0i}, x_{1i}, \ldots, x_{k-1,i}, Y_i)$ is given by substituting the values of x in the prediction equation, i.e.,

$$\hat{Y}_i = \hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_{k-1} x_{k-1,i},$$

and the *residuals* $r_i = Y_i - \hat{Y}_i$ are given by

$$\begin{aligned} r_1 &= Y_1 - \hat{Y}_1 = Y_1 - (\hat{\beta}_0 x_{01} + \hat{\beta}_1 x_{11} + \ldots + \hat{\beta}_{k-1} x_{k-1,1}) \\ r_2 &= Y_2 - \hat{Y}_2 = Y_2 - (\hat{\beta}_0 x_{02} + \hat{\beta}_1 x_{12} + \ldots + \hat{\beta}_{k-1} x_{k-1,2}) \\ &\cdot \\ &\cdot \\ &\cdot \\ r_n &= Y_n - \hat{Y}_n = Y_n - (\hat{\beta}_0 x_{0n} + \hat{\beta}_1 x_{1n} + \ldots + \hat{\beta}_{k-1} x_{k-1,n}) \end{aligned} \tag{6-19}$$

or in matrix notation

$$r = Y - X\hat{\beta} \tag{6-19M}$$

where

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_n \end{bmatrix}.$$

In classical least-squares analysis, $\hat{Y}_i$ is termed the *adjusted value* of the observed value Y_i. It is important to distinguish between the *errors* of the Y_i with respect to the corresponding *true* values y_i, and the *residuals* of the Y_i with respect to their *adjusted* or *predicted* values $\hat{Y}_i$; that is, between the e_i of Equations (6-11) and the r_i of Equations (6-19).

Step 4 Procedure—Estimation of σ^2. The estimate s^2 of σ^2 is computed from

$$\begin{aligned} s^2 &= \frac{1}{n-k} \Sigma r^2 \\ &= \frac{1}{n-k} \left\{ \Sigma Y^2 - \sum_0^{k-1} \hat{\beta}_j q_j \right\} \end{aligned} \tag{6-20}$$

or in matrix notation

$$\begin{aligned} s^2 &= \frac{1}{n-k} (r'r) \\ &= \frac{1}{n-k} \{Y'Y - \beta' Q\}. \end{aligned} \tag{6-20M}$$

Step 3 Example—Calculation of Deviation Between Predicted and Observed Value of the Y's. The *predicted* or *adjusted* values $\hat{Y}_i$ corresponding to the observations Y_i are obtained by substituting the values of the x's into the prediction equation. For the first observation, substituting $x_{11} = 1$, $x_{21} = 8$, $x_{31} = 1$ leads to

$$\begin{aligned}\hat{Y}_1 &= 0.735\ 320\ 652\ (1) + 0.232\ 175\ 526\ (8) + 0.031\ 664\ 286\ (1)\\ &= 2.624\ 389\ 146.\end{aligned}$$

The corresponding residual is

$$\begin{aligned}r_1 &= Y_1 - \hat{Y}_1\\ &= 2 - 2.624\ 389\ 146\\ &= -.624\ 389\ 146.\end{aligned}$$

The full data, the corresponding predicted values ($\hat{Y}_i$) and their residuals (r_i), are:

i	x_{1i}	x_{2i}	x_{3i}	Y_i	$\hat{Y}_i$	Residuals r_i
1	1	8	1	2	2.624 389 146	−.624 389 146
2	2	8	7	4	3.549 695 514	.450 304 486
3	2	6	0	4	2.863 694 460	1.136 305 540
4	3	1	2	4	2.501 466 054	1.498 533 946
5	4	2	7	3	3.627 283 662	−.627 283 662
6	4	5	1	3	4.133 824 524	−1.133 824 524

Step 4 Example—Estimation of σ^2. The estimate s^2 of σ^2 may be computed directly from the sum of squared residuals. Thus,

$$\begin{aligned}s^2 &= \frac{1}{n-k}\,\Sigma r^2\\ &= \frac{1}{3}\,(5.808\ 473\ 047)\\ &= 1.936\ 157\ 682\end{aligned}$$

where n is the number of observational points (here 6) and k is the number of coefficients estimated from the data (here 3). Alternatively, s^2 may be evaluated from

$$\begin{aligned}s^2 &= \frac{1}{n-k}\left\{\Sigma Y^2 - \sum_1^3 \hat{\beta}_j\, q_j\right\}\\ &= \frac{1}{6-3}\,\{70 - (0.735\ 320\ 652)\ (54) - (0.232\ 175\ 526)\ (97) - (0.031\ 664\ 286)\ (62)\}\\ &= \frac{1}{3}\,(5.808\ 473\ 038)\\ &= 1.936\ 157\ 679.\end{aligned}$$

Extracting the square root gives

$$s = 1.391\ 4588.$$

Step 5 Procedure—Estimated Standard Deviations of the Coefficients. The estimated standard deviation of $\hat{\beta}_i$ is given by $s\sqrt{c_{ii}}$, where c_{ii} is the ith diagonal term of the inverse of the matrix of normal equations

$$
\begin{aligned}
&\text{est. s.d. of } \hat{\beta}_0 = s\sqrt{c_{00}} \\
&\text{est. s.d. of } \hat{\beta}_1 = s\sqrt{c_{11}} \\
&\text{est. s.d. of } \hat{\beta}_2 = s\sqrt{c_{22}} \\
&\quad\cdot \\
&\quad\cdot \\
&\quad\cdot \\
&\text{est. s.d. of } \hat{\beta}_{k-1} = s\sqrt{c_{k-1,k-1}}.
\end{aligned}
\tag{6-21}
$$

Step 6 Procedure—Standard Deviation of a Linear Function of the $\hat{\beta}$'s. The standard deviation of $\hat{L} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \ldots + a_{k-1}\hat{\beta}_{k-1}$ is estimated by

$$\text{est. s.d. of } \hat{L} = s\sqrt{\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} a_i a_j c_{ij}} \tag{6-22}$$

or in matrix form

$$\text{est. s.d. of } \hat{L} = s\sqrt{l'(X'X)^{-1}l} \tag{6-22M}$$

where $\qquad l' = (a_0, a_1, \ldots, a_{k-1}).$

Cases of special interest are:

(a) estimate of a single coefficient, i.e., $\hat{L} = \hat{\beta}_i$, in which case Equation (6-22) reduces to Equation (6-21);

(b) estimate of the difference of two coefficients, i.e., $\hat{L} = \hat{\beta}_i - \hat{\beta}_j$, in which case Equation (6-22) becomes

$$\text{est. s.d. of } (\hat{\beta}_i - \hat{\beta}_j) = s\sqrt{c_{ii} + c_{jj} - 2c_{ij}}. \tag{6-23}$$

Step 7 Procedure—Standard Deviation of a Predicted Point. Using the results of Step 6, the predicted *yield* $\hat{Y}_h$, at any chosen point $(x_{0h}, x_{1h}, \ldots, x_{k-1,h})$, is given by

$$\hat{Y}_h = \hat{\beta}_0 x_{0h} + \hat{\beta}_1 x_{1h} + \ldots + \hat{\beta}_{k-1,} x_{k-1,h}$$

which is a linear function of the $\hat{\beta}$'s. Application of Equation (6-22) leads to

$$\text{est. s.d. of } \hat{Y}_h = s\sqrt{\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} x_{ih} x_{jh} c_{ij}} \tag{6-24}$$

or in matrix notation

$$\text{est. s.d. of } \hat{Y}_h = s\sqrt{l'(X'X)^{-1}l} \tag{6-24M}$$

where $\qquad l' = (x_{0h}, x_{1h}, \ldots, x_{k-1,h})$

Step 5 Example—Estimated Standard Deviations of the Coefficients. The values of the estimated standard deviations of the $\hat{\beta}$'s are:

Coefficient	$\sqrt{c_{ii}}$	Estimated Standard Error of Coefficient, $s\sqrt{c_{ii}}$
$\hat{\beta}_1$	.232 581	.323 627
$\hat{\beta}_2$	.099 934	.139 054
$\hat{\beta}_3$	.147 531	.205 283

Step 6 Example—Standard Deviation of a Linear Function of the Coefficients. For illustrative purposes, consider $\hat{L} = \hat{\beta}_2 - 10\,\hat{\beta}_3$.

By Equation (6-22), or Equations (6-22M),

$$\text{est. s.d. of } \hat{L} = s\sqrt{c_{22}} + 100\,c_{33} - 20\,c_{23}$$

in matrix notation

$$\text{est. s.d. of } \hat{L} = s\sqrt{l'(X'X)^{-1}l}$$

with $\qquad l' = (0, 1, -10)$.

Numerical evaluation in this instance gives

$$\text{est. s.d. of } \hat{L} = 1.391\ 4588 \left(\frac{537471}{239418}\right)^{\frac{1}{2}}$$
$$= 2.0848.$$

Step 7 Example—Standard Deviation of a Predicted Point. By Equation (6-24), or Equation (6-24M), the estimated standard deviation of the predicted *yield* $\hat{Y}_h$, corresponding to any chosen point (x_{1h}, x_{2h}, x_{3h}), is given by

$$\text{est. s.d. of } \hat{Y}_h = s\sqrt{\sum_{i=1}^{3}\sum_{j=1}^{3} x_{ih}x_{jh}c_{ij}}$$

or in matrix notation

$$\hat{Y}_h = s\sqrt{l'(X'X)^{-1}l}$$

where $\qquad l' = (x_{1h}, x_{2h}, x_{3h})$.

Thus, the estimated standard error of $\hat{Y}_1$, the predicted or adjusted *yield* corresponding to the first observational point (1, 8, 1), is

$$\text{est. s.d. of } \hat{Y}_1 = s[c_{11} + 8c_{12} + c_{13} + 8c_{21} + 64c_{22} + 8c_{23} + c_{31} + 8c_{32} + c_{33}]^{\frac{1}{2}}$$
$$= 1.391\ 4588 \left(\frac{111420}{239418}\right)^{\frac{1}{2}}$$
$$= 0.949\ 235.$$

Step 8 Procedure—Analysis of Variance Test of Significance of a Group of $p < k$ of the Coefficients. To test the statistical significance of a set of p of the β's (for simplicity the last p), start with a reduced set of normal equations, omitting the last p rows and columns, and repeat Steps 2, 3, and 4, as a problem with $(k - p)$ variables:

(a) The equations in Step 2 then are reduced to

$$\begin{aligned} &\Sigma x_0^2 \hat{\beta}_0 + \ldots + \Sigma x_0 x_{k-p-1} \hat{\beta}_{k-p-1} = q_0 \\ &\quad\vdots \\ &\Sigma x_0 x_{k-p-1} \hat{\beta}_0 + \ldots + \Sigma x_{k-p-1}^2 \hat{\beta}_{k-p-1} = q_{k-p-1} \end{aligned} \tag{6-25}$$

and its solution becomes

$$\begin{aligned} &\hat{\beta}_0^* = c_{00}^* q_0 + \ldots + c_{0,k-p-1}^* q_{k-p-1} \\ &\quad\vdots \\ &\hat{\beta}_{k-p-1}^* = c_{1,k-p-1}^* q_0 + \ldots + c_{k-p-1,k-p-1}^* q_{k-p-1}. \end{aligned} \tag{6-26}$$

(b) These values c_{ij}^* will, in general, be different from the c_{ij} for the original equations, so that new coefficients

$$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_{k-p-1}^*$$

will result.

(c) A new value of s^2, say s^{*2}, is computed from

$$s^{*2} = \frac{1}{n - (k - p)} \left\{ \Sigma Y^2 - \sum_0^{k-p-1} \hat{\beta}_i^* q_i \right\}. \tag{6-27}$$

These operations can be handled conveniently by matrix methods. Paragraph 6-9 contains a further discussion of "Matrix Methods."

An *Analysis of Variance* table is formed as follows:

	d.f.	Sum of Squares	Mean Square
Total	n	ΣY^2	
Reduction due to k constants	k	$\sum_0^{k-1} \hat{\beta}_i q_i$	K
Residual (after k constants)	$n - k$	$\Sigma Y^2 - \sum_0^{k-1} \hat{\beta}_i q_i$	s^2
Reduction due to $k - p$ constants	$k - p$	$\sum_0^{k-p-1} \hat{\beta}_i^* q_i$	A
Residuals after $k - p$ constants	$n - (k - p)$	$\Sigma Y^2 - \sum_0^{k-p-1} \hat{\beta}_i^* q_i$	s^{*2}
Reduction due to additional p constants	p	$\sum_0^{k-1} \hat{\beta}_i q_i - \sum_0^{k-p-1} \hat{\beta}_i^* q_i$	P

If the y's are normally distributed about their expected values, then

Step 8 Example—Analysis of Variance Test of Significance of Last Coefficient. The required *Analysis of Variance* table is:

	d.f.	Sums of Squares	Mean Square
Total	6	70.000 000	
Reduction due to 3 constants ($\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$)	3	64.191 527	21.397 176 = K
Residuals (after 3 constants)	3	5.808 473	1.936 158 = s^2
Reduction due to $\hat{\beta}_1$ and $\hat{\beta}_2$ only	2	64.145 461	32.072 730 = A
Residuals (after $\hat{\beta}_1$ and $\hat{\beta}_2$)	4	5.854 539	1.463 635 = s^{*2}
Reduction due to $\hat{\beta}_3$	1	.046 066	.046 066 = P

As implied by Equation (6-27), the *sum of squares for the reduction due to* $\hat{\beta}_1$ *and* $\hat{\beta}_2$ *only* = $\hat{\beta}_1^* q_1 + \hat{\beta}_2^* q_2$, where $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are the estimates of β_1 and β_2 that are obtained when β_3 is taken equal to zero; i.e., when the underlying functional relation is taken to be $y = \beta_1 x_1 + \beta_2 x_2$.

The steps required to evaluate $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are

$$(X_1'X_1) = \begin{bmatrix} 50 & 67 \\ 67 & 194 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{5211}\begin{bmatrix} 194 & -67 \\ -67 & 50 \end{bmatrix}$$

$$[X_1'Y] = \begin{bmatrix} 54 \\ 97 \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \end{bmatrix} = (X'X)^{-1}[X_1'Y]$$

$$= \frac{1}{5211}\begin{bmatrix} 194 & -67 \\ -67 & 50 \end{bmatrix}\begin{bmatrix} 54 \\ 97 \end{bmatrix}.$$

They yield

$$\hat{\beta}_1^* = 0.763\ 193\ 245$$
$$\hat{\beta}_2^* = 0.236\ 422\ 951.$$

Hence, reduction due to $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ only is given by

$$(0.763\ 193\ 245)(54) + (0.236\ 422\ 951)(97) = 64.145\ 461$$

as shown in the *Analysis of Variance* table.

Step 8 Procedure (Cont)

(a) $F = \frac{K}{s^2}$ is distributed as F with d.f. $= k, n - k$, and serves as a test of whether all k constants account for a significant reduction in the error variance.

(b) $F = \frac{P}{s^2}$ is distributed as F with d.f. $= p, n - k$, and serves as a test of whether the addition of the p coefficients accounts for a significant reduction in the error variance over that accounted for by the first $k - p$ constants.

NOTE: In cases where a constant term is involved (i.e., $x_{01} = 1$) we would use

$$F = \frac{\left\{ \sum_{0}^{k-1} \hat{\beta}_i q_i - \frac{(\Sigma Y)^2}{n} \right\} / (k-1)}{s^2} .$$

which is distributed as F with $(k - 1)$ and $(n - k)$ degrees of freedom as a test for the efficacy of the prediction equation.

Step 9 Procedure—Confidence Interval Estimates. L_1 and L_2 constitute a 100 $(1 - \alpha)$ % confidence interval estimate for:

(a) a coefficient β_i,

when $L_1 = \hat{\beta}_i - t_{n-k,\alpha}$ (est. s.d. of $\hat{\beta}_i$)
$L_2 = \hat{\beta}_i + t_{n-k,\alpha}$ (est. s.d. of $\hat{\beta}_i$);

(b) a predicted point on the curve $\hat{Y}_i$,

when $L_1 = \hat{Y}_i - t_{n-k,\alpha}$ (est. s.d. of $\hat{Y}_i$)
$L_2 = \hat{Y}_i + t_{n-k,\alpha}$ (est. s.d. of $\hat{Y}_i$);

(c) a difference of two coefficients $\hat{\beta}_i - \hat{\beta}_j$,

when $L_1 = (\hat{\beta}_i - \hat{\beta}_j) - t_{n-k,\alpha}$ (est. s.d. of $\hat{\beta}_i - \hat{\beta}_j$)
$L_2 = (\hat{\beta}_i - \hat{\beta}_j) + t_{n-k,\alpha}$ (est. s.d. of $\hat{\beta}_i - \hat{\beta}_j$).

In the above, $t_{n-k,\alpha}$ is the value of Student's t for $(n - k)$ degrees of freedom exceeded with probability $\frac{\alpha}{2}$.

Step 8 Example (Cont)

The test of significance for $\hat{\beta}_3$ is

$$F = \frac{P}{s^2}$$

$$= \frac{.046\ 060}{1.936\ 158}$$

$$= .024, \text{ d.f.} = 1, 3.$$

The value of $F(1, 3)$ exceeded with probability .05 is 10.13. The observed F does not exceed this critical value, so that β_3 is not regarded as being statistically significantly different from zero.

Step 9 Example—Confidence Interval Estimates. For β_1, the 95% confidence interval estimate $L_1 \leq \beta_1 \leq L_2$ is determined by

$$L_1 = \hat{\beta}_1 - t_{3,.05} \text{ (est. s.d. of } \hat{\beta}_1)$$
$$= 0.735\ 320\ 652 - 3.182\ (.323\ 627)$$
$$= -.294\ 460$$

$$L_2 = \hat{\beta}_1 + t_{3,.05} \text{ (est. s.d. of } \hat{\beta}_1)$$
$$= 0.735\ 320\ 652 + 3.182\ (.323\ 627)$$
$$= 1.765\ 102$$

where $t_{3,.05} = 3.182$ is the value of Student's t distribution for three degrees of freedom exceeded with probability .025 (or exceeded in absolute value with probability .05).

6-4 MULTIPLE MEASUREMENTS AT ONE OR MORE POINTS

More than one measurement may be made at some or at all of the values of the independent variable x. This usually is done when the random errors are suspected of being composed of two components—one component associated with the variation of the points about the curve, and the other component associated with the variation of repeat determinations. The jth measurement at the ith point then can be represented as

$$Y_{ij} = \beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_{k-1} x_{k-1,i} + \epsilon_i + \eta_{ij} \tag{6-28}$$

where the ϵ's and η's are independent and have variances σ^2 and σ_0^2, respectively.

If a number p_i of repeat determinations are made at each of the n points, the estimation of σ^2 and σ_0^2 follows from a modification of the *Analysis of Variance* table:

	Sum of Squares	d.f.	Mean Square
Total	$\sum_{i=1}^{n}\sum_{j=1}^{p_i} Y_{ij}^2 = T$	$\sum_{1}^{n} p_i$	
Reduction due to fitted constants	$\sum_{i=0}^{k-1} \hat{\beta}_i (\Sigma x_i Y) = C$	k	C/k
Residual (after fitted constants)	$T - C = R$	$\Sigma p_i - k$	$R/(\Sigma p_i - k)$
Repeat determinations	$\sum_{i=1}^{n}\sum_{j=1}^{p_i} (Y_{ij} - \bar{Y}_i)^2 = E_0$	$\Sigma p_i - n$	$E_0/(\Sigma p_i - n)$
Variations of averages about the curve	$R - E_0 = E_1$	$n - k$	$E_1/(n - k)$

The expected value of $E_0/(\Sigma p_i - n)$ is σ_0^2, and the expected value of $E_1/(n - k)$ is $\sigma_0^2 + p\sigma^2$, where all the p_i are equal to p.

The quantity $[E_1/(n - k)]/[E_0/(\Sigma p_i - n)] = F$ is (under the assumption of a normal distribution for the ϵ's and η's) distributed as F, if $\sigma^2 = 0$, with $n - k$ and $\Sigma p_i - n$ degrees of freedom, and may be used to test the statistical significance of the component of variance associated with the ϵ's by comparing the observed F value with tables of the F distribution.

If all the p_i are equal, the proper variance estimate to use in calculating the standard errors, or confidence intervals of the estimated constants, is $E_1/(n - k)$.

6-5 POLYNOMIAL FITTING

If it can be assumed that the relation between the *dependent variable* Y and the *independent variable* x is

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_{k-1} x_i^{k-1} + \epsilon_i \tag{6-29}$$

and that the errors of measurement ϵ_i are independent and have the same variance σ^2, then the techniques for multiple regression carry over without change, by setting:

$$x_{0i} = 1;\ x_{1i} = x_i;\ x_{2i} = x_i^2;\ \ldots;\ x_{k-1,i} = x_i^{k-1}.$$

The normal equations are

$$\begin{aligned} n\beta_0 + \Sigma x\beta_1 + \Sigma x^2\beta_2 + \ldots + \Sigma x^{k-1}\beta_{k-1} &= \Sigma Y \\ \Sigma x\beta_0 + \Sigma x^2\beta_1 + \Sigma x^3\beta_2 + \ldots + \Sigma x^k\beta_{k-1} &= \Sigma xY \\ \ldots & \\ \Sigma x^{k-1}\beta_0 + \Sigma x^k\beta_1 + \Sigma x^{k+1}\beta_2 + \ldots + \Sigma x^{2k-2}\beta_{k-1} &= \Sigma x^{k-1}Y. \end{aligned} \tag{6-30}$$

Note that if the constant term is assumed to be zero, variable x_0 is dropped, and the first row and column are dropped from the normal equations.

In using a polynomial as an approximation to some unknown function, or as an interpolation formula, the correct degree for the polynomial usually is not known. The following procedure usually is applied:

(a) Carry through the steps in fitting polynomials of 2nd, 3rd, 4th, 5th, . . . , degrees.

(b) If the reduction in the error sum of squares due to fitting the kth degree term is statistically significant on the basis of the F-test, whereas the similar test for the $(k + 1)$ degree term is not, then the kth degree polynomial is accepted as the *best* fitting polynomial.

In this procedure, the degree of the polynomial is a random variable, and repetitions of the experiment will lead to different degree polynomials. When the law is *truly polynomial,* the computed curve will either be of correct degree and hence will give unbiased estimates of the coefficients or, *if not* of correct degree, will lead to biased estimates.

When the law is not exactly a polynomial, the error distribution for the Y's will be centered around a value off the curve, and it will be difficult to assess the effect of such systematic errors. In the limiting case, where the variance of the Y's is nearly zero, these systematic errors will be treated as the random error in the measurements. Usually, it will not be valid to assume that these systematic errors are uncorrelated. On the other hand, if these systematic errors are small relative to the measurement error, their effect probably can be neglected.

6-6 INEQUALITY OF VARIANCE

6-6.1 DISCUSSION OF PROCEDURES AND EXAMPLES

When the measurements Y_i have different precision, i.e., when $V(Y_i) = \sigma_i^2$ and $\sigma_{i_1} \neq \sigma_{i_2}$ for at least one pair of subscripts $1 \leq i_1 < i_2 \leq n$, the conditions of the least squares theorem of Paragraph 6-2 are not satisfied. However, the transformed variates

$$Y_i' = \frac{Y_i}{\sigma_i}$$

have a common variance $V(Y_i') = 1$. Often, we have information on the relative magnitudes of the variances σ_i^2 only, and not on their absolute magnitudes. If the variances σ_i^2 are expressed in the form

$$\sigma_i^2 = \frac{\sigma_0^2}{w_i}, \tag{6-31}$$

then w_i is termed the *relative weight** of the measurement Y_i, and the quantities $Y_i^* = \sqrt{w_i}\, Y_i$ have common variance σ_0^2, the magnitude of which may be unknown. In other words, equality of variance is achieved through *weighting* the observations by quantities proportional to the reciprocals of their standard deviations.

* The *absolute weight* of a measurement is, by definition, the reciprocal of its variance.

6-6.2 PROCEDURES AND EXAMPLES

Procedures—The equations representing the expected values of the Y_i^* are

$$\begin{aligned} E(Y_i^*) &= \sqrt{w_i}\, E(Y_i) \\ &= \beta_0 \sqrt{w_i}\, x_{0i} + \beta_1 \sqrt{w_i}\, x_{1i} + \ldots + \beta_{k-1} \sqrt{w_i}\, x_{k-1,i} \\ &= \beta_0\, x_{0i}^* + \beta_1\, x_{1i}^* + \ldots + \beta_{k-1}\, x_{k-1,i}^* \end{aligned} \qquad (6\text{-}32)$$

where $x_{ij}^* = \sqrt{w_i}\, x_{ij}$.

The normal equations for the estimation of the β's are

$$\begin{aligned} &\Sigma\, w\, x_0^2\, \beta_0 + \Sigma\, w\, x_0 x_1 \beta_1 + \ldots + \Sigma\, w\, x_0 x_{k-1} \beta_{k-1} = \Sigma\, w\, x_0 Y \\ &\Sigma\, w\, x_0 x_1 \beta_0 + \Sigma\, w\, x_1^2 \beta_1 + \ldots + \Sigma\, w\, x_1 x_{k-1} \beta_{k-1} = \Sigma\, w\, x_1 Y \\ &\ldots \\ &\Sigma\, w\, x_0 x_{k-1} \beta_0 + \Sigma\, w\, x_1 x_{k-1} \beta_1 + \ldots + \Sigma\, w\, x_{k-1}^2 \beta_{k-1} = \Sigma\, w\, x_{k-1} Y \end{aligned} \qquad (6\text{-}33)$$

The estimate s^2 of σ_0^2 is given by the usual formula

$$s^2 = \frac{\sum_{i=1}^{n} Y_i^{*2} - \sum_{j=1}^{k-1} \hat{\beta}_j \left(\sum_{i=1}^{n} x_{ji}^*\, Y_i^* \right)}{n - k} \qquad (6\text{-}34)$$

which may be written, in terms of the original variables, as

$$\begin{aligned} s^2 &= \frac{\Sigma\, r_i^2\, w_i}{n-k} \\ &= \frac{\Sigma\, w_i Y_i^2 - \Sigma\, \hat{\beta}_j \left(\sum_{i=1}^{n} w_i\, x_{ij}\, Y_i \right)}{n-k}. \end{aligned} \qquad (6\text{-}35)$$

Note that in the case where the value of σ_0^2 is known, we may perform a test of significance of the closeness of the observed estimate to the known value by forming the ratio $F = \frac{s^2}{\sigma_0^2}$ and comparing this value with the 100 $(1 - \alpha)$ percentage point of the F distribution for $n - k$ and ∞ degrees of freedom; or, equivalently, we may compare $\chi^2 = \frac{(n-k)\, s^2}{\sigma_0^2}$ with the 100 $(1 - \alpha)$ percentage point of the χ^2 distribution for $n - k$ degrees of freedom. Restatement of the foregoing, using matrix notation, goes as follows:

If $\text{Var}\,(Y_i) = \text{Diag}\,(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$

$$= \sigma_0^2\, \text{Diag} \left(\frac{1}{w_1}, \frac{1}{w_2}, \ldots, \frac{1}{w_n} \right),$$

then the transformed variates

$$\begin{aligned} Y^* &= \text{Diag}\,(\sqrt{w_1}\, \sqrt{w_2} \ldots \sqrt{w_n}) Y = WY \quad \text{and} \\ X^* &= \text{Diag}\,(\sqrt{w_1}\, \sqrt{w_2} \ldots \sqrt{w_n}) X = WX \end{aligned}$$

satisfy the requirements of the least squares theorem of Paragraph 6-2, and the normal equations are

$$\begin{aligned} &(X^*)'(X^*)\hat{\beta} = (X^*)'Y^* \quad \text{or,} \\ &X'\, W^2\, X\, \hat{\beta} = X'\, W^2\, Y. \end{aligned} \qquad (6\text{-}33M)$$

The estimate of σ_0^2 is given by

$$s^2 = \frac{1}{n-k}\{Y^{*\prime}\, Y^* - \hat{\beta}'\,(X^*)'Y^*\} \quad \text{or,}$$
$$s^2 = \frac{r'\, W^2\, r}{n-k} \tag{6-34M}$$

Examples—Fitting Straight Line Relation (Variance of Y Proportional to Abscissa). Consider the estimation of the coefficients of a line where

$$y = \alpha + \beta x_i,$$

and where $\mathrm{Var}(Y_i) = \sigma^2 x_i$, $i = 1, 2, \ldots, n$. The equations of expectation are

$$\begin{aligned} E(Y_1) &= \alpha + \beta x_1 \\ E(Y_2) &= \alpha + \beta x_2 \\ &\ldots \\ E(Y_n) &= \alpha + \beta x_n \end{aligned} \tag{6-36}$$

Transforming to $Y_i^* = Y_i/\sqrt{x_i}$, gives

$$\begin{aligned} E(Y_1^*) &= \frac{\alpha}{\sqrt{x_1}} + \beta\sqrt{x_1} \\ E(Y_2^*) &= \frac{\alpha}{\sqrt{x_2}} + \beta\sqrt{x_2} \\ &\ldots \\ E(Y_n^*) &= \frac{\alpha}{\sqrt{x_n}} + \beta\sqrt{x_n} \end{aligned} \tag{6-37}$$

and the normal equations for estimating α and β become

$$\begin{aligned} \hat{\alpha}\,\Sigma\frac{1}{x_i} + \hat{\beta}\, n &= \Sigma\, Y_i/x_i \\ \hat{\alpha}\, n + \hat{\beta}\,\Sigma\, x &= \Sigma\, Y_i \end{aligned} \tag{6-38}$$

Direct solution of these equations gives

$$\hat{\beta} = \frac{n\,\Sigma\frac{Y_i}{x_i} - \Sigma\, Y_i\left(\Sigma\frac{1}{x_i}\right)}{n^2 - (\Sigma x_i)\left(\Sigma\frac{1}{x_i}\right)} \tag{6-39}$$

$$\hat{\alpha} = \frac{n\,\Sigma\, Y_i - \Sigma x_i\,\Sigma Y_i/x_i}{n^2 - (\Sigma x_i)\,\Sigma\left(\frac{1}{x_i}\right)} \tag{6-40}$$

and for the estimate of σ^2,

$$s^2 = \frac{1}{n-2}\left\{\Sigma\frac{1}{x_i}\,(Y - \hat{\alpha} - \hat{\beta}\, x_i)^2\right\}. \tag{6-41}$$

6-7 CORRELATED MEASUREMENT ERRORS

6-7.1 DISCUSSION OF PROCEDURES AND EXAMPLES

If the errors of measurement are not independent but instead are correlated so that they have covariances

$$\text{Covar}(Y_i, Y_j) = \sigma_{ij} = \sigma_{ji} \tag{6-42}$$

and variances

$$\text{Var}(Y_i) = \sigma_i^2,$$

then a transformation of the variables $Y_1, Y_2, \ldots, Y_n$, to new variables $Y_1^*, Y_2^*, \ldots, Y_n^*$, is required so that the method of least squares may be applied. In some simple cases, a transformation in the form of sums and differences of the original variables immediately suggests itself, and the expected values of the new variables are computed easily. The example used to illustrate the techniques presented in this Paragraph is such a case.

6-7.2 PROCEDURES AND EXAMPLES

Procedures—The variances and covariances may be represented by the $n \times n$ variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdot & \cdot & \cdot & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdot & \cdot & \cdot & \sigma_{2n} \\ \cdot \;\cdot & \cdot & & & & \\ \sigma_{n1} & \sigma_{n2} & \cdot & \cdot & \cdot & \sigma_n^2 \end{bmatrix} \tag{6-42M}$$

Assuming V to be of full rank, i.e., determinant of V is not zero, it is possible to factor V into the product

$$V = T\,T' \tag{6-43M}$$

where T is lower triangular and T' is the transpose of T. The required transformation then is given by

$$Y^* = T^{-1}\,Y \qquad \text{and} \qquad X^* = T^{-1}\,X \tag{6-44M}$$

where $(Y^*)' = (Y_1^*, Y_2^*, \ldots, Y_n^*)$ is the vector of transformed variables and $Y' = (Y_1, Y_2, \ldots, Y_n)$ is the vector of original variables. X^* and X are the matrices representing the equations of expected values of the transformed variables and of the original variables, respectively. (See Paragraph 6-9 for the method of computing T and T^{-1}).

The normal equations then are

$$(X^*)'(X^*)\,\hat{\beta} = (X^*)'\,Y^* \tag{6-45M}$$

or, in terms of the original variables,

$$X'V^{-1}X\,\hat{\beta} = X'V^{-1}\,Y, \tag{6-46M}$$

and the estimates of the β's are given by

$$\begin{aligned} \hat{\beta} &= [(X^*)'(X^*)]^{-1}(X^*)'Y^* \\ &= (X'V^{-1}X)^{-1}X'V^{-1}Y. \end{aligned} \tag{6-47M}$$

The variance estimate

$$s^2 = \frac{1}{n-k}\left\{\Sigma\, Y^{*2} - \sum_{i=1}^{k} \hat{\beta}_i \left(\sum_{j=1}^{n} x_{ij}^*\;\; Y_j^*\right)\right\} \tag{6-48}$$

Procedures (Cont)

is an estimate of unity when the variances and covariances are known. This may be written as

$$s^2 = \frac{1}{n-k}\{Y'V^{-1}Y - \hat{\beta}'X'V^{-1}Y\}$$
$$= \frac{1}{n-k}\{r'V^{-1}r\} \qquad (6\text{-}48M)$$

where r is the column vector of deviations, $r = Y - X\hat{\beta}$.

If, instead of V, a matrix with entries proportional to the variances and covariances is used, say $W = \frac{1}{\sigma_0^2}V$, then s^2 is an estimate of σ_0^2.

Examples—Parabolic Relationship with Cumulative Errors. If the errors of measurements of Y at successive x values in a case of a parabolic law $Y = \beta_0 + \beta_1 x + \beta_2 x^2$ are cumulative, i.e.,

$$Y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$
$$Y_2 = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_1 + \epsilon_2$$
$$\cdots$$
$$Y_n = \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \sum_1^n \epsilon_i$$

then $E(Y) = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = X\beta.$

If all the ϵ's are from the same distribution,

then
$$\text{Var}(Y_i) = i \cdot \sigma^2$$
$$\text{Covar}(Y_i\, Y_j) = \begin{cases} i \cdot \sigma^2 & i < j \\ j \cdot \sigma^2 & j < i \end{cases}$$

and the variance covariance matrix becomes

$$V = \sigma^2 \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 2 & 2 & & & & 2 \\ 1 & 2 & 3 & & & & 3 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 1 & 2 & 3 & & & & n \end{bmatrix}.$$

Taking $W = \frac{1}{\sigma^2}V$, the necessary transformation is given by factoring W into $W = T\,T'$.

6-7.2 PROCEDURES AND EXAMPLES (CONT)

Examples (Cont)

A little computation gives

$$W = \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 1 & 1 & & \\ & & & \cdot & \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ & 1 & 1 & & 1 \\ & & 1 & & 1 \\ & & & \cdots & \\ & & & & 1 \end{bmatrix} = T\,T'$$

$$W^{-1} = (T')^{-1}T^{-1} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ & 1 & -1 & 0 & 0 \\ & & 1 & -1 & 0 \\ & & \cdot & \cdot & \cdot \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ 0 & -1 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix}$$

which, for the transformed variate, gives

$$Y^* = T^{-1}\,Y = \begin{bmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ 0 & -1 & 1 & & & \\ \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 - Y_1 \\ Y_3 - Y_2 \\ \cdot \quad \cdot \quad \cdot \\ Y_n - Y_{n-1} \end{bmatrix}$$

Examples (Cont)

and

$$X^* = T^{-1}X = \begin{bmatrix} 1 & & & & & & \\ -1 & 1 & & & & & \\ 0 & -1 & 1 & & & & \\ \cdot & \cdot & \cdot & & & & \\ 0 & 0 & 0 & \cdot\ \cdot\ \cdot & & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_1 & x_1^2 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 \\ 0 & x_3 - x_2 & x_3^2 - x_2^2 \\ \cdot & \cdot & \cdot \\ 0 & x_n - x_{n-1} & x_n^2 - x_{n-1}^2 \end{bmatrix}.$$

Note that $Y_i^* = Y_i - Y_{i-1}$

$= (x_i - x_{i-1})\beta_0 + (x_i^2 - x_{i-1}^2)\beta_1 + \epsilon_i$ for $i \geq 2$

and $Y_1^* = Y_1$;

hence, the Y^*'s have the same variance, and have zero covariances.

The normal equations become

$$X^{*\prime}X^*\beta = \begin{bmatrix} 1 & x_1 & x_1^2 \\ x_1 & x_1^2 + \sum_1^n (x_i - x_{i-1})^2 & x_1^3 + \sum_2^n (x_i - x_{i-1})(x_i^2 - x_{i-1}^2) \\ x_1^2 & x_1^3 + \sum_2^n (x_i - x_{i-1})(x_i^2 - x_{i-1}^2) & x_1^4 + \sum_2^n (x_i^2 - x_{i-1}^2)^2 \end{bmatrix}\beta$$

$$= \begin{bmatrix} Y_1 \\ x_1Y_1 + \sum_2^n (x_i - x_{i-1})(Y_i - Y_{i-1}) \\ x_1^2Y_1 + \sum_2^n (x_i^2 - x_{i-1}^2)(Y_i - Y_{i-1}) \end{bmatrix}$$

6-7.2 PROCEDURES AND EXAMPLES (CONT)

Examples (Cont)

or, in terms of the original matrices $X'W^{-1}X\beta = X'W^{-1}Y$, give

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & \cdots & \cdots & & & & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \cdot & \cdot & \\ 1 & x_n & x_n^2 \end{bmatrix} \beta$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_n \\ x_1^2 & x_2^2 & & x_n^2 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \cdots & & \\ & & & -1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$$

which, upon multiplication, will be seen to give the same normal equations as above. If the analysis is carried out in terms of the transformed variables, σ^2 is estimated by

$$s^2 = \frac{\Sigma Y^{*2} - \Sigma \hat{\beta}_i \left(\sum_j x_{ji}^* Y_j^* \right)}{n - 3}$$

or equivalently, in terms of the original variables, by

$$s^2 = \frac{1}{n-3} \{ Y'W^{-1}Y - \hat{\beta}'X'W^{-1}Y \} .$$

6-8 USE OF ORTHOGONAL POLYNOMIALS WITH EQUALLY SPACED x VALUES

6-8.1 DISCUSSION OF PROCEDURES AND EXAMPLES

The fitting of a polynomial

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_{k-1} x^{k-1} \qquad (6\text{-}49)$$

to observations at n equally-spaced values of x (spaced a distance D apart) can be simplified by transforming the x's to new variables $\xi_0', \xi_1', \ldots, \xi_{k-1}'$, which are orthogonal to each other.

The variables then become

$$\xi'_0 = \quad \xi_0 = 1$$

$$\xi'_1 = \lambda_1\xi_1 \quad \text{where } \xi_1 = \frac{x_i - \bar{x}}{D}$$

$$\xi'_2 = \lambda_2\xi_2 \quad \text{where } \xi_2 = \left(\frac{x_i - \bar{x}}{D}\right)^2 - \frac{n^2 - 1}{12}$$

$$\xi'_3 = \lambda_3\xi_3 \quad \text{where } \xi_3 = \left(\frac{x_i - \bar{x}}{D}\right)^3 - \left(\frac{x_i - \bar{x}}{D}\right)\left(\frac{3n^2 - 7}{20}\right) \qquad (6\text{-}50)$$

$$\xi'_4 = \lambda_4\xi_4 \quad \text{where } \xi_4 = \left(\frac{x_i - \bar{x}}{D}\right)^4 - \left(\frac{x_i - \bar{x}}{D}\right)^2\left(\frac{3n^2 - 13}{14}\right) + \frac{3}{560}(n^4 - 10n^2 + 9)$$

$$\xi'_5 = \lambda_5\xi_5 \quad \text{where } \xi_5 = \left(\frac{x_i - \bar{x}}{D}\right)^5 - \left(\frac{x_i - \bar{x}}{D}\right)^3\frac{5(n^2 - 7)}{18} + \left(\frac{x_i - \bar{x}}{D}\right)\left(\frac{15n^4 - 230n^2 + 407}{1008}\right)$$

. . .

where $\xi_{k+1} = \xi_1\xi_k - \dfrac{k^2(n^2 - k^2)}{4(4k^2 - 1)}\xi_{k-1}$.

The λ_i are chosen so that the elements of ξ_i are integers.

By fitting Y as a function

$$Y = \alpha_0\xi'_0 + \alpha_1\xi'_1 + \ldots + \alpha_{k-1}\xi'_{k-1}, \qquad (6\text{-}51)$$

the estimation of the α's and the analysis of variance are simplified because the normal equations are in diagonal form.

In order to obtain the estimates of the β's and their associated standard errors, or to use Equation (6-51) for predicting a value for a point not in the original data, an extra calculation but no matrix inversion is required.

Tables of ξ', λ, and $\Sigma(\xi')^2$ are given by Fisher and Yates(6) for $n \leq 75$, and by Anderson and Houseman(7) for $n \leq 104$ for up to 5th degree polynomials; in DeLury(8) for $n \leq 26$ for all powers; and in Pearson and Hartley(9) for $n \leq 52$ for up to 6th degree polynomials. Table 6-1 is a sample from Fisher and Yates.(6)

To illustrate the calculations, consider the fitting of a cubic to the following (x, Y) points:

x	Y
10	3.4
20	11.7
30	37.2
40	80.1
50	151.4
60	253.2
70	392.6

TABLE 6-1. SAMPLE TABLE OF ORTHOGONAL POLYNOMIALS

3

	ξ'_1	ξ'_2
	−1	+1
	0	−2
	+1	+1
$\Sigma(\xi')^2$	2	6
λ	1	3

4

	ξ'_1	ξ'_2	ξ'_3
	−3	+1	−1
	−1	−1	+3
	+1	−1	−3
	+3	+1	+1
$\Sigma(\xi')^2$	20	4	20
λ	2	1	10/3

5

	ξ'_1	ξ'_2	ξ'_3	ξ'_4
	−2	+2	−1	+1
	−1	−1	+2	−4
	0	−2	0	+6
	+1	−1	−2	−4
	+2	+2	+1	+1
$\Sigma(\xi')^2$	10	14	10	70
λ	1	1	5/6	35/12

6

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	−5	+5	−5	+1	−1
	−3	−1	+7	−3	+5
	−1	−4	+4	+2	−10
	+1	−4	−4	+2	+10
	+3	−1	−7	−3	−5
	+5	+5	+5	+1	+1
$\Sigma(\xi')^2$	70	84	180	28	252
λ	2	3/2	5/3	7/12	21/10

7

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	−3	+5	−1	+3	−1
	−2	0	+1	−7	+4
	−1	−3	+1	+1	−5
	0	−4	0	+6	0
	+1	−3	−1	+1	+5
	+2	0	−1	−7	−4
	+3	+5	+1	+3	+1
$\Sigma(\xi')^2$	28	84	6	154	84
λ	1	1	1/6	7/12	7/20

8

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	−7	+7	−7	+7	−7
	−5	+1	+5	−13	+23
	−3	−3	+7	−3	−17
	−1	−5	+3	+9	−15
	+1	−5	−3	+9	+15
	+3	−3	−7	−3	+17
	+5	+1	−5	−13	−23
	+7	+7	+7	+7	+7
$\Sigma(\xi')^2$	168	168	264	616	2184
λ	2	1	2/3	7/12	7/10

9

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	0	−20	0	+18	0
	+1	−17	−9	+9	+9
	+2	−8	−13	−11	+4
	+3	+7	−7	−21	−11
	+4	+28	+14	+14	+4
$\Sigma(\xi')^2$	60	2,772	990	2,002	468
λ	1	3	5/6	7/12	3/20

10

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	+1	−4	−12	+18	+6
	+3	−3	−31	+3	+11
	+5	−1	−35	−17	+1
	+7	+2	−14	−22	−14
	+9	+6	+42	+18	+6
$\Sigma(\xi')^2$	330	132	8,580	2,860	780
λ	2	1/2	5/3	5/12	1/10

11

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	0	−10	0	+6	0
	+1	−9	−14	+4	+4
	+2	−6	−23	−1	+4
	+3	−1	−22	−6	−1
	+4	+6	−6	−6	−6
	+5	+15	+30	+6	+3
$\Sigma(\xi')^2$	110	858	4,290	286	156
λ	1	1	5/6	1/12	1/40

12

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	+1	−35	−7	+28	+20
	+3	−29	−19	+12	+44
	+5	−17	−25	−13	+29
	+7	+1	−21	−33	−21
	+9	+25	−3	−27	−57
	+11	+55	+33	+33	+33
$\Sigma(\xi')^2$	572	12,012	5,148	8,008	15,912
λ	2	3	2/3	7/24	3/20

13

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	0	−14	0	+84	0
	+1	−13	−4	+64	+20
	+2	−10	−7	+11	+26
	+3	−5	−8	−54	+11
	+4	+2	−6	−96	−18
	+5	+11	0	−66	−33
	+6	+22	+11	+99	+22
$\Sigma(\xi')^2$	182	2,002	572	68,068	6,188
λ	1	1	1/6	7/12	7/120

14

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	+1	−8	−24	+108	+60
	+3	−7	−67	+63	+145
	+5	−5	−95	−13	+139
	+7	−2	−98	−92	+28
	+9	+2	−66	−132	−132
	+11	+7	+11	−77	−187
	+13	+13	+143	+143	+143
$\Sigma(\xi')^2$	910	728	97,240	136,136	235,144
λ	2	1/2	5/3	7/12	7/30

15

	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	0	−56	0	+756	0
	+1	−53	−27	+621	+675
	+2	−44	−49	+251	+1000
	+3	−29	−61	−249	+751
	+4	−8	−58	−704	−44
	+5	+19	−35	−869	−979
	+6	+52	+13	−429	−1144
	+7	+91	+91	+1001	+1001
$\Sigma(\xi')^2$	280	37,128	39,780	6,466,460	10,581,480
λ	1	3	5/6	35/12	21/20

TABLE 6-1. SAMPLE TABLE OF ORTHOGONAL POLYNOMIALS (Continued)

	16					17					18				
	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
	+1	−21	−63	+189	+45	0	−24	0	+36	0	+1	−40	−8	+44	+220
	+3	−19	−179	+129	+115	+1	−23	−7	+31	+55	+3	−37	−23	+33	+583
	+5	−15	−265	+23	+131	+2	−20	−13	+17	+88	+5	−31	−35	+13	+733
	+7	−9	−301	−101	+77	+3	−15	−17	−3	+83	+7	−22	−42	−12	+588
	+9	−1	−267	−201	−33	+4	−8	−18	−24	+36	+9	−10	−42	−36	+156
	+11	+9	−143	−221	−143	+5	+1	−15	−39	−39	+11	+5	−33	−51	−429
	+13	+21	+91	−91	−143	+6	+12	−7	−39	−104	+13	+23	−13	−47	−871
	+15	+35	+455	+273	+143	+7	+25	+7	−13	−91	+15	+44	+20	−12	−676
						+8	+40	+28	+52	+104	+17	+68	+68	+68	+884
$\Sigma(\xi')^2$	1,360	5,712	1,007,760	470,288	201,552	408	7,752	3,876	16,796	100,776	1,938	23,256	23,256	28,424	6,953,544
λ	2	1	$\frac{10}{3}$	$\frac{7}{12}$	$\frac{1}{10}$	1	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{20}$	2	$\frac{3}{2}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{3}{10}$

Note: For n > 8, only the values for positive $\xi'_1 = \frac{x - \bar{x}}{D}$ are given.

Note: In Table 6-1, only the values for positive $\xi'_1 = \frac{x - \bar{x}}{D}$ are given for $n > 8$. The missing values ($n/2$ rows for n even and $(n-1)/2$ ws for n odd) must be supplied by using the given rows in reverse order, changing the sign for odd-numbered ξ'. See $n = 7$ and $= 8$ for example of this rule.

From Table 6-1, for $n = 7$ we copy out:

ξ'_0	ξ'_1	ξ'_2	ξ'_3	Y
1	−3	5	−1	3.4
1	−2	0	1	11.7
1	−1	−3	1	37.2
1	0	−4	0	80.1
1	1	−3	−1	151.4
1	2	0	−1	253.2
1	3	5	1	392.6

where

$$\xi'_0 = 1 \qquad \lambda_0 = 1$$

$$\xi'_1 = \frac{x - \bar{x}}{10} \qquad \lambda_1 = 1$$

$$\xi'_2 = \xi_1^2 - 4 \qquad \lambda_2 = 1$$

$$\xi'_3 = \frac{\xi_1^3 - 7\xi_1}{6} \qquad \lambda_3 = 1/6$$

with $\bar{x} = 40$, $D = 10$.

6-8.2 PROCEDURES AND EXAMPLES

Step 1 Procedure—Form the quantities

$$\begin{array}{l} \Sigma Y \\ \Sigma \xi_1' Y \\ \Sigma \xi_2' Y \\ \cdot \\ \cdot \\ \cdot \\ \Sigma \xi_{k-1}' Y \end{array} \tag{6-52}$$

and, using the values of $\Sigma \xi_1'^2$, $\Sigma \xi_2'^2$, . . . , given in Table 6-1 form the estimates of the parameters, α_0, α_1, . . . , as follows:

$$\begin{aligned} \hat{\alpha}_0 &= \frac{\Sigma Y}{n} = \bar{y} \\ \hat{\alpha}_1 &= \frac{\Sigma \xi_1' Y}{\Sigma \xi_1'^2} \\ \hat{\alpha}_2 &= \frac{\Sigma \xi_2' Y}{\Sigma \xi_2'^2} \\ &\cdots \\ \hat{\alpha}_{k-1} &= \frac{\Sigma \xi_{k-1}' Y}{\Sigma \xi_{k-1}'^2} \end{aligned} \tag{6-53}$$

Step 2 Procedure—Calculate the deviations r_i from

$$r_i = Y_i - \bar{y} - \hat{\alpha}_1 \xi_{1,i}' - \hat{\alpha}_2 \xi_{2,i}' - \ldots - \hat{\alpha}_{k-1} \xi_{k-1,i}'. \tag{6-54}$$

Step 1 Example—Using the values copied from Table 6-1, the following calculations are made:

$$\begin{aligned} \Sigma\xi_0'Y &= 929.6 & \Sigma\xi_0'^2 &= 7 \\ \Sigma\xi_1'Y &= 1764.8 & \Sigma\xi_1'^2 &= 28 \\ \Sigma\xi_2'Y &= 1093.8 & \Sigma\xi_2'^2 &= 84 \\ \Sigma\xi_3'Y &= 33.5 & \Sigma\xi_3'^2 &= 6. \end{aligned}$$

The estimates of the coefficients in the representation of y as a function of the ξ_i', i.e., as

$$y = \alpha_0\xi_0' + \alpha_1\xi_1' + \alpha_2\xi_2' + \alpha_3\xi_3'$$

are given by

$$\hat{\alpha}_0 = \Sigma\xi_0'Y/\Sigma\xi_0'^2 = 929.6/7 = 132.8$$

$$\hat{\alpha}_1 = \Sigma\xi_1'Y/\Sigma\xi_1'^2 = 1764.8/28 = 63.0285\ 7143$$

$$\hat{\alpha}_2 = \Sigma\xi_2'Y/\Sigma\xi_2'^2 = 1093.8/84 = 13.0214\ 2857$$

$$\hat{\alpha}_3 = \Sigma\xi_3'Y/\Sigma\xi_3'^2 = 33.5/6 = 5.5833\ 3333.$$

Step 2 Example—The predicted value for the point $x = 10$ is given by substituting its corresponding values of the ξ's ($\xi_0' = 1$, $\xi_1' = -3$, $\xi_2' = 5$, and $\xi_3' = -1$) in the equation

$$\hat{Y}_x = 132.8 + 63.028\ 5714\ \xi_1' + 13.021\ 4286\ \xi_2' + 5.583\ 3333\ \xi_3',$$

i.e., $\hat{Y}_{10} = 132.8 + 63.028\ 5714(-3) + 13.021\ 4286(5) + 5.583\ 3333(-1)$

$$= 3.238\ 0955$$

leading to a deviation between observed and calculated of

$$r_{10} = 3.4 - 3.238\ 0955$$
$$= 0.161\ 9045.$$

For the entire set of points, we get:

Observed Y	Calculated $\hat{Y}$	Residual $r = Y - \hat{Y}$
3.4	3.238 0955	0.161 9045
11.7	12.326 1905	−0.626 1905
37.2	36.290 4761	0.909 5239
80.1	80.714 2856	−0.614 2856
151.4	151.180 9523	0.219 0477
253.2	253.273 8095	−0.073 8095
392.6	392.576 1905	0.023 8095

Step 3 Procedure—The estimate of σ^2 is given by

$$s^2 = \frac{1}{n-k}\{\Sigma Y^2 - \hat{\alpha}_0\Sigma Y - \hat{\alpha}_1\Sigma\xi_1' Y - \ldots - \hat{\alpha}_{k-1}\Sigma\xi_{k-1}' Y\}. \qquad (6\text{-}55)$$

Step 4 Procedure—The estimate of the standard deviations of the $\hat{\alpha}$'s is given by

$$\text{s.d. } (\hat{\alpha}_j) = \frac{s}{\sqrt{\sum_{i=1}^{n} \xi_{ji}'^2}}. \qquad (6\text{-}56)$$

Step 5 Procedure—The *Analysis of Variance* table becomes:

	d.f.	Sum of Squares
Total	n	ΣY^2
Reduction due to fitting α_0	1	$\hat{\alpha}_0\ (\Sigma Y) = R_0$
Deviations from fit with α_0	$n-1$	$\Sigma Y^2 - R_0$
Reduction due to fitting α_1	1	$\hat{\alpha}_1\ (\Sigma\xi_1' Y) = R_1$
Deviations from fit with α_0, α_1	$n-2$	$\Sigma Y^2 - R_0 - R_1$
. . .		
Reduction due to fit of α_{k-1}	1	$\hat{\alpha}_{k-1}\ (\Sigma\xi_{k-1}' Y) = R_{k-1}$
Deviations from fit with $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}$	$n-k$	$\Sigma Y_0 - R_0 - R_1 \ldots - R_{k-1}$

Step 3 Example—The estimate of σ^2 is given by

$$s^2 = \frac{1}{7-4} \Sigma r^2$$

$$= \frac{1}{3} (1.676\ 9048)$$

$$= .558\ 9683$$

$$s = \sqrt{.558\ 9683}$$

$$= .7476.$$

Step 4 Example—The standard deviations of the coefficients are given by

$$\text{s.d.}\ (\hat{\alpha}_i) = s/\sqrt{\Sigma \xi_i'^2}$$

$$\text{s.d.}\ (\hat{\alpha}_0) = .7476/\sqrt{7} = .2826$$

$$\text{s.d.}\ (\hat{\alpha}_1) = .7476/\sqrt{28} = .1413$$

$$\text{s.d.}\ (\hat{\alpha}_2) = .7476/\sqrt{84} = .0816$$

$$\text{s.d.}\ (\hat{\alpha}_3) = .7476/\sqrt{6} = .3052.$$

Step 5 Example—The Analysis of Variance table becomes:

	d.f.	Sum of Squares	Mean Square
Total	7	249 115.26	
Reduction due to coef. of ξ_0'	1	123 450.88	123 450.88
Residuals from $\hat{\alpha}_0\xi_0'$	6	125 664.38	20 944.06
Reduction due to coef. of ξ_1'	1	111 232.822 86	111 232.82
Residuals from $\hat{\alpha}_0\xi_0' + \hat{\alpha}_1\xi_1'$	5	14 431.557 14	2 886.31
Reduction due to coef. of ξ_2'	1	14 242.838 57	14 242.84
Residuals from $\hat{\alpha}_0\xi_0' + \hat{\alpha}_1\xi_1' + \hat{\alpha}_2\xi_2'$	4	188.718 57	47.18
Reduction due to coef. of ξ_3'	1	187.041 67	187.04
Residuals from $\hat{\alpha}_0\xi_0' + \ldots + \hat{\alpha}_3\xi_3'$	3	1.676 90	.5590

Step 6 Procedure—Convert to an equation in the original x units by substituting the expressions in Equations (6-50) into Equation (6-51). By writing the $\hat{\beta}$'s as linear functions of the $\hat{\alpha}$'s, say

$$\hat{\beta}_k = \sum_{i=0}^{k} b_i \hat{\alpha}_i ,$$

the standard deviation can be computed from

$$\text{s.d. of } \hat{\beta}_k = \sqrt{\sum_{0}^{k} b_i^2 \, (\text{s.d. of } \hat{\alpha}_i)^2} .$$

The following Equations (6-57) show the $\hat{\beta}$'s as a function of the $\hat{\alpha}$'s for polynomials up to 5th degree. (If a polynomial of 4th degree is used, simply disregard the terms involving $\hat{\alpha}_5$; if 3rd degree, disregard the terms involving $\hat{\alpha}_4$ and $\hat{\alpha}_5$; etc.)

As an example, if a 4th degree polynomial is fitted, the estimate $\hat{\beta}_3$ is given by

$$\hat{\beta}_3 = \frac{\lambda_3}{D^3} \hat{\alpha}_3 - 4 \frac{\lambda_4}{D^3} \left(\frac{\bar{x}}{D}\right) \hat{\alpha}_4$$

and the s.d. of $\hat{\beta}_3$ is estimated by

$$s \sqrt{\left(\frac{\lambda_3}{D^3}\right)^2 \frac{1}{\Sigma \, (\xi_3')^2} + \left(\frac{4\lambda_4}{D^3}\right)^2 \left(\frac{\bar{x}}{D}\right)^2 \frac{1}{\Sigma \, (\xi_4')^2}} .$$

See Equations (6-57) on page 6-36.

Step 6 Example—To obtain the equation in terms of the original x variable, i.e., expressing y as $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, we substitute as follows:

$$y = \alpha_0 (1) + \alpha_1 \left(\frac{x-40}{10}\right) + \alpha_2 \left[\left(\frac{x-40}{10}\right)^2 - 4\right] + \frac{\alpha_3}{6}\left[\left(\frac{x-40}{10}\right)^3 - 7\left(\frac{x-40}{10}\right)\right]$$

$$= (\alpha_0 - 4\alpha_1 + 12\alpha_2 - 6\alpha_3) + \left(\frac{\alpha_1}{10} - \frac{8}{10}\alpha_2 + \frac{41}{60}\alpha_3\right) x + \left(\frac{\alpha_2}{100} - \frac{2\alpha_3}{100}\right) x^2 + \frac{\alpha_3}{6000} x^3 .$$

Substituting the estimated values for the α's gives

$$Y = 3.4428\ 5714 - .299\ 007\ 9375\, x + .018\ 547\ 6191\, x^2 + .000\ 930\ 5556\, x^3 .$$

The standard deviations of the $\hat{\beta}$'s are given by

$$\text{s.d. of } \hat{\beta}_0 = \text{s.d. of } (\hat{\alpha}_0 - 4\hat{\alpha}_1 + 12\hat{\alpha}_2 - 6\hat{\alpha}_3)$$

$$= s\sqrt{\frac{1}{7} + \frac{(-4)^2}{28} + \frac{(12)^2}{84} + \frac{(-6)^2}{6}}$$

$$= s\sqrt{\frac{59}{7}}$$

$$= 2.170$$

$$\text{s.d. of } \hat{\beta}_1 = \frac{s}{60}\sqrt{\frac{(6)^2}{28} + \frac{(-48)^2}{84} + \frac{(41)^2}{6}}$$

$$= .2190$$

$$\text{s.d. of } \hat{\beta}_2 = \frac{s}{100}\sqrt{\frac{1}{84} + \frac{(-2)^2}{6}}$$

$$= .006\ 158$$

$$\text{s.d. of } \hat{\beta}_3 = \frac{s}{6000}\frac{1}{\sqrt{6}}$$

$$= .0000\ 5087.$$

$$
\begin{aligned}
\hat{\beta}_0 &= \hat{\alpha}_0 - \lambda_1\left[\frac{x}{D}\right]\hat{\alpha}_1 + \lambda_2\left[\left(\frac{x}{D}\right)^2 - \frac{n^2-1}{12}\right]\hat{\alpha}_2 - \lambda_3\left[\left(\frac{x}{D}\right)^3 - \left(\frac{x}{D}\right)\left(\frac{3n^2-7}{20}\right)\right]\hat{\alpha}_3 + \lambda_4\left[\left(\frac{x}{D}\right)^4 - \left(\frac{x}{D}\right)^2\left(\frac{3n^2-13}{14}\right) + \frac{3}{560}\left(n^4-10n^2+9\right)\right]\hat{\alpha}_4 \\
&\quad - \lambda_5\left[\left(\frac{x}{D}\right)^5 - \left(\frac{x}{D}\right)^3\frac{5\left(n^2-7\right)}{18} + \left(\frac{x}{D}\right)\left(\frac{15n^4-230n^2+407}{1008}\right)\right]\hat{\alpha}_5 \\
\hat{\beta}_1 &= \frac{\lambda_1}{D}\hat{\alpha}_1 - \frac{2\lambda_2}{D}\left(\frac{x}{D}\right)\hat{\alpha}_2 + \frac{\lambda_3}{D}\left[3\left(\frac{x}{D}\right)^2 - \frac{3n^2-7}{20}\right]\hat{\alpha}_3 - \frac{\lambda_4}{D}\left[4\left(\frac{x}{D}\right)^3 - 2\left(\frac{x}{D}\right)\left(\frac{3n^2-13}{14}\right)\right]\hat{\alpha}_4 \\
&\quad + \frac{\lambda_5}{D}\left[5\left(\frac{x}{D}\right)^4 - 3\left(\frac{x}{D}\right)^2\frac{5\left(n^2-7\right)}{18} + \frac{15n^4-230n^2+407}{1008}\right]\hat{\alpha}_5 \\
\hat{\beta}_2 &= \frac{\lambda_2}{D^2}\hat{\alpha}_2 - \frac{3\lambda_3}{D^2}\left(\frac{x}{D}\right)\hat{\alpha}_3 + \frac{\lambda_4}{D^2}\left[6\left(\frac{x}{D}\right)^2 - \left(\frac{3n^2-13}{14}\right)\right]\hat{\alpha}_4 - \frac{\lambda_5}{D^2}\left[10\left(\frac{x}{D}\right)^3 - 3\left(\frac{x}{D}\right)\frac{5\left(n^2-7\right)}{18}\right]\hat{\alpha}_5 \qquad (6\text{-}57) \\
\hat{\beta}_3 &= \frac{\lambda_3}{D^3}\hat{\alpha}_3 - \frac{4\lambda_4}{D^3}\left(\frac{x}{D}\right)\hat{\alpha}_4 + \frac{\lambda_5}{D^3}\left[10\left(\frac{x}{D}\right)^2 - \frac{5\left(n^2-7\right)}{18}\right]\hat{\alpha}_5 \\
\hat{\beta}_4 &= \frac{\lambda_4}{D^4}\hat{\alpha}_4 - \frac{5\lambda_5}{D^4}\left(\frac{x}{D}\right)\hat{\alpha}_5 \\
\hat{\beta}_5 &= \frac{\lambda_5}{D^5}\hat{\alpha}_5
\end{aligned}
$$

6-9 MATRIX METHODS

6-9.1 FORMULAS USING TRIANGULAR FACTORIZATION OF NORMAL EQUATIONS

The matrix for the left-hand side of normal equations can be factored into $(X'X) = TT'$ where T is lower triangular, so that $(X'X)^{-1} = (T')^{-1} T^{-1} = (T^{-1})' T^{-1}$.

Thus, $\hat{\beta} = (T^{-1})' (T^{-1}Q)$ where $Q = X'Y$.

Denote the column vector $T^{-1}Q$ by

$$g = T^{-1}Q = \begin{bmatrix} g_1 \\ g_2 \\ \cdot \\ \cdot \\ \cdot \\ g_k \end{bmatrix}.$$

Therefore, $\hat{\beta} = (T^{-1})'g$.

This representation leads to certain simplifications, e.g.:

(a) The estimate of σ^2 is given by

$$\begin{aligned} s^2 &= \frac{1}{n-k} (Y'Y - \hat{\beta}'Q) \\ &= \frac{1}{n-k} (Y'Y - g'T^{-1}Q) \\ &= \frac{1}{n-k} (Y'Y - g'g) \\ &= \frac{1}{n-k} \left(\Sigma Y_i^2 - \sum_1^k g_i^2 \right). \end{aligned}$$

(b) The variance of a linear function, $L = a'\hat{\beta}$ of the $\hat{\beta}$'s is given by

$$\begin{aligned} s^2 \{a' (T^{-1})' T^{-1} a\} &= s^2 (T^{-1}a)' (T^{-1}a) \\ &= s^2 \Sigma h_i^2 \end{aligned}$$

when $h = \begin{bmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ \cdot \\ h_k \end{bmatrix} = T^{-1}a.$

(c) The reduction in sum of squares due to fitting the last p constants is

$$\sum_{k-p+1}^{k} g_i^2.$$

This formulation also permits us to make a detailed *Analysis of Variance* table. An important caution is in order. The reduction due to the addition of $\hat{\beta}_i$ is the reduction *given that* $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{i-1}$, *have been fitted to the data.* The reduction due to $\hat{\beta}_i$ given that any other set of coefficients have been fitted will be different.

The *Analysis of Variance* table becomes:

	d.f.	Sum of Squares
Reduction due to fitting $\hat{\beta}_1$	1	g_1^2
Residual (after fitting $\hat{\beta}_1$)	$n - 1$	$\Sigma Y^2 - g_1^2$
Additional reduction fitting $\hat{\beta}_2$	1	g_2^2
Reduction due to fitting $\hat{\beta}_1$ and $\hat{\beta}_2$	2	$g_1^2 + g_2^2$
Residual (after fitting $\hat{\beta}_1$ and $\hat{\beta}_2$)	$n - 2$	$\Sigma Y^2 - g_1^2 - g_2^2$
.		
.		
.		
Additional reduction due to fitting $\hat{\beta}_k$	1	g_k^2
Reduction due to fitting $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$	k	$\sum_1^k g_i^2$
Residual (after fitting $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$)	$n - k$	$\Sigma Y^2 - \sum_1^k g_i^2$

This form of analysis is especially useful in the analysis for polynomials where the ordering is by powers of x. In the multiple regression case, the reduction attributed to $\hat{\beta}_i$ is dependent upon the ordering of the parameters $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{i-1}$, and will be different for different orders.

6-9.2 TRIANGULARIZATION OF MATRICES

The real symmetric matrix

$$N = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}$$

can, if N is non-singular (i.e., if $|N| \neq 0$), be factored into the product of two triangular matrices so that $N = TT'$, i.e.,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix} = \begin{bmatrix} c_{11} & & & \\ c_{21} & c_{22} & & \\ \cdots & & & \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ & c_{22} & & c_{n2} \\ & & & \cdots \\ & & & c_{nn} \end{bmatrix}$$

The elements c_{ij} are computed from the following (note that $c_{ij} = 0$ for $j > i$):

$$c_{11} = \sqrt{a_{11}}$$
$$c_{21} = a_{21}/c_{11}$$
$$\vdots$$
$$c_{n1} = a_{n1}/c_{11}$$
$$c_{22} = \sqrt{a_{22} - c_{21}^2}$$
$$c_{32} = (a_{32} - c_{31}\, c_{21})/c_{22}$$
$$\vdots$$
$$c_{n2} = (a_{n2} - c_{n1}\, c_{21})/c_{22}$$

- - - - - - - - -

$$c_{jj} = \sqrt{a_{jj} - c_{j,j-1}^2 - c_{j,j-2}^2 - \ldots - c_{j,1}^2}$$
$$\vdots$$
$$c_{ij} = (a_{ij} - c_{i,j-1}c_{j,j-1} - c_{i,j-2}c_{j,j-2} - \ldots - c_{i,1}c_{j,1})/c_{jj}.$$

As an example, consider

$$N = \begin{bmatrix} 4 & 6 & 8 & 10 \\ 6 & 25 & 20 & 27 \\ 8 & 20 & 36 & 30 \\ 10 & 27 & 30 & 36 \end{bmatrix}$$

Applying the formulas for c_{ij}, we get

$$c_{11} = \sqrt{4} = 2$$
$$c_{21} = 6/2 = 3$$
$$c_{31} = 8/2 = 4$$
$$c_{41} = 10/2 = 5$$
$$c_{22} = \sqrt{25 - (3)^2} = 4$$
$$c_{32} = [20 - 4(3)]/4 = 2$$
$$c_{42} = [27 - 5(3)]/4 = 3$$
$$c_{33} = \sqrt{36 - 2^2 - 4^2} = 4$$
$$c_{43} = [30 - 3(2) - 5(4)]/4 = 1$$
$$c_{44} = \sqrt{36 - 1^2 - 3^2 - 5^2} = 1.$$

This gives

$$N = \begin{bmatrix} 2 & & & \\ 3 & 4 & & \\ 4 & 2 & 4 & \\ 5 & 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 & 5 \\ & 4 & 2 & 3 \\ & & 4 & 1 \\ & & & 1 \end{bmatrix}.$$

The inverse of a triangular matrix

$$T = \begin{bmatrix} c_{11} & & & \\ c_{21} & c_{22} & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

is given by

$$T^{-1} = \begin{bmatrix} b_{11} & & & \\ b_{21} & b_{22} & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

where

$$b_{11} = \frac{1}{c_{11}}$$

$$b_{21} = -(b_{11}\,c_{21})/c_{22}$$

$$b_{31} = -(c_{31}\,b_{11} + c_{32}\,b_{21})/c_{33}$$

$$\vdots$$

$$b_{n1} = -(c_{n1}\,b_{11} + c_{n2}\,b_{21} + \ldots + c_{n,n-1}\,b_{n-1,1})/c_{nn}$$

$$b_{22} = \frac{1}{c_{22}}$$

$$b_{32} = -c_{32}\,b_{22}/c_{33}$$

$$b_{42} = -(c_{42}\,b_{22} + c_{43}\,b_{32})/c_{44}$$

$$\vdots$$

$$b_{n2} = -(c_{n2}\,b_{22} + c_{n3}\,b_{32} + \ldots + c_{n,n-1}\,b_{n-1,2})/c_{nn}$$

$$\cdots$$

$$b_{jj} = \frac{1}{c_{jj}}$$

$$\vdots$$

$$b_{ij} = -(c_{ij}\,b_{jj} + c_{i,j+1}\,b_{j+1,j} + \ldots + c_{i,i-1}\,b_{i-1,j})/c_{ii}.$$

Example:

$$\text{For } T = \begin{bmatrix} 2 & & & \\ 3 & 4 & & \\ 4 & 2 & 4 & \\ 5 & 3 & 1 & 1 \end{bmatrix}.$$

The elements of T^{-1} are

$$b_{11} = \frac{1}{2}$$

$$b_{21} = -\frac{1}{2}\cdot\frac{3}{4} = -\frac{3}{8}$$

$$b_{31} = -\left[4\left(\frac{1}{2}\right) + 2\left(-\frac{3}{8}\right)\right]/4 = -\frac{5}{16}$$

$$b_{41} = -\left[5\left(\frac{1}{2}\right) + 3\left(-\frac{3}{8}\right) + 1\left(-\frac{5}{16}\right)\right]/1 = -\frac{17}{16}$$

$$b_{22} = \frac{1}{4}$$

$$b_{32} = -2\left(\frac{1}{4}\right)/4 = -\frac{1}{8}$$

$$b_{42} = -\left[3\left(\frac{1}{4}\right) + 1\left(-\frac{1}{8}\right)\right]/1 = -\frac{5}{8}$$

$$b_{33} = \frac{1}{4}$$

$$b_{43} = -1\left(\frac{1}{4}\right) = -\frac{1}{4}$$

$$b_{44} = 1.$$

Thus, $T^{-1} = \frac{1}{16}\begin{bmatrix} 8 & & & \\ -6 & 4 & & \\ -5 & -2 & 4 & \\ -17 & -10 & -4 & 16 \end{bmatrix}$

and $N^{-1} = (TT')^{-1} = (T'^{-1})\,(T^{-1})$ gives

$$N^{-1} = \frac{1}{16}\begin{bmatrix} 8 & -6 & -5 & -17 \\ & 4 & -2 & -10 \\ & & 4 & -4 \\ & & & 16 \end{bmatrix}\frac{1}{16}\begin{bmatrix} 8 & & & \\ -6 & 4 & & \\ -5 & -2 & 4 & \\ -17 & -10 & -4 & 16 \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 414 & 156 & 48 & -272 \\ 156 & 120 & 32 & -160 \\ 48 & 32 & 32 & -64 \\ -272 & -160 & -64 & 256 \end{bmatrix}.$$

6-9.3 REMARKS

By forming the matrix product

$$\begin{bmatrix} X' \\ Y' \end{bmatrix}(X,Y) = \begin{bmatrix} X'X & X'Y \\ Y'X & Y'Y \end{bmatrix}$$

and replacing $Y'X$ by 0 (a null matrix) and $Y'Y$ by I (the identity matrix), we obtain

$$N = \begin{bmatrix} X'X & X'Y \\ 0 & I \end{bmatrix}.$$

In this form, Y may be a single vector of observations $Y' = (Y_1 Y_2 \ldots Y_n)$, or a set of p vectors

$$Y = \begin{bmatrix} Y_{11} & \cdots & Y_{1p} \\ Y_{21} & & Y_{2p} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ Y_{n1} & & Y_{np} \end{bmatrix}.$$

Then,

$$N^{-1} = \begin{bmatrix} (X'X)^{-1} & -\beta \\ 0 & I \end{bmatrix},$$

where I is $p \times p$ and 0 is $p \times k$, gives all the values needed for the computations of this Paragraph.

REFERENCES

1. M. G. Kendall and W. R. Buckland, *A Dictionary of Statistical Terms*, Oliver and Boyd, Ltd., Edinburgh, 1954.
2. R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., New York, N.Y., 1952.
3. O. L. Davies, (ed.), *The Design and Analysis of Industrial Experiments*, Oliver and Boyd, Ltd., Edinburgh, 1954.
4. O. Taussky, (ed.), *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, (National Bureau of Standards Applied Mathematics Series, No. 39), U. S. Government Printing Office, Washington, D. C., 1954.
5. P. S. Dwyer, *Linear Computations*, John Wiley & Sons, Inc., New York, N.Y., 1951.
6. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (5th edition), Oliver and Boyd, Ltd., Edinburgh, 1957.
7. R. L. Anderson and E. E. Houseman, *Tables of Orthogonal Polynomial Values Extended to N = 104*, Research Bulletin No. 297, Agricultural Experiment Station, Iowa State College, Ames, Iowa, 1942.
8. D. B. DeLury, *Values and Integrals of the Orthogonal Polynomials up to n = 26*, University of Toronto Press, 1950.
9. E. S. Pearson and H. O. Hartley, (eds.), *Biometrika Tables for Statisticians*, Vol. I (2d edition), Cambridge University Press, Cambridge, England, 1958.

(AMCRD-TV)

FOR THE COMMANDER:

LEO B. JONES
Major General, USA
Chief of Staff

OFFICIAL:

P. R. HORNE
COL, GS
Chief, Administrative Office

DISTRIBUTION:
Special

ENGINEERING DESIGN HANDBOOKS

Listed below are the Handbooks which have been published or are currently under preparation. Handbooks with publication dates prior to 1 August 1962 were published as 20-series Ordnance Corps pamphlets. AMC Circular 310-38, 19 July 1963, redesignated those publications as 706-series AMC pamphlets (e.g., ORDP 20-138 was redesignated AMCP 706-138). All new, reprinted, or revised Handbooks are being published as 706-series AMC pamphlets.

No.	Title
100	*Design Guidance for Producibility
104	*Value Engineering
106	Elements of Armament Engineering, Part One, Sources of Energy
107	Elements of Armament Engineering, Part Two, Ballistics
108	Elements of Armament Engineering, Part Three, Weapon Systems and Components
110	Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data
111	Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data
112	Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments
113	Experimental Statistics, Section 4, Special Topics
114	Experimental Statistics, Section 5, Tables
115	Basic Environmental Concepts
116	*Basic Environmental Factors
120	*Design Criteria for Environmental Control of Mobile Systems
121	Packaging and Pack Engineering
123	*Hydraulic Fluids
125	Electrical Wire and Cable
127	*Infrared Military Systems, Part One
128(S)	*Infrared Military Systems, Part Two (U)
130	Design for Air Transport and Airdrop of Materiel
134	Maintainability Guide for Design
135	Inventions, Patents, and Related Matters
136	Servomechanisms, Section 1, Theory
137	Servomechanisms, Section 2, Measurement and Signal Converters
138	Servomechanisms, Section 3, Amplification
139	Servomechanisms, Section 4, Power Elements and System Design
140	Trajectories, Differential Effects, and Data for Projectiles
145	*Dynamics of a Tracking Gimbal System
150	Interior Ballistics of Guns
160(S)	Elements of Terminal Ballistics, Part One, Kill Mechanisms and Vulnerability (U)
161(S)	Elements of Terminal Ballistics, Part Two, Collection and Analysis of Data Concerning Targets (U)
162(S-RD)	Elements of Terminal Ballistics, Part Three, Application to Missile and Space Targets (U)
165	Liquid-Filled Projectile Design
170(C)	Armor and Its Application to Vehicles (U)
175	Solid Propellants, Part One
176(C)	Solid Propellants, Part Two (U)
177	Properties of Explosives of Military Interest
178(C)	†Properties of Explosives of Military Interest, Section 2 (U)
179	Explosive Trains
180	*Principles of Explosive Behavior
185	Military Pyrotechnics, Part One, Theory and Application
186	Military Pyrotechnics, Part Two, Safety, Procedures and Glossary
187	Military Pyrotechnics, Part Three, Properties of Materials Used in Pyrotechnic Compositions
188	*Military Pyrotechnics, Part Four, Design of Ammunition for Pyrotechnic Effects
189	Military Pyrotechnics, Part Five, Bibliography
190	*Army Weapon System Analysis
195	*Development Guide for Reliability, Part One
196	*Development Guide for Reliability, Part Two
197	*Development Guide for Reliability, Part Three
198	*Development Guide for Reliability, Part Four
199	*Development Guide for Reliability, Part Five
200	*Development Guide for Reliability, Part Six
201	*Rotorcraft Engineering, Part One, Preliminary Design
202	*Rotorcraft Engineering, Part Two, Detail Design
203	*Rotorcraft Engineering, Part Three, Qualification Assurance
205	*Timing Systems and Components
210	Fuzes
211(C)	Fuzes, Proximity, Electrical, Part One (U)
212(S)	Fuzes, Proximity, Electrical, Part Two (U)
213(S)	Fuzes, Proximity, Electrical, Part Three (U)
214(S)	Fuzes, Proximity, Electrical, Part Four (U)
215(C)	Fuzes, Proximity, Electrical, Part Five (U)
235	*Hardening Weapon Systems Against RF Energy
239(S)	*Small Arms Ammunition (U)
240(C)	Grenades (U)
241(S)	*Land Mines (U)
242	Design for Control of Projectile Flight Characteristics
244	Ammunition, Section 1, Artillery Ammunition--General, with Table of Contents, Glossary and Index for Series
245(C)	Ammunition, Section 2, Design for Terminal Effects (U)
246	†Ammunition, Section 3, Design for Control of Flight Characteristics
247	Ammunition, Section 4, Design for Projection
248	†Ammunition, Section 5, Inspection Aspects of Artillery Ammunition Design
249	Ammunition, Section 6, Manufacture of Metallic Components of Artillery Ammunition
250	Guns--General
251	Muzzle Devices
252	Gun Tubes
255	Spectral Characteristics of Muzzle Flash
260	*Automatic Weapons
270	Propellant Actuated Devices
280	Design of Aerodynamically Stabilized Free Rockets
281(S-RD)	Weapon System Effectiveness (U)
282	†Propulsion and Propellants
283	Aerodynamics
284(C)	Trajectories (U)
285	Elements of Aircraft and Missile Propulsion
286	Structures
290(C)	Warheads--General (U)
291	Surface-to-Air Missiles, Part One, System Integration
292	Surface-to-Air Missiles, Part Two, Weapon Control
293	Surface-to-Air Missiles, Part Three, Computers
294(S)	Surface-to-Air Missiles, Part Four, Missile Armament (U)
295(S)	Surface-to-Air Missiles, Part Five, Countermeasures (U)
296	Surface-to-Air Missiles, Part Six, Structures and Power Sources
297(S)	Surface-to-Air Missiles, Part Seven, Sample Problem (U)
327	Fire Control Systems--General
329	*Fire Control Computing Systems
331	Compensating Elements
335(S-RD)	*Nuclear Effects on Weapon Systems (U)
340	Carriages and Mounts--General
341	Cradles
342	Recoil Systems
343	Top Carriages
344	Bottom Carriages
345	Equilibrators
346	Elevating Mechanisms
347	Traversing Mechanisms
350	*Wheeled Amphibians
355	The Automotive Assembly
356	Automotive Suspensions
357	*Automotive Bodies and Hulls

* *UNDER PREPARATION--not available*
† *OBSOLETE--out of stock*