WAPD-TM-1292
DOE RESEARCH AND
DEVELOPMENT REPORT

STATISTICS FOR NUCLEAR ENGINEERS AND SCIENTISTS
PART 1: BASIC STATISTICAL INFERENCE

FEBRUARY 1981

CONTRACT DE-AC11-76PN00014

BETTIS ATOMIC POWER LABORATORY
WEST MIFFLIN, PENNSYLVANIA

Operated for the U. S. Department of Energy by
WESTINGHOUSE ELECTRIC CORPORATION

DISCLAIMER

WAPD-TM-1292

STATISTICS FOR NUCLEAR ENGINEERS AND SCIENTISTS

PART 1: BASIC STATISTICAL INFERENCE

William J. Beggs

February, 1981

Contract No. DE-AC11-76PN00014

Printed in the United States of America
Available from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151

This document is an interim memorandum prepared primarily for internal reference and does not represent a final expression of the opinion of Westinghouse. When this memorandum is distributed externally, it is with the express understanding that Westinghouse makes no representation as to completeness, accuracy, or usability of information contained therein.

BETTIS ATOMIC POWER LABORATORY
West Mifflin, Pennsylvania

Operated for the U.S. Department of Energy by
WESTINGHOUSE ELECTRIC CORPORATION

79-778/0152L

NOTICE

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

PREFACE

When one draws conclusions from data, he is knowingly or unknowingly using statistics. How good these conclusions are will depend upon how good were the statistical techniques used in working up the data. Similarly, a person who plans an experiment or any collection of data is also using statistics in an area known as the design of experiments. Here again, how good the experiment is will depend upon the statistical design techniques used. The purpose of this text is to provide the nuclear engineer or scientist with a basic course in statistical design and analysis. Sufficient extra advanced material is included to enable a student who has completed the basic course to deal with problems in selected subject areas. Specifically, this text has served as a basis for several training courses at the Bettis Atomic Power Laboratory.

The text is divided into two parts. Part 1, entitled Basic Statistical Inference, deals with the basic language and concepts of statistical analysis. It covers in seven chapters descriptive statistics, probability, simple inference for normally distributed populations, and for non-normal populations as well, comparison of two populations, the analysis of variance, quality control procedures, and linear regression analysis. Chapters 1, 2, 3, and 6 have been used for a short (20 hours) course in basic inference at Bettis Atomic Power Laboratory. Chapters 4 and 5 can be used to present a short course on quality control, and sampling plans. Chapter 7 has been used to present a short course on regression analysis and model building. A semester length course (32 hours) is presented which includes material from all 7 chapters of Part 1, with emphasis on Chapters 3, 6, and 7, and with selected additional material on experimental designs from Part 2.

Part 2, Design of Experiments, will contain material on the philosophy of experimental designs, completely randomized designs, balanced block designs, incomplete block designs, nested or hierarchical designs, factorial designs, and response surface methodology. Two or more short courses or a semester course could be presented from this material.

In both parts of the text, sections which are either more complex or theoretical than that usually covered in a first course are indicated with an asterisk so that the reader may skip over them on first reading. Five appendices are also presented to provide additional theoretical information to the interested student. Furthermore, a particularly useful feature of this text is the collection of 17 tables of various types included with the text material. These tables should satisfy the majority of applications that an engineer or scientist faces.

INTENTIONALLY BLANK

STATISTICS FOR NUCLEAR ENGINEERS AND SCIENTISTS

CONTENTS

*Reader may prefer to skip these sections on first reading.

CONTENTS (cont'd)

CONTENTS (cont'd)

CONTENTS (cont'd)

ABSTRACT

This report is intended for the use of engineers and scientists working in the nuclear industry, especially at the Bettis Atomic Power Laboratory. It serves as the basis for several Bettis in-house statistics courses. The objectives of the report are to introduce the reader to the language and concepts of statistics and to provide a basic set of techniques to apply to problems of the collection and analysis of data. Part 1 covers subjects of basic inference.

Part 1: BASIC STATISTICAL ANALYSIS

CHAPTER 1. INTRODUCTION: STATISTICS AND DATA

1.0 INTRODUCTION

What is statistics? To some engineers and scientists it is an esoteric or suspicious subject. This attitude results from a lack of understanding of the purpose and usefulness of statistical analyses and is due, no doubt in part, to the misleading claims made by some advertisers and salesmen in their attempt to impress their clientele with graphical displays and scientific terminology. It is hoped that this text will help reduce or eliminate suspicion as to the usefulness of statistics and, at the same time, develop in the reader a healthy skepticism as to the real meaning of all reported values.

In particular, it will be shown that statistics is not simply a tedious chore of number manipulation; but that, in deed, the smaller the data base, the more important and valuable is the proper application of statistics. To accomplish this will require the achievement of three basic objectives:

1. To introduce the reader to the language of statistics.
2. To introduce and have the reader understand the basic underlying concepts of statistical analysis.
3. To provide the reader with a basic set of techniques to use on his own problems of planned collection and analysis of data.

Of course, one cannot become a statistician simply by mastering the material in this book. Many sophisticated techniques have been developed that are beyond the scope of this text. However, it is hoped that this text lays the statistical foundation from which an interested reader can build his knowledge in the direction in which his work leads him.

1.1 Statistics Defined

To return to the question, statistics is many things to many people. The renowned statistician, Dr. George E.P. Box, has been heard to say "Statistics is what statisticians do", and he was only half joking. The idea behind such a circular definition is that statisticians do get involved in many areas and cross many scientific, economic, and social disciplines in performing their tasks. To some of the above, the statistician may be a mathematician; to the mathematician, he may appear to be a follower of recipes (an "engineer of numbers"). To the small daughter of a statistician, her daddy was "a doctor of sick numbers". In fact, he is none and all.

More formally statistics can be defined as "A SCIENCE WHICH GAINS KNOWLEDGE OF PHENOMENA BY INFERENCE IN THE PRESENCE OF UNCERTAINTY." The three key words in this definition are underlined. Science implies a specific body of logical laws or theorems and a specific body of methodology. Statistics has such laws and methodology. Although it may be argued that the laws are those of mathematics, it is the application of the laws to real-life situations that separates statistics from mathematics. This application is inference. Statistics infers a state of nature based on observations and conjectures in order to aid in a decision-making process. This is opposed to probability, a branch of mathematics, which deduces certain events given an assumed state of nature. The third key word in the definition of statistics is uncertainty. It is the position of this text that nature is not

deterministic. Although it is possible and even very likely that some true physical relationships exist, man is either not allowed to observe them or is unable to observe them, or both. As the history of science has shown over and over again, we can never be sure we have absolute knowledge. There is always a chance of arriving at a wrong conclusion regardless of how strong the evidence appears. Statistics deals with this chance of erroneous conclusions. One skeptic has sarcastically defined statistics as "a science in which you are wrong 5 percent of the time." Actually there is some wisdom in that definition, although not in the context in which it was meant. The point is that there is uncertainty in all human observations, and statistics attempts to deal with this uncertainty in a quantified manner and with a specialized body of knowledge.

Having established that statistics is a science, let it be now known that statistics is also an art, or rather, that the application of statistics is an art. In dealing with any problem there is difficulty in defining the real issue and of determining what techniques to apply in the solution to the problem. Thus, the statistician in consultation with his client or the engineer or scientist himself must develop the art of asking the right questions and of providing the proper tools to solve the problem.

1.2 The Scope of Statistics

Statistics, like mathematics, is unique in that it is applied only in reference to some other discipline or field of endeavor. Physical and biological scientists, engineers of all types, sociologists, psychologists, economists, agronomists, and pollsters all need to deal with statistical information at one time or another. Statistics is even used to decide the authorship of historical papers and the age and origin of bone fossils. Actual cases from these and other disciplines can be found in an excellent book for the layman entitled "Statistics: A Guide to the Unknown" [31].*

In fact, whenever one deals with data of any kind, he is dealing with statistics. It is not a question of whether or not to use statistics; it is a question of whether one uses good and proper statistical techniques. To most people, statistics is synonymous with data analysis. It is certainly true that data analysis is a major part of what statisticians do. We will in Part 1 of this text discuss analytical techniques to describe a collection of data, to test hypotheses about the source of that data, to compare two or more collections of data for equivalence, to fit curves and multivariable functions, and to establish limits for the control of production processes. These techniques are useful to scientists from the basic researcher to the production manager.

In addition to data analysis and the inferences drawn from the data, a second equally important aspect of statistics exists which is not readily recognized as being in the domain of statistics. This area is that of data collection, or more accurately, the planning of the collection of data. This area is generally known as the Design of Experiments, where "experiments" is taken in a general sense to mean any collection of data. As with computers, the cliche "garbage in-garbage out" applies to scientific studies as well.

*Number in brackets refer to reference.

Design of Experiments deals with the efficient utilization of experiments to obtain the maximum information. It is in the efficient use of experiments, and therefore in the efficient use of time and money, in which many scientists fail to apply good statistics. Basic techniques of good "experimental" design will be discussed as necessary in Part 1 of this text, since there is often a direct connection between proper design and proper analysis. Part 2 will be devoted to more detailed discussion of the concepts, techniques, and applications of the area of design of experiment.

In summary, then, statistics can be described as the proper planning and analysis of a collection of data or experiment from any source. It incorporates mathematics, scientific theory, and empiricism. It fits well with the spiraling path of knowledge gained known as the Scientific Method (see Figure 1.1). First a conjecture or hypothesis about the state of nature is made based on theory or past observations or both. Then a plan of investigation, i.e., an experiment, is developed and carried out. The data obtained are analyzed in a proper manner and inferences made. This leads to a new conjecture, which hopefully is a more accurate description of the true state of nature under examination. Statistical analysis, though not a

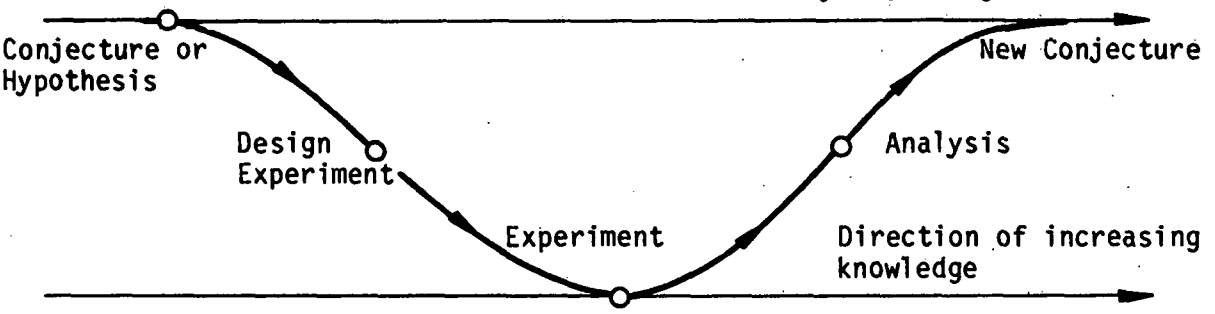


Figure 1.1., The Scientific Method

decision-making process by itself, is then an extremely useful tool for making decisions in a quantitative manner.

1.3 Statistics in the Nuclear Industry

In this section some simple but typical examples of the application of statistics in the nuclear industry will be discussed. It should be noted that the techniques used are not unique to the nuclear industry, but are, in general, equally applicable in any other industry. Many of the examples that appear throughout this text are based on the commercial applications of nuclear energy; i.e., pressurized water reactors (PWR).

1. Comparison of Coolant Chemistries on Material Corrosion

In the area of basic research, a typical task is the study of the effect of different primary or secondary coolant chemistries on material corrosion. Several different chemical compositions of coolants may be tested in a test facility for a given period of time such that equal information about all chemistry compositions is obtained. The data for each coolant tested may be characterized in a statistical sense by computing the arithmetic average and a measure of spread, called the standard deviation. A confidence interval based on the available data may be constructed which is said to

contain the true coolant corrosion value with some level of probability or confidence. Comparisons among the averages of the coolants may be made using a technique known as the analysis of variance and using an F-test. Finally, the experiment may be repeated over longer time periods and the relationship of corrosion as a function of time is estimated.

2. Combining of Uncertainties to Obtain a Design Limit

In the area of developing design limits for plant assembly the uncertainty involved in several components and assembly operations may be combined to produce a limit which can be expected to be exceeded with a satisfactorily small probability. The individual factors may be manufacturing tolerances on diameters of particular components, the eccentricity of the centers of components meant to be concentric, and the misalignment possible or likely to occur during plant assembly.

Physicists are concerned about setting design limits on a variety of variables that could affect core performance. An entire chapter in nuclear design manuals may be used to discuss the handling of uncertainties of manufacturing and inspection data, operational uncertainties, and design model uncertainties.

3. Qualification of Tools, Materials, and Vendors

In developmental work it is often necessary to determine the right tools or materials to use or to establish the capability of a technician or vendor to perform an assigned task. Tests can be planned to evaluate such things that eliminate undesired sources of variability and allow clear analysis of the object of interest.

An example of a program to evaluate materials and processing variables is the analysis of the strength of zircaloy ingots. One or more vendors may need to be evaluated and distinction between their respective capabilities to produce acceptable material may be required. The composition of the ingots may be varied to try to obtain the best possible ingot characteristics. Process variables may include the number of rolls performed and the force exerted on each roll, the temperatures at which various operations are performed, and the time for which the ingot is kept at each temperature. All of these variables could be analyzed in one carefully planned multifactor experiment and could include other response variables, besides strength, as well.

4. Evaluation of Product Lot

In production activities it is essential to produce and continue producing acceptable material. One way to achieve this is through a sampling plan. Each lot of product, such as fuel elements, is sampled, and each sample element is tested for acceptability. If too many rejectable elements are found, the entire lot of elements is rejected, and appropriate action must be taken to adjust the production process. Another procedure for assuring continuing acceptable product is a control chart. Once a production process is determined to be in control, i.e., producing acceptable product, control limits are developed such that if any future observation falls outside these limits, the process is declared out-of-control and corrective action is

taken. The control limits may be based on the variability of the average of several elements or on the availability of individual elements. Process variables of interest may include such things as element dimensions, loading and cracks.

5. Use of Operating Plant Data

During the actual operation of a plant, data can and should be taken and analyzed to monitor the behavior of all operating systems. This data can be used to detect potential difficulties before they become serious, to identify the source of abnormalities, to evaluate the effectiveness of specific programs or techniques, and to make adjustments in operation procedures as required. Inferences from operating plant data are particularly difficult to make because the data does not, in general, come from carefully planned and executed experiments. Thus, unidentified sources of variability may exist in the data which confound the effects of the variable under study.

1.4 Descriptive Statistics

In all cases of statistical analysis, regardless of how well the experiment was planned, a sensible first step is to summarize the data in some way so that the main characteristics of the data are apparent. The natural step to most engineers or scientists is to plot the data in some way. In addition to an informative plot of the data it is useful to quantify some of the characteristics of the data, such as the center, midpoint or most frequently observed point of the data, the spread of the data, and perhaps the degree of asymmetry and peakedness of the data. All of these data characteristics are in the domain of statistical analysis known as descriptive statistics.

In the remaining sections of this chapter some procedures for plotting the data and for obtaining descriptive statistics will be discussed. It is advantageous at this point, however, to distinguish between a population and a sample. A population is a collection of all possible units having a certain attribute, such as living in the United States, or having been produced by a certain vendor under a certain contract. Some populations are finite in number, whereas some are infinite in size. The possible outcomes from the measurements of the length of a fuel element is an example of an infinite population, since there are an infinite number of positions between any two points on a continuous scale. (In practice, however, since all observations are read to the nearest unit, the collection of observations is discrete.) A sample is a subset of a population, is necessarily finite, and consists of specific values for each individual in the sample. It is significant to remember, however, that before a sample is actually taken, its value is unknown.

1.5 Observations Vary

Flip a coin. Ignoring the possibility that it lands on edge, what do you observe? Either a head or a tail. Suppose the result of the first flip was a head. Flip again. Head or tail? Repeat several times. Almost surely you will observe some heads and some tails. That is, the result of flipping a coin varies from one trial to another (unless you are using a two-headed or two-tailed coin!). Moreover, the outcome, head or tail, varies in a random

fashion, i.e., in an unpredictable way. Thus, the outcome of a coin flip is called a random variable, since the realization of the act of flipping varies in a random fashion. Once observed, the result is a fixed quantity, a data point.

How do we characterize the random variable resulting from flipping a coin? We would intuitively expect that for a fair coin, half of an even number of trials will result in a head and half in a tail. Suppose we flipped a coin 100 times and observed 43 heads and 57 tails. Is the coin not fair? In fact, obtaining 43 heads in 100 flips of a fair coin is a reasonable result. A second experiment of 100 trials may result in 54 heads. Observations vary from one trial to another; so the results of 100 trials can also be expected to vary from one time to another. One of the important questions to be answered is how do these observations vary? In what manner are the results of random variables distributed along the line of possible results?

One attempt to answer this question is to take many observations and let your data tell you how they are distributed. This empirical approach is simply to plot the observations.

1.5.1 A Dot Design

Let us consider the following example of observations from a delayed neutron gage on 100 pellets of uranium fuel for a commerical power reactor.

Table 1.1. Delayed Neutron Counts* of Fuel Pellets (counts/gm)

96	88	93	112	102	116	119	93	116	119
89	98	92	97	121	117	108	104	113	94
103	98	106	118	120	105	100	102	106	105
109	97	95	102	104	122	94	109	113	111
100	96	114	98	121	109	110	97	92	106
109	101	121	99	94	113	107	114	96	92
101	103	101	95	98	101	98	106	102	96
95	104	102	116	96	100	107	99	95	110
113	100	97	102	112	97	106	91	100	97
91	100	92	97	80	101	99	99	104	101

*Coded to make computation easier by subtracting 24,000 from the actual count of each pellet.

One way to examine the variability of the data is to plot each observation in a dot diagram (see Figure 1.2). For a sufficiently large number of observations a dot diagram may leave the viewer with a good understanding of the nature of the response being plotted. However, more often a distortion of the true state of nature may be caused by the discreteness of the data and the dot diagram procedure when in fact the true response is a continuous variable. That is, theoretically, the response may take on any value in a specified range of values, but the dot diagram only represents a part of the total possible outcomes. In particular, two things may happen. First, the dot diagram may be scattered over a wide range with only one or two observations at any one value. This may lead to the false impression that any one value is as likely to occur as any other, an

assumption which in most situations is far from the truth. The second possibility of deception in a dot diagram is the tendency for people to record the same value repeatedly. What this phenomenon means is that many observations are in the vicinity of these values but are really spread about these values. Thus, a more realistic picture of the true distribution of a set of data would be given by smoothing out of the dot diagram. This step is accomplished by a histogram.

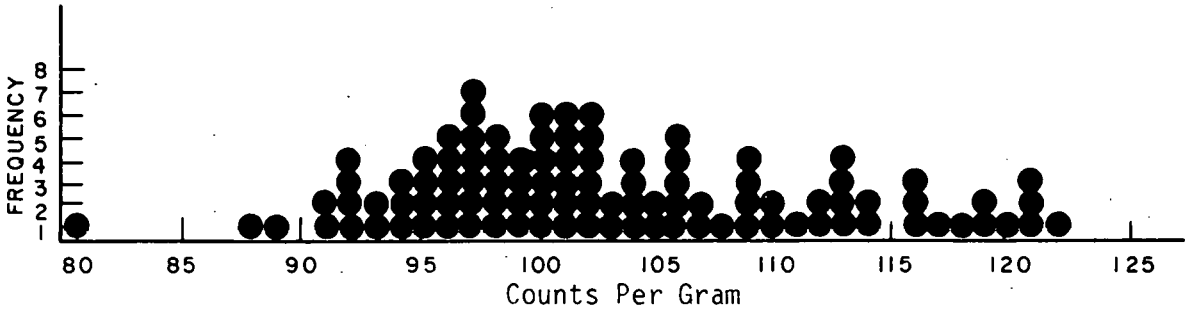


Figure 1.2. Dot Diagram: Delayed Neutron Counts of Fuel Pellets

1.5.2 Plotting a Histogram

A histogram is simply a bar chart of frequencies of occurrence of values in specified intervals.

The first step in constructing a histogram is to decide on the number and width of intervals. Miller and Freund [25] recommend $5 \leq k \leq 15$ intervals, where k is the number of intervals. Other authors may suggest slightly different numbers of intervals. The decision on the number of intervals is usually based on the range and number of data, i.e., the maximum observation minus the minimum observation. In the delayed neutron count data, the range is 122 - 80 = 42. Using the above rule of thumb, we could use 11 intervals with a width of 4 counts, or 9 intervals with a width of 5 counts. The choice of interval limits or boundaries may also be made in several ways. To ensure nonoverlapping intervals we could for the delayed neutron example choose intervals such as 89.5-94.5, or 89.9-94.9 or 90-94. Table 1.1 gives the frequencies for the latter choice of interval limits, reflecting the discreteness of the recorded data.

The class mark $\bar{x}_i$ is the midpoint of the interval and can be used to represent the entire interval, as we shall see shortly. Figure 1.3 shows the histogram for the frequency data given in Table 1.2. The straight lines connecting the midpoints of the intervals are called the frequency polygon.

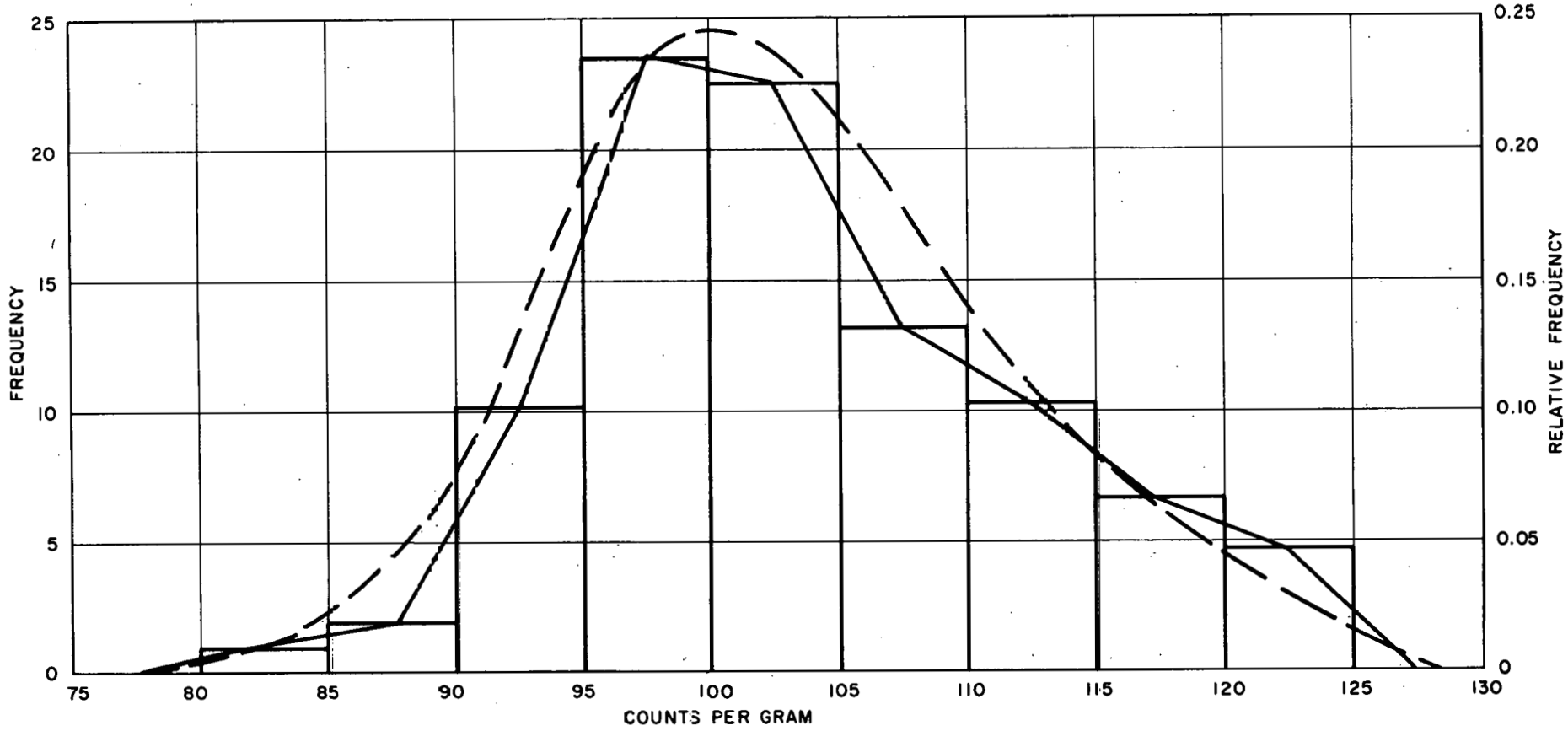


Figure 1.3. Histogram: Delayed Neutron Counts of Fuel Pellets

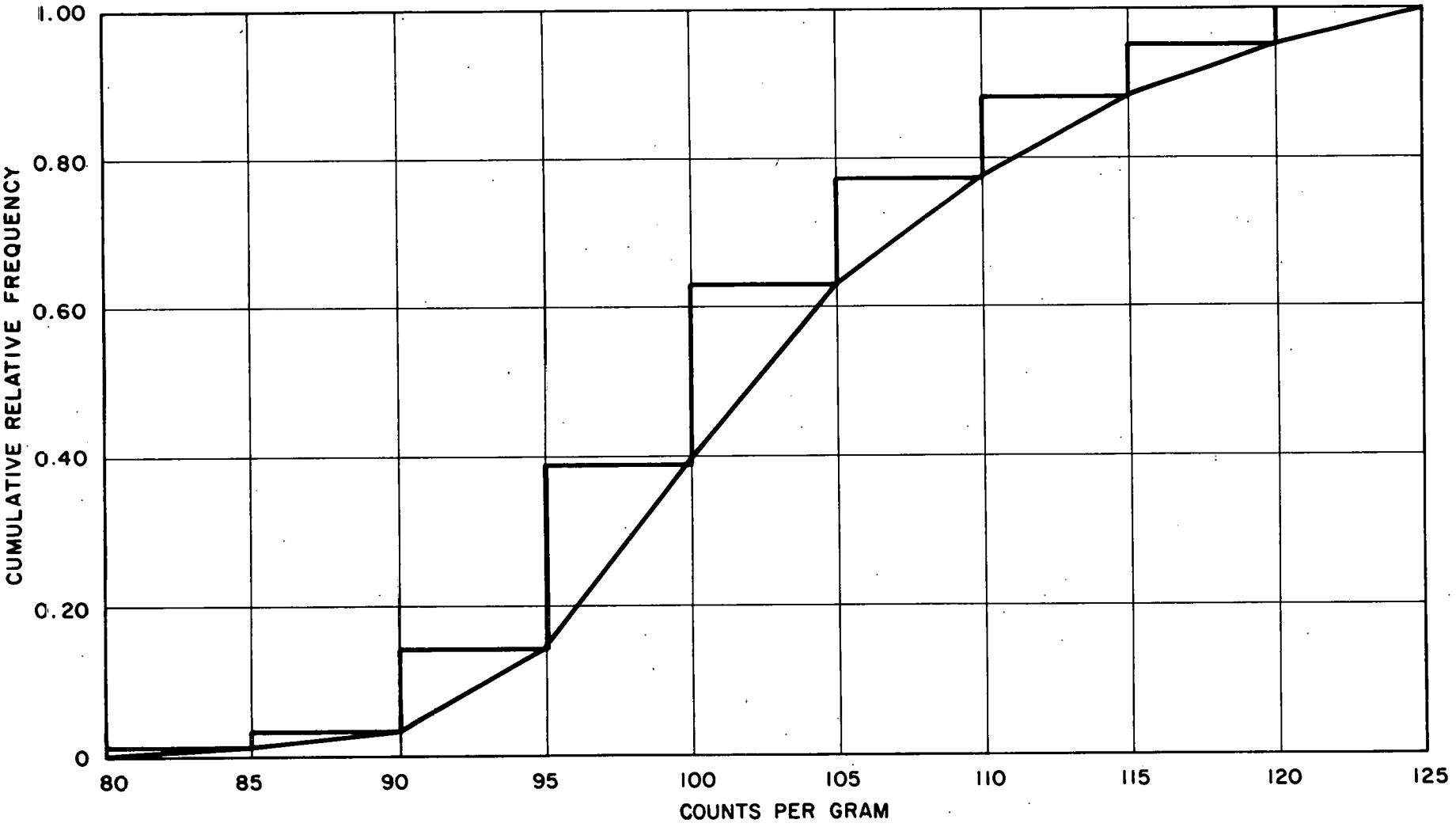


Figure 1.4. Cumulative Distribution for Delayed Neutron Counts

Table 1.2. Frequency Table of Delayed Neutron Counts

Interval	Class Mark x_i	Frequency	Relative Frequency	Cumulative Relative Frequency
80-84	82	1	0.01	0.01
85-89	87	2	0.02	0.03
90-94	92	11	0.11	0.14
95-99	97	25	0.25	0.39
100-104	102	24	0.24	0.63
105-109	107	14	0.14	0.77
110-114	112	11	0.11	0.88
115-119	117	7	0.07	0.95
120-124	122	5	0.05	1.00

If we record the relative frequency of each interval as n_i/N, where n_i is the frequency in the ith interval and $N = \sum_{i=1} n_i$ is the sum total of points, we may plot the cumulative relative frequency diagram. A series of straight lines connecting the interval limits of the cumulative relative frequency diagram is called an ogive. The cumulative relative frequency diagram and ogive is shown in Figure 1.4.

1.6 The Characterization of Data

What is of more interest than a frequency polygon or an ogive, however, is the kind of smooth, continuous curve which may be drawn through the histogram. The histogram gives us an idea as to the distribution of values being measured. Its main job, however, is simply to point out the central tendencies in the data and to indicate the degree of variation inherent in the data. If we believe that Mother Nature has a game plan for delayed neutron counts/gram (and everything else for that matter), then there exists a true distribution for these counts/gram. The 100 observations we recorded are a sample from all conceivable observations of delayed neutron counts/gram of fuel pellets which we may make if we continued to observe a large number of pellets from the population of such pellets.

Thus, we have sampled from a population and constructed a histogram or picture of the distribution of events recorded. This histogram is an estimate of the real distribution of events. How do we characterize the distribution of this population?

1.6.1 Location Parameters

The term parameter refers to characertistics of the true underlying population whose distribution we are trying to describe. Information about these parameters is available through functions, called statistics, of the sample data. A statistic summarizes the information about a parameter of a distribution that is available in a collection of sample data.

Since we are after the central tendencies, the first step in

characterizing the distribution is to report its location or center value. Actually, there are several possible choices for determining this location parameter. In every histogram there is one value which divides the observations in half. The median of the sample is that value for which half of the observations is above and half below. If there are an odd number of observations, say $2n + 1$, this median value is unique, having exactly n observations on either side. If there is an even number, the median may be the common value of the two middle observations, or the midpoint between the two middle values.

A second location parameter is the mode. This is the value having the greatest frequency of occurrences. The sample mode is the highpoint in the histogram. For the delayed neutron gage data, a quick glance at Figure 1.3 and 1.4 tells us that the sample median is in the interval 100-104 and the mode is in the interval 95-99. For the sample mode, it is usually sufficient to use the midpoint or class-mark of the interval, 97, as the representative value. The sample median can be determined more precisely by counting the observations and averaging the 50th and 51st largest values. Or, we can approximate it by interpolation. From the frequency table, we see that 39 observations lie below 100 and 24 lie in the interval 100-104. Eleven of these lie below the sample median and 13 above. Thus the median could be represented conceptually by the average of the 11th and 12th observation in this interval, assuming all 24 observations were equally spaced within the interval. Since the interval is 5 units in width, the sample median can be determined as follows:

$$\frac{11 + 12}{2} \times \frac{5}{24} = 2.4$$

$$100 + 2.4 = 102.4$$

[Note from the dot diagram, that both the 50th and 51st observation are equal to 102, so that the median of the data is exactly 102.]

The median and mode are interesting and useful location parameters, but the most valuable location parameter is the mean, or center of gravity of the distribution. The mean is the first moment of the distribution and is estimated by the arithmetic average of the observations. Using the frequency table constructed for the histogram, however, we can provide an alternative estimate of the mean. Let $\bar{x}_i$ be the classmark of the ith interval and n_i the number of observations lying in that interval, then

$$\bar{\bar{x}} = \frac{1}{\Sigma n_i} \sum_{i=1}^{k} n_i \bar{x}_i \qquad (1.1)$$

where $\bar{\bar{x}}$ is the average of all observations. It sometimes occurs that the end interval in a frequency table or histogram is open ended; e.g., instead of an observation being between 80 and 84, inclusive, we group several intervals having none or very few observations together and simply record it as representing values less than 80. These open-ended intervals have arbitrary classmarks.

1.6.2 Measuring the Spread

Knowing the location of the central value (mean, mode, or median) of the histogram is not sufficient information for making inferences about the true distribution. We also need to know how varied or spread out the observations are. A measure of variation is called the variance and is usually estimated by

$$s^2 = \sum_{u=1}^{n} \frac{(x_u - \bar{x})^2}{n-1}$$

where the x_u are the actual observations, $\bar{x}$ is the average, and n is the number of observations. For large n the calculation is obviously cumbersome. A quicker calculation which makes use of the data already tabulated in Table 1.2 is to calculate

$$s^2_g = \frac{\Sigma n_i (\bar{x}_i - \bar{\bar{x}})^2}{n-1} \qquad (1.2)$$

$$= \frac{\Sigma n_i \bar{x}_i^2 - [(\Sigma n_i \bar{x}_i)^2/n]}{n-1}$$

where $\bar{x}_i$ is the class mark of the ith interval and the summation is over the number of intervals.

A measure of the spread of a distribution is the standard deviation, which is simply the square root of the variance and is estimated by $s = \sqrt{s^2}$.

1.6.3 Other Characteristics of Interest

There are two other characteristics of a distribution which are of general interest in describing a set of data: skewness and kurtosis.

Skewness is a measure of the asymmetry of the distribution, i.e., the tendency of a population toward high or low values. The usual measure of skewness can be written in terms of data from a histogram as

$$\frac{1}{n-1} \Sigma_i n_i (\bar{x}_i - \bar{\bar{x}})^3 / s_g^3$$

For a symmetric distribution, this measure should be zero.

Kurtosis is often described as the measure of peakedness of the distribution and can be calculated by

$$\frac{1}{n-1} \Sigma_i \, n_i (\bar{x}_i - \bar{\bar{x}})^4 / s_g^4$$

The flatter the distribution, the smaller the value becomes. However, the concept of kurtosis is not frequently applied when dealing with most sets of data.

All of the characteristics discussed in this section will be described more formally in the section on probability distribution functions.

1.7 Estimating the Mean and Variance: The Delayed Neutron Gage Example

The mean and variance of the delayed neutron gage data could be estimated by using all data points and calculating $\bar{x}$ and s^2 directly. The results of such a calculation are

$$\bar{x} = 102.99$$

$$s^2 = 75.53$$

Using the grouping of Table 1.2, we may estimate the grouped mean, $\bar{x}_g$, and variance, s_g^2, as follows:

$\bar{x}_i$	n_i	$\bar{x}_i n_i$	$\bar{x}_i^2 n_i$
82	1	82	6724
87	2	174	15138
92	11	1012	93104
97	25	2425	235225
102	24	2448	249696
107	14	1498	160286
112	11	1232	137984
117	7	819	95823
122	5	610	74420
	100	10300	1068400

From (1.1) $\bar{x}_g = 103.00$

From (1.2) $$s_g^2 = \frac{[1068400 - [(10300)^2/100]}{99}$$

$$= 75.76$$

An easier approach for calculating the estimate of variance is to code the data. That is, let u_i be the coded value obtained by

$$u_i = \frac{\bar{x}_i - x_o}{c} = \frac{\bar{x}_i - 102}{5}$$

where a convenient center value x_o is $\bar{x}_5 = 102$, and c is a convenient scale value.

Thus, $$x_i = 5\, u_i + 102.$$

Then

u_i	n_i	$u_i n_i$	$u_i^2 n_i$
-4	1	-4	16
-3	2	-6	18
-2	11	-22	44
-1	25	-25	25
0	24	0	0
1	14	14	14
2	11	22	44
3	7	21	63
4	5	20	80
	100	20	304

and $\bar{u} = 20/100 = 0.2$

$$s_u^2 = \frac{304 - (20)^2/100}{(99)} = 3.03$$

Then $\bar{\bar{x}} = c\bar{u} + x_o \quad , c = 5$

$= 5 \times 0.2 + 102$

$= 103.0$

and $s_x^2 = c^2 s_u^2$

$= (5)^2(3.03)$

$= 75.75$

(The interested reader will want to read Section 2.3.3 on the basic properties of means and variances in order to understand the relationships between $\bar{\bar{x}}$ and $\bar{u}$ and s_x^2 and s_u^2.)

1.8 Correlation and Independence

Before turning to the more formal mathematical development of statistics in Chapter 2, let us first consider some intuitive explanation of some important concepts: correlation and independence. Correlation is a measure of how much one random variable moves as another random variable moves. Figure 1.5a shows a distinct pattern of observations for two highly correlated variables, say the length and weight of fuel elements. It is obvious that as the length increases, the weight also increases. The correlation coefficient, ρ, between two random variables must be between -1 and +1. If the random variables are independent, i.e., there is no dependency of one variable upon another, then the correlation coefficient is zero. The plot of observations from independent random variables will resemble a shotgun pattern. For example, the uranium content in a fuel element is independent of the amount of an impurity present in the nonfuel portion of the element.

In Section 2.1 a more formal probabilistic definition of independence will be presented. It is sufficient here just to recognize that some random variables are independent and some are correlated, and how we deal with data will depend on the relationship between variables.

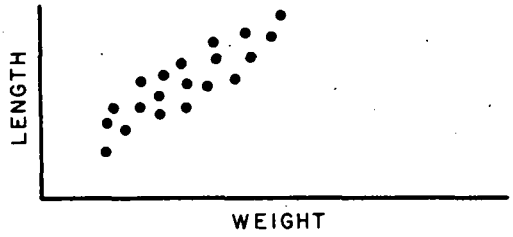


Figure 1.5a. Correlated Variables

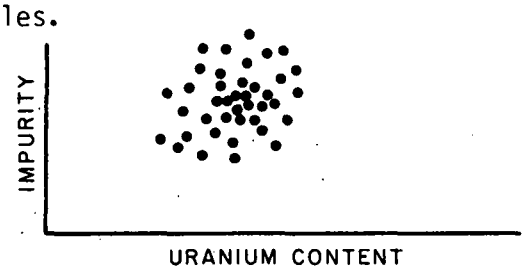


Figure 1.5b. Uncorrelated or Independent Variables

CHAPTER 2. PROBABILITY AND PROBABILITY FUNCTIONS

2.0 INTRODUCTION

How often will a certain value occur for a given process? In the delayed neutron gage data, how often can we expect to find a value of 98 or, more reasonably, how often should a value between 95 and 102 occur? The relative frequency is a measure of the proportion of the total possible outcomes of a process which have a certain common attribute. The relative frequency of the interval ($95 < x < 102$) is 0.33. In the limit, as the number of observations tend toward infinity, the relative frequency of an event coverges to the actual probability of the event. Although we seldom actually know the probabilities of real events, we must still be able to calculate probabilities of certain events as if we did know the true probabilities of other events. Hence, we need to study the basic rules of probability theory.

2.1 Definitions and Basic Laws

Let us begin by distinguishing between probability and statistics. Probability is that branch of mathematics which deals with the assignment of relative frequencies of occurrence of the possible outcomes of a process or experiment according to some mathematical function. Statistics is the theory and practice of drawing inferences about a response or process based on assumptions, data, and the laws of probability. Statistics properly includes the problems of the collection as well as the analysis of data. Although both fields operate in the face of uncertainty due to random error, statistics is inductive in nature, while probability is deductive. One thing we would like to do, then, is to deduce the probability of a complex event based on the probabilities of certain simple events, such as the probabilities of component parts operating successfully at some time T. To do this we need to know the basic laws of probability and the combination of events.

We shall define an event E as some particular occurrence from a set of many possible occurrences (the set of all possible occurrences is called a sample space). An event may occur in several ways, so that we can consider it as a collection of elements all of which have a common attribute. For example, the occurrence of an odd face on a throw of a die is an event which has three elements 1, 3, and 5. The probability of an event is the relative frequency of occurrences of that event with respect to all occurrences. The assignment of a probability value to an event is determined by assumption, by prior knowledge of the event, or by observation of a large number of trials or experiments. We say, for example, that the probability of obtaining an odd face on a die is 1/2, assuming that the die is a fair one. This assignment of probability Pr(E) to that event is verified by experimentation.

Let E_1 and E_2 be two events. The following are definitions of events based on E_1 and E_2:

$\bar{E}$	means	not E (i.e., E has not occurred)
$E_1 + E_2$	means	either E_1 or E_2 or both occur
E_1E_2	means	both E_1 and E_2 occur
$E_1 \mid E_2$	means	E_1 occurs given that E_2 has occurred.

Given the above definition of events, the basic laws or axioms of probability are:

1. Basic Laws of Probability

   a. $0 \leq Pr(E) \leq 1$

   b. $Pr(E + \bar{E}) = 1$

   c.1 $Pr(E_1 + E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1E_2)$

   c.2 If E_1 and E_2 are mutually exclusive,
   $Pr(E_1 + E_2) = Pr(E_1) + Pr(E_2)$

   d.1 $Pr(E_1\ E_2) = Pr(E_1|E_2)\ Pr(E_2)$

   d.2 If E_1 and E_2 are statistically independent, then
   $Pr(E_1E_2) = Pr(E_1)\ Pr(E_2)$

2. Proofs

   a. Let S be the event that something occurs, i.e., any event which can occur is a part of S, and let Φ be an event that never occurs. By the definition of probability we define $Pr(\Phi) = 0$ and $Pr(S) = 1$. Thus, if an event cannot occur, we say it occurs with probability zero. On the other hand, something always occurs with certainty. Thus, for any event E, it has the bounds of 0 and 1 in probability.

   b. If $\bar{E}$ is "not $\bar{E}$", then E and $\bar{E}$ make up all events which could possibly occur. That is, $E + \bar{E} = S$. Hence, $Pr(E + \bar{E}) = Pr(S) = 1$ (see Figure 2.1).

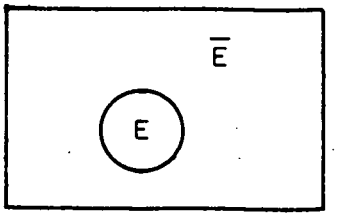


Figure 2.1. $Pr(E + \bar{E}) = Pr(S) = 1$

   c.1 Suppose there are n_{12} ways that events E_1 and E_2 may occur together, n_{10} ways for E_1 to occur alone, and n_{20} ways for E_2 to occur alone (see Figure 2.2).

If the total number of possible outcomes is N, then

$$\Pr(E_1+E_2) = \frac{n_{10} + n_{20} + n_{12}}{N}$$

However, E_1 occurs in a total of $n_{10} + n_{12}$ ways, and E_2 in $n_{20} + n_{12}$ ways. Thus

$$\Pr(E_1) = \frac{n_{10}+ n_{12}}{N}$$

$$\Pr(E_2) = \frac{n_{20}+ n_{12}}{N}$$

Summing $\Pr(E_1)$ and $\Pr(E_2)$, we see that the term $\Pr(E_1E_2) = n_{12}/N$ occurs twice.

Subtracting this term from $\Pr(E_1) + \Pr(E_2)$, we see the result is $\Pr(E_1+E_2)$, i.e.,

$$\Pr(E_1 + E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1E_2)$$

$$= \frac{n_{10} + n_{12}}{N} + \frac{n_{20} + n_{12}}{N} - \frac{n_{12}}{N}$$

$$= \frac{n_{10} + n_{20} + n_{12}}{N} .$$

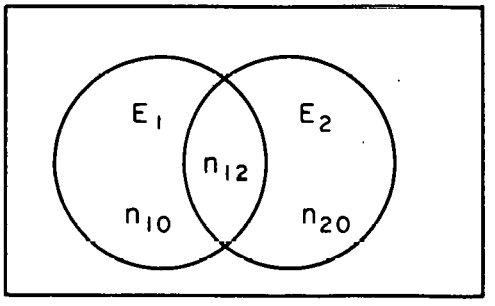


Figure 2.2. $\Pr(E_1 + E_2) = \frac{n_{10} + n_{20} + n_{12}}{N}$

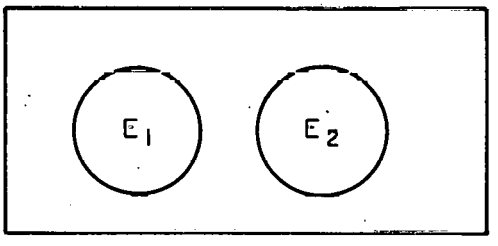


Figure 2.3. $\Pr(E_1E_2) = 0$

If E_1 and E_2 are mutually exclusive events, then E_1 and E_2 cannot occur together (see Figure 2.3). Thus, $Pr(E_1E_2) = 0$ and c.2 follows. The obvious extension holds for the probability of k mutually exclusive events,

$$Pr(E_1 + E_2 + \ldots + E_k)$$
$$= Pr(E_1) + Pr(E_2) + \ldots + Pr(E_k).$$

The extension of the general case c.1 to k events is left as an exercise for the reader.

d.1 As above, let N be the total number of possible occurrences, E_1 and E_2 together occur in n_{12} ways, E_1 alone in n_{10} ways, and E_2 alone in n_{20} ways.

Thus, $Pr(E_2) = \frac{n_{20} + n_{12}}{N}$ and $Pr(E_1E_2) = \frac{n_{12}}{N}$.

How many ways can E_1 occur, given that E_2 has occurred?

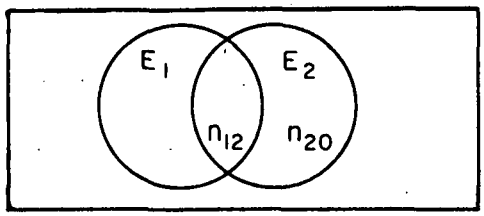


Figure 2.4. $Pr(E_1|E_2) = \frac{n_{12}}{n_{20} + n_{12}}$

E_1 may occur in only n_{12} ways if E_2 has already occurred (see Figure 2.4). E_2 may have occurred in n_{20} ways, so that

$$Pr(E_1|E_2) = \frac{n_{12}}{n_{20} + n_{12}}.$$

That is, of the total number of ways in which E_2 occurs, n_{12} of them contain the occurrence of E_1 also. Now multiplying $Pr(E_1|E_2)$ by $Pr(E_2)$ we obtain

$$Pr(E_1|E_2)\ Pr(E_2) = \frac{n_{12}}{n_{20} + n_{12}} \times \frac{n_{20} + n_{12}}{N}$$
$$= \frac{n_{12}}{N} = Pr(E_1E_2).$$

If E_1 and E_2 are pairwise or statistically independent then the occurrence of E_2 does not affect the probability of the occurrence of E_1, i.e., $Pr(E_1\ E_2) = Pr(E_1)$. Thus, d.2 follows. The extension of d.2 is not as obvious as the extension for c.2. The result is that if $E_1, E_2, \ldots, E_k$ are mutually, or statistically, independent, then

$$Pr(E_1E_2 \ldots E_k) = Pr(E_1)Pr(E_2) \ldots Pr(E_k).$$

However, note that mutual independence requires not only pairwise independence, $Pr(E_iE_j) = Pr(E_i)\ Pr(E_j)$ but also independence at every other stage, i.e.,

$$Pr(E_iE_jE_l) = Pr(E_i)Pr(E_j)Pr(E_l)$$

$$Pr(E_lE_jE_lE_m) = Pr(E_i)Pr(E_j)Pr(E_l)Pr(E_m)$$

2.1.1 An Illustration of the Basic Laws of Probability

To illustrate these basic laws of probability consider the event E that a control rod is successfully driven into the core when properly triggered. Suppose independent testing of the control rod mechanism has shown that the probability of successful operation of the rod is 0.99, i.e.,

$$Pr(E) = 0.99$$

The probability of failure, $\bar{E}$, is, therefore

$$Pr(\bar{E}) = 0.01$$

since E (success) and $\bar{E}$ (failure = not success) are mutually exclusive events and cover all possible outcomes of events which could occur, and

$$Pr(E + \bar{E}) = 1.$$

If there are 2 such rods in a core and they act independently when triggered by the same signal, the probability that both rods operate successfully ("scram") is

$$Pr(E_1E_2) = Pr(E_1)Pr(E_2)$$

$$= (0.99)^2$$

$$= 0.9801$$

If one rod successfully scramming can shut down the reactor, the probability of shutdown is the probability that either one or both rods scram,

$$Pr(E_1 + E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1E_2)$$

$$= 0.99 + 0.99 - 0.9801$$

$$= 0.9999$$

Suppose further that there is a probability that the triggering signal for the control rods will not operate properly. The rods themselves will not scram unless the signal is triggered. Thus, the successful operation of the control rods is conditional upon the successful operation of the triggering device. Suppose the probability of successful operation is $Pr(E_3) = 0.95$. Thus, the successful shutdown of the reactor requires both the successful operation of the triggering device and the successful scram of at least one of the two control rods, $E_4 = E_1 + E_2$, conditional on the first event taking place. This probability is found by

$$\Pr(E_3E_4) = \Pr(E_3)\Pr(E_4|E_3)$$
$$= (0.95)(0.9999)$$
$$= 0.949905$$

2.1.2 An Application to System Evaluation

It often occurs that the components of a system have been tested separately and the component reliabilities, i.e., probability of successful operation, are known. The system itself, however, may not be subjected to tests, either due to time, expense, or the destructive nature of such a test. Suppose we are dealing with a two-stage system for which the first stage has a backup system (see Figure 2.5). Let the probability of components A_1, A_2, and B operating successfully be $P_{A_1} = 0.90$, $P_{A_2} = 0.80$ and $P_B = 0.95$, respectively, where these probabilities were determined by extensive and separate tests. We want to know the probability of the system operating successfully.

Let S mean the system operates successfully, and S_I and S_{II} mean that stages I and II, respectively, operate successfully. Then

$$\Pr(S) = \Pr(S_{II}|S_I)\ \Pr(S_I)$$

Now, stage I operates successfully if A_1 or A_2 operates successfully.

$$\Pr(S_I) = \Pr(A_1 + A_2)$$
$$= \Pr(A_1) + \Pr(A_2) - \Pr(A_1A_2).$$

Since A_1 and A_2 were tested separately, $\Pr(A_1A_2) = \Pr(A_1)\Pr(A_2)$. Hence,

$$\Pr(S_I) = 0.90 + 0.80-(0.90 \times 0.80)$$
$$= 1.70 - 0.72$$
$$= 0.98$$

Finally, the system will operate if S_{II} operates successfully, given S_I has.

$$\Pr(S) = \Pr(S_{II}|S_I)\ \Pr(S_I)$$
$$= \Pr(B)\ \Pr(S_I)$$
$$= 0.95 \times 0.98$$
$$= 0.9310$$

The probability that the system will operate successfully, given the reliability of the three component engines as 0.90, 0.80, and 0.95, is 0.9310.

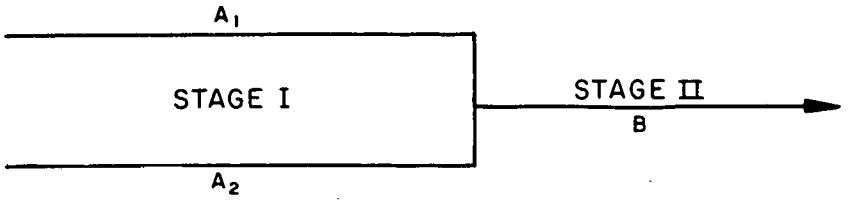


Figure 2.5. Two-Stage System

*2.1.3 Bayes' Theorem. An Application of Conditional Probabilities

A very basic and important application of probability law d.1 was discovered by the Reverend Thomas Bayes in the late nineteenth century. It is simply a double application of law d.1 which allows us to make probability statements about one event E_2, given E_1, when we have knowledge of the probabilities of E_1 given E_2.

Let E_1 and E_2 be two simple events not independent of each other. Then

$$Pr(E_1E_2) = Pr(E_1|E_2)Pr(E_2).$$

We can also write it

$$Pr(E_1E_2) = Pr(E_2|E_1)\ Pr(E_1).$$

Thus, equating the two, we have

$$Pr(E_2|E_1) = \frac{Pr(E_1|E_2)\ Pr(E_2)}{Pr(E_1)}$$

Expanding on this, consider E to occur in k mutually exclusive and exhaustive (i.e., there are no other) ways, E_1, 1 = 1, 2, ..., k. Similarly, H may occur in ℓ mutually exclusive and exhaustive ways H_j, j = 1, 2, ..., ℓ . Then E_i may occur with any of the mutually exclusive and exhaustive events H_j. That is,

$$E_i = E_iH_1 + E_iH_2 + \ldots + E_iH_\ell .$$

Now

$$Pr(H_j|E_i) = \frac{Pr(E_i|H_j)\ Pr(H_j)}{Pr(E_i)}, \quad Pr(E_i) > 0.$$

But

$$\begin{aligned} Pr(E_i) &= Pr(E_iH_1 + \ldots + E_iH_\ell) \\ &= Pr(E_iH_1) + Pr(E_iH_2) + \ldots + Pr(E_iH_\ell) \\ &= \sum_{j=1} Pr(E_iH_j) \end{aligned}$$

*Sections marked by * are advanced topics and may be skipped by first-time readers

In turn, however, $Pr(E_iH_j) = Pr(E_i|H_j)\ Pr(H_j)$. Thus,

$$Pr(E_i) = \sum_{j=1} Pr(E_i|H_j)\ Pr(H_j)$$

and finally,

$$Pr(H_j|E_i) = \frac{Pr(E_i|H_j)\ Pr(H_j)}{\sum_{j=1} Pr(E_i|H_j)\ Pr(H_j)}$$

That is, we can deduce the probability of H_j, given E_i, from knowledge of the probabilities of H_j and $E_i|H_j$.

Example 2.0

Suppose three furnaces are used to sinter fuel pellets. It is observed that the furnaces handle 40,40, and 20 percent of the pellet blends, respectively. The probabilities for a furnace producing unacceptable grain sizes for pellet blends are thought to be 0.05, 0.05, 0.10, respectively, based on previous studies. If the information that a blend was sintered in a particular furnace is not kept, it may be of interest to know the probability that an unacceptable blend came from furnace #3. Let

E_1 = unacceptable grain size for blend
H_1 = blend sintered in furnace #1
H_2 = blend sintered in furnace #2
H_3 = blend sintered in furnace #3

and

$$Pr(E_1|H_1) = 0.05,\ Pr(E_1|H_2) = 0.05,\ Pr(E_1|H_3) = 0.10$$

$$Pr(H_1) = 0.40 \quad Pr(H_2) = 0.40 \quad Pr(H_3) = 0.20$$

Then,
$$Pr(E_1) = Pr(E_1H_1) + Pr(E_1H_2) + Pr(E_1H_3)$$
$$= Pr(E_1|H_1)Pr(H_1) + Pr(E_1|H_2)Pr(H_2) + Pr(E_1|H_3)Pr(H_3).$$

The probability that furnace #3 sintered a particular blend, given that it was defective, is

$$Pr(H_3/E_1) = \frac{Pr(E_1H_3)}{Pr(E_1)} = \frac{Pr(E_1|H_3)Pr(H_3)}{Pr(E_1)}$$

$$= \frac{(0.10)\ (0.20)}{(0.05)\ (0.40) + (0.05)\ (0.40) + (0.10)\ (0.20)}$$

$$= \frac{0.02}{0.06} = 0.33$$

Thus, the probability that a given defective blend was sintered in furnace #3 is 0.33, which is considerably higher than the 0.20 probability that any blend was sintered in furnace #3.

The corresponding probabilities for the other furnaces are also 0.33. Thus, although furance #3 produced only one-fifth of the pellet blends, it is equally likely that a defective blend came from any one of the furnaces.

This example was straightforward and the result reasonably intuitive. The power of the Bayes' theorem technique becomes more appreciable in more complex applications.

2.2 Probability Functions

Suppose we repeatedly run trials of an experiment. Again we interpret experiment rather loosely here as an occurrence or activity which can be observed. The length of time it takes to dial a certain telephone number is an experiment as well as a large-scale investigation of the properties of a certain chemical. The common element is some characteristic of the outcome which can be quantified. For example, the result of a coin flip can be assigned a value 1 for a head or 0 for a tail. Furthermore, the outcome may vary from one trial to another in a random, i.e., unpredictable, manner. We can define an event E as the result that the characteristic value x of the outcome is less than or equal to some fixed value X. If the frequency of occurrence of the event E tends to a limit, we call x a random variable. In practical situations, the assumption of a limiting frequency is reasonable.

Definition 2.1

The real-valued quantity x assigned to the outcome of an experiment which varies in an unpredictable manner is called a random variable.

2.2.1 Cumulative Distribution Function

The limiting frequency of an event E, defined by $x \leq X$, is the probability of the event, $Pr(x \leq X)$. The set of probabilities for an experiment can be formalized by a mathematical model which defines the possible values for $Pr(x \leq X)$ for all possible X. By convention, we say that probability accumulates as we move from left to right on the real axis x. Thus, $Pr(x \leq X)$ is called the cumulative distribution function.

Definition 2.2

The cumulative distribution function (cdf) of a random variable x is

$$F(X) = Pr(x \leq X).$$

To illustrate what we mean by cumulates, consider the following:

Example 2.1

Assign x = 0 if a tail results from the tossing of a coin, and x = 1, if a head results. The statement $x \leq 1$, (X=1), means that either a tail (0) or a head (1) occurs. This is certainty, i.e., Pr(Head or Tail) = 1; i.e., $Pr(x \leq 1) = 1$.

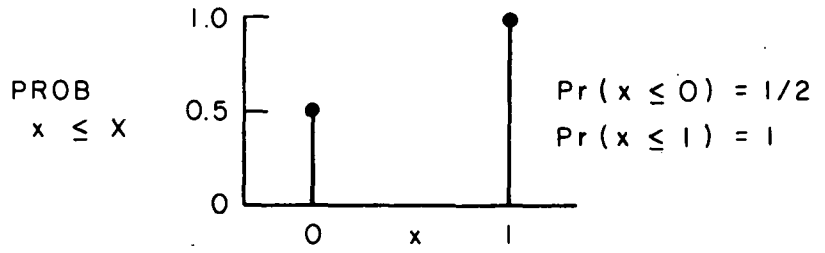


Figure 2.6. Cumulative Distribution of A Coin Toss

Example 2.2

The length of a beam is measured. The beam is one of a large number of beams of various lengths. Theoretically, a beam could be of zero or infinite length. Thus, $F(0) = Pr(x \leq 0) = 0$, $Pr(x < \infty) = 1$. That is, the probability that the beam has no length is zero, and the probability that the beam has some length is 1 (certainty). The probability that the beam is no more than X units in length is $F(X) = Pr(x \leq X)$.

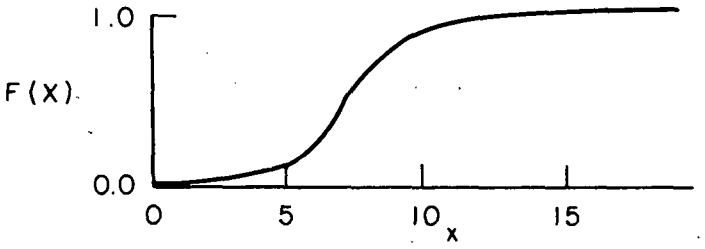


Figure 2.7. Cumulative Distribution Function of the Length of a Beam

2.2.2 A Discrete Random Variable

The above examples illustrate that as we move to the right along X we include more and more of the sample space and hence the probability of the event $x \leq X$ increases or accumulates as x increases. Examples 2.1 and 2.2 also demonstrate the two types of random variables: discrete and continuous.

Definition 2.3

A discrete random variable takes on only a countable number of possible outcomes.

There may be an infinite number of outcomes, but the outcomes can be ordered in a one-to-one correspondence to the real integers. An example of this type of discrete random variable is the number of gamma emissions from a radioactive source over a specified period of time. In fact, all count data

are discrete random variables. Other examples are the number of aces in a hand of bridge, the number of multiple birthdays in a roomful of people, and the number of rejects from a lot of items subject to a quality test.

The probability of obtaining any one of the possible outcomes can be computed and described for all values of the random variable.

Definition 2.4

The probability function of a discrete random variable is

$$p(x) = Pr(x = X),\ x = 0, 1, 2, \ldots, k.$$

Figure 2.8 shows the probability function for a toss of one die. Figure 2.9 shows the probability function for a typical emission count variable.

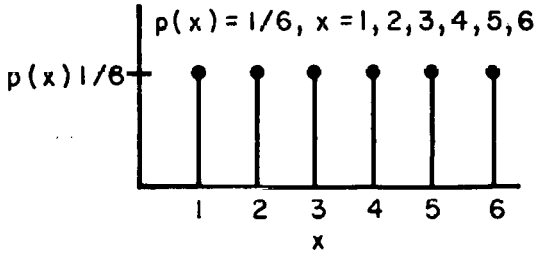


Figure 2.8. Probability Function for a Die

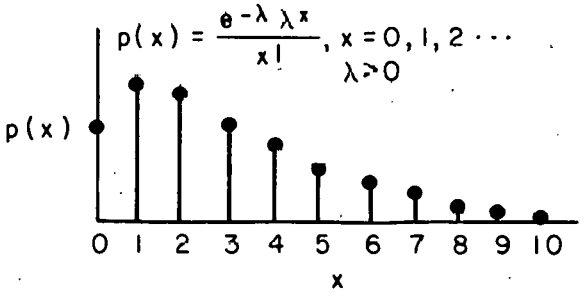


Figure 2.9. Probability Function For Emissions Count

We should note that the probability function accumulates, i.e.,

$$\sum_{x=0}^{X} Pr(x = X) = Pr(x \leq X)$$

or

$$\sum_{x=0}^{X} p(x) = F(X)$$

and if the sum is over the entire range of x, $\sum_{x} p(x) = 1$.

2.2.3 A Continuous Random Variable

A variable such as a length measurement may take on an infinite number of possible values. Such a variable is a continuous random variable.

Definition 2.5

A random variable which may take on any value in a continuum is called a continuous random variable.

The range of possible values for the random variable x may be from 0 to 1 or from 0 to ∞, or most generally, from $-\infty$ to ∞. In any case, there are an infinite number of values for x. Of course, since x is continuous, the cumulative distribution function F(X) is continuous (see Figure 2.7). However, because of the infinity of possible values for x, the Pr(x = X), X being one specified value, is zero. Hence, the probability function as defined for a discrete random variable is inapplicable. We may, however, talk about the probability of x being in a small interval about X and define the density of probability in that interval. As the interval width is made small, we in fact define the derivative of the cdf F(X) at the point X.

Definition 2.6

The probability density function (pdf) for a continuous random variable x is

$$f(x) = \frac{dF(x)}{dx}$$

(NOTE: F(x) and f(x) represent the function describing the properties of the random variable x. The symbol F(X) represents the value of F(x) at x = X.

The probability of the event $X_1 < x < X_2$ is then

$$F(X_1<x<X_2) = Pr(X_1 < x < X_2)$$
$$= \int_{X_1}^{X_2} f(x)dx \quad .$$

In words, the probability of the event is the area under the pdf of x bounded by X_1 and X_2.

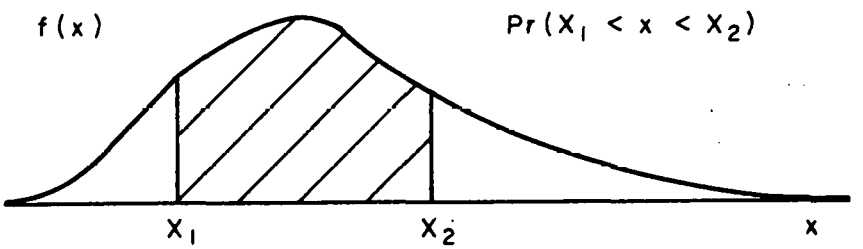


Figure 2.10. Probability Density Function for Continuous X

A very typical continuous distribution is called the uniform distribution which assigns equal probability density to all values in a specified range.

$$f(x) = \begin{cases} 0 , & x < 0 \\ 1/\theta , & 0 \le x \le \theta \\ 0 , & x > \theta \end{cases}$$

and

$$F(x) = \begin{cases} 0 , & x < 0 \\ \int_0^x f(x')dx' , & 0 \le x \le \theta \\ 1 , & x > \theta \end{cases}$$

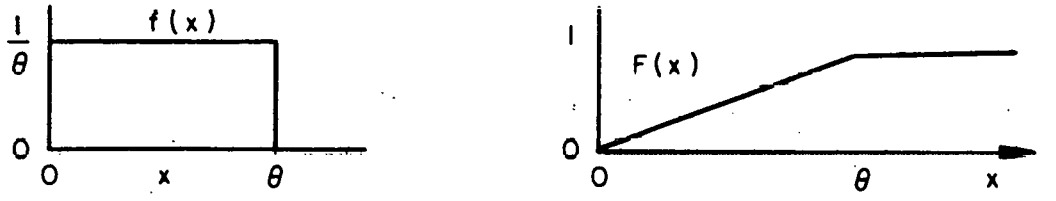


Figure 2.11. pdf and cdf for Uniform Distribution

Physical measurements such as length, height, weight, etc., are continuous random variables. Other continuous random variables are time elapsed between successive events, or the elapsed time until a certain event. Examples of several commonly occurring distributions both discrete and continuous will be discussed in Sections 2.4 and 2.5.

2.3 Characteristics of Probability Functions

Thus far we have discussed probability functions, probability distribution functions, and cumulative distribution functions, mathematical expressions of the ways outcomes of a variable are distributed according to frequency of occurrence. Every random variable has a distribution function. Every distribution function can be characterized in several ways. Knowledge of the form of the mathematical expression and of the constants called parameters involved in these expressions completely describes the distribution function of a random variable. However, knowledge of the characteristics called moments of the distribution also provides a great deal of useful information.

Moments help describe such attributes of the distribution as the central location, spread, asymmetry, and peakedness of a distribution. Generally, a moment is the expected value of a function of the random variable, and there are several kinds of moments. The most important and typical moments are called central moments.

2.3.1 Moments

First, let us consider the most basic moment, the first moment about the origin. This we call the mean and is the mean value or expected value of the distribution. That is, it is the point of balance of the distribution or its center of gravity. (For simplicity, we will only discuss continuous random variables here. For discrete random variables, we need only replace the $\int$ sign by the summation sign Σ and drop the differential.)

Definition 2.7

The First Moment or Mean

$$\mu = \int_x xf(x)dx$$

$$= E(x) \quad ,$$

where the integral is over the range of x.

For sake of generality, we will use $-\infty$ to ∞. The symbol E(x) means the expected value of x. The mean is called the location parameter of the distribution of x, since it locates the center of gravity of the distribution.

We can define other moments, called raw moments, or simply the moments about the origin, in a similar way:

$$\mu'_2 = \int_{-\infty}^{\infty} x^2 f(x)dx = E(x^2)$$

$$\mu'_3 = \int_{-\infty}^{\infty} x^3 f(x)dx = E(x^3)$$

etc.

Definition 2.8

The moments about the origin μ'_r, or $E(x^r)$ are obtained by

$$\mu'_r = \int_{-\infty}^{\infty} x^r f(x)dx \quad .$$

Although these moments are useful, more informative type of moments are the moments about the mean, called central moments.

Definition 2.9

Central Moments

$$\mu_r = E(x-\mu)^r = \int_{-\infty}^{\infty} (x-\mu)^r \, f(x)dx$$

We should recognize that μ_1, the first moment about the mean, is always 0. Following the rules of integration we see

$$\mu_1 = \int_{-\infty}^{\infty} (x-\mu) \, f(x)dx = \int_{-\infty}^{\infty} xf(x)dx - \mu \int_{-\infty}^{\infty} f(x)dx$$

$$= \mu'_1 - \mu = \mu - \mu = 0$$

since $\int_{-\infty}^{\infty} f(x)dx = F(\infty) = Pr(x < \infty) = 1,$

$\int_{-\infty}^{\infty} xf(x)dx \equiv \mu'_1, \; \mu'_1 = \mu.$

The next most useful moment is the second moment about the mean, called the variance σ^2. The second central moment $\mu_2 = E(x-\mu)^2$ is the moment of inertia about the mean. That is, it measures the degree to which the frequency of a variable spreads out as it moves away from its mean or central value.

Definition 2.10

Variance of a Distribution

$$\sigma^2 = \mu_2 = E(x-\mu)^2$$
$$= \int_{-\infty}^{\infty} (x-\mu)^2 \, f(x)dx$$

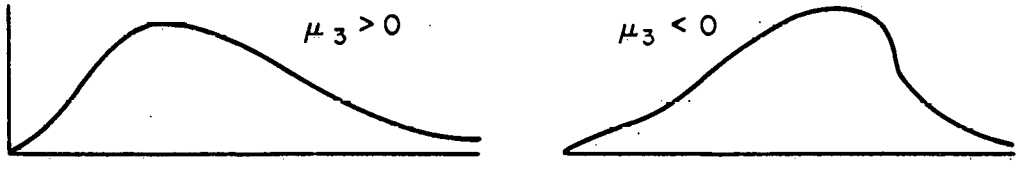
This is always non-negative and only for degenerate distributions is it zero. The square root of the variance is called the standard deviation σ.

Other important central moments are:

1. Third central moment:

$$\mu_3 = E(x-\mu)^3 = \int_{-\infty}^{\infty} (x-\mu)^3 \, f(x)dx$$

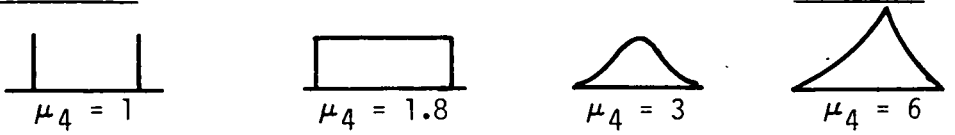
μ_3 measures the asymmetry of the distribution. A perfectly symmetric distribution has $\mu_3 = 0$. If a distribution has a long tail to the right, $\mu_3 > 0$, and if the long tail is to the left, $\mu_3 < 0$. This is called skewness.



2. Fourth central moment:

$$\mu_4 = E(x-\mu)^4 = \int_{-\infty}^{\infty}(x-\mu)^4 f(x)dx$$

μ_4 measures the degree of peakedness, i.e., the degree to which a distribution bunches up to form a peak. Flat distributions are called platykurtic and peaked distributions are called leptokurtic. The property of peakedness is known as kurtosis.



(Normal Distribution)

These moments of skewness and kurtosis are often measured in other terms:

$$\lambda 3 = \frac{\mu_3}{\sigma 3} \text{ and } \lambda 4 = \frac{\mu_4}{\sigma 4} - 3 \quad (\lambda 4 \geq -2).$$

Higher order moments do exist, but their physical interpretation is difficult and are not of great practical importance. However, knowledge of all the moments is equivalent to knowledge of the form of the distribution itself. That is, the moments completely specify a distribution.

The relationship between the central moments and the raw moments is obvious especially since the central moment is defined in terms of μ_1 or μ. In fact, the moments of one type can be obtained from knowledge of the moments of the other.

*2.3.2 Relationship between Moments

Consider moments about any two arbitrary points, a and b. The moments will be denoted μ_r (a) and μ_r (b). By use of the binomial expansion theorem,

Theorem 2.1

For $\mu_r(a) = \int_{-\infty}^{\infty}(x-a)^r f(x)dx$, and $\mu_r(b) = \int_{-\infty}^{\infty}(x-b)^r f(x)dx$,

$$\mu_r(a) = \sum_{j=0}^{r}\binom{r}{j}(b-a)^{r-j}\mu_j(b)$$

Proof

Consider $(q+w)^r$. Expand by Taylor's series about $q = 0$. Then

$$(q+w)^r = w^r + rqw^{r-1} + \frac{r(r-1)}{2}q^2 w^{r-2} + \cdots + \frac{r(r-1)}{2}q^{r-2}w^2 + rq^{r-1}w + q^r$$

$$= \sum_{j=0}^{r}\binom{r}{j}q^j w^{r-j} \qquad (2.1)$$

Now substitute for q and w, q = x-b and w = b-a

Then (2.1) becomes

$$(x-b + b-a)^r = (x-a)^r = \sum_{j=0}^{r} \binom{r}{j} (x-b)^j (b-a)^{r-j}$$

Multiply both sides by f(x) and taking the integral on both sides over x, we get

$$\int_x (x-a)^r f(x)dx = \int_x \sum_j \binom{r}{j} (x-b)^j (b-a)^{r-j} f(x)dx$$

Since $(b-a)^{r-j}$ is a constant with respect to x and the integral is independent of the summation, we obtain

$$E(x-a)^r = \sum_{j=0}^{r} \binom{r}{j} (b-a)^{r-j} (x-b)^j f(x)dx$$

$$\mu_r(a) = \sum_{j=0}^{r} \binom{r}{j} (b-a)^{r-j} \mu_j(b) \quad .$$

We may now substitute for a and b as follows: a = 0 and b = μ or a = μ, b = 0. Then, since $\mu'_1 = \mu$, we can write the first three moments of one type as a function of the other type:

$\mu'_1 = \mu$ $\qquad$ $\mu_1 = 0$

$\mu'_2 = \mu_2 + \mu^2$ $\qquad$ $\mu_2 = \mu'_2 - \mu^2$

$\mu'_3 = \mu_3 + 3\mu\mu_2 + \mu^3$ $\qquad$ $\mu_3 = \mu'_3 - 3\mu'_2\mu + 2\mu^3$

The relationship between moments for r > 3 is left as an exercise for the reader.

2.3.3 Basic Properties of Expected Values

Let y be a random variable and let a be a constant. Denote the variance of y by Var(y). The following are basic properties or rules for manipulating means and variances.

1) $E(ay) = a E(y)$

2) $E(y + a) = E(y) + a$

3) $Var(ay) = E(ay - E(ay))^2 = a^2 Var(y)$

4) $Var(a + y) = Var(y)$

The proofs for these results consist of simply looking at the integral (or summation) expressions and applying the basic rules of integration.

In general, we may write down the expected value of any function of a random variable, say g(y) as

5) $E\ (g(y)\ ) = \int_{-\infty}^{\infty} g(y)\ f(y)dy$,

where f(y) is the probability density function of y.

Of particular interest in statistics is the mean and variance of a linear combination of independent random variables. Let

$$y = a_1x_1 + a_2 \times x_2 + \ldots + a_nx_n$$

be a linear statistic, then

6) $E(y) = E(\sum_i a_i x_i) = \sum_i a_i E(x_i)$

If $E(x_i)$ is the same for all i, then

$$E(y) = \sum_i a_i\ E(x)$$

7) $Var(y) = \sum_i Var(a_i x_i) = \sum_i a^2_i\ Var(x_i)$

$= Var(x) \sum_i a^2_i$, if $Var(x_i) = Var(x)$, all i.

Example 2.3

Consider $y = a_1x_1 + a_2x_2$, $E(x_1) = E(x_2) = \mu$ and $Var\ (x_i) = Var(x_2) = \sigma^2$, then for $a_1=a_2 = 1$,

$$E(y) = (a_1+a_2)\ E(x) = 2\mu$$

$$Var(y) = (a_1^2 + a_2^2)\ Var(x)$$

$$= 2\sigma^2 \quad .$$

For x_1 and x_2 not independent, but correlated+,

$$E\ [(x_1 - \mu\ )\ (x_2 - \mu\ )] = \rho\sigma^2 \quad .$$

Then $E(y) = 2\mu$ as before, but

$$Var(y) = Var(a_1x_1 + a_2x_2)$$

$$= a_1^2\ Var(x)_1) + a_2^2\ Var(x_2) + 2a_1a_2\ E\ [(x_1 - \mu\ )\ (x_2 - \mu\ )]$$

$$= 2\sigma^2 + 2\rho\sigma^2$$

$$= 2\sigma^2\ (1 + \rho\ ) \quad .$$

+If two random variables are correlated, then a change in one variable is associated with a change in the other; e.g., for a positive correlation coefficient ρ , an increase in x_1 will result in an increase in x_2. For negative ρ , an increase in x_1 will result in a decrease in x_2.

Example 2.4

For $y = x_1 - x_2$, i.e., $a_1 = 1$, $a_2 = -1$,

$$E(y) = 0$$

$$\text{Var}(y) = (a_1^2 + a_2^2 + 2a_1a_2\rho)\sigma^2$$

$$= 2\sigma^2(1 - \rho)$$

$$= 2\sigma^2 \quad \text{, if } x_1 \text{ and } x_2 \text{ are independent.}$$

Linear statistics will be very important in making inferences about a distribution. Appendix A, Uncertainty Analysis deals with the evaluation of linear and more general functions of random variables.

2.4 Discrete Distributions

A discrete distribution function $p(x)$ is the mathematical description of the distribution of probability of the outcomes for some discrete random variable; i.e., the response x may take on any of a set of distinct values, often finite in number but possibly an infinite number of countable values which can be placed in a one-to-one relationship with the positive real integers.

There are many discrete distributions. We will discuss some of the more important ones. Discrete distributions usually deal with count data or attributes of a variable. Some typical variables described by discrete distributions are (1) the success or failure of an outcome to meet or attain some specific attribute, (2) the number of successes in a series of trials, (3) the number of defects in a product, (4) the number of defective pieces in a lot, and (5) the number of counts or emissions from a source.

2.4.1 Point Binomial

The simplest example of a discrete distribution is the success or failure of some event or trial. The best example of this is the toss of a coin: a head is a success and a tail is a failure (or vice versa). Let p be the probability of success. Then

$$p(x) = \begin{cases} p & , x = 1 \text{ (success)} \\ 1-p & , x = 0 \text{ (failure)} \\ 0 & , \text{otherwise.} \end{cases}$$

A trial of this kind resulting in success or failure is called a "Bernoulli trial." The expected value of x is

$$E(x) = \sum_{x=0}^{1} xp(x) = 0(1-p) + 1.p = p \quad .$$

Other moments:

$$E(x^2) = \sum_{x=0}^{1} x^2p(x) = 0(1-p)+1^2.p = p$$

$$\text{Var}(x) = E[x - E(x)]^2 = E(x-p)^2$$

$$= E(x^2) - 2p\,E(x) + p^2$$

$$= E(x^2) - p^2$$

$$= p - p^2 = p(1-p)$$

2.4.2 Binomial Distribution

Let x be the number of items in a sample of n from a population of such items that have a common attribute and let p be the proportion of items in the population which have that attribute, then

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad , \quad x = 0, 1, 2, \ldots, n, \quad 0 < p < 1,$$

$$\text{where} \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

Obtaining an item from the sample which has the desired attribute is often called a success. Thus, x is the number of successes in n trials. Note that the binomial is a two-parameter distribution with parameters n and p. The mean of a binomial distribution is

$$E(x) = np,$$

and the variance can be shown to be

$$\text{Var}(x) = np(1-p)$$

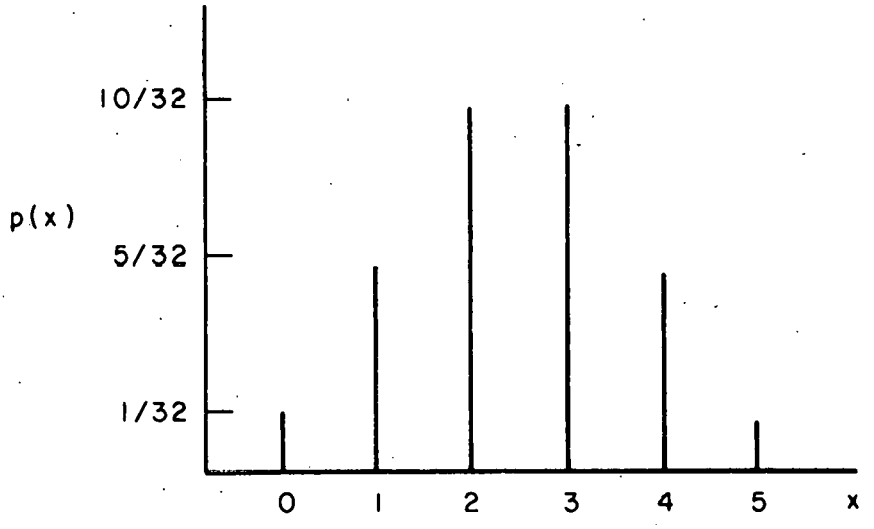


Figure 2.12. Binomial Distribution for n=5, p=0.50

Data which can be classified as success or failure according to whether or not they have a certain attribute can usually be described by a binomial distribution.

Example 2.5

Consider sampling n fuel rods. These rods either meet length specification or not. The variable x is the number of fuel rods found in the n rods examined that meet specifications. Suppose n = 5 and p, the probability of obtaining a "good" fuel rod, is 0.6. Then, the probability of finding at least 4 rods that meet specifications is

$$Pr(x \geq 4) = \sum_{x=4}^{5} \binom{5}{x} (0.6)^x (0.4)^{5-x}$$

$$= \binom{5}{4}(0.6)^4(0.4)^1 + \binom{5}{5}(0.6)^5 (0.4)^0$$

$$= 0.3370$$

Extension: Multinomial Distribution

The binomial can be extended to cover classification of outcomes into k categories rather than 2. This distribution is known as the multinomial distribution.

Let x_i be the number of trials resulting in category i, and p_i be the probability on any one trial of obtaining that result. Then the distribution of the number of results of n trials into k categories is given by

$$p(x_1,x_2,\ldots,x_k) = (x_1,x_2^n,\ldots,x_k)\ p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},\ x_i = 0,1,\ldots,n$$

where $\sum_{i=1}^{k} x_i = n$ (hence $x_k = n-x_1-x_2\ldots-x_{k-1}$), and $\sum_{i=1}^{k} p_i = 1$,

$$E(x_i) = np_i$$

$$Var(x_1) = np_1\ (1-p_1)$$

$$and (x_1,\ ^n x_2,\ldots x_k) = \frac{n!}{x_1!\ x_2!\ldots x_k!}$$

Example 2.6

Suppose the probability that a certain kind of valve will wear out in fewer than 500 hours is 0.50, the probability that it will wear out in fewer than 800 but more than 500 hours is 0.30, and the probability that it will last more than 800 hours is 0.20. To find the probability that among 10 such valves 4 will wear out in fewer than 500 hours, 4 will wear out in fewer than 800 but more than 500 hours, while 2 will last more than 800 hours, we have only to substitute x_1 =n 4, x_2 = 4, x_3 = 2, n= 10, p_1 = 0.50, p_2 = 0.30, p_3 = 0.20, and we get

$$p\ (4,\ 4,\ 2) = \frac{10!}{4!4!2!}\ (0.50)^4(0.30)^4(0.20)^2$$

$$= 0.064$$

2.4.3 Hypergeometric Distribution

The binomial distribution is an example of sampling with replacement. On each trial the probability of a success is the same. In some cases sampling is performed without replacement. The distribution for this case is known as the hypergeometric distribution.

Let N be the population size, D the number of items having the attribute of interest, n the sample size, and x the number of items in the sample having the attribute of interest.

Then,

$$p(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, 2, \ldots, \min(D,n)$$

(i.e., x cannot be greater than the number of items in the sample or the number having the desired attribute)

or

$$p(x) = \frac{\binom{n}{x}\binom{N-n}{D-x}}{\binom{N}{D}}$$

Furthermore,

$$E(x) = \frac{nD}{N} = np, \quad p = \frac{D}{N},$$

$$Var(x) = np(1-p)\,\frac{N-n}{N-1}.$$

We note that as N, the population size, becomes large, sampling without replacement becomes more and more like sampling with replacement. Thus, the hypergeometric approaches the binomial distribution as $N \to \infty$.

Example 2.7

If we are sampling from a lot of N = 10 pieces which contains 2 defective pieces, what is the probability of obtaining zero defectives in a sample of size n = 4? [D = 2, N = 10, n = 4]

$$Pr(x = 0) = \frac{\binom{2}{0}\binom{8}{4}}{\binom{10}{4}} = \frac{8!}{4!4!}\,\frac{6!4!}{10!} = \frac{6 \cdot 5}{10 \cdot 9}$$

$$= 0.333$$

2.4.4 Poisson Distribution

Let x be the number of occurrences of some event in a given interval of time or space, and let the parameter λ be the mean number of such events in the interval. Then,

$$p(x) = \frac{e^{\lambda}\,\lambda^{x}}{x!}, \quad \lambda > 0, \; x = 0, 1, 2, \ldots.$$

The Poisson distribution has the interesting property that its one parameter is both the mean and variance of the distribution, i.e.,

$$E(x) = \lambda$$

$$Var(x) = \lambda$$

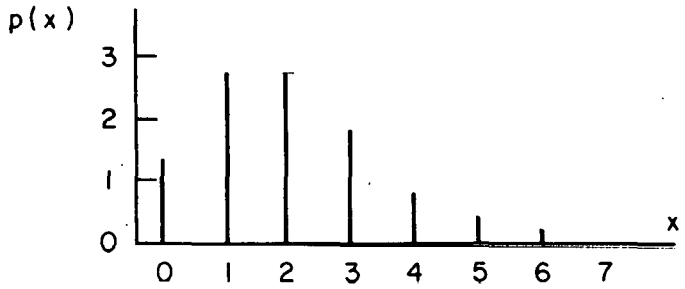


Figure 2.13. Poisson Distribution ($\lambda = 2$)

The Poisson distribution can usually be applied to count-type data such as the number of white cells in a drop of blood or the number of particles emitted from a radioactive source, or to approximate the binomial distribution (see Section 2.4.2) when n is large and p small but np is constant.

Example 2.8

Consider the number of alpha particles being emitted from a radioactive source during a period of 10 seconds. If the average rate of emissions is 5 per 10-second interval, the probability of getting 7 emissions is

$$Pr(x = 7) = \frac{e^{-5}5^7}{7!} = 0.104 \quad .$$

*2.4.5 A Comparison

Since the binomial, hypergeometric, and Poisson distributions are all used in quality control type problems, a comparison of the three is useful.

Table 2.1. Comparison of Binomial, Hypergeometric, and Poisson Distributions

	Parameters	Probability of Success p	Sample Size n	Population Size N
Binomial (with replacement)	p,n	fixed	specified	infinite
Hypergeometric (w/o replacement)	D,n,N	$p_{n+1}=\frac{D-x}{N-n}$	specified	finite
Poisson	$\lambda = np$	Very small	Unknown but large	unknown but larger than n

The following table and figure illustrate the relationship between the binomial and Poisson distribution:

Table 2.2. Binomial Versus Poisson

	Binomial			Poisson
x	p = 0.2, n=5	p = 0.1, n=10	p = .05, n = 20	λ = 1
0	0.3277	0.3487	0.3585	0.3679
1	0.4096	0.3874	0.3774	0.3679
2	0.2048	0.1937	0.1887	0.1839
3	0.0512	0.0574	0.0596	0.0613
4	0.0064	0.0112	0.0133	0.0153
5	0.0003	0.0015	0.0022	0.0031
6		0.0001	0.0003	0.0005
7		0.0000	0.0000	0.0001
8				0.0000

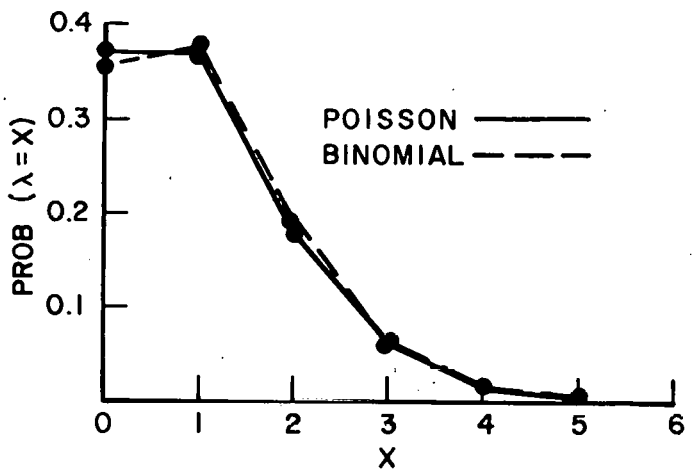


Figure 2.14. Binomial (p = 0.05, n=20) versus Poisson (λ = 1)

Applications of the binomial, hypergeometric, and Poisson distributions are discussed in more detail in Chapters 4 and 5.

2.4.6 Negative Binomial Distribution

Another very useful discrete distribution is the negative binomial. This distribution is used to determine, for example, how many trials will be required before the rth success is obtained. The derivation is as follows:

The probability of getting the rth success in exactly the xth trial is

$$\Pr\begin{pmatrix}\text{rth success on}\\ \text{xth trial}\end{pmatrix} = \Pr\begin{pmatrix}\text{r-1 success in}\\ \text{x-1 trials}\end{pmatrix} \Pr(\text{success})$$

$$= \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r} p$$

$$= \binom{x-1}{r-1} p^{r} (1-p)^{x-r}, \quad x \geq r \geq 0, \quad r = 1, 2, \ldots.$$

If we now let $k = x-r$, $x = k+r$, we have

$$p_{(k)} = \Pr(\text{rth success on k+rth trial}) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots.$$

To see that this is a legitimate probability function, we need to recognize the fact that

$$\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} (1-p)^k = p^{-r},$$

so that the probabilities $p_{(k)}$ add to 1.

The mean and variance of the negative binomial distribution for the random variable $k = x-r$ can be shown to be

$$\mu_k = r(1-p)/p$$

$$\mu_k^2 = r(1-p)/p^2 .$$

Thus, the mean and variance of x, the total number of trials required to obtain the rth success, are obtained using the basic properties of expected values (Section 2.3.3),

$$\mu_x = \mu_k + r = r/p$$

$$\sigma^2_x = \sigma^2_k = r(1-p)/p^2$$

2.5 Continuous Distributions

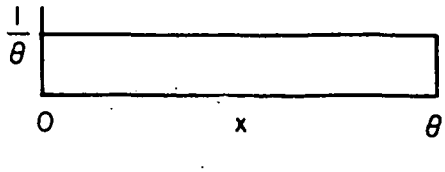
Most measurement data of physical properties, such as length, width, weight, etc., are continuous. A brief description of some of the more important continuous distributions which do describe real variables follows.

2.5.1 Rectangular Distribution

The rectangular or uniform distribution is

$$f(x) = \frac{1}{\theta}, \quad 0 < x < \theta,$$

$$E(x) = \frac{\theta}{2}, \quad Var(x) = \frac{\theta^2}{12}$$



The primary uses of the uniform distribution is in situations where any value in a range is equally likely to occur. It is also used to approximate a relatively small range of values for another continuous distribution, i.e., in drawing histograms.

More generally, x could range from a to b, where a and b could be any values such that b >a. Then the distribution function becomes

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

and

$$E(x) = \frac{b + a}{2} \text{ and } Var(x) = \frac{(b-a)^2}{12}.$$

In the notation above, $b = \theta$, $a = 0$.

The next two distributions are used extensively in reliability and life-testing problems.

2.5.2 Gamma Distribution

Let x be the time to failure of some component that follows a gamma distribution,

$$f(x) = \frac{1}{\Gamma_\beta \lambda^\beta} x^{\beta-1} e^{-x/\lambda}, \quad x > 0, \quad \lambda > 0, \beta \geq 1,$$

where β is a shape parameter, λ is a scale parameter and Γ_β is a gamma function. For integer β, $\Gamma_\beta = (\beta - 1)!$. The mean and variance of x are

$$E(x) = \lambda\beta,$$
$$Var(x) = \lambda^2\beta.$$

The gamma distribution has found wide usage in reliability work for describing the mean time to failure, the mean time between failures, and for other life-testing types of problems.

2.5.3 Exponential Distribution

The exponential distribution, a single parameter special case of the gamma distribution, has found wide usage also. The exponential distribution is

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} , \quad x>0, \ \lambda>0,$$

where $E(x) = \lambda$,

$$Var(x) = \lambda^2 \ .$$

Some uses of the exponential distribution are in describing failure times of electron tubes, resistors, and capacitors.

Example 2.9

Suppose the mean time to failure of a particular type of capacitor is known to be 5 years. If the lifetime follows an exponential distribution, the probability that a given capacitor lasts less than 10 years is

$$Pr(x<10) = \int_0^{10} \frac{1}{5} e^{-t/5} \, dt = 1 - e^{-2}$$

$$= 0.865$$

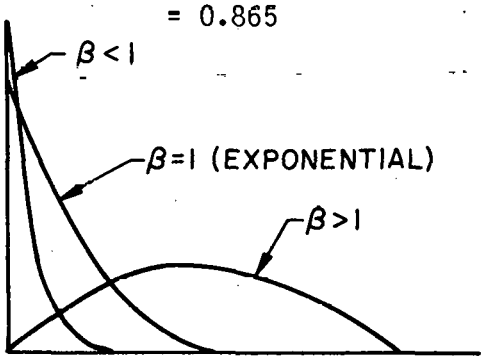


Figure 2.15. Gamma Distributions

2.5.4 Weibull Distribution

A three-parameter distribution which is a further generalization of the gamma distribution is the Weibull,

$$f(x) = \left(\frac{\beta}{\lambda}\right)\left(\frac{x-\gamma}{\lambda}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\lambda}\right)^{\beta}\right] , \quad x \geq \gamma$$

where the additional parameter is a location parameter. The mean and variance are

$$E(x) = \gamma + \lambda \, (1/\beta)!$$

$$Var(x) = \lambda^2 \left[\Gamma(2/\beta) - \Gamma(1/\beta)\right] \ .$$

Since the form of the probability density function is so complicated, the cumulative distribution function

$$F(x) = \int_0^x f(x') \, dx'$$

$$= 1 - \exp\left[-\left(\frac{x-\gamma}{\lambda}\right)^{\beta}\right] \ ,$$

is often used.

The Weibull distribution has been used successfully to describe the behavior of the lifetime of certain mechanical parts, ball bearings, electronic components, and subassemblies.

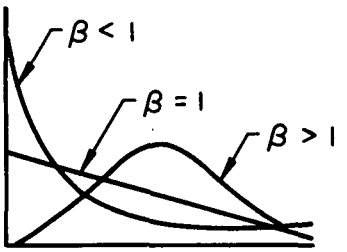


Figure 2.16. Weibull Distribution

2.5.5 Normal Distribution

Let x be a response being measured, such that,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}(x-\mu)^2\right] \quad ,-\infty<x<\infty, \quad -\infty<\mu<\infty, \quad \sigma^2>0 \ ,$$

where μ is the mean, and σ^2 is the variance of the distribution,

$$E(x) = \mu \ ,$$

$$Var(x) = \sigma^2 \ .$$

The normal distribution is perhaps the one that occurs most often in practice. Most physical measurements such as length, height, and weight tend to be normally distributed. Beside this fact, however, the normal distribution is extremely valuable because of a result known as a central limit theorem. The central limit theorem states that the average $\bar{x}$ of n independent observations which follows some distribution with finite mean μ and variance σ^2 will tend to have a normal distribution with mean μ and variance σ^2/n for sufficiently large n. By "tend", it is meant that as n becomes larger, the approximation of the exact distribution by the normal becomes better and better. Fortunately, "sufficiently large n" is relatively small for most cases,even as low as 3 or 4 for nicely behaved, symmetric distributions. The central limit theorem will be discussed in more detail in connection with inferences on the mean of a distribution in Chapter 3.

A third reason for the prominence of the normal distribution in statistics is that many other distributions converge to a normal distribution under appropriate conditions. For example, the binomial and Poisson distributions both tend toward a normal distribution as their means np and λ, respectively, get large.

There are an infinite number of normal distributions that can be represented by the two parameters μ and σ^2. Since it is difficult to integrate the normal distribution, it was necessary to compute tables of a normal variable. To avoid a multitude of different normal distribution tables, a standard normal variable was defined and tabled.

Let x be a normal distribution with mean μ and variance σ^2 and denoted by $N(\mu, \sigma^2)$. Then

$$z = \frac{x-\mu}{\sigma}$$

is a standard normal variable and has a mean of 0 and a variance of 1, i.e., z is distributed as N(0,1). All probability statements about a normal distribution can be answered, then, by referring to the standardized normal.

Example 2.10

Suppose we wish to determine the probability that the length of a rod is less than 40.2 inches when it is known that the distribution of the length of such manufactured rods is normal with mean of 40 inches and a variance of 0.04 $inch^2$. Thus, $\sigma = 0.2$ inch and

$$z = \frac{x-\mu}{\sigma} = \frac{40.2-40}{0.2} = 1.$$

Then,

$$\begin{aligned} Pr(x<40.2) &= Pr(z<1) \\ &= 1-Pr(z>1) \\ &= 1-0.1587 = 0.8413 \end{aligned}$$

where Pr(z>1) is from Table III, Normal Distribution.

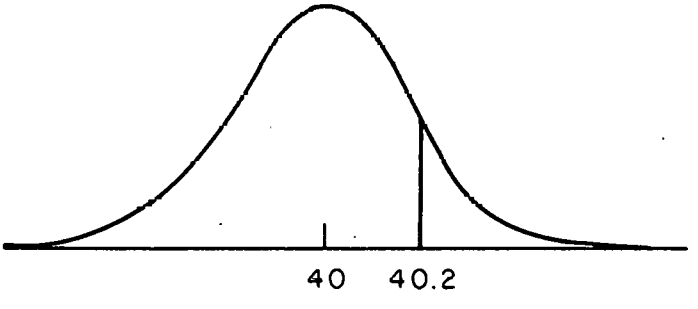


Figure 2.17. Normal Distribution

CHAPTER 3
INFERENCE ABOUT A SINGLE POPULATION: NORMAL DISTRIBUTION

3.0 INTRODUCTION

The object of statistics is to infer the properties of some measurement or response for use in describing or predicting that measurement or response.

We know that no matter what we are measuring or controlling, and even under the best of conditions, observations

1. vary from one trial to another,
2. follow some distribution function,
3. have moments, particularly a mean and variance.

These facts we summarize in the notation of a model,

$$x_i = \mu + \epsilon_i .$$

where x_i is the ith observation of some phenomenon whose true value is μ, and ϵ_i is a random error for the ith observation which we assume to have some distribution function with a mean and variance, usually 0 and σ^2, respectively, and, very importantly, each error is <u>independent</u> of every other.

For the moment consider μ fixed and known. Then x_i is a random variable consisting of a constant μ and a random error term ϵ. Thus,

$$E(x_i) = E(\mu + \epsilon_i)$$

$$= \mu + E(\epsilon_i).$$

If $E(\epsilon_i) = 0$, μ is the expected value or mean of the distributionof x's, and

$$Var(x_i) = Var(\epsilon_i) \quad .$$

That is, since μ is a constant, the variance of the observation x is the same as the variance of the error. It is in these properties of the distribution, the mean and variance, that we are most interested. Estimation theory is the body of procedures that lends itself to answering questions about parameters of a distribution. It provides us ways of determining best guesses which are optimum in some sense (see Appendix B). The idea is to <u>infer</u> the value of important parameters based on the information that is available in a set of data. Inference may take the form of

1. a test of hypothesis of a particular value of a parameter,
2. a confidence interval of feasible values of a parameter,
3. other types of intervals concerning a single future observation or concerning the population as a whole, or
4. a test for a distribution which properly describes a set of data.

In this chapter we shall begin with a simple test of hypothesis of the mean of a normal distribution and then proceed to estimation of the mean and variance and their associated tests of hypotheses and intervals. We conclude

the chapter with an examination of the adequacy of the assumption of a normal population for describing a set of data, and a section on the detection and treatment of outliers.

3.1 A Simple Test of Hypothesis

Suppose we are sampling from a normal distribution whose mean is unknown but whose variance is known and denoted by σ^2. Suppose the response being measured is the number of hours per week put in by a member of first-line management and further suppose $\sigma^2 = 16\ (\text{hr.})^2$. If you believe the mean number of hours worked per week to be 50 hours, what can we say if we obtain a sample value of 44 hours for one individual for one week? Thus, we would like to test the hypothesis that the mean μ of the distribution of hours per week worked by first-line management personnel is 50 hours, given that the distribution is normal with variance $\sigma^2 = 16$.

The test is carried out by converting the distribution to a standard normal z (sometimes called a normal deviate) and determining the probability of obtaining the given data point under the assumption that the hypothesis (usually referred to as the null hypothesis) is correct. If this probability is not too small, we would tentatively accept the null hypothesis as a likely value of μ. If it is too small, we would reject the hypothesized value as reasonable, for if it were correct, the event recorded would be a "rare" event.

A rare event is by definition something which occurs rarely, i.e., with small probability. How small? In many practical situations, a rare event is defined as something having less than 0.05 probability of occurring. On other occasions a value of 0.025, 0.01, or even 0.10 is used. This probability value for a rare event is called the significance level α of a test of hypothesis. Although $\alpha = 0.05$ has become a standard value, it is by no means magical. In fact, the size of α is arbitrary and should be determined by the experimenter for the situation at hand. Since we reject a hypothesized value if the probability of its occurrence is less than α, we see that α is in fact the probability of erroneously rejecting a true hypothesis. What should determine your choice of significance level is your willingness, or unwillingness, to make such an error.

Besides determining the size of the significance level of a test of hypothesis, we must also indicate in what direction or directions we may observe an error. In other words we need to specify an alternative hypothesis to accept if we reject the original or null hypothesis. Let us denote the null hypothesis by $H_0: \mu = \mu_0$. Then the following alternatives are possible:

1. $H_A: \mu = \mu_A$

   A specific alternative value is considered. This would in general require a lot of previous information and a well defined problem. Practical situations would not usually be so clear cut.

2. $H_A: \mu < \mu_0$ (or $\mu > \mu_0$)

This is a <u>one-sided</u> test of hypothesis. We suspect that if H_0 is false, then the true value is less than (or greater than) μ_0 but not greater (less). More data is necessary to determine a specific value. Problems concerning product improvement usually involve a one-sided alternative hypothesis.

3. $H_A: \mu \neq \mu_0$

This is a <u>two-sided</u> test of hypothesis. If H_0 is false we do not have any indication before the test which direction the true value is likely to be from the hypothesized value. Comparisons of products usually involve two-sided alternatives.

Returning to the hours worked per week problem, we may formalize the decision process as follows:

<u>Example 3.1</u> Assume x follows a $N(\mu, \sigma^2 = 16)$ distribution

$H_0: \mu = 50$, $\alpha = 0.05$

$H_A: \mu \neq 50$

$$x_0 = 44, \text{ so } z_0 = \frac{x_0 - \mu_0}{\sigma} = \frac{44-50}{4} = -1.5$$

$$Pr(x < 44) = Pr(z < -1.5) = Pr(z > 1.5)$$

$$= 0.0668$$

Since H_A is a two-sided alternative, we divide α into two equal parts, allocating 2-1/2% of the total distribution to each tail. Since 0.0668 > 0.025, we <u>do not reject</u> $H_0: \mu = 50$ based on the single observation x = 44. We <u>tentatively accept</u> $\mu = 50$ until evidence to the contrary appears.

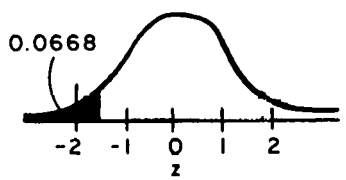


Figure 3.1.a. Pr(z <-1.5)

<u>Example 3.2</u> Normal Distribution

$H_0: \mu = 400$, $\sigma^2 = 36$, $\alpha = 0.05$

$H_A: \mu > 400$

Observed: $x_0 = 415$ $z_0 = \frac{x_0 - \mu_0}{\sigma} = \frac{415 - 400}{6} = 2.5$

$$Pr(x > 415) = Pr(z > 2.5) = 0.00621$$

Since $0.00621 < 0.05$, reject H_0: $\mu = 400$. The mean is evidently <u>greater</u> than 400.

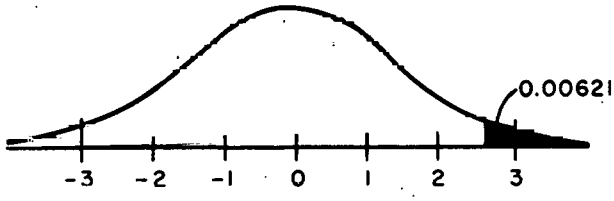


Figure 3.1.b. $Pr(z > 2.5)$

3.2 Confidence Interval for the Mean

Suppose $x_1, x_2, \ldots, x_n$ are independent measurements of some physical characteristic, such as length or weight. If $E(x_i) = \mu$, for all i, then

$$E(\bar{x}) = E\left[\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\right] = \mu$$

Thus, the mean of the average is the same as the mean of the individual variables. Hence, $\bar{x}$ is called an <u>unbiased</u> estimate of μ. Any single observation can be used as a point estimate of μ but we greatly prefer $\bar{x}$. Why? Consider the variance of $\bar{x}$. If $Var(x_i) = \sigma^2$ for all i, and since all x are independent of each other as well as all distributed in the same manner, then

$$Var(\bar{x}) = Var\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n^2}\Sigma Var(x_i) = \frac{n\sigma^2}{n^2} = \frac{1}{n}\sigma^2 .$$

That is, $\bar{x}$ has a variance which is 1/n as large as the variance of a single observation. Thus, the variability of $\bar{x}$ is much smaller than the variability of a single observation. We are more sure or confident of the location of the mean μ.

However, if only the average $\bar{x}$ is reported, we still know nothing about the variance of the estimate and do not know whether to place much faith in $\bar{x}$ as a good estimate of μ or not. To describe whether or not $\bar{x}$ is a good estimate we need to know its variance, or an estimate of the variance. For the moment we shall assume σ^2 is known and proceed to define an interval estimate for μ.

3.2.1 Interval Estimate for Mean μ

To decide how good $\bar{x}$ is as an estimate for μ, we need to know how likely or probable $\bar{x}$ is as a value for α compared to other values. This type of information is provided in an interval statement for μ. If $\bar{x}$ is the average of the n observations, and σ^2 is the known variance of x_i, then an interval estimate for μ is

$$\bar{x} \pm \sigma/\sqrt{n}$$

That is, $\bar{x}$ is reported along with its standard deviation, sometimes called the standard error. Of course, another interval is $\bar{x} \pm 2\sigma/\sqrt{n}$, and yet another is $\bar{x} \pm 3\sigma/\sqrt{n}$. In fact, there are any number of possible interval estimates. Which one should we use?

Let us first recognize that the larger n is, i.e., the more observations taken, the smaller will be the interval. The smaller the interval estimate is, the more confident we are in the estimated value of μ. But what is the probability that the interval even contains the true value of μ? That depends on the multiple of σ in the interval expression. It also depends on the probability distribution of $\bar{x}$. The distribution of $\bar{x}$ may be of the same general form as the distribution of the individual x's but with a variance σ^2/n. This distribution could be normal, binomial, Poisson, exponential, etc. But some of these distributions are or can be asymmetric, and we have presented a symmetric interval.

Of course, we are not restricted to a symmetric interval. If we know the exact distribution of $\bar{x}$, we can find values μ_L and μ_U, a lower and upper limit of μ, which would encompass μ with a high probability called the confidence level, γ. That is, we could calculate two values based on $\bar{x}$ and $\sigma/\sqrt{n}$ such that 95% of the time when we take observations and construct the interval, the interval will contain the true mean μ. In other words, 95% (or 0.95 probability) is the measure of our confidence in the location of μ. We call such an interval a 95% confidence interval. We can, of course, use any probability value instead of 0.95, e.g., 0.90, 0.99, or even 0.67. The interval measures our degree of belief for μ. Our belief is greatest at $\mu = \bar{x}$, our best estimate of μ. As we move away from $\bar{x}$, our degree of belief in

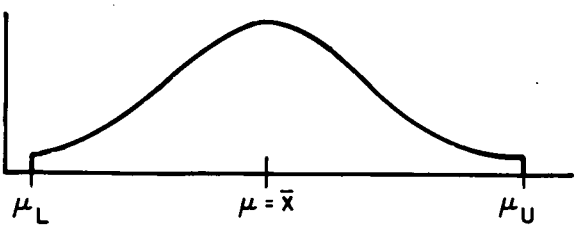


Figure 3.2. Degree of Belief

μ decreases, but we are willing to accept these values of μ as possible only up to a point. Those points beyond which we will no longer believe as possible values of μ are the interval boundaries μ_L and μ_U. The degree of our belief in the value of μ becomes so small (i.e., the probability is so small) at μ_L or μ_U that we say we cannot believe it. But 95% of the distribution lies between the boundaries.

The above has been a general discussion of confidence intervals for μ. In the next section we will see that one distribution usually suffices for all confidence intervals for means.

3.2.2 The Central Limit Theorem

In the above section it seems that we are faced with the problem of determining the distribution of the individual measurements, and then, solving that, must find in some cases asymmetric limits around $\bar{x}$. In practice, for the problem of providing confidence intervals for a mean, we do not have all these problems because of a simple but powerful theorem known as the central limit theorem.

Theorem 3.1

If $x_1, x_2, \ldots, x_n$ are identically and independently distributed with mean μ and variance σ^2, then $\bar{x}$ is distributed approximately as a normal distribution with mean μ and variance σ^2/n for sufficiently large n (see Appendix C).

In practice n does not have to be very large for the approximation of the normal distribution to the true distribution to be good. For symmetric distributions, n = 3 or 4 is usually sufficient. For asymmetric distributions, n depends on the degree of asymmetry but, again, moderate size such as n = 6 to 8 is usually large enough.

Now, given that $\bar{x}$ is, for practical purposes, distributed normally with mean μ and variance σ^2/n, we can write down a symmetric 95% confidence interval for μ, which is in fact the shortest possible 95% confidence interval,

$$x \pm 1.96 \ \sigma/\sqrt{n} \quad .$$

For 90% confidence the value is $\pm 1.645\,\sigma/\sqrt{n}$, or in general

$$\bar{x} \pm z_{\alpha/2}\ \sigma/\sqrt{n},$$

where $\gamma = 1 - \alpha = 0.90$ is the confidence level. Note that for repeated samples, $\bar{x}$ will vary, but the width of the interval is always the same, since σ is assumed known. An interpretation of a 95% confidence interval, then, is that for repeated samples of size n, 95 out of 100 (or 19 of 20) intervals properly constructed will contain the correct value μ.

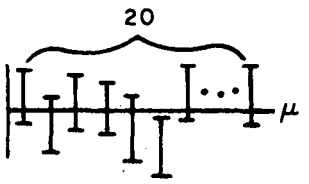


Figure 3.3. 95% Confidence Intervals: $\bar{x} \pm 1.96\ \sigma/\sqrt{n}$

Another interpretation is that our belief in the value of μ is a random variable with a normal probability density function with mean $\bar{x}$ and variance σ^2/n. That is, $\bar{x}$ is the most believable value of μ, but other values are also probable. We stretch our degree of belief so that 95% of the values in the distribution are considered possible, but anything beyond $\pm 1.96\ \sigma/\sqrt{n}$, we will declare unbelievable.

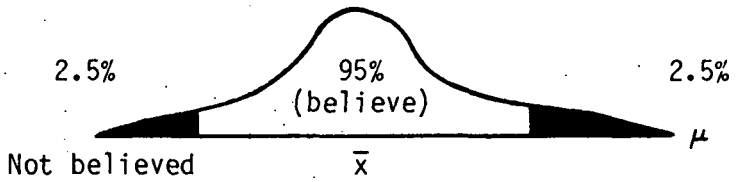


Figure 3.4. Degree of Belief in μ

As a result of the central limit theorem, then, we have an unified approach to the problem of estimating the mean, the location parameter of a distribution, and a procedure for evaluating the usefulness, or precision, or reproducability of that estimate. This approach is the confidence interval for the mean,

$$\bar{x} \pm z_{\alpha/2}\ \sigma/\sqrt{n}.$$

Example 3.3

Twelve readings of fuel concentration are taken on a measuring device which is known through a long history of observations to have a variance σ^2 of 81 $(\text{ppm})^2$. The average of the n = 12 readings was $\bar{x}$ = 93 ppm. The two-sided 90% confidence interval for mean fuel concentration is

$$\bar{x} - z_{0.05}\,\sigma/\sqrt{n}, \ < \mu < \ \bar{x} + z_{0.05}\,\sigma/\sqrt{n}\ .$$

Interpolating from Table III, $z_{0.05} = 1.645$; from above, $\bar{x} = 93$ and the standard deviation is $\sigma = 9$. Thus,

$$93 \pm [1.645\ (9/\sqrt{12})]$$

$$93 \pm 4.3 \text{ ppm}$$

Thus, with 90% confidence (1-0.05-0.05) and variance known, the true mean fuel concentration as measured by a particular instrument is between 88.7 and 97.3 ppm.

Example 3.4

Eight observations are taken on the green density of fuel pellets after compaction. The results are 0.54, 0.48, 0.59, 0.52, 0.46, 0.53, 0.45, and 0.43. Assume they follow a normal distribution

$$\bar{x} = 0.50,\ n = 8$$

Then, if $\sigma^2 = 0.0025$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{0.50 - \mu}{0.05/\sqrt{8}}$$

The 95% confidence interval for the mean value of the distribution of green densities of fuel pellets is

$$\mu : \quad [0.50 \pm 1.96 \times 0.05/\sqrt{8}] \quad ,$$

$$0.465 < \mu < 0.535$$

*3.2.3 A Second Central Limit Theorem

There is a second, more general, central limit theorem (CLT-II) which we will state without proof. If we consider that most observations x are a representative of some deterministic function μ perturbed by some random error ϵ, we have the model

$$x = \mu + \epsilon .$$

In many situations this error term ϵ is assumed to be normally distributed. The second central limit theorem gives a justification for this assumption. In reality most errors ϵ are not single errors but an accumulation or net result of many errors from known sources and unknown sources, identifiable and unidentifiable. For example, in recording the percent concentration of a certain molecule during reaction, the reading may be subject to errors regarding

1. exact time of reading
2. initial concentrations
3. temperature
4. pressure
5. atmospheric conditions
6. recording devices
7. operator's attitude and condition

The point is that many factors contribute to the final error; but, according to CLT-II as long as all errors follow certain basic conditions (i.e., all distributions have mean, variance, and third absolute moment) and no one error dominates all the others, then the distribution of the total error quite likely has a normal distribution regardless of the distributions of the individual errors.

Formally, the theorem states:

Theorem 3.2

Let x_i be independently distributed random variables with distribution functions $f_i(x_i)$, means μ_i, variances σ^2_i, and finite third absolute moments ω^3_i,

$$\omega^3_i = \int |x_i - \mu_i|^3 f_i(x_i)\, dx_i,$$

$$i = 1, 2, 3, \ldots, n.$$

Let $\mu = \Sigma\mu_i$, $\sigma^2 = \Sigma\sigma^2_i$, and $\alpha = (\Sigma\omega^3_i)^{1/3}$.

If $\lim_{n\to\infty} \frac{\alpha}{\sigma} = 0$, then Σx_i is distributed as a normal distribution with mean μ and variance σ^2.

3.3 Hypothesis Test for Mean, Variance Known

In Section 3.1 we discussed the basic logic behind a test of hypothesis. In Section 3.2 we found that because of the central limit theorem, the average $\bar{x}$ is normally distributed with mean μ and variance σ^2/n. We say then that the sampling distribution of $\bar{x}$ is $N(\mu, \sigma^2/n)$ or that the normal is the reference distribution for $\bar{x}$. Thus, using

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

we may obtain a much more precise test of hypothesis by using n observations rather than 1.

Example 3.5

Suppose a sample of 36 observations were taken from a population with a known variance of $\sigma^2 = 9$. The average was $\bar{x} = 6$. If we hypothesize that the mean is $H_0: \mu = 5$ against the alternative $H_A: \mu > 5$, what is the probability of obtaining $\bar{x} = 6$ or higher?

$$H_0: \mu = 5, \sigma^2 = 9, n = 36$$

$$H_A: \mu > 5$$

$$\bar{x} = 6$$

Then the random variable $\bar{x}$ is N ($\mu = 5, \frac{\sigma^2}{n} = \frac{9}{36}$). The standard normal deviate statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-5}{1/2}$, and

$$\Pr(x > 6 | \mu = 5, \sigma^2 = 9, n = 36)$$

$$= \Pr(z > \frac{6 - 5}{1/2} = 2)$$

$$= \Pr(z > 2) = 0.0228 \text{ (from Table III)}$$

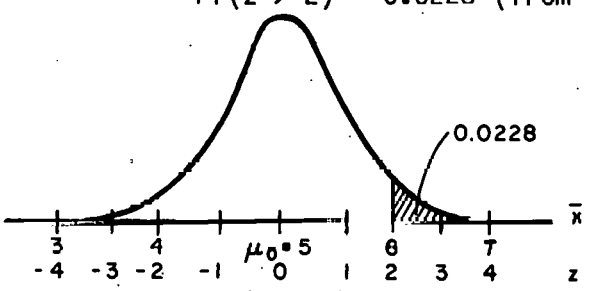


Figure 3.5. $\Pr(\bar{x} > 6) = \Pr(z > 2)$

Since 0.0228 is less than the significance level of $\alpha = 0.05$, we may declare $\mu_0 = 5$ to be an unreasonable quess of μ, since it leads to a "rare event" for $\bar{x}$.

Example 3.6

$H_0: \mu = 10, \sigma^2 = 16, \quad n = 16$

$H_A: \mu \neq 10$, (A two-sided alternative hypothesis)

$\bar{x} = 8$

$$\Pr(\bar{x} > 8 | \mu_0 = 10, \sigma^2 = 16, n = 16)$$

$$= \Pr(z > \frac{8 - \mu_0}{\sigma/\sqrt{n}} = \frac{8 - 10}{1} = -2) = 0.9772$$

or $\Pr(z < -2) = 0.0228$ - A rare event, Reject $H_0: \mu = 10$.

Another way to view hypothesis testing is by examining the confidence interval. For a particular confidence level $1 - \alpha$ all values of μ which are between the limiting values are not contradicted by the data. Thus, a confidence interval serves as a multiple test of hypothesis.

Example 3.7

For Example 3.5, we had $\bar{x} = 6$, $\sigma^2 = 9$, $n = 36$. A 95% two-sided confidence interval for μ is $\bar{x} \pm z_{0.025}\, \sigma/\sqrt{n}$,

i.e., $\quad 6 \pm 1.96 \cdot (3/6)$

or $\quad 5.0 < \mu < 7.0$.

Thus, all values for μ between 5.0 and 7.0 are not contradicted by the data. There is no evidence to reject any of these values, based on the 95% confidence interval.

3.4 An Estimate of the Variance

In the previous sections we have assumed the variance σ^2 of the distribution of observations to be known. In most cases, this is an unrealistic situation. We require an estimate of the variance from the same data from which we estimate the mean. Of course, an estimate of σ^2 from another set of data may be used, if available, as long as there is some assurance that the variance is the same for both sets.

Assuming that the observations $x_1, x_2, \ldots, x_n$ come from a normal distribution with mean μ and variance σ^2, we will use the estimate s^2 for σ^2, where

$$s^2 = \frac{1}{n-1} \Sigma(x_i - \bar{x})^2$$

or equivalently,

$$s^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - (\Sigma x_i)^2 / n\right] \quad .$$

Note that s^2 is not the maximum likelihood estimate for σ^2 given in Appendix B.1.c, but is an <u>unbiased</u> estimate of σ^2.

3.4.1 The Chi-Square (χ^2) Distribution for s^2

To show that s^2 is unbiased for σ^2, we need to give its distribution. Let x_i be a normally distributed variable with mean μ and variance σ^2; then

$$z_i = \frac{x_i - \mu}{\sigma} \qquad \sim \qquad N(0, 1).*$$

The distributional form is

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right]$$

$$f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \quad , \; z_i = \frac{x_i - \mu}{\sigma};$$

$$-\infty < x_i < \infty \; , \; -\infty < z_i < \infty$$

Let $u_i = z_i^2$, $u_i > 0$. Then, applying the proper transformation, we have

$$f(u_i) = \frac{1}{\sqrt{2\pi}} u^{-1/2} e^{-u_i/2}$$

*The symbol, $\sim$ means "is distributed as".

which is known as a chi-square distribution with one degree of freedom, χ^2_1. That is, a squared normal deviate has a χ^2_1 distribution. Now, consider n independent variables x_i, i = 1, 2, ...,n, identically distributed as N (μ, σ^2). Then, each u_i, where

$$u_i = z_i^2 = \left(\frac{x_i - \mu}{\sigma}\right)^2$$

has a χ^2_1 distribution, and

$$u = \sum_{i=1}^{n} u_i$$

has a χ^2_n distribution, where

$$f(u) = \frac{1}{2^{\frac{n}{2}} \Gamma\frac{n}{2}} u^{\frac{n}{2}-1} e^{-\frac{1}{2}u}, \quad u > 0.$$

Thus, the sum of n independent squared normal deviates has a χ^2_n distribution, where the parameter n, or degrees of freedom, is the number of independent normal deviates involved in u. The mean of χ^2_n distribution is n, and its variance is 2n, an important fact.

Now, returning to s^2,

$$s^2 = \frac{1}{(n-1)} \Sigma(x_i - \bar{x})^2 = \frac{1}{n-1} \Sigma\left[(x_i - \mu) - (\bar{x} - \mu)\right]^2$$

$$= \frac{1}{n-1} \left[\Sigma(x_i - \mu)^2 + n(\bar{x} - \mu)^2\right]$$

But, $\sum_i \frac{(x_i - \mu)^2}{\sigma^2}$ is distributed as χ^2_n with mean n, and

$$\frac{n(\bar{x} - \mu)^2}{\sigma^2} = \frac{(\bar{x} - \mu)^2}{\sigma^2/n}$$

is distributed as χ^2_1 with mean 1. Thus, $\Sigma(x_i - \mu)^2$ follows a $\sigma^2 \chi^2_n$ distribution, and $n(\bar{x} - \mu)^2$ follows a $\sigma^2 \chi^2_1$ distribution and, by Cochran's theorem,

$$\Sigma(x_i - \mu)^2 - n(\bar{x} - \mu)^2 = \Sigma(x_i - \bar{x})^2$$

follows a $\sigma^2\chi^2_{n-1}$ distribution with mean $(n-1)\sigma^2$. Hence, $(n-1)s^2 \sim \sigma^2\chi^2_{n-1}$

$$E(s^2) = \frac{1}{n-1} E\left[\sigma^2 \chi^2_{n-1}\right] = \frac{(n-1)\sigma^2}{n-1} = \sigma^2 \quad .$$

Thus, we have shown that s^2 follows a χ^2 distribution and is an unbiased estimated of σ^2. In general we denote this fact by $\nu s^2/\sigma^2 \sim \chi^2_\nu$, where ν is the degreee of freedom for s^2. The χ^2 distribution is called the reference distribution for s^2 and is used in making inferences about a single variance. Note that although the n observations are n independent normal variables, we have n-1 degrees of freedom for s^2 here in the situation of sampling from a single population, since we subtract $n\bar{x}^2$ from Σx^2_i. We lose one degree of freedom due to estimating μ by $\bar{x}$.

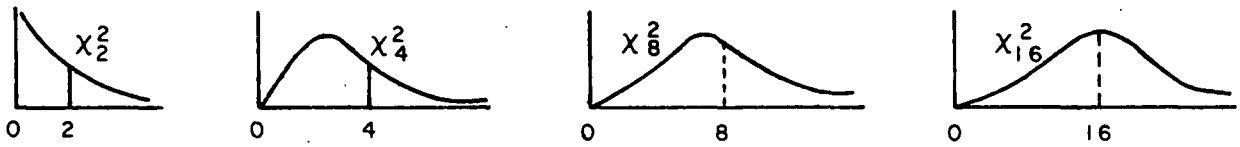


Figure 3.6. Chi-Square Distributions

In other words, a linear constraint has been placed on the n individual observations; that is, by definition of $\bar{x}$, the term $\Sigma(x_i-\bar{x})$ must sum to zero. This fact reduces the freedom of the n observations by 1 dimension. Consider that prior to observation, $x_1, x_2, \ldots, x_n$ may be free to define any point in n-dimensional space. If a restriction is applied, such as $\Sigma(x_i-\bar{x}) = 0$, an anchor has been placed on the observations, limiting their freedom to n-1 dimensions. In general, for every parameter that is estimated, such as the mean, μ, a new restriction is placed on the data. As a result, the degrees of freedom left to estimate the variance from the observations are reduced from n independent observations, if all parameters are known to $\nu = n-p$, where p parameters (i.e., linear restrictions) are estimated.

3.4.2 Inference on the Variance

Now that we have an estimate s^2 for the variance, we may ask if the estimate of σ^2 is any good, or if some hypothesized value for σ^2 is reasonable. Let us do two examples.

Example 3.8

The thickness of nine flat metal components are measured to determine the variance of the manufacturing process. That is, it is desired to determine how precisely the process can reproduce a component of average thickness. The variance includes component to component differences as well as measurement uncertainty. Assuming a normal distribution for these measurements, an unbiased estimate of the variance based on the data below is found to be 8.75 $(\text{mils})^2$.

Component Thickness (mils)

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
156	6	36
150	0	0
152	2	4
148	-2	4
151	1	1
151	1	1
146	-4	16
148	-2	4
148	-2	4

$$s^2 = \sum_{i=1}^{9} \frac{(x_i-\bar{x})^2}{n-1} = \frac{70}{8} = 8.75, \quad \nu = n-1=8$$

$s = 2.96$ mils

Alternate methods of calculation:

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

$$= \Sigma x_i^2 - (\Sigma x_i)^2/n$$

$\Sigma x_i = 1350 \quad \Sigma(x_i-\bar{x}) = 0 \quad \Sigma(x_i-\bar{x})^2 = 70$

$\bar{x} = 150$, $n=9$

1. Test of Hypothesis

A one-sided test of hypothesis that the true variance is $(2.5 \text{ mils})^2$ is based on a chi-square distribution with $\nu = n-1=8$ degrees of freedom. For a 5% significance level, the test is as follows:

$H_0: \sigma^2 = (2.5)^2$, $n=9$, $\alpha=0.05$

$H_A: \sigma^2 > (2.5)^2$

Test Statistic = $\chi_8^2 = \nu s^2/\sigma^2$

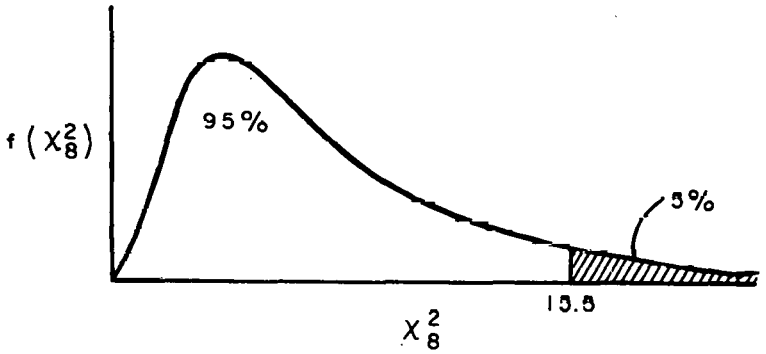


Figure 3.7. A χ_8^2 Distribution

The value of a χ_8^2 distribution that allows 5% of the distribution in the right hand tail is approximately 15.5 (Table IV). Thus, if the calculated value of χ_8^2 based on $\sigma^2 = (2.5)^2$ exceeds 15.5, the observation of $s^2=8.75$

can be considered a rare event and the null hypothesis $\sigma^2 = (2.5)^2$ can be rejected at the 5% significance level.

$$\chi^2_8 = \frac{\nu s^2}{\sigma 2} = \frac{8(8.75)}{6.25} = \frac{70}{6.25} = 11.2.$$

Since 11.2 < 15.5, there is no reason to reject H_0.

2. Confidence Interval

A two-sided 95% confidence interval for σ^2 is obtained by considering the probability statement

$$\Pr(\chi^2_{8,0.975} < \chi^2_8 = \frac{\nu s^2}{\sigma 2} < \chi^2_{8,0.025}) = 0.95$$

where

$$\Pr(\chi^2_8 > \chi^2_{8,0.025}) = 0.025$$

$$\Pr(\chi^2_8 > \chi^2_{8,0.975}) = 0.975 \text{ (i.e., } \Pr(\chi^2_8 < \chi^2_{8,0.975}) = 0.025)$$

Solving the inequality for σ^2 results in

$$\frac{\nu s^2}{\chi^2_{8,0.025}} < \sigma^2 < \frac{\nu s^2}{\chi^2_{8,0.975}}$$

From the above, $\nu s^2 = 70$, and from Table IV,

$\chi^2_{8,0.025} = 17.5$ and $\chi^2_{8,0.975} = 2.18$, we get

$$\frac{70}{17.5} < \sigma^2 < \frac{70}{2.18}$$

$$4.0 < \sigma^2 < 32.1 .$$

Note that $\sigma^2 = (2.5)^2$ is within the interval, so that a 95% two-sided test would not reject this hypothesis. Also, note that the interval is not symmetric about the estimate $s^2 = 8.75$. This is because the distribution of a χ^2 with 8 degrees of freedom is itself highly asymmetric.

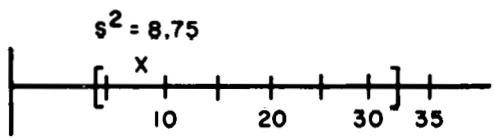


Figure 3.8. A 95% Confidence Interval for σ^2 based on χ^2_8

Example 3.9

Consider obtaining 120 observations on beginning of life loading of a type of fuel rod. Suppose the estimate of variance of these observations was determined to be $s^2 = 205.8\ (10^{-4}\ gm)^2$ with n-1 = 119 degrees of freedom. Thus, $\frac{\nu s^2}{\sigma^2} = \frac{119 s^2}{\sigma^2}$ is a χ^2_{119} variable. Is 205.8 an unusual value to obtain from χ^2_{119} distribution? Using the reference distribution χ^2_{119}, we need to calculate

$$\Pr\left(\chi^2_{119} > \frac{\nu s^2}{\sigma^2} \,|\sigma^2\right)$$

$$= \Pr\left(\chi^2_{119} > \frac{119\ 205.8}{\sigma^2} \,|\sigma^2 = \sigma_0^2\right),$$

where σ_0^2 is the hypothesized value being examined and the symbol $|\sigma^2$ means "given the value of σ^2." Suppose we test $H_0: \sigma^2 = 225$. Then,

$$\frac{119 \times 205.8}{225} = 108.8 \quad .$$

From a table of χ^2 distributions we find that for $\nu = 120$ (our case has $\nu = 119$), $\Pr\ (\chi^2_{120} > 108.8)$ is about 0.75. However, most tables do not go as high as $\nu = 120$. In fact, many tables stop at $\nu = 30$. The reason is that as ν gets large, the χ^2 distribution approaches a normal distribution. Thus, standardizing χ^2 by subtracting the mean ν and dividing by the standard deviation $\sqrt{2\nu}$,

$$z = \frac{\chi^2 - \nu}{\sqrt{2\nu}} \text{ is an N(0, 1) variable .}$$

(Other approximations of a chi-square to a standardized normal distribution exist which are slightly more accurate than the above approximation for moderate ν. The approximation used here is the most straightforward.)

In our case, $\chi^2_{119} = \frac{\nu s^2}{\sigma_0^2} = 108.8$.

$$z = \frac{108.8 - 119}{\sqrt{238}} = \frac{-10.2}{15.43} = -0.66$$

$$\Pr\ (z > -0.66) = \Pr\ (\chi^2_{119} > 108.8) = 0.7454 \approx 0.75$$

This implies that we accept $H_0: \sigma^2 = 225$.

A 95% confidence interval for σ^2 may also be calculated using a normal approximation. For a two-sided, 95% interval, we need to determine $z_{\alpha/2}$ from

$$\Pr(-z_{\alpha/2} < z < z_{\alpha/2}) = 0.95 \quad .$$

This implies $$-1.96 < z = \frac{\frac{119 \times 205.8}{\sigma^2} - 119}{15.43} < 1.96,$$

which in turn implies

$$119 - 1.96 \times 15.43 < \frac{119 \times 205.8}{\sigma^2} < 119 + 15.43 \times 1.96$$

$$119 - 30.24 < \frac{24490.2}{\sigma^2} < 119 + 30.24$$

$$88.76 < \frac{24490.2}{\sigma^2} < 149.24 \quad \text{(See Figure 3.9.)}$$

$$\frac{24490.2}{149.24} < \sigma^2 < \frac{24490.2}{88.76}$$

$$164.1 < \sigma^2 < 275.9 \quad .$$

Thus, as Figure 3.10 shows, a 95% confidence interval for σ^2 on the loading data is [164,276]. (Note that 225 is well within the interval and should be accepted.)

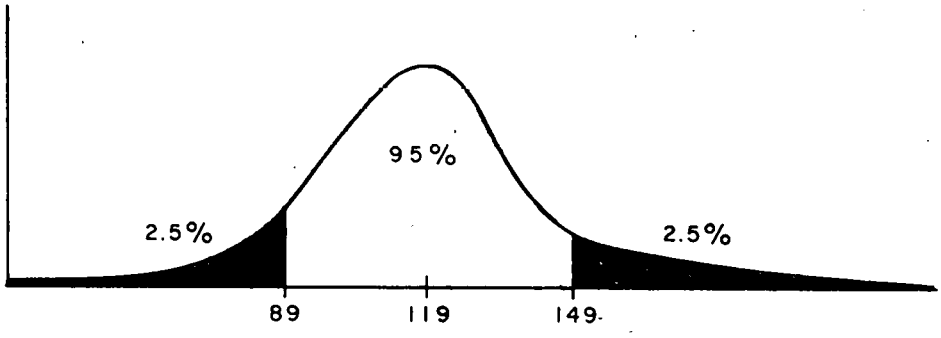


Figure 3.9. A χ^2_{119} Distribution

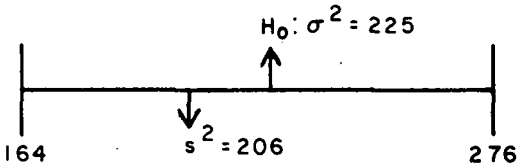


Figure 3.10. 95% Confidence Interval for σ^2 (Normal Approximation)

3.5 Inference on the Mean, Variance Unknown

It is most common that we do not know the variance of the population from which we sample. Hence we need to estimate the variance by s^2. Then, to test hypotheses on the mean of a distribution, we cannot use the normal deviate statistic z. Instead, we use an approximation to the N(0, 1) variable which depends on the estimate s^2 of σ^2 as well as n and μ_0. This distribution is called the t, or more precisely, the Student-t distribution.

3.5.1 The t-Distribution

Like the standard normal distribution, the t-distribution is also a bell-shaped curve with mean 0, but it is a more spread out distribution. It has one parameter, ν, the degrees of freedom of the estimate s^2. Thus, the t-distribution is an approximation to the standard normal distribution which depends on how good the estimate of variance is. If n were large, the estimate of σ^2 would be quite good and the t-distribution would be very close to the normal. For small n, the fact that the variance estimate is not as accurate results in the t-distribution being flatter and more spread than the standard normal.

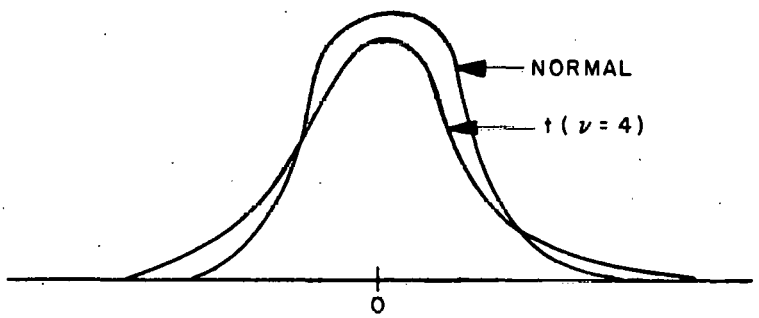


Figure 3.11. t_4 vs N(0, 1)

By definition, a t-variable is a function of a normal deviate z and a χ^2_ν distribution which is independent of z and estimates σ^2. Specifically,

$$t = \frac{z}{\sqrt{\chi^2_\nu / \nu}}$$

For inference about the mean μ of a normal distribution, we know that

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

and $\frac{\nu s^2}{\sigma^2} = \chi^2_\nu$, where $\nu = n - 1$.

$$\text{Thus,} \qquad t = \frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}} = \sqrt{n}\,\frac{\bar{x} - \mu}{s} \quad .$$

Thus, the t is like the normal deviate z that is used in tests of hypotheses for the mean when the variance is known. It is approximated for the case when the variance is unkown by substituting the estimate s for the standard deviation σ. The functional form of the t-distribution is

$$f(t\nu) = \frac{1}{\sqrt{\nu\pi}} \; \frac{\left[\frac{\nu+1}{2}\right]}{\left[\frac{\nu}{2}\right]} \; \left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}} \qquad -\infty < t < \infty$$

and has a mean of 0 and variance $\frac{\nu}{\nu - 2}$.

The value for which 5% of the distribution is in one tail of the distribution is not 1.645 as for the normal distribution, but is something larger, depending on ν; e.g.,

$$\begin{aligned} t_{5,0.05} &= 2.015 \\ t_{10,0,05} &= 1.812 \\ t_{20,0.05} &= 1.725 \\ t_{60,0.05} &= 1.671 \\ t_{\infty,0.05} &= 1.645 \quad . \end{aligned}$$

3.5.2 Confidence Interval for the Mean

A confidence interval for the mean when the variance is unknown and estimated by s^2 is constructed in exactly the same way as in the variance known case, where we now replace σ by s and $z_{\alpha/2}$ by $t_{\nu,\alpha/2}$; i.e., a 95% confidence interval is

$$\bar{x} - t_{\nu,.025}\, s/\sqrt{n} < \mu < \bar{x} + t_{\nu,0.025}\, s/\sqrt{n} \quad .$$

Example 3.10

The observations on the volume percent of a solid lubricant added to fuel compositions before compaction into green pellets are 1.20, 1.27, 1.33, 1.19, 1.09, and 1.24. Assume they follow a normal distribution:

$$n = 6,\ \bar{x} = 1.22,\ s^2 = 66.4 \times 10^{-4}$$

$$t_5 = \frac{\bar{x} - \mu}{s/\sqrt{n}}\ ,\quad \Pr\left(-t_{5,0.025} < t < t_{5,0.025}\right) = 0.95$$

From the table, $t_{5,0.025} = 2.571$

and

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.22 - \mu}{\sqrt{\frac{664}{6} \times 10^{-5}}} = \frac{1.22 - \mu}{3.33 \times 10^{-2}}$$

Thus,

$$-t_{5,0.025} < t < t_{5,0.025}$$

is

$$-2.571 < \frac{1.22 - \mu}{3.33 \times 10^{-2}} < 2.571\ ,$$

which gives $1.22 - 2.571 \times 3.33 \times 10^{-2} < \mu < 1.22 + 2.571 \times 3.33 \times 10^{-2}$

$$1.22 - 0.086 < \mu < 1.22 + 0.086$$

Thus, with 95% confidence μ lies within

$$[1.134,\ 1.306]$$

A 99% confidence interval would use $t_{5,0.005} = 4.032$ and would include

$$[1.086,\ 1.354]$$

That is, with 99% confidence, μ is captured by the interval.

A test of hypothesis would be performed just as before, also with s and t_ν replacing σ and z. In the above example, any value of μ which may be hypothesized between 1.134 and 1.306 would not be rejected by the data for a 5% two-sided test, since they lie within the 95% confidence interval.

3.6 Determining Sample Size n

We recognize that decisions based on observed data which are subject to errors of a random nature are, in turn, subject to error. We define these errors in the decision process of a test of hypothesis. Controlling the size of the error in the decision making process, however, is the object of statistical theory, particularly the area we will later call the design of experiments. In the present context of dealing with a single population, the selection of a sample size is the controlling factor in the size of the errors of inference.

3.6.1 Determination of n for Given Confidence Interval Width

The single most important aspect in determing the size of inference errors is the sample size n. In some instances in which the experimenter is seeking to gain information rather than test any particular hypothesis, the sample size may be chosen to assure a confidence interval of a specified width.

Suppose that we wish to determine the mean value of a process with a high degree of precision. More precisely, suppose we desire that the estimated value $\bar{x}$ be within $\pm$ q units of the true mean, i.e.,

$$|\bar{x} - \mu| < q \quad ,$$

with a high degree of confidence. Equivalently, we desire the confidence interval for μ to be no longer than 2q. If we standardize the statistic $\bar{x}$, we find

$$q = \frac{\sigma z_{\alpha/2}}{\sqrt{n}}$$

where σ is the known standard deviation of x_i and $z_{\alpha/2}$ is the appropriate normal deviate value. Thus, to assure with $100(1-\alpha)\%$ confidence that $\bar{x}$ is within q units of μ, when σ is known, we need only to solve for n,

$$n = \left(\frac{\sigma z_{\alpha/2}}{q}\right)^2$$

Example 3.11

Suppose that you want to determine how many observations would be required to estimate the mean center grain size of fuel pellets to within q = 0.50 ASTM numbers with 95% confidence. For σ = 0.66 ASTM No.:

$$n = (.66 \times 1.96/0.50)^2 = (2.59)^2,$$

$$= 7 \text{ observations},$$

where the value of n is rounded up to the nearest integer to ensure that the confidence level is at least the value stated.

The above is a well-defined problem, since q is specified, σ is known, and $z_{\alpha/2}$ easily obtained given the confidence level desired. If σ is unknown, replacing σ by s and $z_{\alpha/2}$ by $t_{\nu,\,\alpha/2}$, we need to solve

$$n = \left(\frac{s t_{\nu,\,\alpha/2}}{q}\right)^2$$

where for estimating the mean of a population, $\nu = n - 1$. This is an iterative process, however, since ν and n must be in close agreement. Alternatively, instead of specifying an absolute value of q, we can define q in terms of the number of standard deviations $\bar{x}$ is to be from μ. That is, since s is an estimate of σ, let q = Ds.

Then, $$n = \left(\frac{s t_{\nu, \alpha/2}}{Ds}\right)^2$$
$$= \left(\frac{t_{\nu, \alpha/2}}{D}\right)^2$$

As a first approximation, we can use z instead of t to solve for n.

3.6.2 Errors of a Test of Hypothesis

We have seen that we can perform statistical tests for the purpose of accepting or rejecting as reasonable a hypothesized value of a parameter. We assign a level of significance α to the test, indicating that we are not positive in our judgment. We can, in fact, make two errors in judgment:

1. Reject a true hypothesis (Type I error)
2. Accept a false hypothesis (Type II error)

We would like to make these errors as infrequently as possible. We determine the frequency of these errors by their probability of occurrence.

Pr (Reject a true hypothesis) = α
Pr (Accept a false hypothesis) = β

We should see that Type I error is, in fact, the definition of a significance test, i.e., for a one-sided test of $H_0: \mu = \mu_0$ against $H_A: \mu > \mu_A$,

Pr (Reject a true hypothesis)

$$= \Pr(\text{reject } H_0 \mid H_0 \text{ true})$$

$$= \Pr(\bar{x} > \bar{x}_{crit} \mid H_0: \mu = \mu_0) = \alpha \quad .$$

Thus, in testing for a mean, rejecting $H_0: \mu = \mu_0$ infers finding a value of $\bar{x}$ which is greater than (or less than) some critical value $\bar{x}_{crit}$ defined to be the boundary of reasonableness for $\bar{x}$ coming from a distribution with a mean of μ_0. In other words, $\bar{x}$ is in the rejection region, $\bar{x} > \bar{x}_{crit}$, and we reject μ_0 as a reasonable value of μ at the significance level α used. However, it is possible that $\bar{x}$ comes from the distribution with $\mu = \mu_0$. If so, we would be making an error of Type I. The probability of rejecting a true hypothesis is denoted by α.

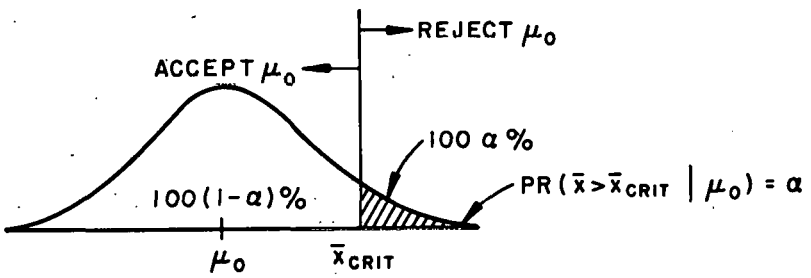


Figure 3.12. Type I Error

On the other hand, we could accept $H_0: \mu = \mu_0$ falsely. If we found $\bar{x} < \bar{x}_{crit}$, and accepted $\mu = \mu_0$, we could also be in error. Suppose $\mu \neq \mu_0$ but $\mu = \mu_0 + 2$ or $\mu = \mu_0 + 3$, etc. Then we commit Type II error. What are the chances of doing this? The probability of committing Type II error depends on μ, the correct value, or, as we shall see, some alternative value for μ which we would like to uncover. We would like to keep the probability of this type of error as low as possible. In fact, it is a measure of a good hypothesis test that the probability of Type II error is small. For a one-sided test, then,

$$\Pr(\text{Type II error}) = \Pr(\text{Accept } H_0 \mid H_0 \text{ false})$$

$$= \Pr(\bar{x} < \bar{x}_{crit} \mid \mu = \mu_A) = \beta$$

where, for testing means, $\bar{x}_{crit}$ is the critical value dividing the acceptance and rejection regions for $\mu = \mu_0$. However, the probability of Type II error depends not on μ_0 but on some other μ.

For tests of means, the general procedure is often as follows:

a) Choose a sample size n.

b) Choose a significance level α (the size of Type I error you will allow) and hypothesize $H_0: \mu = \mu_0$.

c) Calculate $\bar{x}_{crit}$: for a one-sided test,

$$\bar{x}_{crit} = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x}_{crit} = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

or, for a two-sided test,

$$\bar{x}_{crit,U} = \mu_0 + z_{\alpha/2}\, \sigma / \sqrt{n}$$

$$\bar{x}_{crit,L} = \mu_0 - z_{\alpha/2}\, \sigma / \sqrt{n}$$

d) Determine $\beta = \Pr(\text{Type II error})$ for various other values of μ

e) Take sample and calculate $\bar{x}$ from data, either accepting or rejecting μ_0.

For a given μ, if β is large, the test is not as good as it should be.

In section 3.6.3 we will see that the Type II error size can be specified for a given μ of concern, and used along with the Type I error size to determine the sample size n required to achieve the required hypothesis test. This illustrates the interdependency of α, β, and n in the construction of proper tests of hypotheses on means. Given any two, the third value can be determined.

A plot of the probability of Type II error is called the Operating Characteristic Curve (OC Curve) of the test. From it we can see how good a test is for given alternative values of μ, and compare it to OC curves for other tests (different n or α). The other side of the OC coin is called the power of the test. That is, the power is the probability of rejecting H_0 when H_0 is false, a correct judgment, but for a specified alternative hypothesis. Thus, the power measures the capability of a test to reject $H_0: \mu = \mu_0$ in favor of correctly accepting an alternative $H_A: \mu = \mu_A$. The OC and power curves for a one-sided test are shown in Figure 3.13.

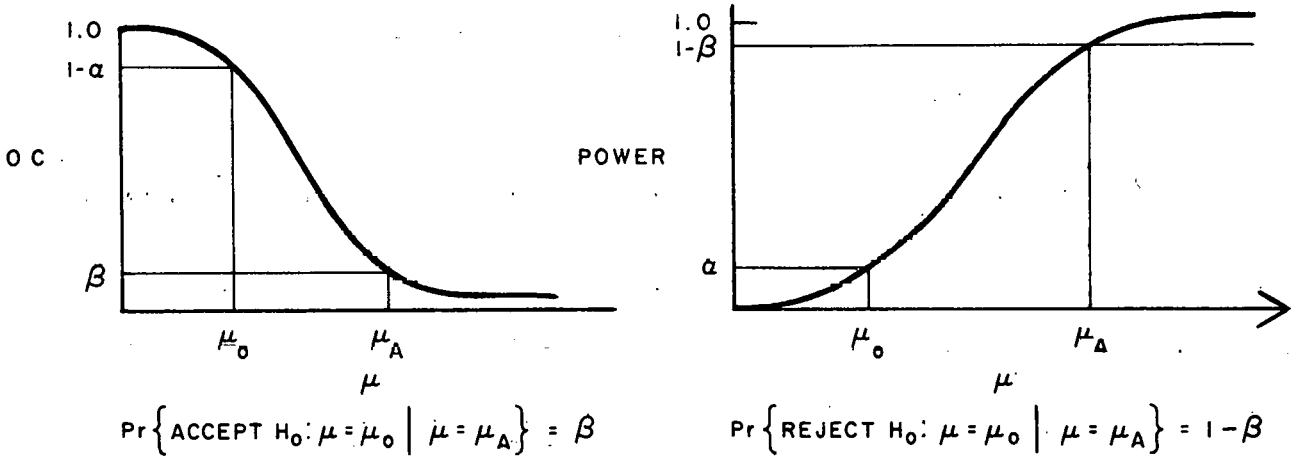


Figure 3.13. OC Curve and Power Curve for One-Sided Test of Hypothesis

Note that when $\mu = \mu_0$, $\beta = 1 - \alpha$ and $\alpha = 1 - \beta$. That is, when $\mu = \mu_0$, $\Pr(\text{reject } H_0: \mu = \mu_0 \mid \mu = \mu_0) = \alpha$, the Type I error. For a two-sided test of hypothesis, the OC and power curves are as shown in Figure 3.14.

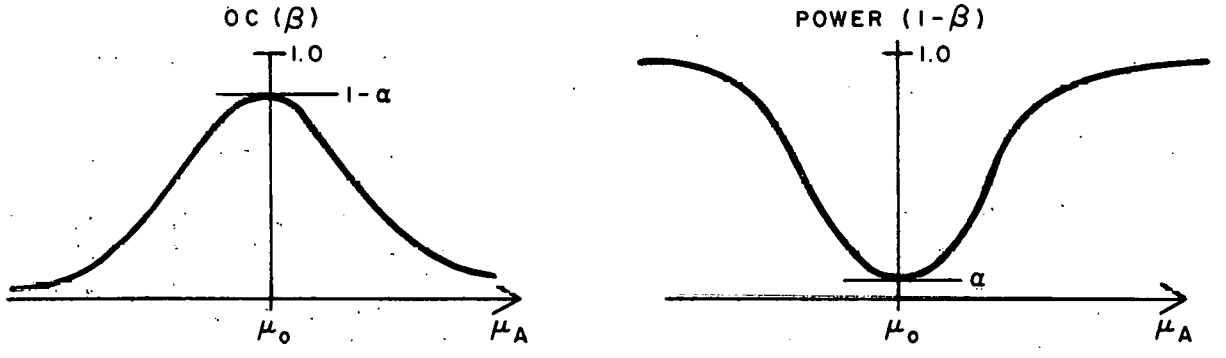


Figure 3.14. OC Curve and Power Curve for Two-Sided Test of Hypothesis

Example 3.12

Consider observing the center grain size of 25 fuel pellets. The target value for center grain size is ASTM No. 5.5 and it is desirable to detect a mean center grain size of ASTM 6.0 or higher. Suppose past analysis has indicated a standard deviation for center grain size of 0.66 (ASTM No.). Thus, for a 5% one-sided significance test for the mean, we have

Significance level: $\alpha = 0.05$

$H_0: \mu = 5.5$

$H_A: \mu = 6.0$

, $\sigma = 0.66$, $n = 25$, $\sigma/\sqrt{n} = 0.132$

1. Determine $\bar{x}_{crit}$:

Type I error: $Pr(\bar{x} > \bar{x}_{crit} | \mu = 5.5)$

$$= Pr\left(z > \frac{\bar{x}_{crit} - \mu_0}{\sigma/\sqrt{n}} = z_\alpha\right) = 0.05$$

so: $z_{0.05} = 1.645 = \dfrac{\bar{x}_{crit} - 5.5}{0.132}$

$$\bar{x}_{crit} = 5.5 + 0.132 \times 1.645 = 5.5 + 0.2171$$

$$= 5.72$$

i.e., $Pr(\bar{x} > 5.72 | \mu = 5.5) = 0.05$

$= Pr(\text{Reject } H_0 | H_0 \text{ true})$

2. Type II Error:

For $\mu = 6.0$, what is probability of falsely accepting $H_0: \mu = 5.5$?

$$Pr\left(\bar{x} \overset{(\text{Accept } H_0)}{<} 5.72 \mid \mu = 6\right) = ?$$

$$z = \frac{\bar{x} - 6}{0.132}, \quad z_{crit} = \frac{5.72 - 6}{0.132} = \frac{-0.28}{0.132} = -2.12$$

$Pr(z < -2.12) = 0.017$

i.e., probability of falsely accepting $\mu = 5.5$ when μ really equals 6.0, is only 0.017, when $n = 25$ and $\bar{x}_{crit} = 5.72$.

Or power of test to determine that $\mu = 6$ and not equal to 5.5 is 0.983, a powerful test for distinguishing between 5.5 and 6.0!

Other values of μ may be postulated and an OC curve or power curve drawn.

Table 3.1.
OC and Power for Test on Mean Grain Size

μ	Pr (Accept)=β	Power=Pr (Reject H_0)=1 - β
5.5	0.95	0.05
5.6	0.82	0.18
5.7	0.56	0.44
5.8	0.27	0.73
5.9	0.09	0.91
6.0	0.017	0.983

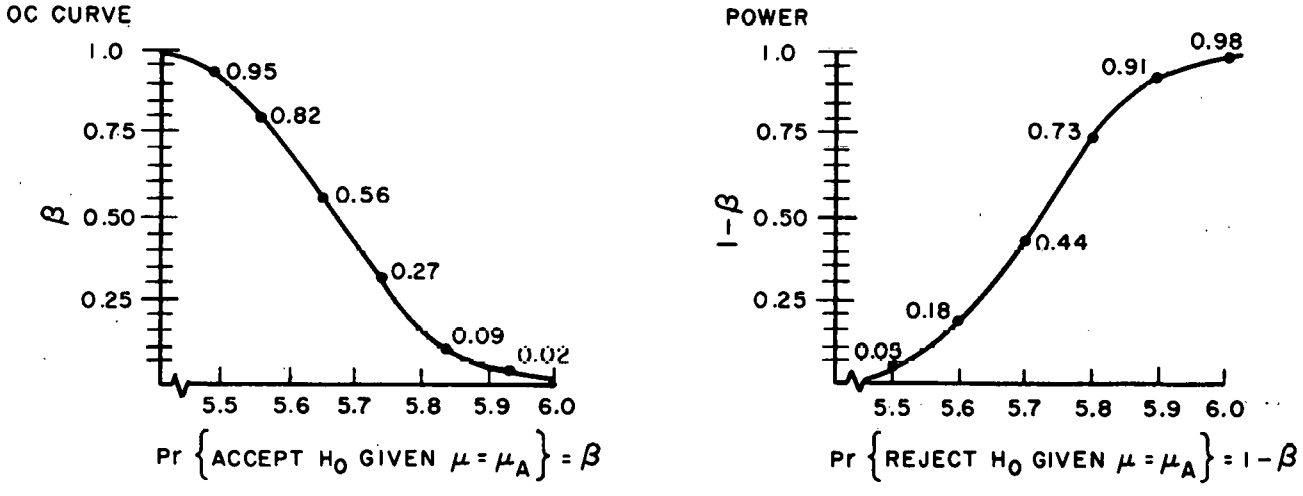


Figure 3.15. OC and Power Curves for Test on Mean Grain Size

Thus, with a power of 0.983 the test on $\bar{x}$ with n = 25 and $\bar{x}_{crit}$ = 5.72 will <u>detect</u> μ = 6.0 and reject μ = 5.5; or, in other words, the error we make in rejecting μ = 5.5 falsely is at most 0.05, and the error of accepting H_0: μ = 5.5 falsely when μ = 6.0 is at most 0.017.

3.6.3 Determination of Sample Size Using Type I and Type II Error, Variance Known

The most useful aspect of Type II error is not in determining the size of the error of falsely accepting H_0 after the test has been determined, but in determining the test itself; that is, in determining the sample size n required and the critical value required in order to produce a test which will give both a low Type I error probability and a low Type II error probability for some particular alternative hypothesis which you have in mind.

Suppose you are in charge of accepting or rejecting a lot of steel rods. The manufacturer of the rods claims the bolts to have a diameter of 20 mils. (i.e., $H_0: \mu = 20$). You know that if the average bolt diameter is as large as 24 mils (H_0: $\mu = 24$), you will have to throw out the lot. So, you are in a position of wanting to reject the lot, if it is good, with only a small probability, but if it is bad (i.e., $\mu \geq 24$) you want a high probability of detecting it. Thus, fix α, the significace level for your test of $\mu = 20$, at a pre-determined level, such as 0.05 or 0.01, and select a second probability level, β, for making the second type error of accepting a bad lot, i.e., accepting a lot whose real mean diameter is 24 mils.

Example 3.13

Consider again the center grain size problem in Example 3.12. You would like to construct a test to accept the hypothesis that $\mu = 5.5$ (ASTM No.) with a 0.05 significance level. On the other hand, you want to be relatively sure that if the true mean center grain size for this batch of pellets is as high as 6.0 (ASTM No.), you do not accept the null hypothesis that $\mu = 5.5$. Suppose you assign a probability level of $\beta = 0.10$ for making this Type II error. Furthermore, since time spent means money spent, you want to perform this test as cheaply as possible in order to achieve your goals. In Example 3.12, a test using n = 25 samples gave $\beta = 0.017$ for H_A: $\mu = 6.0$. To determine how many fewer observations are required to give $\beta = 0.10$, we proceed as follows:

a. $\Pr(\text{reject } H_0 \mid \mu = 5.5) = 0.05.$

This implies $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$ [Suppose σ is known to be 0.66]

Then

$$z_{0.05} = 1.645 = \frac{\bar{x}_{crit} - 5.5}{0.66/\sqrt{n}}$$

$$\bar{x}_{crit} = 5.5 + \frac{1.64 \times 0.66}{\sqrt{n}}$$

$$= 5.5 + 1.086/\sqrt{n}$$

b. $\Pr(\text{accept } H_0 \mid \mu = 6.0) = 0.10$

This implies $\Pr(\bar{x} < \bar{x}_{crit} \mid \mu = 6.0)$.

Then $$z = \frac{\bar{x} - 6.0}{0.66/\sqrt{n}} < z_{1-\beta} \frac{\bar{x}_{crit} - 6.0}{0.66/\sqrt{n}}, \quad z_{0.90} = -1.282$$

$$\bar{x}_{crit} = 6.0 - \frac{1.282 \times 0.66}{\sqrt{n}}$$

$$= 6.0 - \frac{0.846}{\sqrt{n}}$$

From a. and b. we have two equations in the two unknowns n and $\bar{x}_{crit}$. Solving for n, we have

$$6.0 - \frac{.846}{\sqrt{n}} = 5.5 + \frac{1.086}{\sqrt{n}}$$

or

$$(6.0 - 5.5)\sqrt{n} = 1.086 + 0.846$$

$$n = \left(\frac{1.932}{0.5}\right)^2 = 3.864^2 = 14.9$$

$$n = 15.$$

That is, in order to have <u>at most</u> 0.05 probability of rejecting $\mu = 5.5$ falsely <u>and</u> 0.10 probability of accepting $\mu = 5.5$ when μ really equals 6.0, we need n = 15 observations. Substituting back into a. we obtain

$$\bar{x}_{crit} = 5.5 + \frac{1.086}{\sqrt{n}} = 5.5 + \frac{1.086}{\sqrt{15}}$$

$$= 5.78.$$

Thus, the test is to take 15 observations with a decision line at 5.78, i.e.,

if $\bar{x} > 5.78$, reject $H_0: \mu = 5.5$, accept $\mu = 6.0$

if $x \leq 5.78$, accept $H_0: \mu = 5.5$

For the test, $\alpha \leq 0.05$, $\beta \leq 0.10$.

The procedure is general. In general terms, for $H_A: \mu > \mu_0$

a. $\frac{\bar{x}_{crit} - \mu_0}{\sigma / \sqrt{n}} = z_\alpha$; b. $\frac{\bar{x}_{crit} - \mu_A}{\sigma / \sqrt{n}} = z_{1-\beta}$

Solving for n,

$$n = (z_\alpha - z_{1-\beta})^2 \frac{\sigma^2}{(\mu_A - \mu_0)^2} ,$$

where z_α and $z_{1-\beta}$ are normal deviate values having $100\alpha\%$ and $100(1-\beta)\%$ of the distribution to the right of the critical value, μ_0 is the original hypothesis, μ_A is the alternative hypothesis under consideration, and σ^2 is the <u>known</u> variance of the measurement.

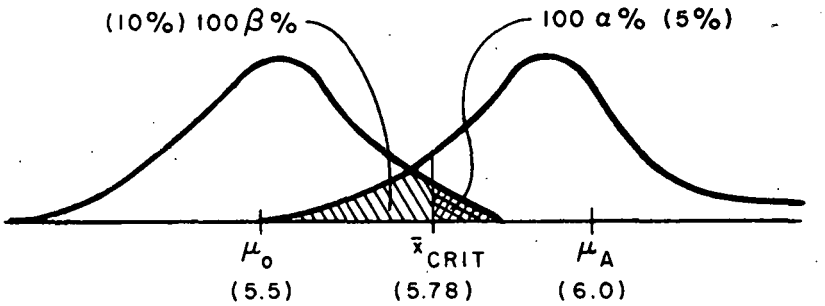


Figure 3.16. Type I and Type II Error: One Sided Alternative

Note:

1. if $\beta = 0.50$, $z_\beta = z_{1-\beta} = 0$. This is effectively what happens when Type II error considerations are ignored:

$$n = z_\alpha^2 \frac{\sigma^2}{(\mu_A - \mu_0)^2}$$

2. For a two-sided test for μ_0, just reduce α to $\alpha/2$ in each tail of the distribution; but since the null hypothesis can only be in error on one side of μ_0, there is no need to partition β into two parts.

3. It is required that σ^2 be known.

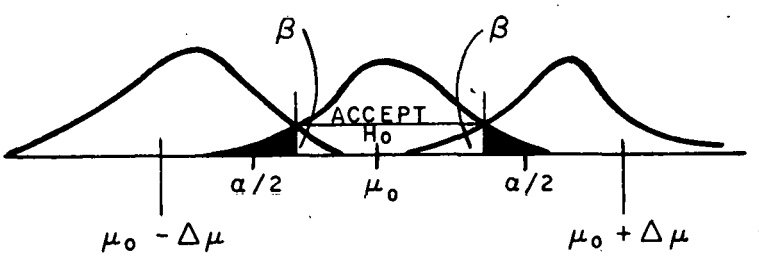


Figure 3.17. Type I and Type II Error: Two Sided Alternative

3.6.4 Sample Size Determination: Variance Unknown

Previously we assumed that σ^2 was known in order to determine the sample size required to assure a test with $\alpha = 0.05$ (probability of Type I error) and $\beta = 0.10$ (Probability of Type II error). We may still determine n using an estimate of σ^2 but we can expect n to be large due to the uncertainty involved in estimating the variance.

We could begin by making a rough guess of σ, hopefully based on some prior information. Even so, we do not know n and hence do not know which t-distribution to use. We could proceed as follows:

1. Either guess n and look up $t_{n-1,\alpha}$, or use a ballpark value of t such as 2 for $t_{0.025}$, 1.3 for $t_{0.10}$ and 1.7 for $t_{0.05}$.

2. Using these approximate t values and our guess $\hat{\sigma}$ of σ, solve for $n_{(1)}$

$$n_{(1)} = (\tilde{t}_{\alpha} - \tilde{t}_{1-\beta})^2 \frac{\hat{\sigma}^2}{(\mu_A - \mu_0)^2}$$

where $\tilde{t}_{\alpha}$ and $\tilde{t}_{1-\beta}$ are the approximate t-values for Type I and Type II errors respectively, μ_0 is the hypothesized value for μ and μ_A is the alternative value we are guarding against.

3. If $n_{(1)}$ is not what was guessed previously, repeat the process by using $t_{n_{(1)}-1,\alpha}$ and $t_{n_{(1)}-1,\,1-\beta}$.

$$n_{(2)} = (t_{n_{(1)}-1,\alpha} - t_{n_{(1)}-1,\,1-\beta})^2 \frac{\hat{\sigma}^2}{(\mu_A - \mu_0)^2}$$

4. Continue until two successive n's agree. Check to see that the corresponding critical values $\bar{x}_{crit}$ obtained from the two equations

Type I a. $\bar{x}_{crit} = \mu_0 + t_{n-1,\alpha}\hat{\sigma}/\sqrt{n}$

Type II b. $\bar{x}_{crit} = \mu_A - t_{n-1,\,1-\beta}\hat{\sigma}/\sqrt{n}$

are in reasonable agreement.

Remember that these values, n and $\bar{x}_{crit}$ for α and β consideration, depended on a guess of σ.

An alternative approach is to consider an alternative hypothesis for μ in terms of the standard deviation, σ. That is, instead of guessing σ, choose μ_A to be $\mu_0 + D\sigma$, where D is some constant. The previous procedure then only changes in that in place of

$$\frac{\sigma}{\mu_A - \mu_0}, \text{ we have } \frac{\sigma}{\mu_0 + D\sigma - \mu_0} = \frac{1}{D}.$$

then $$n_{(i)} = \frac{(t_{i,\alpha} - t_{i,1-\beta})^2}{D^2} .$$

Example 3.14

Consider again the center grain size data. Suppose we want to test $H_0: \mu = 5.5$ and want to detect a mean of $5.5 + 2\sigma$, if it exists. Thus, $H_A: \mu = 5.5 + 2\sigma$. Using $\alpha = 0.025$ and $\beta = 0.10$, and roughly $\tilde{t}_{0.025} = 2$ and $\tilde{t}_{0.90} = -1.3$, we get on the first iteration

1. $$n_{(1)} = \frac{(2 + 1.3)^2}{2^2} = \frac{(3.3)^2}{4} = \frac{10.9}{4} = 2.7 \rightarrow 3$$

2. for $n = 3$, $t_{2,\ 0.025} = 4.303$, $t_{2,0.90} = -1.886$

$$n_{(2)} = \frac{(4.303 + 1.886)^2}{4} = \frac{(6.189)^2}{4} = 9.57 \rightarrow 10$$

3. for $n=10$, $t_{9,0.025}=2.262$, $t_{9,0.90} = -1.383$

$$n_{(3)} = \frac{(2.262 + 1.383)^2}{4} = \frac{(3.645)^2}{4} = 3.3 \rightarrow 4$$

4. for $n = 4$, $t_{3,0.025} = 3.182$, $t_{3,0.90} = -1.638$

$$n_{(4)} = \frac{23.2}{4} = 5.8 \rightarrow 6$$

5. $n = 6$, $t_{5,0.025} = 2.571$, $t_{5,0.90} = -1.476$

$$n_{(5)} = \frac{16.4}{4} = 4.1 \rightarrow 5$$

6. $n = 5$, $t_{4,0.025} = 2.776$, $t_{4,0.90} = -1.533$

$$n_{(6)} = \frac{18.7}{4} = 4.7 \rightarrow 5$$

7. Stop

To test $\mu = 5.5$ with a probability of falsely rejecting H_0 of $\alpha = 0.025$, if $\mu > 5.5$ and to detect a true mean of 2σ from 5.5 with probability of $1-\beta = 0.90$, we need 5 observations. The test is,

accept $H_0: \mu = 5.5$ if

$$t = \frac{\bar{x} - 5.5}{s/\sqrt{5}} < 2.776$$

and accept $H_A: \mu = 5.5 + 2\sigma$ and reject H_0: if

$$t = \frac{\bar{x} - 5.5}{s/\sqrt{5}} > 2.766.$$

This is a lot of work if your initial guess for the t value is far off. Fortunately, there is a table, Table IX, which gives sample sizes required for various α, β, and D. We see for D = 2, α = 0.025, and β = 0.10, that n = 5.

Table 3.2
Summary of Computation

ν_{i-1}	$t_{\nu, i-1}$	$t_{1-\beta, \nu_{i-1}}$	n_i	
1. x	2.0	-1.3	3	α = 0.025
2. 2	4.303	-1.886	10	β = 0.10
3. 9	2.262	-1.383	4	$t_{\alpha, (0)}$=2.0,
4. 3	3.182	-1.638	6	$t_{1-\beta, (0)}$ = -1.3
5. 5	2.571	-1.476	5	
6. 4	2.776	-1.533	5	

Note that if we are only concerned with the width of the confidence interval for μ, set β = 0.50 and $t_{1-\beta}$ = 0.

3.7 Tolerance Intervals for a Normal Population

We are all familiar by now with a $100\gamma\%$ = 100% $(1-\alpha)$ confidence interval. The confidence interval is a statement about the location of a parameter. It gives the confidence that we place in our estimate of the parameter by placing bounds on the spread of values which may be plausible as values for that parameter, based on the data at hand. In terms of estimating the mean, we are sure, at least 95% of the time that when we calculate a confidence interval $\bar{x} \pm t_{n-1,0.025}s/\sqrt{n}$, the interval will actually contain the true mean μ.

There are other kinds of intervals, however, which are of great importance. One interval places bounds on the proportion of the sampled population contained within it at a certain degree of confidence. This is known as a tolerance interval, or to distinguish it from other type of intervals which also go by the name of tolerance, we may refer to these intervals as statistical tolerance content intervals. We shall discuss first normal tolerance intervals and later proceed to distribution-free tolerance intervals in Chapter 4.

3.7.1 Construction of Two-Sided Tolerance Intervals

For a two-sided tolerance interval, a proportion P of the population is said to be within two limits, the lower limit determined by $x_L = \bar{x} - Ks$, and the upper limit determined by $x_U = \bar{x} + Ks$, where $\bar{x}$ is the average of the n observation in the sample, s is the estimated standard deviation, and K is a tabulated value dependent on the sample size, the proportion of the population in the interval, and the confidence level. Table VII(a) gives values of K for various sample sizes, proportions and confidence levels. Note four things about this interval:

1. It deals only with a normal population,
2. K depends on 3 values, n, γ, and P,
3. It is a two-sided interval,

4. It determines the bounds of a proportion of the population, not just the sampled data. A population consists of all values, past, present, and future.

For a sample size of 10 observations, a proportion of P = 0.99 of the population and a confidence level of $\gamma = 0.95$, we find K = 4.433. Thus, at least 99% of the population should fall between $\bar{x} - 4.433\ s$ and $\bar{x} + 4.433\ s$ with 95% confidence. This interval is commonly denoted as 95/99 tolerance interval.

In the above discussion and for the values in Table VII(a), it has been assumed that the standard deviation s is based on the data on hand. Weissberg and Beatty [32] have developed tables for the construction of two-sided tolerance intervals based on a normal distribution that allow for the estimate of variance to be based on an independently obtained set of data from that used to estimate the mean.

Example 3.15

Consider again the grain size data of Example 3.12. Assuming grain sizes follow a normal population, suppose the estimate of the mean is $\bar{x} = 5.92$ and $s = 0.66$, with $n = 25$.

A 95/99 Tolerance Interval is

$$\bar{x} \pm K(n = 25,\ \gamma = 0.95,\ P = 0.99)s$$

$$5.92 \pm 3.457\ (0.66)$$

$$5.92 \pm 2.28$$

That is, with 95% confidence, at least 99% of the population of grain sizes can be expected to lie within (3.64, 8.20). [Note: The usual notation is γ/P Tolerance Interval, where γ, the confidence level is the first number to appear, and P, the proportion, the second number.] Other intervals are given below:

			Limits	
$\gamma = 1 - \alpha$	P	K(n = 25, γ, P)	Lower	Upper
95	95	2.631	4.20	7.64
95	99	3.457	3.64	8.20
99	95	2.972	3.96	7.88
99	99	3.904	3.34	8.50

Let us consider another example.

Example 3.16

A large shipment of 0 gauge wire is received. It is desired that these wires meet upper and lower specification limits (i.e., fall between limits) on the diameter. A sample of 15 observations were taken, the results being $\bar{x} = 0.338$, $s = 0.012$.

A 99/99 Tolerance Interval is calculated.

$$0.338 \pm K(n = 15, \gamma = 0.99, P = 0.99)s$$

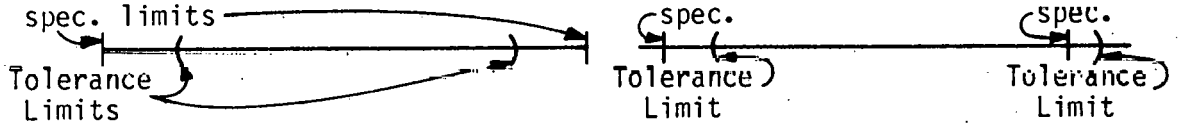
$$0.338 \pm 4.605\ (0.012)$$

$$0.338 \pm 0.05526$$

$$0.282,\ 0.394$$

[Note: Round lower limit down, upper limit up]

Thus, with 99% confidence, we can expect that at least 99% of the wires will have a diameter between 0.282 and 0.394 inches. The important question, then, is how do these values compare with the specification limits? If the calculated statistical tolerance or content interval is within the specification limits,



then all is well. If one or both of the calculated tolerance limits are outside the specification limits, then the quality of the lot is suspect. Perhaps the lot is of insufficient quality for use, or perhaps, a smaller tolerance interval, e.g., 95/90, would fall within the specs which would be acceptable to you. Finally, it is possible that the specifications are too tight. It is up to the subject experts to clarify this problem.

3.7.2 Construction of a One-sided Tolerance Limit

For a one-sided statistical tolerance content limit, a proportion of the population will lie below a certain upper limit, $x_U = \bar{x} + Ks$, or above a certain lower limit, $x_L = \bar{x} - Ks$, where K is the tabulated values for one-sided limits found in Table VII(b). These values of K differ from those of Table VII(a) for a two-sided interval. For example, for 10 observations, a proportion of 0.99 and a confidence level of 0.95, we find K = 3.981.

Example 3.17

One-sided Tolerance Interval, Normal Distribution

The n = 25 rotor shaft diameters had an average of $\bar{x} = 0.249$ in. and an estimated variance of 0.000009 $in.^2$. A one-sided 90/95 tolerance limit for rotor shaft diameters is

$$\bar{x} + Ks$$

$$0.249 + 2.132\ (0.003)$$

$$0.249 + 0.006.$$

Thus, with 90% confidence, at least 95% of all rotor shafts in this population will have diameters under 0.255. (K obtained from Table VII(b).)

3.8 Prediction Interval

Suppose a statistical tolerance interval has been calculated. Now,how many items in a future lot of items may we expect to fall within these limits? (We may generalize a tolerance interval to consist of not just a symmetric interval about the average, but any interval containing a desired proportion of the population, such as the upper quarter of values.) One way to answer the question is to say that if a 95/95 interval were presented and n = 100 new items were in question, we could expect (with 95% confidence) that 100 P% of the N items would fall within the calculated limits, i.e., 0.95 x 100 = 95 of the 100 new observations will be within the previously calculated interval. Although this is a reasonable approach, there is yet another type of interval assigned specifically to answer this question.

A prediction interval tells us within what limits we may expect to find one or more future observations from a normal distribution. The general theory can be seen from examining the problem of predicting an interval for one future observation.

Let $\bar{x}$ be the estimate of the mean μ of a $N(\mu, \sigma^2)$ population. Since $x_i = \mu + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$ we see that for a future observation

$$\hat{x}_i = \bar{x} + \epsilon \quad , \text{ where } \bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \text{ and } \epsilon \sim N(0, \sigma^2).$$

Thus, $\hat{x}$ follows a normal distribution with mean μ and variance $(\sigma^2 + \frac{\sigma^2}{n})$. Thus,

$$t_{n-1} = \frac{\hat{x} - \bar{x}}{\sqrt{Var(\hat{x})}} = \frac{\hat{x} - \bar{x}}{\sqrt{(1 + \frac{1}{n})\, s^2}}$$

where s^2 is the estimate of σ^2. Then

a. A 95% prediction interval for a single future observation is

$$\bar{x} - t_{n-1,0.025}\, s\sqrt{1 + 1/n} < \hat{x} < \bar{x} + t_{n-1,\,0.025}\, s\sqrt{1 + 1/n}.$$

b. For an average of q future observations, we would have

$$\bar{x} - t_{n-1,\,0.025}\, s\sqrt{\frac{1}{q} + \frac{1}{n}} < \hat{\bar{x}}_q < \bar{x} + t_{n-1,0.025} s\sqrt{\frac{1}{q} + \frac{1}{n}}.$$

[Note degrees of freedom for t is still the degrees of freedom of s^2, i.e., n-1]

c. For a prediction interval containing all of k future observations, we really need to integrate a multivariate t-distribution over the appropriate range, i.e., find or solve

$$\Pr\{-L < t_1 < L;\ -L < t_2 < L;\ \ldots;\ -L < t_k < L\} = 1 - \alpha$$

where

$$t_i = \frac{\hat{x}_i - \bar{x}}{s\sqrt{1 + \frac{1}{n}}}, \quad i = 1, 2, \ldots, k, \text{ and } \pm L \text{ are the limits.}$$

These values were developed by J. Hahn and given in Table VIII. An approximation which always over-estimates the width of the interval is

$$\bar{x} \pm (1 + 1/n)^{1/2}\, s\, t_{n-1,\, \alpha/2k}$$

i.e., instead of using $t_{n-1,\, \alpha/2}$, reduce α by a factor of k

and, use $\alpha/2k$.

This is the result of assuming each t_i is independent of each other (which they are not since each $\hat{x}_i$ depends on $\bar{x}$) and

$$\Pr\{-L<t_1<L;\ldots;-L<t_k<L\} = \Pr\{-L<t_1<L\}\ \Pr\{-L<t_2<L\}\ldots\Pr\{-L<t_k<L\}$$
$$\geq (1 - k\alpha),$$

if each individual interval is at the $(1-\alpha)$ level.

[Proof: if $\Pr(S_1) = 1 - \alpha_1$, $\Pr(S_2) = 1 - \alpha_2$

$$\Pr(S_1 S_2) = \Pr(S_1) + \Pr(S_2) - \Pr(S_1 + S_2)$$
$$= 1 - \alpha_1 + 1 - \alpha_2 - \Pr(S_1 + S_2)$$

but $\max \Pr(S_1 + S_2) = 1$. Therefore,

$$\Pr(S_1 S_2) \geq 1 - \alpha_1 + 1 - \alpha_2 - 1$$
$$\geq 1 - \alpha_1 - \alpha_2$$
$$= 1 - 2\alpha \text{ if } \alpha_1 = \alpha_2$$

Thus, if we make each test at the α/k level (use $t_{\nu,\ \alpha/2k}$ for two-sided test), the total probability will be $\geq 1 - \alpha$.]

The problem is that for some k, the t value is difficult to find in tables and needs to be interpolated. Exact prediction intervals are available from Hahn's tables, where the prediction intervals is of the form $\bar{x} + rs$, and r is provided in Table VIII. Approximate prediction intervals for large enough n and k are always available for other distributions via a normal approximation.

Example 3.18

Suppose the average radius of n = 15 fuel pellets was determined to be $\bar{x}$ = 0.338 in. with an estimated standard deviation of s = 0.012 in. With 95% confidence, the prediction intervals for k = 1, 5, and 10 future pellets being in the interval are

$$n = 15,\ x = 0.338,\ s = 0.012,\ \gamma = 0.95$$

k	r	rs	$\bar{x} \pm r(k, n, \gamma)s$
1	2.22	0.02664	0.311, 0.365
5	3.04	0.03648	0.301, 0.375
10	3.37	0.04044	0.297, 0.379

The approximate prediction interval is

$$\bar{x} \pm r's$$

where $r' = (1 + 1/n)^{1/2}\, t_{n-1,\, \alpha/2k}$.

For the above data, the approximate intervals are

k	$t_{\alpha/2k}$	r'	$\bar{x} \pm r's$
1	2.145	2.215	0.311, 0.365
5	2.977	3.075	0.301, 0.375
10	3.326	3.435	0.296, 0.380

Approximate Prediction Intervals

The prediction interval for the average of the k future observations is $\bar{x} \pm r''s$, where

$$r'' = (1/k + 1/n)^{1/2}\, t_{n-1,\, \alpha/2}$$

For the above data, the intervals are:

k	$(1/k + 1/n)^{1/2}$	r"	$\bar{x} \pm r''s$
1	1.033	2.215	0.311, 0.365
5	0.516	1.108	0.326, 0.350
10	0.408	0.876	0.327, 0.349

3.9 Inference About the Distribution

We now turn to a very fundamental question. How do we know whether the distribution we have assumed is appropriate? Implicitly, in all other statistical inferences we make we assume we know the distribution, usually a normal distribution. It would be reassuring to know that the distributional assumption is correct. For a sufficiently large sample size, it may well be that a histogram adequately describes a particular distribution, whose parameters may then be estimated. For a small sample size, however, a histogram would not yield sufficient information about the distribution. What

is needed is an analytical approach. Such an approach is the goodness of fit test for a distribution. The object is to hypothesize a distribution and then test the data for the adequacy of the fit or agreement between data and the hypothesized probability distribution.

In fact, a histogram or some other kind of plot is an essential first step in determining the distribution to be hypothesized. In addition, one's knowledge of the physical situation should be utilized in arriving at a null hypothesis. The test compares the observed data at each data point if discrete, or in each histogram cell if continuous, with the expected number of observations predicted by the hypothesized distribution.

3.9.1 The Chi-Square Goodness of Fit Test

The objective here is to test the null hypothesis that a collection of data follows a certain distribution function, f(x). Thus,

$$H_0: \; x_i \sim f(x)$$

$$H_A: \; x_i \text{ not distributed as } f(x).$$

The data can be considered to be outcomes from a multinomial distribution: exactly, if the data is discrete; approximately by use of discrete intervals as in the construction of a histogram, if the data is continuous. The form on the multinomial distribution is

$$p(n_1, n_2, \ldots, n_k) = \binom{N}{n_1, n_2, \ldots, n_k} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

where n_i are the number of observations found in the ith interval, $x_{i-1} < x < x_i$, p_i is the true probability of an observation x from the hypothesized distribution f(x) falling into the ith interval, i.e.,

$$p_i = \begin{cases} Pr(x = x_i), & \text{if } f(x) \text{ discrete} \\ Pr(x_{i-1} < x < x_i), & \text{if } f(x) \text{ continuous} \end{cases}$$

and also

$$\sum_{i=1}^{k} p_i = 1 \qquad \sum_{i=1}^{k} n_i = n$$

$$E(n_i) = np_i$$

$$\text{Var}(n_i) = np_i(1 - p_i)$$

It makes sense, then, that if f(x) is the correct hypothesized distribution for the data obtained, the observed number of observations n_i in each interval should agree with the expected number of observations, $E_i = np_i$, in each interval. Of course, the agreement is not expected to be perfect due to random variation. Thus, the hypothesis test must answer the question, "Does the data confirm or contradict the assumption that f(x) is the underlying distribution?"

The Chi-Square goodness-of-fit test has been devised to answer this question. It is based on a approximate distribution for the likelihood ratio test, a discussion of which is beyond the scope of this text. However, the application of the test is as follows: The test is to compute the following statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i are the observed frequencies in each interval, i = 1, 2, ..., k,

$E_i = np_i$ are the expected frequencies based on an assumption of f(x).

The test statistic is approximately a χ^2_ν, where the degrees of freedom ν is the number of outcomes k minus the number of linear relationships satisfied by $O_i - E_i$ (e.g., $\Sigma(O_i - E_i) = n - n = 0$). In general, if the distribution is completely hypothesized, then $\nu = k - 1$, but if parameters of the distribution need to be estimated, a degree of freedom is lost for each such parameter estimated from the data. How good the approximation is depends on the expected frequency in the interval or cell with smallest probability of occurring. A good rule of thumb to follow is to require $E_i \geq 5$ for all i.

Thus, since we never know the exact distribution from which we are sampling with 100% assurance, we may test our assumption of a particular distribution by constructing a histogram, postulating a distribution, calculating χ^2_ν and comparing to a critical value. If it is larger than the critical value at the 95% level, we say that the data contradicts the assumption of the distribution postulated.

3.9.2 Examples of the Goodness-of-Fit Test

We present here two examples, both dealing with the normal distribution, the first assuming a specific value for the mean and variance, the second using estimates of these parameters. To obtain E_i, the expected frequencies, we need to find the probability of an event $x_{i-1} < x < x_i$. To facilitate the computation we standardize the hypothesized normal distribution, $f(x) \sim N(\mu, \sigma^2)$, and evaluate the cumulative distribution as follows:

$$F(z_i) = F\left(\frac{x_i - \mu}{\sigma}\right),$$

$$F(z_i) = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

To get the frequencies in the intervals $z_{i-1} < z < z_i$, we take the difference

$$p_i = F(z_i) - F(z_{i-1}).$$

If the mean and variance are estimated, we replace μ by $\bar{x}$ and σ by s.

Example 3.19

Consider the delayed neturon count data on 100 fuel pellets given in Table 1.1. It has been hypothesized that the data follows a normal

distribution with a mean of 100 and a standard deviation of 9.0 counts per gram. (The interval bounds have been redefined slightly for convenience). The data is summarized in Table 3.3.

Table 3.3
Delayed Neutron Counts Per Gram: $\mu = 100, \sigma = 9.0$, n = 100

Interval Upper Bound $(x_{i-1} \le x < x_i)$	$z_i = \frac{x_i - \mu}{\sigma}$	$Pr(z_{i-1} < z < z_i)$	p_i	Expected $E_i = np_i$	Observed O_i	$\frac{(O_i - E_i)^2}{E_i}$
<85	-1.67	0.0475	0.0475	4.75	1	2.96
90	-1.11	0.1335	0.0860	8.60	2	5.07
95	-0.56	0.2877	0.1542	15.42	11	1.27
100	0	0.5000	0.2123	21.23	25	0.67
105	0.56	0.7123	0.2123	21.23	24	0.36
110	1.11	0.8665	0.1542	15.42	14	0.13
115	1.67	0.9525	0.0860	8.60	11	0.67
>115	>1.67	1.0	0.0475	4.75	12	11.07
				100	100	22.20

$$\chi^2 = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = 22.20 > \chi^2_{7,\ 0.005} = 20.278$$

The result of this test is to reject the assumption that the distribution of delayed neutron counts per gram is $N(100, 9^2)$ since the χ^2-test exceeds its critical value. However, note that there are three assumptions implicit in that null hypothesis; i.e., normality, $\mu = 100$, and $\sigma = 9$. In the next example, a test will be made on normality, but the estimates of the mean and variance will be used.

Example 3.20

Test the hypothesis that the underlying distribution can be considered to be normal, but use the estimated mean and variance from Section 1.7; $\bar{x} = 103.0$, $s = 8.7$. The calculations are summarized in Table 3.4. Note that the intervals <85 and $85 \le x_i < 90$ have been combined to conform with the rule-of-thumb that the expected value for an interval should be about 5 or more.

Table 3.4
Delayed Neutron Counts: $\bar{x} = 103.0$, $s = 8.7$, n = 100

Interval Upper Bound $(x_{i-1} \le x < x_i)$	$z_i = \frac{x_i - \bar{x}}{s}$	$Pr(z_{i-1} < z < z_i)$	p_i	Expected $E_i = np_i$	Observed O_i	$\frac{(O_i - E_i)^2}{E_i}$
<90	-1.49	0.0681	0.0681	6.81	3	2.13
95	-0.92	0.1788	0.1107	11.07	11	0.00
100	-0.34	0.3669	0.1881	18.81	25	2.04
105	0.23	0.5910	0.2241	22.41	24	0.11
110	0.80	0.7881	0.1971	19.71	14	1.65
115	1.38	0.9162	0.1281	12.81	11	0.26
120	1.95	0.9744	0.0582	5.82	7	0.24
>120	>1.95	1.0	0.0256	2.56	5	2.33
				100	100	8.76

Since the mean and variance were both estimated, 2 degrees of freedom must be subtracted for the χ^2 distribution. Thus, $\chi^2_{8-1-2} = \chi^2_5 = 8.76$. The 5% value for a Chi-Square with 5 degrees of freedom is 11.07. Thus, we would accept the hypothesis that the delayed neutron data is normally distributed with an estimated mean 103.0 and an estimated standard deviation of 8.7 counts.

In Chapter 4 examples will be given for testing the hypothesis that an underlying distribution for a population is non-normal.

3.9.3 Cumulative Probability Plots

Another technique for testing the distribution of a collection of data is to plot the data on specially constructed probability paper. Such paper is commercially available for many distributions, particularly the normal and the Weibull distributions, and can be readily constructed for any distribution. For a small number of data points (<20), the individual observations themselves can be plotted on a cumulative probability scale. For large data sets, the upper bound of an interval can be plotted on a cumulative probability scale. If the data follows the distribution hypothesized, a straight line can be reasonably drawn through the plotted points. If severe departures from the line exists, this is considered evidence that the hypothesized distribution is inadequate. This procedure is approximate in that no quantitative measure of "goodness" is available to determine if the distribution fits the data. Hence, whenever possible, an analytical test, such as the Chi-Square test discussed above, should be applied. However, the cumulative probability plot on special paper is an excellent visual aid, and with some practice, good judgment can quickly be developed.

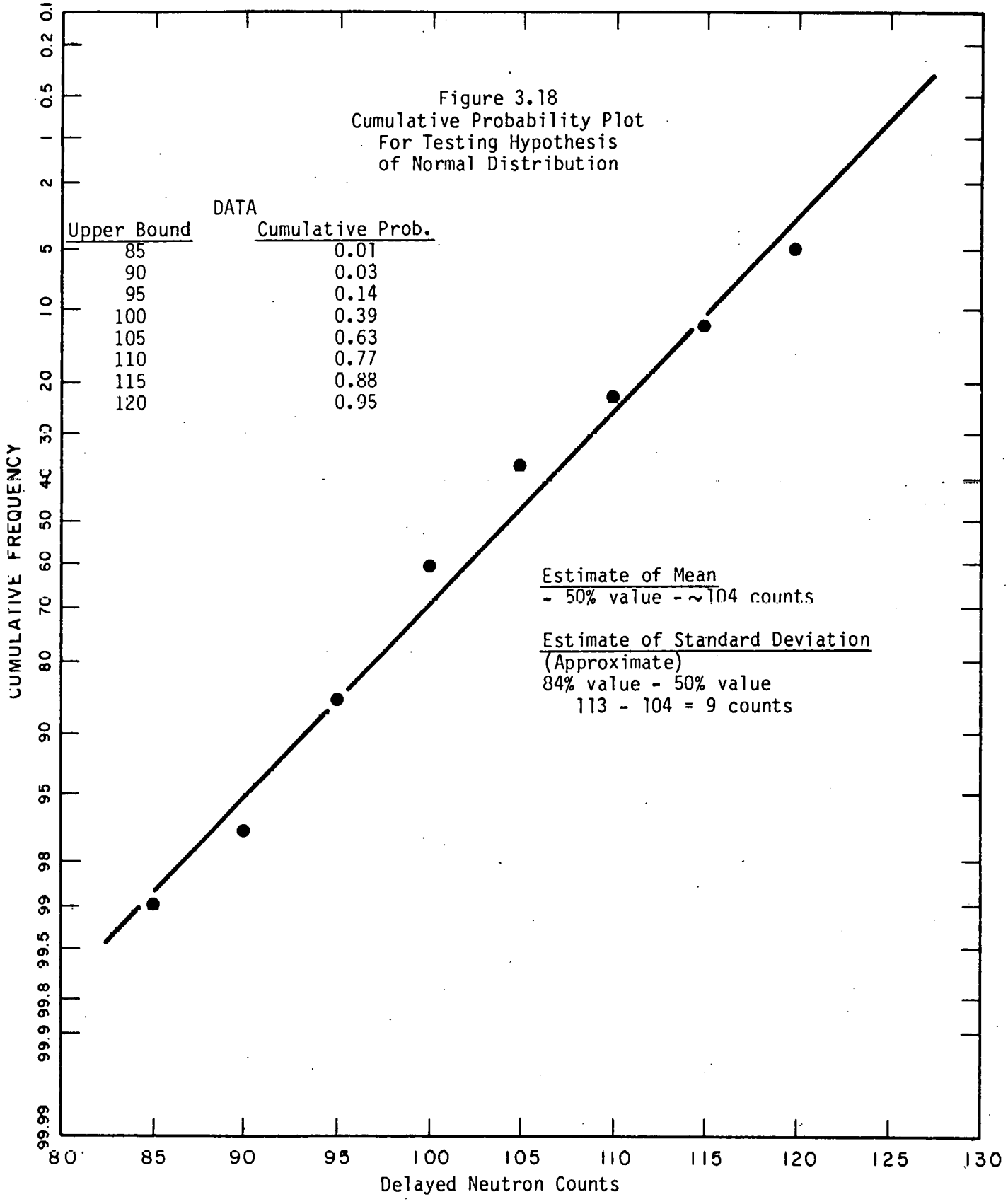
To illustrate this technique, consider again the delayed neutron data in Table 1.1. In Figure 3.18 the upper bounds of the intervals (taken here as in Examples 3.19 and 3.20 to be 85, 90, etc.) are plotted on the horizontal axis, and the observed cumulative frequencies are plotted on the vertical axis on a special probability scale. A hand-drawn line has been passed through the plotted points. With experience you will find that these points are well situated about the line so that the judgment here is that the distribution does appear to be normal. This, of course, agrees with the results from Example 3.20.

An additional feature is that because we know much about the normal distribution, once we have accepted it as the distribution, we can also obtain estimates of the mean and standard deviation from this plot. The 50% value, $X_{0.50}$ (i.e., the value of counts per gram that corresponds to the point on the drawn line that intersects the 0.50 probability value) is an estimate of the mean. Here is it found to be 104 counts per gram, compared to 103 used in Example 3.20. The standard deviation can be estimated by finding the value which is 1 σ from the mean. From Table III, we find ($Pr(z<1) = 0.8413$. Thus, from Figure 3.18, $X_{0.84} = 113$ and σ is estimated by $X_{0.84}-X_{0.50}=113-104 = 9$.

Figure 3.18
Cumulative Probability Plot
For Testing Hypothesis
of Normal Distribution

DATA

Upper Bound	Cumulative Prob.
85	0.01
90	0.03
95	0.14
100	0.39
105	0.63
110	0.77
115	0.88
120	0.95



Thus, a cumulative probability plot can be used to test the hypothesis of normality of a collection of data; and if judged acceptable rough estimates can be obtained for μ and σ of this distribution.

This compares well with the usual estimate of s = 8.7.

*3.10 Outliers

An outlier is an observation that is significantly different from the rest of the sample. An observation that comes from a sufficiently different distribution than the rest of the observations should appear as an outlier, but all observed outliers are not necessarily maverick observations. A true outlier may come from a distribution with a different mean or a different variance, or both, than the rest of the data, or even have a different functional form for the probability density function.

Usually true outliers result from mistakes rather than random errors; mistakes such as the transposition of digits, a misplaced decimal, or use of the wrong standard or measuring device or procedure. These types of true outliers are often easily detected and their omission from the data leads to more precise and accurate analyses. On the other hand, some observed outliers may be the most important information in the sample if it truly indicates the potential variability of the process from which the data comes.

For example, consider a corrosion test of an element in which the depth of corrosion is recorded for each of 200 sites. The occurrence of one observed outlier may be indicative of the difficulty of producing corrosion resistant elements, rather than indicating a bad observation caused by mistakes in data taking or transmission.

In the next section, three tests for detecting an outlier will be presented. Once having flagged an observation as an outlier, there is still the question of what to do with it. Section 3.10.2 briefly discusses some possibilities, but it should be noted here that the recommended procedure for handling outliers is to keep it, unless a reason can be established that justifies its dismissal or modification.

3.10.1 Tests for Outliers

Many tests for the detection of outliers exist, but generally require the assumption of normality for the distribution of the underlying population. The procedures given below assume a normal distribution and test for the existence of a single outlier. Hence, the tests are one sided hypothesis tests for the null hypothesis that the extreme observation x_e belongs to this population.

A. The r-test

Order the $n \leq 7$ observations such that $x_1 < x_2 < \ldots < x_n$. Then to test for the extreme value x_1, let

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1} \quad .$$

If r_{10} is greater than the critical value for r_{10}, α, found in Table XV, at the desired significance level α, then x_1 is declared to be an outlier.

To test for an extreme value that is the largest observation, use

$$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}, \quad n \leq 7 \quad .$$

Equivalent formulae for $8 \leq n \leq 10$, $11 \leq n \leq 13$, and $14 \leq n \leq 30$ are given in Table XV.

B. The T-Test: Standard Deviation Obtained from Same Sample

A test known as the T-test can be used in which the standard deviation of the underlying distribution is estimated from the available sample; i.e.,

$$s = \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / n-1 \right]^{1/2} \quad .$$

This test calculates the extreme studentized deviate as

$$T = \frac{|x_e - x|}{s}$$

If T exceeds the critical value, $T_{n,\alpha}$ given in Table XVI, at the desired significance level, the extreme observation is considered to be an outlier.

C. The T-Test: Standard Deviation Obtained From an Independent Sample

This test is the same as the T-test above except that here the standard deviation of the underlying distribution is estimated from an independent sample, such as might be obtained in a study to establish the precision of the process. The degrees of freedom of this estimate, $\hat{\sigma}$, are not dependent on the number of observations used to estimate the mean. Now if

$$T = \frac{x_e - \bar{x}}{\hat{\sigma}} > T_{n,\nu,\alpha} \quad \text{(from Table XVII)}$$

the observation is declared an outlier.

Example 3.21

Three observations on the counts per gram for a single fuel pellet are obtained by a delayed neutron gage:

x_i = 24610.4, 24623.0, 24933.2

$\bar{x}$ = 24722.2, s = 182.84, n = 3

Test A:

$$r_{10} = \frac{x_3 - x_2}{x_3 - x_1} = \frac{24933.2 - 24623.0}{24933.2 - 24610.4} = 0.961$$

This is less than the $\alpha = 0.01$ value for n = 3 of 0.988 obtained from Table XV. Thus, the r-test does not declare the observation 24933.2 to be an outlier at the 1% level of significance.

Test B:

$$T = \frac{\bar{x}_e - x}{s} = \frac{24933.2 - 24722.2}{182.84} = \frac{211.0}{182.84} = 1.15$$

We see from Table XVI that this value is right at the critical value at the α =0.05, 0.025 and 0.01 levels. This indicates that this test using s is not sufficiently sensitive to detect an outlier from this small group of data.

Test C:

It is known from 3 observations on each of 20 other pellets that the standard deviation can be estimated as

$\hat{\sigma}$ = 79.03 with 40 degrees of freedom.

Then

$$T = \frac{x_e - \bar{x}}{\hat{\sigma}} = \frac{24933.2 - 24722.2}{79.03} = 2.67$$

The critical value from Table XVII for α = 0.01 is 2.34. Thus, with more information supplied on σ from an independent sample, we declare the observation 24933.2 in this sample of 3 to be an outlier.

The fact that each of the three tests yields different results at the 1% significance level illustrates the difference in the sensitivity of the tests for detecting outliers due to the amount of information contained in the data. The r-test, which simply uses relative distances between specified

values, contains the least information of the three tests and fails to detect an outlier. The T-test using the estimate of the standard deviation from the same sample incorporates more information with the result that the extreme value is now found to be on the border line between being called an outlier and not being called an outlier. The T-test using an independent sample to estimate the standard deviation utilizes information in addition to the sample being examined. Because of this additional information it becomes a more sensitive test and is thus able to detect an outlier. Which test should be applied in a given situation depends upon the amount of information available.

The above tests are for the detection of a single outlier. To detect more than one outlier, the above tests can be performed sequentially by throwing out the first detected outlier and testing the remaining sample as the original sample was tested. Unless the sample is quite large, however, it is unlikely that more than one true outlier can be detected.

3.10.2 The Treatment of an Outlier

Having detected an outlier, what do we do with it? We don't want to ignore it completely if it carries important information about the process. On the other hand, we are penalizing ourselves with an erroneous mean and an inflated standard deviation if we keep a completely spurious observation.

Three suggestions are considered below:

A. The Anscombe Rule:

Delete a detected outlier. This is a severe step to take. To reduce the risk of erroneously throwing out valuable information, it is suggested that the significance level to use in the test for outliers should be 0.01 or less.

B. The Winsorization (W) Rule:

Replace the detected outlier by its nearest neighbor.

C. The Semi-Winsorization (S) Rule:

Replace the detected outlier by the critical test value,

$$\text{new } x_e = \bar{x} \pm T_\alpha \hat{\sigma}$$

where $\bar{x}$ is the original average,

T_α is the σ-level value for either T-test, and

$\hat{\sigma}$ is the estimated standard deviation for the test used.

Rules W and S are attempts to protect against falsely declaring an extreme value an outlier. Thus, an σ value of 0.05 or 0.01 may be satisfactory. Many alternatives are possible and the best procedure for a specific case may change from case to case.

CHAPTER 4
INFERENCES ON NON-NORMAL POPULATIONS

Thus far we have assumed that the distribution of the sampled observations is a normal distribution, or at least that the distribution of an average of n observations is normal. In this chapter we will briefly examine the problem of estimation and inference for parameters of some commonly occurring non-normal distribution. After a review of the maximum likelihood principle of estimation we will deal specifically with the binomial, Poisson, and exponential distributions.

4.1 Review of Maximum Likelihood Criteria

To obtain the most likely value of a parameter given a set of data, we maximize the joint distribution function $f(x_1, x_2, \ldots, x_n|\theta)$. For n given observations randomly taken, we write the same function as a likelihood function

$$L(\underline{\theta} \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\underline{\theta}),$$

where $\underline{\theta}$ represents a set of one or more parameters. If the range of x_i does not depend on $\underline{\theta}$, then we make maximize $L(\underline{\theta} \mid \underline{x})$ by solving the derivative of the likelihood function for θ_j,

$$\frac{d}{d\theta_j} L(\underline{\theta} \mid \underline{x}) = 0$$

where $\underline{x}$ represents the set of n x's. If there are more than one θ_j, we must solve the resulting differential equations simultaneously. The reader is referred again to Appendix B. for more details.

4.2 Inference on a Binomial Distribution

The distribution known as a binomial has the following form

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

where p may be (1) the probability of success; i.e., a selection of an item which meets some specifications, or (2) the proportion of defectives in a lot, or (3) the percentage of certain components that meet some specification.

Suppose we sample from a binomial population with parameters p and n a total of k times. Thus, $x_1, x_2, \ldots, x_k$ are a random sample from a binomial (n, p). The maximum likelihood estimate of p is obtained as follows:

$$L(p \mid x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}, \quad x_i = 0, 1, 2, \ldots, n,$$

$$= \left[\prod_{i=1}^{k} \binom{n}{x_i}\right] p^{\Sigma x_i} (1-p)^{nk - \Sigma x_i}$$

$$\ell nL(p) = \sum_{i=1}^{k} x_i \ell n\, p + (nk - \Sigma x_i)\, \ell n\, (1-p) + \sum_{i=1}^{k} \ell n \binom{n}{x_i}$$

$$\frac{d\, \ell nL\ (p)}{dp} = \frac{\Sigma x_i}{p} + \frac{(nk - \Sigma x_i)}{1-p}\ (-1) = 0$$

$$\hat{p}_{ML} = \sum_{i=1}^{k} x_i / nk$$

Thus the estimate of p is the average $\bar{x} = \sum_{i=1}^{k} x_i/k$ divided by n, the sample size. Since $E(x_i) = np$, the expected value and the variance of $\hat{p}$ are

$$E(\hat{p}) = E\, \frac{1}{nk} \sum_{i=1}^{k} x_i = \frac{k\ np}{nk} = p$$

$$Var(\hat{p}) = \frac{1}{(nk)^2} \Sigma Var(x_i) = \frac{k}{(nk)^2} [np(1-p)] = \frac{p(1-p)}{nk}\ .$$

We can quickly see, however, that sampling k times from a binomial population (n, p) is equivalent to sampling from a <u>single</u> binomial population with parameters (N, p), where N = nk. Then,

$$\hat{p} = \frac{\sum_{i=1}^{N} x_i}{N},\ E\ (\hat{p}) = p,\ Var\ (\hat{p}) = \frac{Np(1\ -\ p)}{N^2} = \frac{p(1\ -\ p)}{N}\ .$$

To obtain a confidence interval for p, we proceed as follows:

Find p_1 and p_2 such that

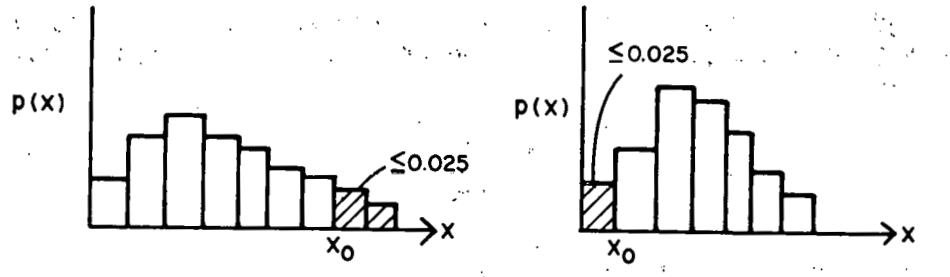
1. p_1 is the largest value of p for which

$$Pr(x \geq x_0) = \sum_{x=x_0}^{n} \binom{n}{x} p^x\ (1\ -\ p)^{n\ -\ x} \leq \alpha/2$$

2. p_2 is the smallest value of p for which

$$Pr(x \leq x_0) - \sum_{x=0}^{x_0} \binom{n}{x} p^x\ (1\ -\ p)^{n\ -\ x} \leq \alpha/2.$$

Then, $p_1 < p < p_2$ is a $100(1 - \alpha)\%$ confidence interval for p, where x_0 is an observed value. This choice of p_1 and p_2 gives the shortest possible interval.



$$\sum_{x=x_0}^{n} \binom{n}{x} p_1^x (1-p_1)^{n-x} \leq 0.025 \qquad \sum_{x=0}^{x_0} \binom{n}{x} p_2^x (1-p_2)^{n-x} \leq 0.025$$

Figure 4.1 Tails of a Binomial Distribution $p_1 < p < p_2$

Note the inequality signs. This is due to the discrete nature of the binomial variable. For a given probability level, there may not be a discrete x value to correspond to it. Conventionally, we find the value of x which gives $\leq \alpha/2$ in probability in the tail regions.

Example 4.1

Suppose n = 20 and x_0 = 2 defective pieces were found. Construct a 95% confidence interval for p. We need to solve

1. $\sum_{x=2}^{20} \binom{20}{x} p_1^x (1-p_1)^{20-x} \leq \alpha/2 = 0.025$

and 2. $\sum_{x=0}^{2} \binom{20}{x} p_2^x (1-p_2)^{20-x} \leq \alpha/2 = 0.025$

Fortunately, tables and charts are available for constructing intervals for proportions. Table XI gives curves for $\gamma = 1-\alpha = 0.95$ and 0.99 Reading these curves indicate that with 95% confidence $p_1 = 0.01$, $p_2 = 0.32$. Using a

table of binomial probability values (Table I) and interpolating we can verify these values. Thus, we can say with 95% confidence or better that p is contained within $0.01 < p < 0\ 32$, where $n = 20$, $x_0 = 2$, and $\hat{p} = 0.1$.

4.2.1 A Normal Approximation for the Binomial

If n is sufficiently large and p is not extremely small or extremely large, an approximating confidence interval for a porportion may be obtained from a normal approximation. The procedure is to standardize the binomial variable

$$z = \frac{x - np}{\sqrt{np(1-p)}} .$$

Dividing through both sides by n, we have $\hat{p} = x/n$,

$$z = \frac{x/n - p}{\sqrt{\frac{p\ (1-p)}{n}}} .$$

The two-sided interval would be

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{1}{n}p(1-p)} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{1}{n}p(1-p)} .$$

However, this contains p. Replace p by its estimate $\hat{p} = x/n$ and we obtain

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{1}{n}\frac{x}{n}\left(1-\frac{x}{n}\right)} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{1}{n}\frac{x}{n}\left(1-\frac{x}{n}\right)}$$

For $\hat{p}$ very small or very large, this approximate interval could result in values less than 0 or larger than 1.

Example 4.2

For $n = 20$, $x_0 = 2$, $\hat{p} = 0.1$ as before, a 95% confidence interval for p using a normal approximation is obtained as follows:

$$\frac{x}{n} \pm z_{0.025}\sqrt{\frac{1}{20}\ (0.1)(0\ 9)}$$

then $$0.1 \pm 1.96\ (0.067)$$

and the interval is (-0.03, 0.23) (compared to (0.01, 0.32) found previously).

Although -0.03 is impossible, using 0 for a lower bound is not a bad approximation. However, 0.23 is far from 0.32. The normal approximation can be improved somewhat by adding a correction factor of $\pm \frac{1}{2}$ to x:

$$\frac{x \pm \frac{1}{2}}{n} \pm z_{.025}\sqrt{\frac{1}{n}p(1-p)}$$

$$0.1 \pm .025 \pm 0.13$$

$$(-0.055 < p < 0.255)$$

This is better in that it gives a wider interval, although here the negative value is inappropriate. The problem with the intervals suggested here is the sample size. For a confidence interval based on a normal distribution to be adequate, the distribution of x/n must approach a normal distribution. For this, n must be sufficiently large. Miller and Freund [25] suggest $n > 100$. Also note that $\frac{x/n - p}{\sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}}$ is <u>not</u> a t-distribution since $\frac{1}{n}p(1-p)$ is <u>not</u> an estimate of $Var(\hat{p})$ that is statistically <u>independent</u> of $\hat{p}$, as $\bar{x}$ and s^2 are statistically independent.

4.2.2 Comparing Two Proportions

Consider comparing the proportion of defectives from two lots of material. Sample n_1 observations from one population and n_2 from the other. To test

$$H_o: \ p_1 = p_2 \text{ or } p_1 - p_2 = 0$$

against $H_A: \ p_1 \neq p_2$,

we consider the test statistic

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

where n_1 and n_2 are both large and under the null hypothesis, $p_1 = p_2 = p$,

$$Var\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) = \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2} = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

The combined estimate of p is $\hat{p} = \frac{\Sigma x_i}{\Sigma n_i} = \frac{x_1 + x_2}{n_1 + n_2}$.

Example 4.3

Given $n_1 = 400$, $x_1 = 128$, $n_2 = 500$, $x_2 = 115$. Then

$$\hat{p} = \frac{128 + 115}{900} = 0.27$$

and

$$z = \frac{\frac{128}{400} - \frac{115}{500}}{\sqrt{(0.27)\ (0.73)\left(\frac{1}{400} + \frac{1}{500}\right)}} = 3.02 > z_{0.025} = 1.96.$$

Therefore we reject $p_1 = p_2$ based on this data.

4.2.3 Comparing k Proportions

Consider the problem of comparing several proportions for differences. The null hypothesis is

$$H_o: \ p_1 = p_2 = \ldots = p_k$$

and the alternative hypothesis is that at least two of these proportions are unequal. We again make use of the normal approximation when n_i, $i = 1, 2, \ldots, k$ are large:

$$z_i = \frac{x_i - n_i p_i}{\sqrt{n_i p_i (1-p_i)}}$$

where x_i is the number of defectives found in a sample of size n_i, and p_i is the proportion defective. Since each population from which we sample provides an independent normal deviate z_i, we can define χ^2_k distribution

$$\chi^2_k = \sum_{i=1}^{k} \frac{(x_i - n_i p_i)^2}{n_i p_i (1-p_i)} \quad . \qquad (4.1)$$

Under the null hypothesis, the pooled estimate of the common proportion p is

$$\hat{p} = \frac{\Sigma x_i}{\Sigma n_i} \quad .$$

Substituting into (4.1) for each p we have

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i \hat{p})^2}{n_i \hat{p}(1-\hat{p})} \quad . \qquad (4.2)$$

We may then test H_0 by comparing the value obtained above with the critical value (right-hand tail) of $\chi^2_{k-1,\alpha}$. We have k-1 degrees of freedom since we had to estimate the common value p.

Example 4.4

It is desired to test whether the proportion of supplies brought back and exchanged by a certain vendor is subject to seasonal variations. The quarterly data is

	First Quarter	Second Quarter	Third Quarter	Fourth Quarter	Total
# exchanged	29	12	8	21	70
# not exchanged	81	118	92	139	430
TOTAL	110	130	100	160	500

Under H_0: $p_1=p_2=p_3=p_4=p$, $n\hat{p} = 70, \hat{p} = 70/500 = 0.14$.

$$\chi^2_3 = \sum_{i=1}^{4} \frac{(x_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})} = \frac{(29 - 110(0.14))^2}{110(0.14)(0.86)} + \frac{(12 - 130(0.14))^2}{130(0.14)(0.86)}$$

$$+ \frac{(8 - 100(0.14))^2}{100(0.14)\ (0.86)} + \frac{(21 - 160(0.14))^2}{(160(0.14)(0.86)}$$

$$= \frac{184.96}{13.244} + \frac{38.44}{15.652} + \frac{36}{12.04} + \frac{1.96}{19.264}$$

$$= 13.97 + 2.46 + 2.99 + 0.10$$

$$= 19.52 \quad .$$

This exceeds $\chi^2_{3,.01} = 11.345$, so we conclude that there is evidence of seasonal differences.

The Chi-Square test (4.2) can be viewed in a somewhat different manner. Looking at the data as shown in the example above, we can identify 8 "cells", where each entry is the observed frequency for that cell O_{ij}, $i = 1, 2, 3, 4$; $j = 1, 2$. We then view the problem as a goodness of fit test for a multinomial distribution with 8 classifications of cells. For Example 4.4.,

$$\chi^2_3 = \sum_{j}^{2} \sum_{i}^{4} \frac{(O_i - E_i)^2}{E_i} = 19.5 \qquad (4.3)$$

where E_i are the expected frequencies $n_i p$. The distribution is still a χ^2_3 because knowledge of any three of the proportions will enable us to determine all of them. The advantage to this approach is that it will allow generalization to problems of r x k layouts where r > 2.

4.3 Inference on a Poisson Distribution

In situations in which it is of interest to know how many events of a certain kind occur in a given period of time, we know that a Poisson distribution usually applies. Recall that for a Poisson variable x,

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

where λ is the mean rate of occurrence for the unit of time being used. The maximum likelihood estimate of λ is $\bar{x}$ (see Appendix B.) based on n observations.

As for the binomial parameter p, an interval for which we have at least 95% confidence of containing λ based on a single observation x can be obtained by solving for λ_1 and λ_2:

$$1. \quad Pr(x \geq x_0 | \lambda_1) = \sum_{x=x_0}^{\infty} e^{-\lambda} \lambda^{x_1}/x! \leq 0.025$$

and

$$2. \quad Pr(x \leq x_0 | \lambda_2) = \sum_{x=0}^{x_0} e^{-\lambda} \lambda^{x_2}/x! \leq 0.025$$

Example 4.5

Suppose 2 accidents occurred in a 12-hour period along a particular production line. The confidence interval for the mean rate of accidents is such that

$$1. \quad Pr(x \geq 2 | \lambda_1) \leq .025$$

$$2. \quad Pr(x \leq 2 | \lambda_2) \leq .025 \quad .$$

From Table II, we see that the largest λ, satisfying 1. is 0.24, and the smallest λ_2 satisfying 2. is 7.2. Thus a 95% confidence interval for λ based on $x_0 = 2$ is $0.24 < \lambda < 7.2$.

This is verified in Table XII, a table of confidence limits for a Poisson parameter.

4.3.1 k > 1 Observations from a Poisson

For more than one observation from the same Poisson distribution, it can be shown that the sum of Poisson variables is again a Poisson with a parameter $n\lambda$ instead of λ.

$$p\left(\sum_{i=1}^{n} x_i\right) = e^{-n\lambda} (n\lambda)^{\Sigma x_i} / (\Sigma x_i)! \text{ or } e^{-\theta} \theta^X / X!$$

where $\theta = n\lambda$, $X = \sum_{i=1}^{n} x_i$

The maximum likelihood estimate of λ is Σx_i, so the maximum likelihood estimate of λ is

$$\hat{\lambda}_{ML} = \frac{1}{n}\Sigma x_i \qquad \text{as found previously.}$$

To find a confidence interval for λ then

1. $Pr(X \geq X_0 | \theta_1) = Pr(\Sigma x_i \geq X_0 |\ n, \lambda_1) \leq 0.025, \lambda_1 = \theta_1/n,$
2. $Pr(X \leq X_0 | \theta_2) = Pr(\Sigma x_i \leq X_0 |\ n, \lambda_2) \leq 0.025, \lambda_2 = \theta_2/n,$

$\theta_1 < \theta < \theta_2$

Example 4.6

For one observation $X_0 = 2$, a 95% confidence interval for θ is

$$0.24 < \theta < 7.2$$

Thus, for n observations, the 95% confidence interval for $\lambda = \frac{\theta}{n}$ is

$$\frac{0.24}{n} < \lambda < \frac{7.2}{n} .$$

For $n = 4$, $\hat{\lambda} = \frac{\Sigma x_i}{n} = \frac{2}{4} = 0.5$

$$\left(\frac{0.24}{4}, \frac{7.2}{4}\right) = (0.06, 1.8)$$

is a 95% confidence interval for λ when $n = 4$ observations were taken, each over an interval length of 12 hours, $\Sigma x_i = 2$ accidents being recorded.

4.3.2 Normal Approximation for a Poisson

An approximate interval can be obtained for large enough n, or equivalently if λ is quite large. The standardized Poisson variable is

$$z = \frac{x - E(x)}{\sqrt{Var(x)}} = \frac{x - \lambda}{\sqrt{\lambda}} \quad \text{for a single observation on } x,$$

and

$$z = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}}, \quad \hat{\lambda} = \bar{x}, \text{ for } n > 1 \text{ observations.}$$

Using the estimate of $\bar{x}$ for λ as its estimated variance, a two-sided 95% confidence interval is

$$x - 1.96\sqrt{x} < \lambda < x + 1.96\sqrt{x}, \quad \text{for } n = 1,$$

$$\bar{x} - 1.96\sqrt{\bar{x}/n} < \lambda < \bar{x} + 1.96\sqrt{\bar{x}/n}, \quad \text{for } n > 1.$$

Example 4.7

If 50 accidents were recorded over a 120-hour period, then from Table XII, a 95% confidence interval for the mean rate θ of occurrences over a 120-hour period is (37.0, 65.9). Using a normal approximation, the interval is

$$\begin{aligned} [\ 50 \pm 1.96\sqrt{50}\ ] &= 50 \pm 1.96\ (7.07) \\ &= 50 \pm 13.9 \\ &= (36.1, 63.9)\ . \end{aligned}$$

The agreement between the exact and normal approximation will, of course, improve as λ gets larger.

To obtain a confidence interval for the mean number λ of occurrences per 12-hour period, we divide the 120 hours into ten 12-hour segments. Then $\hat{\lambda} = \bar{x} = 5.0$ and the exact interval is (3.7, 6.6) and for the normal approximation

$$5.0 \pm 1.96\sqrt{5.0/10}$$

$$5.0 \pm 1.96\ \ (0.707)$$

$$5.0 \pm 1.39$$

$$(3.6, 6.4)$$

4.4 Inference on an Exponential Distribution

The lifetime of many electrical and mechanical components and systems tend to follow an exponential or related distribution. Let x be the lifetime or time-to-failure, then

$$f(x) = \frac{1}{\lambda}\, e^{-x/\lambda}\ , \ x > 0$$

is an exponential distribution with λ the mean time to failure. The variance of an exponential can be shown to be λ^2.

If $x_1, x_2 \ldots, x_n$ are n independent observations from f(x), the maximum likelihood estimate of λ is $\bar{x}$. (Appendix B).

4.4.1 An Exact Confidence Interval for λ, n = 1

For a single observation, we could obtain a confidence interval for λ as we have done for the binomial and Poisson distributions, by

1. finding largest λ_1 such that $Pr(x > x_0|\lambda_1) = 0.025$ and
2. find the smallest λ_2 such that $Pr(x < x_0|\lambda_2) = 0.025$.

This will yield an <u>exact</u> 95% confidence interval $\lambda_1 < \lambda < \lambda_2$ since this is a continuous distribution.

4.4.2 A Normal Approximation for the Exponential

An approximate 95% confidence interval can be obtained by a normal approximation if n, the number of observations, is sufficiently large. Like a Poisson distribution with small λ, an exponential is a very asymmetric distribution, and hence a large sample size is required. Since the variance of x is λ^2, the standardized exponential variable is

$$z = \frac{x - \lambda}{\lambda}$$

and for n observations, $\hat{\lambda} = \bar{x}$,

$$z = \frac{\bar{x} - \lambda}{\lambda/\sqrt{n}} .$$

Replacing λ by its estimate $\bar{x}$ in the denominator of z, we have a 95% confidence interval for λ, the mean time-to-failure,

$$\bar{x} - 1.96\frac{\bar{x}}{\sqrt{n}} < \lambda < \bar{x} + 1.96\frac{\bar{x}}{\sqrt{n}}$$

$$\bar{x}\left(1 - \frac{1.96}{\sqrt{n}}\right) < \lambda < \bar{x}\left(1 + \frac{1.96}{\sqrt{n}}\right) .$$

4.4.3 An Exact Interval, n > 1

An exact two-sided confidence interval for n observations can be constructed by using the fact that twice the sum of n exponential variables is distributed as a χ^2 variable with 2n degrees of freedom, scaled by λ; i.e., $2\Sigma x_i \sim \chi^2_{2n}$. Thus

$$Pr\left(\chi^2_{2n,0.975} < \chi^2_{2n} = \frac{2\Sigma x_i}{\lambda} < \chi^2_{2n,0.025}\right) = 0.95$$

where $Pr(\chi^2_{2n} < \chi^2_{2n,0.975}) = 0.025$ and $Pr(\chi^2_{2n} > \chi^2_{2n,0.025}) = 0.025$.

Thus,

$$\chi^2_{2n,0.025} < \frac{2\Sigma x_i}{\lambda} < \chi^2_{2n,0.975}$$

is an exact 95% confidence interval for λ based on n observations from an exponential distribution.

Example 4.8

Ten observations were made on the lifetime (in minutes) of batteries with the sum of failure times being 54 minutes.

$$\Sigma x_i = 54.0, \quad \chi^2_{20,0.975} = 9.591, \quad \chi^2_{20,0.025} = 34.17 \quad .$$

Then,

$$\frac{2(54.0)}{34.17} < \lambda < \frac{2(54.0)}{9.591}$$

Thus,

$317 < \lambda < 1128$ is a 95% confidence interval for the mean lifetime of batteries.

*4.5 Distribution-Free Tolerance Intervals

In Chapter 3 we discussed tolerance intervals assuming the population being sampled to be normally distributed.

If the data being sampled can neither be assumed normally distributed nor transformed by some simple mathematical function into a normal distribution, one- or two-sided tolerance intervals which are completely free of any distributional assumptions may be made. A one-sided 95/99 distribution-free tolerance interval says that with 95% confidence, 99% of the population being sampled will lie above the minimum observed value in the sample (or below the maximum value). A two-sided 95/99 distribution-free tolerance interval contains 99% of the population between the minimum and maximum sample values with 95% confidence. Intervals found by sample values other than the minimum or maximum are also permissible, but we will restrict our discussion to these limits since they are typical in many examples. For further information on this topic, see Natrella [26].

There are three factors which determine distribution-free tolerance intervals: confidence level, proportion of population, and sample size. Given any two of these factors, the third may be found in one of the Tables XIII(a) through XIII(f), or from Figures XIII(c) or XIII(g). Particularly useful is the capability of determining the sample size required to make a tolerance interval statement of specified confidence level and population proportion. The examples below illustrate these three situations.

Example 4.9

Determination of Sample Size: Proportion and Confidence Level Given

If we desired to obtain an interval which contains at least 90% of the population of fissile loadings of fuel pellets between the minimum and maximum observed values of the sample taken with a confidence level of 0.99, we see from Table XIII(a) that 64 observations are required. A one-sided 99/90 tolerance interval may also be constructed for which 90% of the population will be above the minimum (or below the maximum) value. From Table XIII(e) we see that 44 pellets would be required.

Example 4.10

Determination of Proportion: Confidence Level and Sample Size Given

The maximum fissile loading for a sample of 10 fuel pellets was reported to be 2.630 grams U^{235} and the minimum observation was 2.604 grams U^{235}. From Table XIII(b), we find that for 10 observations a two-sided tolerance interval at a 95% confidence level will contain 61% of the population of fissile loadings between the values of 2.605 and 2.630 grams U^{235} for that type of pellet. Alternatively, from Table XIII(f) we see that 74% of the loading values will be above 2.605 (or below 2.630) with 95% confidence.

Example 4.11

Determination of Confidence Level (Two-Sided): Proportion and Sample Size Given

Of 12 readings taken of U^{235} impurity concentration in zirconium, the maximum observed value was 0.0099 ppm. and the minimum was 0.0088 ppm. From Table XIII(d), we see that 75% of the population will lie between 0.0088 and 0.0099 ppm. with 84% confidence, and 90% of the population will fall within the interval with only 34% confidence.

Example 4.12

Interpolation of Proportion or Sample Size: Confidence Level Given, From Figures XIII (c) and XIII(g)

Alternative to the tables, Figures XIII(c) and XIII(g) may be used to find the proportion of the population in an interval for a given sample size, or for finding the sample size for a given proportion for confidence levels of 0.90, 0.95, and 0.99. The figures allow interpolation of results which do not

appear in the tables. For example, from Figure XIII(c) we see that for 95% confidence, the sample size required for a two-sided interval to contain 96% of the population, is approximately 120 observations. For 100 observations, approximately 93% of the population will fall in the interval with 99% confidence. From Figure XIII(g) we see that 14 observations are required for a 90/85 one-sided distribution-free tolerance interval and for 75 observations and 95% confidence, the one-sided interval will contain better than 96% of the population.

*4.6 An Approximate One-Sided Tolerance Interval for an Exponential Distribution

When the population being sampled is not normal, we cannot, of course, use the results of Section 3.7.2. In fact, exact tolerance intervals for distributions other than the normal are not available. In general, the best that can be obtained are the distribution-free or non-parametric tolerance intervals. These tolerance intervals are valid for any type of distribution but do not use any information which may be available about the distribution of the population being sampled. In general, because of the broad application of this approach, the length of the interval for given n, P, and γ will be wider than would be the case if the distribution were taken into account.

In the case of an exponential distribution, however, an approximate one-sided tolerance interval has been developed. Suppose we would like to assert with a degree of confidence $\gamma = 1 - \alpha$ that at least 100 P% of the components being sampled have a lifetime greater than χ^*. The following approximation has been developed:

$$\chi^* = -2 \sum_{i=1}^{n} x_i \, (\ln P) \, / \chi^2_{2n,\alpha}$$

where x_i are the recorded lifetimes of the sampled components and

$$\Pr\left(\chi^2_{2n} > \chi^2_{2n,\alpha} \right) = \alpha \quad .$$

Coincidentally,

$$2 \Sigma x_i / \chi^2_{2n,\alpha}$$

is the expression used for constructing a confidence interval for the mean lifetime λ, but here it is modified by $\ln P$.

Example 4.13

Suppose 10 observations from a component lifetime study give a total of lifetimes of

$$\Sigma x_i = 5410 \quad , \quad n = 10 \quad .$$

Let $\gamma = 0.95$, $\alpha = 0.05$, and $P = 0.95$. Then $\ell n\ P = \ell n(0.95) = -0.051$

and $\chi^2_{20,\ 0.05} = 31.41$

Thus,

$$\chi^* = -\frac{2(5410)\ (-0.051)}{31.41} = \frac{551.82}{31.41} = 17.6 \quad .$$

That is, with 95% confidence, 95% of the components under investigation can be expected to last longer than 17.6 hours before falling. This is a lower 95/95 tolerance limit.

To obtain an upper tolerance limit, the values P and α must be replaced by $1 - P$ and $\gamma = 1 - \alpha$. Consider the following example:

Example 4.14

34 observations are obtained on wear depth on fuel rods on the point of spring contact. A 95/99 upper tolerance limit for wear depth is obtained assuming an exponential distribution and

$$\sum_{i=1}^{34} x_i = 16 \text{ mils},\ n = 34,\ 1 - P = 0.01,\ \gamma = 0.95$$

$$\ell n\ (0.01) = -4.605,\ \chi^2_{68,0.95} = 50,\ \Pr(\chi^2_{68} > \chi^2_{68,0.95}) = 0.95$$

$$\chi^* = \frac{-2(16)(-4.605)}{50.0} = 2.95 \text{ mils.}$$

That is, with 95% confidence, at least 99% of all rods will have wear of less than 2.95 mils in depth at the point of spring contact.

(Note: using a normal approximation for $\chi^2_{68,0.05}$ gives

$$\chi^2_{68.0.05} = 68 - 1.645\ \sqrt{2 \times 68} = 48.8 \text{ and } \chi^* = 3.02 \text{ mils})$$

4.7 The Chi-Square Goodness-of-Fit Test

In Section 3.9.1 the Chi-Square goodness-of-fit test was discussed and illustrated for testing a normal distribution hypothesis. In this section two examples are given in which the distribution is not hypothesized to be normal. It should be recognized that for any given set of data, more than one distribution may be found to be reasonable. Acceptance of an hypothesis does not mean that the data definitely is that distribution, but only that there is no reason to discount that distribution and it may be used.

Example 4.15

It is desired to determine if the number of defectives found in lots of 500 items can be described by a Poisson distribution with mean = 2. A total of 50 lots were checked with the results given below:

H_0: x Poisson ($\lambda = 2$), n = 50

# defective found, x_i	$Pr(x \le x_i \mid \lambda) = 2$	p_i	$E_i = np_i$	O_i	$(O_i - E_i)^2/E_i$
0	0.135	0.135	6.75	10	1.56
1	0.406	0.271	13.55	15	0.16
2	0.677	0.271	13.55	12	0.18
3	0.857	0.180	9.00	11	0.44
>3	1.000	0.143	7.15	2	3.71
			50.00	50	6.05

With 5 intervals used the observed test value $\chi^2_4 = 6.05$ is found to be less than $\chi^2_{4,0.05} = 9.488$. Thus, a Poisson with $\lambda = 2$ is a reasonable distribution for describing the number of defectives per lot.

Example 4.16

Consider the wear depth data from Example 4.14. To test the hypothesis that an exponential distribution describes this data, consider estimating the mean value λ by $\hat{\lambda} = \bar{x} = 16/34 = 0.471$.

H_0: $f(x) \sim$ exponential (mean $\hat{\lambda}$)

Upper bound on interval $(x_{i-1} < x < x_i)$	$Pr(x < x_i) = 1 - e^{-x_i/\hat{\lambda}}$	$\hat{p}_i$	$E_i = n\hat{p}_i$	O_i	$(O_i - E_i)^2/E_i$
0.25	0.412	0.412	14.0	15	0.07
0.50	0.654	0.242	8.2	7	0.18
0.75	0.797	0.143	4.9	3	0.74
>0.75	1.000	0.203	6.9	9	0.64
			34.0	34	1.63

$\chi^2_{4-1-1=2} = 1.63$ [$\chi^2_{2,0.95} = 0.103$, $\chi^2_{2,0.05} = 5.99$]

An exponential distribution is accepted.

CHAPTER 5
QUALITY CONTROL STATISTICS

Whether producing or receiving manufactured goods, care must be taken to assure that the product meets specified requirements with as little risk and at the lowest cost possible to the producer and/or consumer. This chapter presents some of the techniques utilized to assure all involved of the quality of the product in question.

5.1 Sampling Inspection

Consider the situation in which a manufacturer of a pressurized water reactor is faced with the decision to either accept or reject a large supply or lot of fuel elements. The manufacturer certainly does not want to purchase material from the vendor which will lead to an inferior or defective reactor. Thus, the manufacturer will subject the supply of fuel elements to a test of its quality based on the known manufacturing specification requirements.

The first question to arise is, "What kind of a test?" Should the fuel content be measured precisely, or simply judged to be within or outside the specifications. The former is an example of variable sampling and the latter is an example of attribute sampling Variable sampling deals with any continuous measurement, such as height, weight, length, weight percent, etc., and will be discussed in Sections 5.5--5.7 under the general heading of Control Charts. Attribute sampling deals with discrete data: for example, counts or number of defects, or more usually, go/no-go, and accept/reject data.

Having decided to perform attribute sampling, the sampling inspection procedure must then specify how many items should be examined. Ideally, every item should be inspected for all possible attributes from all possible angles. Practically speaking however, many of the tests required to evaluate attributes are destructive in nature, thus necessitating a sample to be taken. Even in non-destructive testing the cost in both time and money may favor a sampling procedure also

The question of how to sample, i.e., the choice of a sampling plan, is taken up in the next section and is followed by a discussion of the evaluation of sampling plans. For a detailed discussion and compilation of sampling plans, see MIL-STD-105-D [20] and the Dodge-Romig Sampling Tables [9].

5.2 Types of Attribute Sampling Plans

The purpose of a sampling plan is to provide the examiner with a basis for the acceptance or rejection of a lot of material based on a sample of n individual items from the lot. The basis of acceptance is the probability of obtaining a specified number of defectives (i.e., items outside the specification limits) in the sample. In the simplest terms, if the number of defectives, in the sample is too large, the entire lot is rejected. If the number of defectives is less than or equal to the critical number of

defectives, called the acceptance number for the sample, the lot is accepted. The mathematical details of the evaluation of a sampling plan will be given in the next section.

5.2.1 Single Sampling Plan

The most basic sampling plan is the single sampling plan in which a single sample of n items is inspected and the number x of defective items is recorded. If x is greater than the acceptance number c, the lot is rejected. If $x \leq c$, the lot is accepted.

Example 5.1: Single Sampling Plan

Sample Size (n) = 200

Acceptance Number (c) = 3

Accept lot if Number of Defectives $(x) \leq 3$

Reject Otherwise

5.2.2 Double Sampling Plan

In a double sampling plan two samples are planned. After the first sample of size n_1, the sample may be rejected if the number of defectives x_1 is greater than or equal to the rejection number r_1, or it may be accepted if x_1 is less than or equal to the acceptance number c_1. More often, however, $c_1 < x < r_1$ and a second sample of size n_2 is taken. If the total number of defectives $X = x_1 + x_2$ is less than or equal to c_2, the final acceptance number, the lot is accepted. If $X > c_2$, or equivalently $x_2 > c_2 - x_1$, the lot is rejected.

Example 5.2: Double Sampling Plan

Sample Size	Acceptance No. (c_i)	Rejection No. (r_i)
1. $n_1 = 125$	$c_1 = 1$	$r_1 = 4$
2. $n_2 = 125$	$c_2 = 4$	$r_2 = 5$

Accept the lot if $x_1 \leq c_1$, reject if $x_1 \geq r_1$

Take second sample if $c_1 < x_1 < r_1$

Accept the lot if $X = x_1 + x_2 \leq c_2$

Reject otherwise

5.2.3 Multiple Sampling Plans

Multiple sampling plans are simply an extension of the double sampling plan with more than two stages. At each stage a sample of size n_1 is taken

and the number of defectives x_i found is compared to acceptance (c_i) and rejection (r_i) numbers. At each stage, a lot is either accepted, rejected or the decision to take another sample is made. (An exception is that in some plans, no acceptance decision can be made at the first stage.) After a specified number of stages, a final decision is made.

Example 5.3: 7-Stage Sampling Plan

Sample Size	Acceptance No. (c_i)	Rejection No. (r_i)
1. $n_1 = 50$	No Decision	3
2. $n_2 = 50$	0	3
3. $n_3 = 50$	1	4
4. $n_4 = 50$	2	5
5. $n_5 = 50$	3	6
6. $n_6 = 50$	4	6
7. $n_7 = 50$	6	7

Total of N = 350, maximum number

In the long run a multiple sampling plan produces a lower average total sample size N than the equivalent double sampling plan (which in turn has a lower average total sample size than the equivalent single sampling plan). For a given lot, however, the maximum sample size may be required. A drawback to the multiple sampling plans are that they require more careful adherence to the sampling procedures, more careful data handling, and may require samples from other lots to be held in limbo waiting to be tested.

5.2.4 Other Sampling Plans

Two other sampling plans are sequential sampling and continuous sampling. Sequential sampling is the ultimate extension of multiple sampling with the number of stages unspecified. Continuous sampling is a procedure in which 100% inspection is applied until a specified number of consecutive non-defective items is found. Then sampling is performed at some given rate. When a new defective is found, 100% inspection is re-instated and the process is repeated.

5.3 Evaluation of Sampling Plans

Given any single sampling plan (n, c), where n is the sample size and c is the acceptance number, it remains to be determined if the plan provides the desired discrimination between good and bad lots. The statistical tool for evaluating the worth of a sampling plan is the operating characteristic (OC) curve, which was first discussed in connection with hypotheses tests on a mean in Section 3.6.2.

What is required of a sampling plan is a high probability of accepting a good lot (i.e., one with a low proportion of defectives), and a low probability of accepting a bad lot (i.e., one with a high proportion of defectives). More specifically, let's require the high probability of accepting a good lot to be $1 - \alpha = 0.95$. Further, define the proportion of defectives p in a good lot as the Acceptable Quality Level (AQL). Thus, given a sampling plan (n, c),

$$\Pr(\text{Accept lot} \mid n, c, p \leq \text{AQL}) \geq 0.95.$$

If we further define the Rejectable Quality Level (RQL) as the smallest p for which we want to have a low probability, e.g., $\beta = 0.05$, of accepting the lot (equivalently, a high probability of 0.95 of rejecting the lot), then

$$\Pr(\text{Accept lot} \mid n, c, p \geq \text{RQL}) \leq 0.05.$$

To relate this terminology with that used in Section 3.6.2, note that the Type I Error of rejecting a true hypothesis is equivalent to rejecting a lot with p = AQL; i.e.,

$$\begin{aligned}\Pr(\text{Type I Error}) &= \Pr(\text{Reject lot} \mid n, c, p = \text{AQL}) \\ &= 1 - \Pr(\text{Accept lot} \mid n, c, p = \text{AQL}) \\ &\leq \alpha\end{aligned}$$

Type II Error is equivalent to accepting a lot with p = RQL; i.e.,

$$\begin{aligned}\Pr(\text{Type II Error}) &= \Pr(\text{Accept lot} \mid n, c, p = \text{RQL}) \\ &\leq \beta\end{aligned}$$

The operating characteristic curve is then the plot of probabilities for the whole range of possible p values.

5.3.1 O.C. Curve for a Single and Double Sampling Plan

The probability of accepting a lot can be based on the binomial distribution. For a single sampling plan,

$$\Pr(\text{Accept lot} \mid n, c, p) = \sum_{x=0}^{c} \binom{n}{x} p^x (1 - p)^{n - x}$$

where x is the number of defectives found in the sample, p is the proportion assumed to be defective in the lot, and c is the acceptance number for a single sampling plan with sample size n. (For large n (≥ 20) and small p (≤ 0.05), a Poisson distribution may be used to approximate the binomial probabilities, i.e.,

$$\Pr(\text{Accept lot} \mid n, c, p) = \sum_{x=0}^{c} e^{-np} \frac{(np)^x}{x!}.$$

For a double sampling plan, the calculation is a bit more complex. The probability of accepting on the first sample is

$$\Pr(\text{Accept lot} \mid n_1, c_1, r_1, p) = \sum_{x_1=0}^{c_1} \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} .$$

However, if the value of x_1 is between c_1 and r_1 (the rejection number), a second sample is taken. The probability that the lot is accepted after the second sample, given that a second sample was taken, is the sum of probabilities that the second sample will not produce more than $c_2 - x_1$ defectives (for a total of $x_1 + x_2 \leq c_2$) times the probability of getting exactly x_1 defectives in the first sample.

$$\Pr(\text{Accept lot after second sample} \mid n_2, c_2, r_2, p)$$

$$= \sum_{x_1=c_1+1}^{r_1-1} \Pr(x_2 \leq c_2 - x_1 \mid n_2, c_2, r_2, p, x_1) \Pr(x=x_1 \mid n_1, c_1, r_1, p)$$

$$= \sum_{x_1=c_1+1}^{r_1-1} \left[\sum_{x_2=0}^{c_2-x_1} \binom{n_2}{x_2} p^{x_2} (1-p)^{n_2-x_2} \right] \left[\binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \right] .$$

Putting it all together, the probability of accepting a lot with a double sampling plan is the probability of accepting the lot based on the first sample only, plus the probability of accepting the lot after the second sample.

$$\Pr(\text{Accept lot} \mid n_1, n_2, c_1, c_2, r_1, r_2, p)$$

$$= \Pr(\text{Accept lot} \mid n_1, c_1, r_1, p) + \Pr(\text{Accept lot after 2nd sample} \mid n_2, c_2, r_2, p, x_1)$$

$$= \sum_{x_1=x}^{c_1} \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} + \sum_{x_1=c_1+1}^{n-1} \left\{ \left[\sum_{x_2=0}^{c_2-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p)^{n_2-x_2} \right] \left[\binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \right] \right\}$$

To illustrate, let's evaluate the examples of 5.1 and 5.2.

Example 5.4: Single Sampling Plan: $n = 200$ $c = 3$

$$\Pr(\text{Accept lot} \mid n, c, p = 0.05) = \sum_{x=0}^{3} \binom{200}{x} p^{x} (1-p)^{200 - x}$$

$$= 0.009.$$

Double Sampling Plan $n_1 = 125$ $c_1 = 1$, $r_1 = 4$
$n_2 = 125$ $c_2 = 4$, $r_2 = 5$

$$\Pr(\text{Accept lot} \mid n_1, n_2, c_1, c_2, p = 0.05)$$

$$= \sum_{x_1=0}^{1} \binom{125}{x_1} (0.05)^{x_1} (0.95)^{125-x_1} + \left[\sum_{x_2=0}^{2} \binom{125}{x_2} (0.05)^{x_2} (0.95)^{125-x_2} \right] \left[\binom{125}{2} (0.05)^2 (0.95)^{123} \right]$$

$$+ \left[\sum_{x_2=0}^{1} \binom{125}{x_2} (0.05)^{x_2} (0.95)^{125-x_2} \right] \left[\binom{125}{3} (0.05)^3 (0.95)^{122} \right]$$

$$= 0.0124 + 0.0353 \times 0.0477 + 0.0124 \times 0.0761$$

$$= 0.015.$$

5.3.2 Interpretation of O.C. Curves

Many sampling plans may exist whose operating characteristic curves go through one of the specified points ($1 - \alpha$, AQL) or (β, RQL). Thus, to satisfy both type of error requirements, at least two points on the O.C. curve are needed to evaluate the curve. In fact, several sampling plans may have approximately the same O.C. curves. For each single sampling plan, there are approximately equivalent double and multiple sampling plans. Table 5.1 gives a set of O.C. values for the single and double sampling plans and Figure 5.1 shows their O.C. curves.

Table 5.1: Operating Characteristic Values

Percent Defective, p	Single Sample Plan (Ex. 5.1) $n = 200, c = 3$	Double Sample Plan (Ex. 5.2) $n_1 = 125, c_1 = 1, r_1 = 4$ $n_2 = 125, c_2 = 4, r_2 = 5$
0.01	0.858	0.900
0.02	0.431	0.486
0.03	0.147	0.175
0.04	0.040	0.053
0.05	0.009	0.015

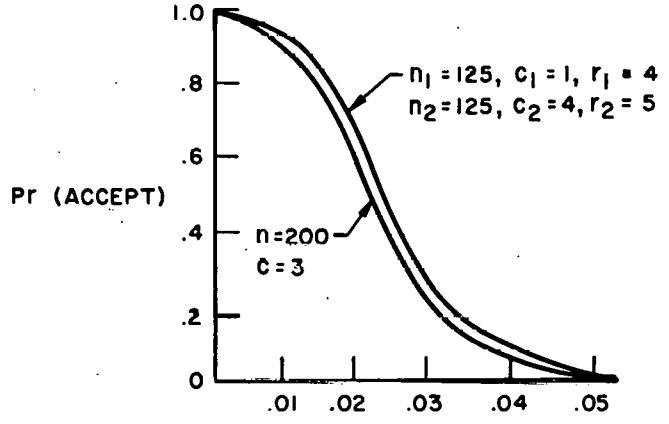
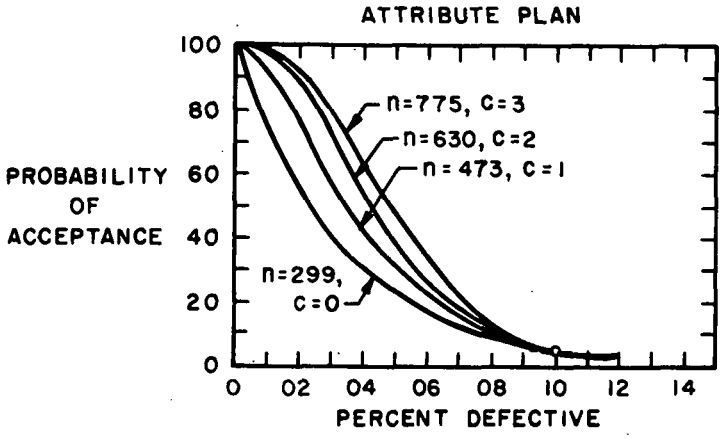


FIGURE 5.1: Operating Characteristic Curve for Nearly Equivalent Single and Double Sampling Plans

The producer of lots of material wants to be sure that good material is not rejected too frequently. Thus, he is concerned primarily about the Type I Error of rejecting a good lot; i.e., p = AQL. The probability of rejecting good lots, Pr(Reject lot | p = AQL) is often called the Producer's Risk. The consumer on the other hand is more concerned about accepting bad lots. The probability of the Type II Error of accepting bad lots, Pr(Accept lot|p=RQL), is often called the Consumer's Risk.

Accordingly, sample plans are sometimes constructed to assure a certain consumer protection. For example, a 90/95 assurance plan requires that the operating characteristic curve of the sampling plan pass through the point $\beta = 0.10$, p = 0.05. That is, with 90% confidence no more than 5% of an accepted lot will be defective. Similarly, a 99/90 consumer assurance plan states that for an accepted lot there is 99% confidence that no more than 10% of the lot is defective. The astute student may recognize the explanation here to be very similar to that of a tolerance interval (Sections 3.7, 4.5), where $\gamma = 1-\beta$, P = 1-p. Figure 5.2 shows four single sampling plans that give 95/99 assurance to the consumer.

FIGURE 5.2: Operating Characteristic Curves Sampling Plans to Meet a 95/99 Assurance Level



*5.4 Average Outgoing Quality Level

Assume that the quality of a lot of incoming material is a certain percent defective and, given a sampling plan, it is advantageous to know the quality of material outgoing. Obviously, a sampling plan which produces a lower average outgoing quality level, AOQL, than does another plan is a preferable sampling plan.

Let p be the proportion of defective pieces in an incoming lot, and Pr(A|p) be the probability of accepting the lot according to a specified sampling plan; e.g., if $x \leq c$ defectives are found in the sample, accept the lot. The average outgoing quality AOQ for the accepted lots is the proportion p of defectives in the lots which were accepted, times the probability of a lot being accepted.

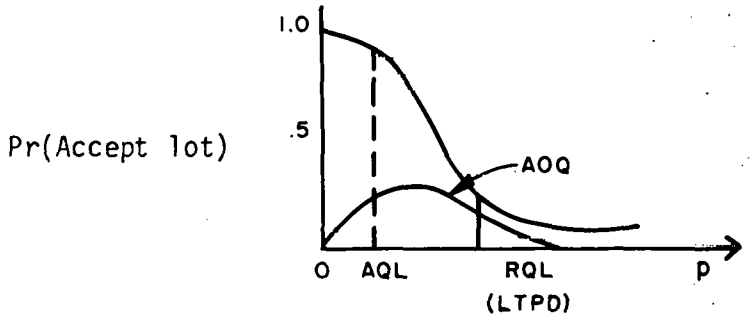
$$\text{AOQ (Accepted lots)} = p \Pr(A|p).$$

If the rejected lots are subjected to 100% inspection and all defectives are replaced with acceptable items, then the total AOQ is

$$\text{AOQ} = p \Pr(A|p) + 0 \Pr(R|p)$$

where Pr(R|p) is the probability of rejecting a lot given the particular sampling plan and proportion defective p, and 0 is the proportion of defects that supposedly is left after 100% inspection.

The average outgoing quality limit or level AOQL, is defined as the limiting quality of the accepted material and is the maximum AOQ over all values of incoming quality, p. If the atual incoming quality p is less than the acceptable quality level, AQL, then few lots are rejected and the AOQ is good. If the incoming quality is greater than the rejectable level, RQL (or LTPD, lot tolerance percent defective) then many lots will be rejected. With 100% inspection of these lots and replacing defectives by good quality material, the AOQ will again be good, even though the incoming quality was poor. When the incoming quality is between the acceptable level and rejectable level, the outgoing quality is the poorest, but still acceptable.



Example 5.5

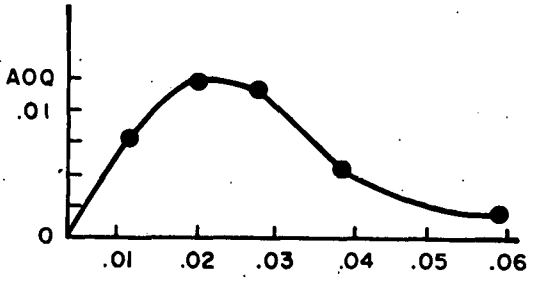
Suppose lots of 1,000 fuel elements for a commercial nuclear reactor are received from the vendor and are tested by sampling 100 elements from each lot. The acceptance criterion is to reject the lot if three or more defectives are found and accept the lot if 2 or fewer are found. The rejected lots are then subjected to 100% inspection and all defectives replaced. The AOQL can be obtained as follows:

$$AOQ = p\ Pr(A|p)$$

using a Poisson approximation for the sampling plan,

$$Pr(A|p) = \sum_{x=0}^{2} e^{-100p}\ (100p)^x/x!$$
$$= e^{-100p}(1 + 100p + (100p)^2/2)$$

p	Pr(A|p)	AOQ
0	1	0
0.01	0.92	0.0092
0.02	0.68	0.0136
0.03	0.42	0.0126
0.04	0.24	0.0096
0.05	0.12	0.0060
0.06	0.06	0.0036



Maximizing p Pr(A|p) with respect to p we obtain a maximum value AOQL of 0.0138 at $p \doteq 0.023$.

Example 5.6

A modification can be made by replacing the defectives found in the accepted lots also. Let D be the number of defectives in the lot and assume a hypergeometric distribution (review Section 2.4.3),

$$Pr(x \le c|D) = \sum_{x=0}^{c} \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=0}^{c} \frac{\binom{D}{x}\binom{1000-D}{100-x}}{\binom{1000}{100}}$$

where x is the number of defectives found in the sample and n-x is the number of acceptable items in the sample. If p is the proportion defective, then $p = \frac{D}{1000}$. The probability of accepting a lot is $Pr(A|p = \frac{D}{1000})$, but

if one defective is found in the sample, it is replaced. Then subtract from Pr(A|p) the probability $\frac{1}{1000}$ Pr(x = 1|D). Similarly for finding two defectives, subtract from Pr(A|p) the value $\frac{2}{1000}$ Pr(x = 2|D). Thus,

$$AOQ = \frac{D\ Pr(x \leq 2|D) - 1\ Pr(x = 1|D) - 2\ Pr(x = 2|D)}{1000} .$$

Values for AOQ, where p = D/1000, D = 1000 p, are:

p	D	AOQ	p	D	AOQ
0.00	0	0.000	0.05	50	0.0060
0.01	10	0.0085	0.06	60	0.0036
0.02	20	0.0127	0.07	70	0.0020
0.03	30	0.0121	0.08	80	0.0011
0.04	40	0.0092	0.09	90	0.0005

The AOQ function can be shown to have a maximum of AOQL = 0.0129 at p=0 023. Note that this is very similar to the previous calculation where we did not replace defectives in the acceptable lots and we used the Poisson approximation.

Example 5.7

Consider the double sampling plan,

	lot size N = 400
	sample size n_1 = 30
First	accept if no defectives found, c_1 = 0
Sample	reject if 2 or more found, r_1 = 2
	if one defect is found, take second sample.
	size of remaining lot = N' = 370
Second	sample size n_2 = 60
Sample	accept if no additional defectives are found, c_2 = 1
	reject if one or more additional defectives are found, r_2 = 2

Suppose D_0 = 5 is an acceptable number of defectives for the lot (i.e., AQL = $\frac{D_0}{N} = \frac{5}{400}$) and D_1 = 10 is a rejectable number of defectives for the lot (i.e., RQL = $\frac{D_1}{N} = \frac{10}{400}$). Then for the first sample, D = D_0 = 5 and D = D_1 = 10, and for the second sample, given that one defective was found in the first sample, D' = D'_1 = 4 and D' = D'_2 = 9. Thus,

$$Pr(Accept|D_0 = 5) = Pr(x_1=0|D_0 = 5) + Pr(x_1= 1|D_0 = 5)\ Pr(x_2=0|D'_0 = 4)$$

$$= \frac{\binom{5}{0}\binom{395}{30}}{\binom{400}{30}} + \frac{\binom{5}{1}\binom{395}{29}}{\binom{400}{30}} \quad \frac{\binom{4}{0}\binom{366}{60}}{\binom{370}{60}}$$

$$= 0.6758 + 0.1361 = 0.8119$$

is the probability of accepting a lot at the AQL value.The probability of accepting a lot at the RQL value, $D_1 = 10$, is

$$\Pr(\text{Accept}|D_1 = 10) = \Pr(x_1=0|D_1=10)+\Pr(x_1=1|D_1=10)\Pr(x_2=0|D_1= 9)$$

$$= \frac{\binom{10}{0}\binom{390}{30}}{\binom{400}{30}} + \frac{\binom{10}{1}\binom{390}{29}}{\binom{400}{30}} \quad \frac{\binom{9}{0}\binom{361}{60}}{\binom{370}{60}}$$

$$= 0.4544 + 0.0753 = 0.5297$$

The AOQL level is obtained by maximizing $pAOQ = \frac{D}{400} \Pr(\text{Accept}|D)$.
For D = 5, AOQ = 0.0101, and for D = 10, AOQL = 0.0133, where we assumed rejected lots are subjected to 100% inspection and all defectives replaced.

5.5 Variable Sampling

Variable sampling deals with continuous measurements, such as length, weight, weight percent, rather than success/failure type data. Many of the principles discussed for attribute data have already been discussed for variable data in Chapter 3. Specifically, the concepts of tests of hypotheses for the mean, operating characteristic curves, and sample sizes determination for specified Type I and Type II error levels were presented earlier. The interested reader should review sections 3.5--3.7 at this time.

In the remainder of this chapter, then, attention will be focused on the quality control techniques of production control charts for continuous variables. An elementary discussion of Shewhart Control Charts for the mean and variation of a production process will be presented. Many other types of control charts exist, including charts for attribute sampling. For more information on control charts, see books on Quality Control, such as by Duncan [11], and basic statistical texts, such as by Johnson and Leone [19], and Miller and Freund [25].

5.6 Shewhart Control Charts

The object of a control chart is to provide an automatic procedure for warning the manufacturer of trouble with his production process. There are several aspects of his product with which he may be concerned. He is

concerned about the average quality of his product, the variability of the product, and the proportion of individual items that do not meet the manufacturing specifications. In addition, he would like a procedure for checking these characteristics of his product that can be performed quickly and easily by the personnel on the line, thus avoiding a possible serious lag time between the occurrence of a problem and its discovery by plant personnel. With these points in minds, three basic types of charts will be discussed: the $\bar{x}$-chart for detecting departure from the process average, the R-chart for detecting problems with process variability, and an individual-measurement chart for monitoring the relationship of individual items to the product specification limits. Finally, an Acceptance Control Chart will be discussed which combines some of the features of an $\bar{x}$-chart and a chart for individuals.

5.6.1 The $\bar{x}$-Chart

The objective of the $\bar{x}$-chart is to keep tabs on the average quality of the product. This is done by taking a sample of n observations at regular intervals. Hopefully, the averages obtained will stay close to the process average established by a collection of previously obtained averages of n observations. If the process drifts or shifts, or simply goes berserk temporarily, the control chart should detect this departure from usual behavior quickly.

The first step in using an $\bar{x}$-control chart is obtaining the process average and control limits. Ideally, the process average will be the same as the nominal or target value for the product characteristic in question. Unfortunately, this is almost never the case in practice, but hopefully it will be close. The process average $\bar{\bar{x}}$ is obtained from the averages of k samples each of size n,

$$\bar{\bar{x}} = \frac{1}{k} \sum_{i=1}^{k} \bar{x}_i, \quad \bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} .$$

At least 20 such averages of about 5 observations each are recommended.

The control limits must be established next. The choice of control limits is arbitrary, but a standard practice is to use 3-sigma limits. That is, the control limits correspond to

$$\mu \pm 3\sigma_{\bar{x}_i}$$

where μ is the true process mean and $\sigma_{\bar{x}_i} = \sigma/\sqrt{n}$ is the true process standard deviation from an average $\bar{x}_i$ of n observations. The 3-sigma limit yields a probability of 0.0027 for an average to exceed either the upper or lower limit by chance alone; i.e., $\Pr(|\bar{x} - \mu| > 3\sigma_{\bar{x}_i}) = 0.0027$. This low chance of exceeding the limits gives a great deal of protection against the costly error of unnecessarily stopping production.

In practice, of course, μ and $\sigma_{\bar{x}_i}$ are unknown. From the same data used to compute the process average $\bar{\bar{x}}$, the estimate of $3\sigma_{\bar{x}_i}$ is obtained. The estimate to be used for this $\bar{x}$-control chart is

$$\bar{\bar{x}} \pm A_2\bar{R}$$

where $\bar{R} = \frac{1}{k}\sum_{i=1}^{k} R_i$, $R_i = \max(x_{ij}) - \min(x_{ij})$, and A_2 is a tabled value (Table XIV) such that $A_2\bar{R}$ is an estimate of $3\sigma_{\bar{x}_i}$. The range R_i is known to be a satisfactory estimate of σ for small sample sizes. It should be noted that it is assumed here that $\bar{x}_i$ follows a normal distribution. The central limit theorem, however, reduces considerably the impact of any departure from normality in the individual observations x_{ij}.

Having established the process average and control limits, one should first plot the original k averages as illustrated in Figure 5.3 to be assured that the process is indeed in control. It is very unlikely, however, that any average $\bar{x}_i$ will exceed the control lines since it was used in establishing those lines. Now succeeding averages may be plotted and the process observed for evidence of being out-of-control. Actually not one, but several, indicators may be used:

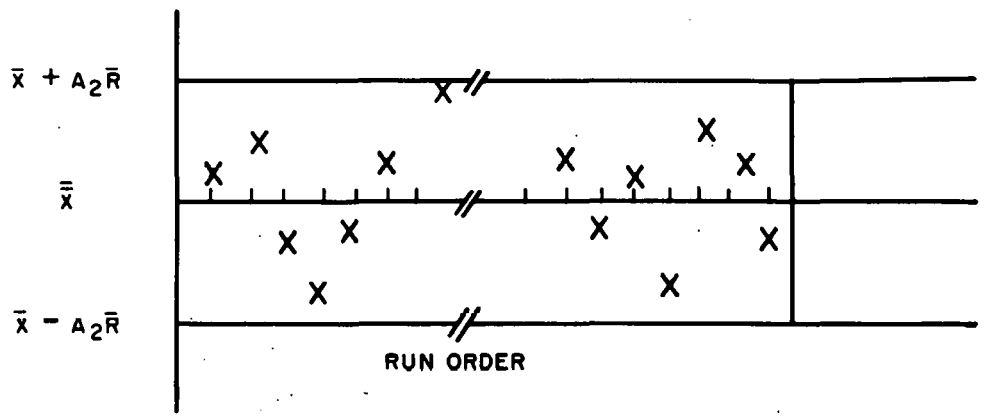


Figure 5.3: $\bar{x}$-Control Chart with 3-Sigma Limits

1. $\bar{x}_i$ exceeds 3-sigma limit.

   This implies a sudden shift in the process has occurred or a trend has developed which may be irreversible if left alone. If the violating value seems isolated, perhaps a data handling error has occurred. The original data for $\bar{x}_i$ should be investigated.

2. $\bar{x}_i$ exceeds 2-sigma limit.

   This is often used as a warning signal, but is not necessarily an out-of-control signal. (2-sigma limits are given by $2 A_2\bar{R}/3$.)

3. A run of considerable length on one side of the target value ($\bar{\bar{x}}$) indicates a possible shift or trend. A run is a series of consecutive averages $\bar{x}_i$ that occur on the same side of the target value. A run of 8 consecutive averages will occur by chance alone approximately 4 times out of a 1,000 compared to the chance of a single average exceeding the 3-sigma limit (3 out of 1,000). Other run indicators with similar probabilities of occurring by chance are

   10 out of 11 on same side

   12 out of 14 on same side

   14 out of 17 on same side

   16 out of 20 on same side.

If evidence is obtained that the process is out-of-control, production is usually stopped, and the process investigated. Upon restarting the process, new control lines should be established.

5.6.2 The R-Chart

Although the process average is important to control, the successive averages could be well within the limits and the process still be out-of-control. This may occur if the process has become extremely erratic. If the standard deviation of the process goes up considerably, the average may appear adequate, but a large proportion of individuals may be exceeding the specification limits. Obviously, then, a control chart monitoring process variability is in order. One such chart would be the s-chart but due to the greater complexity and time required for an operator to calculate $s_i = \sqrt{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2/(n-1)}$ rather than R_i, only the R-chart will be presented here. This is in keeping with the principle of establishing procedures which can be performed quickly and easily by personnel on the production line.

For an R-Chart, the ranges R_i are observed for each of the k samples of n observations used for establishing the limits of the $\bar{x}$-control chart. The 3-sigma control limits for R_i, then are

$$\text{Upper Control Limit} \qquad UCL_R = D_4\bar{R}$$

$$\text{Lower Control Limit} \qquad LCL_R = D_3\bar{R}$$

where D_3 and D_4 are found in Table XIV. If an individual range value R_i exceeds the limits, there is strong evidence of lack of control. The process should be stopped and an investigation into the cause of the change in variation begun.

5.6.3 An Example of an $\bar{x}$-Chart and R-Chart

The $\bar{x}$-chart and the R-Chart go hand-in-hand and should be applied together.

Example 5.7

The data given in Table 5.2 represents coded measurements of the grain size of fuel pellets for the light-water breeder reactor. Samples were taken for 25 consequtive blends of pellet material.

The process average and average range are

$$\bar{\bar{x}} = 13.70, \quad n = 5, \; k = 25$$

$$\bar{R} = 5.88$$

The 3-sigma control limits for $\bar{x}_i$ are

$$\bar{\bar{x}} \pm A_2\bar{R}$$

$$LCL = 13.70 - 0.577\ (5.88) = 10.31$$

$$UCL = 13.70 + 0.577\ (5.88) = 17.09$$

The 3-sigma control limits for R are

$$LCL_R = D_3\bar{R} = 0$$

$$UCL_R = D_4\bar{R} = 2.115 \times 5.88 = 12.44$$

The control charts are shown in Figures 5.4 and 5.5. The process appears to be under control.

Table 5.2: Pellet Grain Size for 25 Consecutive Blends

Sample Number	Sample Measurements (Coded)					Statistics			
	x_1	x_2	x_3	x_4	x_5	$\bar{x}_i$	R_i	s^2_i	s_i
1	11	13	11	9	8	10.4	5	3.80	1.95
2	12	10	13	15	13	12.6	5	3.30	1.82
3	14	11	11	16	11	12.6	5	5.30	2.30
4	10	12	11	12	15	12.0	5	3.50	1.87
5	14	13	12	10	13	12.4	4	2.30	1.52
6	14	17	13	18	8	14.0	10	15.50	3.94
7	18	16	14	11	18	15.4	7	8.80	2.97
8	14	12	9	16	16	13.4	7	8.80	2.97
9	18	13	15	15	14	15.0	5	3.50	1.87
10	12	14	19	16	11	14.4	8	10.30	3.21
11	15	10	18	12	13	13.6	8	9.30	3.05
12	18	13	12	13	14	14.0	6	5.50	2.35
13	15	13	13	14	20	15.0	7	8.50	2.92
14	16	15	14	13	14	14.4	3	1.30	1.14
15	11	9	12	14	12	11.6	5	3.30	1.82
16	13	16	15	19	12	15.0	7	7.50	2.74
17	8	11	12	11	15	11.4	7	6.30	2.51
18	10	11	8	13	13	11.0	5	4.50	2.12
19	14	18	14	20	16	16.4	6	6.80	2.61
20	14	17	16	9	14	14.0	7	9.50	3.08
21	12	16	14	13	13	13.6	4	2.30	1.52
22	21	15	15	15	14	16.0	7	8.00	2.83
23	14	11	12	13	13	12.6	3	1.30	1.14
24	16	22	13	17	17	17.0	9	10.50	3.24
25	14	14	14	16	15	14.6	2	0.80	0.89

$\Sigma R_i = 147$

$\bar{R} = 5.88$

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{25} = 13.70$$

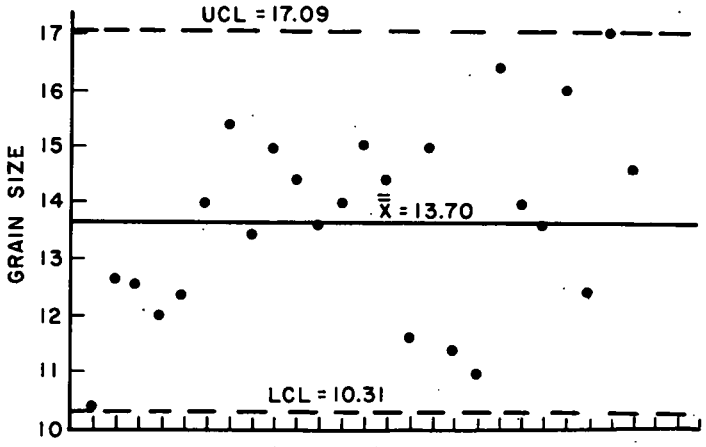


Figure 5.4. $\bar{x}$-Chart on Grain Size

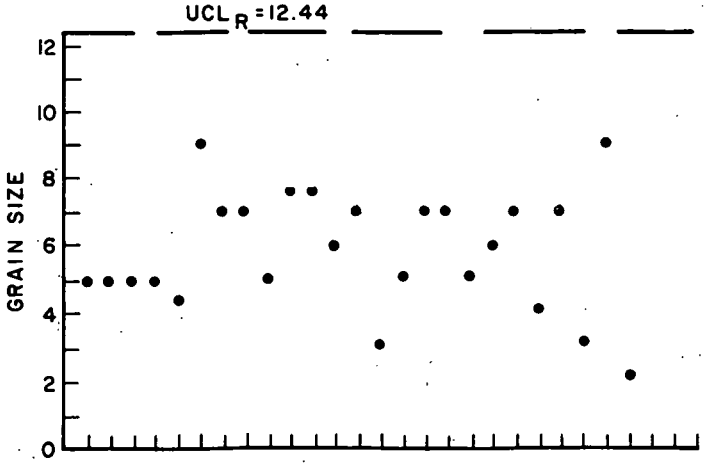


Figure 5.5. R-Chart on Grain Size

5.6.4 Control Charts on Individual Measurements

Neither the $\bar{x}$-chart nor the R-chart ensure the individual observations to be within the specification limits. For this reason a chart for individuals is sometimes desired. This chart would use the process average $\bar{\bar{x}}$ as a center line and the control limits should be the specification limits, or some function of the specification of limits. An ideal arrangement would be for the specification limits to be identical to the 3-sigma limits, $\bar{\bar{x}} \pm 3\sigma$.

The process standard deviation may be estimated in at least two ways: (1) directly, by the usual estimate of the standard deviation of a normal distribution, $s = \sqrt{\Sigma\Sigma(x_{ij} - \bar{\bar{x}})^2/(N - 1)}$ for at least 30 observations; and (2) indirectly, by the estimate of $3\sigma_{\bar{x}}$ for the $\bar{x}$-control chart, $\hat{\sigma} = A_2\bar{R}\sqrt{n}/3$. Alternatively, a moving range could be used to establish the limits

$$\bar{\bar{x}} \pm 3\,\bar{R}/d_2$$

where $\bar{R} = \frac{1}{N-1} \sum_{i=1}^{N-1} | x_{i+1} - x_i |$, and d_2 is given in Table XIV. Since these

successive samples of size 2 used in obtaining $\bar{R}$ are not independent, the use of closer control limits (e.g., $2\,\bar{R}/d_2$) is suggested by some authors (for example, see Johnson and Leone [19]).

5.6.5 Acceptance Control Charts

An acceptance control chart is a control chart for the process average that takes into account the specification limits for individual items. If these specification limits are sufficiently far apart compared to the usual $\bar{x}$-control limits, this procedure may then lead to a relaxation of the $\bar{x}$-control limits.

The procedure is to first determine the process mean that will just meet the specification requirements that not more than 100P% of the individual items will exceed the specification limit (assuming a normal distribution). This process mean is then the rejectable process level, μ_{RPL}. We next wish to establish a test that will give a low probability of accepting a mean at the RPL level or worse. That is, we want to establish a critical value $\bar{x}_{crit}$ such that the Type II error of accepting $\mu \geq \mu_{RPL}$ is small; i.e.,

$$Pr(\bar{x}_i < \bar{x}_{crit} \mid \mu = \mu_{RPL}) \leq \beta, \quad (\beta \text{ small}).$$

Finally, an acceptable process level μ_{APL} can be determined by applying a Type I error calculation for a given α level;

$$Pr(\bar{x}_i > \bar{x}_{crit} \mid \mu = \mu_{APL}) \leq \alpha, \quad (\alpha \text{ small}).$$

Thus, the probability of rejecting an average by chance (and hence illustrating a lack of control), when the process mean is as good or better than μ_{RPL} is small. The control limits for $\bar{x}$ are then the calculated critical values $\bar{x}_{crit}$.

In summary, the acceptance control chart will provide control lines for a sample average such that (1) there is a small probability of stopping production unnecessarily, (2) there is a small probability of continuing production when the process average is indeed out-of-control, and (3) the required proportion of individual items meet the specification limits. This modified control chart for averages should then be used in conjunction with the R-Chart.

Example 5.8

Suppose the nominal stack length of a fuel rod is 84 inches with specification limits of ± 0.5 in. No more than 0.5% of manufactured rods are allowed to exceed each limit. A Type I error probability of 0.05 is to be applied to the process at the APL level and a Type II error probability of $\beta = 0.01$ is to be applied to the process at the RPL level.

Assuming a normal distribution for stack lengths, a sample size of $n = 4$, and a standard deviation of $\sigma = 0.06$ in., the upper acceptance control limit for $\bar{x}$ is obtained as follows.

1. First calculate the rejectable process level μ_{RPL}. The requirement is that at $\mu = \mu_{RPL}$ no more than 0.5% of the population shall exceed the upper specification limit (USL).

$$Pr(\bar{x} > 84.5 \mid \mu = \mu_{RPL}, \sigma = 0.06) = 0.005$$

This yields $\frac{84.5 - \mu_{RPL}}{\sigma} = 2.576 = Pr(z > z_{0.005})$. Then

$$\mu_{RPL} = 84.5 - 2.576\ (0.06)$$
$$= 84.35$$

2. The second step is to calculate the critical value of a test on $\bar{x}$ that will accept $\mu = \mu_{RPL}$ with a probability of 0.01. The upper control limit for $\bar{x}$ is $\bar{x}_{crit}$.

$$Pr(\bar{x} < \bar{x}_{crit} \mid \mu = \mu_{RPL}, \sigma = 0.06, n = 4) = 0.01$$

This gives $\frac{\bar{x}_{crit} - \mu_{RPL}}{\sigma / \sqrt{n}} = -2.326 = Pr(z < z_{0.01})$

$$UCL = \bar{x}_{crit} = \mu_{RPL} - z_{0.01}\ \sigma / \sqrt{n}$$
$$= 84.35 - 2.326\ (0.03)$$
$$= 84.28$$

3. Finally, to establish the process average that will be accepted with a high probability, i.e., the acceptable process level μ_{APL}.

$$Pr(\bar{x} > \bar{x}_{crit} \mid \mu = \mu_{APL}, \sigma = 0.06, n = 4) = 0.05.$$

This gives $\frac{\bar{x}_{crit} - \mu_{APL}}{\sigma / \sqrt{n}} = 1.645 = Pr(z > z_{0.05})$

$$\mu_{APL} = 84.28 - 1.645\ (0.03)$$
$$= 84.23$$

These steps are illustrated in Figure 5.6. The student should perform the calculations required to obtain the lower μ_{RPL}, LCL, and μ_{APL}.

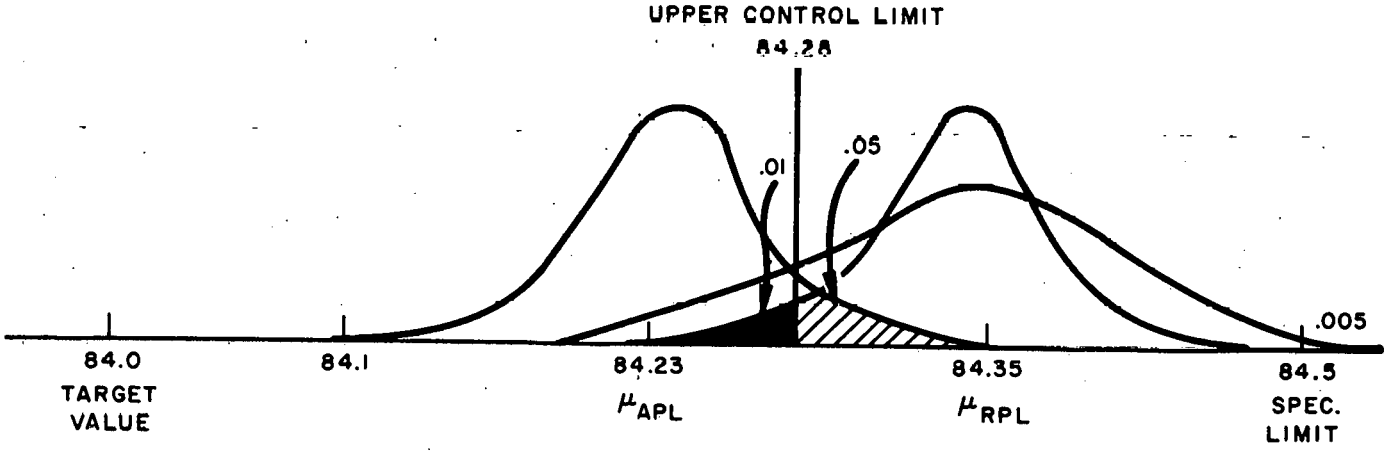


FIGURE 5.6
Upper Acceptance Control Limit for $\bar{x}$: Example 6.8

CHAPTER 6
COMPARISONS OF POPULATIONS

Thus far we have discussed the analysis of data coming from a single population. In many circumstnces we want to compare two or more populations. We may compare the mean value of two production processes or compare their variance to see if one process is more consistent or precise than another. We may compare a new process against a standard or compare many processes to see if any differences exist. In this chapter we shall first examine the problem of comparing two populations and then present a general procedure for comparing any number of populations by an Analysis of Variance (ANOVA) table.

6.1 Comparison of Two Means, Variance Known

Suppose we are interested in comparing the means of two processes, observations from which are considered to have come from the same population. Let x_{ij} be the ith observation from the jth population, μ_j be the mean of that population and ϵ_{ij} be the random error occuring on the ith observation on the jth population. Thus, we may write a model for the problem as

$$x_{ij} = \mu_j + \epsilon_{ij} \quad ,$$

where we usually assume that the errors are independently and normally distributed with mean 0 and variance σ^2_j. The processes under consideration may differ in their means because they represent different treatments, such as two chemical additives, two analytical measuring devices, or two plans of operation. With variances known we can with the use of the Central Limit Theorem compare the means of the treatments by a standard normal variable; let

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}} \qquad (6.1.1)$$

where $\bar{x}_j$ is the average of the n_j observations from the jth treatment and σ^2_j is the known variance. These observations are obtained in a random order to ensure their independence. Thus, the variance of the statistic $\bar{x}_1 - \bar{x}_2$ is $\sigma^2_1/n_1 + \sigma^2_2/n_2$. To test the null hypothesis that $H_0: \mu_1 - \mu_2 = 0$ (or any other value), or to compute a confidence interval for $\mu_1 - \mu_2$, we need only to refer to the table of standard normal deviates.

Example 6.1

Suppose it is thought that by changing the rate of cooling liquid flowing across the cutting edges of a high speed drill the tool life

could be increased. To test this hypothesis, seven observations were obtained using the standard flow rate (process 1) and six observations were obtained using the new (process 2) flow rate, all thirteen trials performed in random order. The results were $\bar{x}_1 = 12.4$ hours $n_1 = 7$, $\bar{x}_2 = 13.6$ hrs., $n_2 = 6$, where it is also assumed that $\sigma_1^2 = 1.0\ (\text{hrs.})^2$, and $\sigma_2^2 = 2.0\ (\text{hrs.})^2$.

$H_0: \mu_1 = \mu_2$ (equivalently $\mu_1 - \mu_2 = 0$)

$H_A: \mu_1 < \mu_2$ $(\mu_1 - \mu_2 < 0)$

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{12.4 - 13.6 - (0)}{\sqrt{\frac{1}{7} + \frac{2}{6}}}$$

$$= \frac{-1.2}{\sqrt{0.48}} = -1.74$$

Since -1.74 is much less than -1.645 ($Pr(z<-1.645) = 0.05$), we reject the hypothesis that the flow rates have no effect on tool life. This inference is based on a one-sided hypothesis test with a 5% significance level (i.e., Type I error probability).

6.2 Comparison of Two Variances

We may also desire to test whether or not two variances are the same. Suppose observations from two sources are normally distributed. Is one process or treatment more variable than another? More precise or reproduceable? To answer the question, we need first estimate the variance from each process as follows:

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (x_{i_1} - \bar{x}_1)^2}{n_1 - 1}, \quad s_2^2 = \frac{\sum_{i=1}^{n_2} (x_{i_2} - \bar{x}_2)^2}{n_2 - 1}$$

Each s^2 follows a χ_ν^2 distribution. More specifically $\frac{(n_1 - 1)\, s_1^2}{\sigma_1^2}$ distributed as $\chi^2_{n_1-1}$ and $\frac{(n_2 - 1)\, s_2^2}{\sigma_2^2}$ is distributed as $\chi^2_{n_2-1}$. To test the null hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we need to develop a test

statistic. In testing a hypothesis about a single variance we used the χ^2 distribution. To test two variances for equality we use the test statistic s^2_1/s^2_2, which follows the F-distribution.

Definition

$$F_{\nu_1, \nu_2} = \frac{\sigma_1^2 \, \chi^2_{\nu_1} / \nu_1}{\sigma_2^2 \, \chi^2_{\nu_2} / \nu_2}$$

Where F_{ν_1, ν_2} is a F-distribution with 2 parameters, ν_1, ν_2 which are the degrees of freedom of the two χ^2 distributions involved in the ratio, ν_1 associated with the numerator and ν_2 with the denominator.

It follows that

$$\frac{s_1^2}{s_2^2} = \frac{\sigma_1^2 \, \chi^2_{n_1 - 1}/(n_1 - 1)}{\sigma_2^2 \, \chi^2_{n_2 - 1}/(n_2 - 1)} = \frac{\sigma_1^2}{\sigma_2^2} F_{n_1 - 1, \, n_2 - 1} .$$

That is, s^2_1/s^2_2 follows a $F_{n_{1-1}, \, n_{2-1}}$ distribution, if $\sigma^2_1 = \sigma^2_2$. Thus, under $H_0: \sigma^2_1 = \sigma^2_2 = \sigma^2$ (i.e., $\sigma^2_1 / \sigma^2_2 = 1$), the test statistic s^2_1 / s^2_2 follows a F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

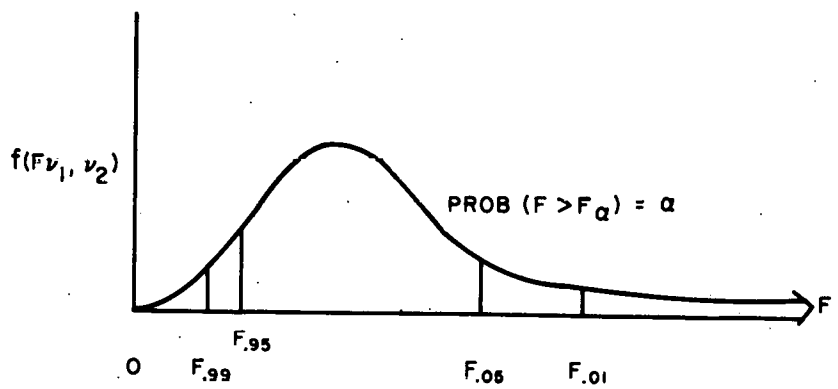


Figure 6.1. F-Distribution

The test of $\sigma_1^2 = \sigma_2^2$ becomes then (for $H_A: \sigma_1^2 \neq \sigma_2^2$)

$$\Pr\left\{F_{\nu_1, \nu_2, 1-\alpha/2} < F_{\nu_1, \nu_2} < F_{\nu_1, \nu_2, \alpha/2}\right\} = 1 - \alpha$$

What are the reasonable values to expect from the distribution? A useful relationship when dealing with the left hand tail of a F-distribution is

$$F_{\nu_1, \nu_2, 1-\alpha} = \frac{1}{F_{\nu_2, \nu_1, \alpha}} .$$

Example 6.2.

Returning to the tool life problem in Example 6.1, construct a 90% confidence interval for the ratio of variances. Suppose the estimates of variances are

$$s_1^2 = 0.74, \; n_1 = 7, \; \nu_1 = 6$$
$$s_2^2 = 2.239, \; n_2 = 6, \; \nu_2 = 5$$

Hypothesis: $H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \qquad \dfrac{s_1^2}{s_2^2} = \dfrac{0.74}{2.239} = 0.3305$

$$H_A: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

$$F_{6,5} = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

$$\Pr\left\{\frac{1}{F_{5,6,0.05}} < F_{6,5} = \frac{s_1^2}{s_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} < F_{6,5,0.05}\right\} = 0.90$$

Then inverting and reversing the signs of the inequalities,

$$\frac{s_1^2}{s_2^2} \frac{1}{F_{6,5,0.05}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} F_{5,6,0.05} .$$

From the table of the F-distribution, $F_{5,6,0.05} = 4.39$, $F_{6,5,0.05} = 4.95$.

$$\frac{0.3305}{4.95} < \frac{\sigma_1^2}{\sigma_2^2} < 0.3305 \times 4.39$$

$$0.067 < \frac{\sigma_1^2}{\sigma_2^2} < 1.45$$

Thus, we can believe with 90% confidence the $\frac{\sigma_1^2}{\sigma_2^2}$ lies between 0.067 and 1.45. Therefore, <u>accept</u> H_0 that σ_1^2 and σ_2^2 <u>could</u> be equal.

A standard approach to comparing two variances is to put the larger estimate in the numerator and test only against $\sigma_A^2 > \sigma_B^2$ (where $s_A^2 > s_B^2$). By this approach, we would test $H_0: \sigma_2^2 / \sigma_1^2 = 1$ against $H_A: \sigma_2^2 > \sigma_1^2$ by comparing $\frac{s_2^2}{s_1^2} = \frac{2.239}{0.74} = 3.03$ against $F_{5,6,.0.05} = 4.39$.

Then we would <u>accept</u> H_0. (This is a one-sided 5% test, or equivalently, a two-sided 10% test).

The assumption of normality of the observations was made in constructing this test. In comparing variances, departures from a normal distribution can cause significant errors in the probability of accepting unequal variances as equal, and vice versa. Thus, care should be taken in comparing variances, particularly in recognizing the distributions involved.

6.3 Comparing Two Means, Variance Unknown

In comparing two processes for differences in their means when the variances were known, it made no difference whether or not these variances were equal. We simply calculated the appropriate standardized normal value and compared it to the critical value. When the variances are unknown, we proceed just as we did when we dealt with a single population; i.e., we replace σ^2 by its estimate s^2 and z by t. As we shall see, however, the inequality of variances causes certain extra difficulties. We shall being with the assumption that the variances are equivalent. This is often a reasonable assumption since the variance of the data is the variance of the errors involved, $\text{Var}(x_{ij}) = \text{Var}(\mu_j + \epsilon_{ij}) = \text{Var}(\epsilon_{ij})$. We often use the same equipment and procedure to obtain data even though the process may have changed, and these errors are sometimes larger than differences in process variability.

6.3.1 t-Test, Variance Equal

In comparing means of two populations where $\sigma_1^2 = \sigma_2^2$, we use the following:

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{1/n_1 + 1/n_2}} \quad , \qquad (6.1.2)$$

where $t_{n_1+n_2-2}$ is a t-distribution with n_1+n_2-2 degrees of freedom since these are the degrees of freedom in the pooled estimate of variance s_p^2,

$$s_p^2 = \frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{n_1-1 + n_2-1} = \frac{\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2i} - \bar{x}_2)^2}{n_1+n_2-2}.$$

If it is true that $\sigma_1^2 = \sigma_2^2$, then the observations about $\bar{x}_1$ can be expected to vary about as much as the observations about $\bar{x}_2$ vary. Thus, two independent estimates of the same parameter σ^2 are available and can be combined. Since there are $n_1 - 1$ and $n_2 - 1$ degrees of freedom (i.e., the means μ_1 and μ_2 were estimated by $\bar{x}_1$ and $\bar{x}_2$) in the individual estimates, there are $n_1 - 1 + n_2 - 1$ degrees of freedom in the pooled or combined estimate of σ^2. If $\bar{x}_1 - \bar{x}_2 > t_{n_1+n_2-2,0.025}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ then the hypothesis $\mu_1 - \mu_2 = 0$ has been rejected at the 5% level(2-sided test). This statistic, however, is completely general for testing any hypothesis $H_o: \mu_1 - \mu_2 = A$. Note that if

$$\sigma_1^2 = \sigma_2^2 \text{, then } \mathrm{Var}(\bar{x}_1 - \bar{x}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2$$

is minimized when $n_1 = n_2$ [try $n_1 + n_2 = 8$: $n_1 = n_2 = 4$ yields

$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$; $n_1 = 3$, $n_2 = 5$ yields $\frac{1}{3} + \frac{1}{5} = 0.533$; $\frac{1}{2} + \frac{1}{6}$ yields 0.667, etc]

This also holds when we estimate σ^2 by s_p^2. Thus, if we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (known or unknown), choose equal sample sizes

$n_1 = n_2$ in order to minimize $\text{Var}(\bar{x}_1 - \bar{x}_2)$ and hence shorten the corresponding confidence interval.

If $n_1 = n_2$, then the test statistic reduces to

$$t_{2(n-1)} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{2s_p^2 / n}} \qquad (6.1.3)$$

where $n_1 = n_2 = n$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and

$$s_p^2 = \frac{s_1^2 + s_2^2}{2} = \frac{\sum_{j=1}^{2} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{2(n-1)} .$$

Example 6.3. Consider again the tool life problem.

Given the following results, test $\mu_1 - \mu_2 = 0$.

$$\bar{x}_1 = 12.4,\ s_1^2 = 0.74,\ n_1 = 7,\ \nu_1 = 6$$

$$\bar{x}_2 = 13.6,\ s_2^2 = 2.239,\ n_2 = 6,\ \nu_2 = 5$$

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

A 95% confidence interval is

$$\bar{x}_1 - \bar{x}_2 \pm t_{11,\,0.025}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$t_{11,.025} = 2.20,\ s_p^2 = \frac{6(0.74) + 5(2.239)}{11} = \frac{15.635}{11} = 1.42$$

$$s_p\sqrt{\frac{1}{7} + \frac{1}{6}} = 0.66.$$

The 95% confidence interval, then, is

$$12.4 - 13.6 \pm 2.20\ (0.66)$$

$$-1.2 \pm 1.45$$

$$(-2.65 < \mu_1 - \mu_2 < 0.25).$$

Since the interval contains the value of zero, we do not reject the hypothesis H_0, that the means are equal.

6.3.2 t-test, Variances Unequal

If $\sigma_1^2 \neq \sigma_2^2$, what happens? Replacing σ_1^2 by s_1^2 and σ_2^2 by s_2^2,

the test statistic becomes

$$t^* = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

but the exact distribution is not clear. The problem of comparing two means when the variances are unknown but unequal is known as the Behrens-Fisher problem and there have been several solutions proposed. Two are suggested here. The basic problem is to correctly specify the critical t-value.

A weighted critical value suggestion is found in Cochran and Cox [5].

$$t^* = \frac{\frac{s_1^2}{n_1} t_{\nu_1, \alpha/2} + \frac{s_2^2}{n_2} t_{\nu_2, \alpha/2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (6.1.5)$$

which reduces to $t_{n-1, \alpha/2}$ if $n_1 = n_2$. It can be shown [26] that the Type I error probability remains reasonably constant over the whole range of possible ratios for σ_1^2 / σ_2^2 if the sample sizes are equal. Thus, when the variances are unknown, choosing equal sample sizes protects against erroneous conclusions.

A second solution makes use of the fact that the sum of two independent χ^2 variables is a weighted χ^2 variable.

That is,

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \sim a \chi_b^2$$

where a is a multiplying constant, $a > 0$, and b is the degree of freedom of the resulting χ^2 distribution. By equating the first two moments of the two sides and solving for b, we can obtain the appropriate critical values for the test of means.

$$E\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right] = E\left[\frac{\sigma_1^2}{n_1 \nu_1} \chi_{\nu_1}^2 + \frac{\sigma_2^2}{n_2 \nu_2} \chi_{\nu_2}^2\right]$$

$$= \frac{\sigma_1^2}{n_1 \nu_1} \nu_1 + \frac{\sigma_2^2}{n_2 \nu_2} \nu_2$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\text{Var}\left[\frac{\sigma_1^2}{n_1 \nu_1} \chi_{\nu_1}^2 + \frac{\sigma_2^2}{n_2 \nu_2} \chi_{\nu_2}^2\right] = \frac{\sigma_1^4}{n_1^2} \frac{2\nu_1}{\nu_1^2} + \frac{\sigma_2^4}{n_2^2} \frac{2\nu_2}{\nu_2^2}$$

$$E(a\chi_b^2) = ab, \quad \text{Var}(a\chi_b^2) = 2a^2b.$$

$$\text{Thus, } ab = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$a^2 b = \frac{\sigma_1^4}{n_1^2 \nu_1} + \frac{\sigma_2^4}{n_2^2 \nu_2}$$

Solving gives

$$b = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\left(\frac{\sigma_1^2}{n_1}\right)^2}{\nu_1} + \frac{\left(\frac{\sigma_2^2}{n_2}\right)^2}{\nu_2}}$$

$$a = \left(\frac{\sigma_1^4}{n_1^2 \nu_1} + \frac{\sigma_2^4}{n_2^2 \nu_2}\right) \Big/ \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Thus, t^* follows a t_b distribution since $\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \sim a\ ^2\chi_b^2$

where b is estimated by replacing σ_1^2 and σ_2^2 by s_1^2 and s_2^2,

$$b = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n-1} + \frac{(s_2^2/n_2)^2}{n-1}} \qquad (6.1.6)$$

Note that $\min(n_1-1, n_2-1) < b \leq n_1 + n_2 - 2$.

The calculation of degrees of freedom in this manner is known as Satterthwaite's Approximation and the method can be applied in general for combining several independent estimates of variance into a single estimate of total variance. However, in some complex cases, care must be taken to assure that one is adding estimates of variances, not subtracting them. In such complex cases, it is recommended that you consult your local statistician!

Example 6.4. Tool life data from Example 6.2.

$n_1 = 7,\ s^2_1 = 0.74,\ \nu_1 = 6$

$n_2 = 6,\ s^2_2 = 2.239,\ \nu_2 = 5$

$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ is to be compared to the critical value of

Cochran-Cox,

$$t^*_{0.025} = \frac{\frac{s_1^2}{7} t_{6,0.025} + \frac{s_2^2}{6} t_{5,0.025}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.546$$

or Satterthwaite's Approximation $t_{b,\ 0.025} = 2.32$

$$\text{where } b = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Big/ \left(\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right)$$

$$= 7.7$$

6.3.3 Determining Sample Size for Comparing Two Means

In Section 3.6 we discussed how to determine the sample size required to test a hypothesis with given Type I and Type II error probabilities. Specifically, we illustrated how to find n the required sample size to assure that we would reject a true null hypothesis $\mu = \mu_0$ about a mean with no more than $100\alpha\%$ error and would accept a false alternative hypothesis $\mu = \mu_A$ with no more than $100\beta\%$ error. If the variance of the distribution were known, the required sample size could be determined directly from

$$n = \frac{(z_\alpha - z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2}$$

If the variance were unknown, an iterative procedure could be used utilizing the t-distribution, or the required sample sizes could be looked up in Table IX.

We can accomplish the same thing in dealing with the problems of comparing two mean. Let

$$H_0: \ \mu_1 - \mu_2 = 0$$

$$H_A: \ \mu_1 - \mu_2 = \delta$$

where δ is a difference in means which we would like to detect with high probability (i.e., accept H_0 when H_A correct with small probability). The

test statistic assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{2\sigma^2 / n}}$$

where for $\sigma_1^2 = \sigma_2^2 = \sigma^2$ the variance of $(\bar{x}_1 - \bar{x}_2)$ is minimized

when the sample sizes for each average are equal. Now following the same procedure as in Section 3.6.3, we find

$$n = \frac{(z_\alpha - z_{1-\beta})^2 \, 2\sigma^2}{\delta^2} .$$

Example 6.5

Suppose we wish to detect a difference between two means of size 2σ, if it exists, where σ is the common standard deviation of the two populations. If we allow $\alpha = 0.05$ probability for mistakenly rejecting the null hypothesis $\mu_1 = \mu_2$, and want to be 95% sure of detecting $\delta = 2\sigma$, i.e., $\beta = 0.05$, then

$$z_{0.05} = 1.645, \; z_{0.95} = -1.645, \; \frac{\delta}{\sigma} = \frac{2\sigma}{\sigma} = 2$$

$$n = \frac{(1.645-(-1.645))^2 \; 2}{(\delta/\sigma)^2}$$

$$= \frac{(3.29)^2 \; 2}{4} = \frac{10.82}{2}$$

$$= 5.41 \rightarrow 6 \quad \text{(rounding up to nearest integer)}$$

That is, we need 6 observations for <u>each</u> average if we are to test $\mu_1 = \mu_2$ with 95% (one-sided) confidence and detect $\mu_1 = \mu_2 = 2\sigma$ with probability 0.95.

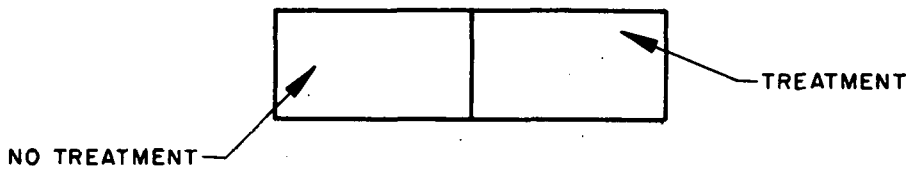
If the variance is unknown and is to be estimated by the data obtained, we replace z by t and must iterate to obtain n. Table X provides the required sample size in this case. For $\delta = 2$, $\alpha = \beta = 0.05$, we find from the table n = 7 (for each average).

Fortunately, the condition of equal variance is not an overriding consideration provided that the sample size from each population is the same or nearly the same. Then, the probability that a $100\gamma\%$ confidence interval contains the true value, $\mu_1 - \mu_2$, is approximately $\gamma = 1 - \alpha$. For equal sample sizes, then, the t-test is insensitive to departures from the assumption of equal variances. It is <u>robust</u> with respect to this assumption.

6.4 The Paired-t test

It is not unusual for an experiment for distinguishing between means to have a very large variance estimate, thus making it extremely difficult to find differences even when they do exist. The situation is not unlike finding a golf ball in the rough! If the grass (variance) is high, we have to look hard and long (large sample size) to detect the ball ($\mu_1 - \mu_2$). If we can "cut the grass", we have a better chance of finding what we are looking for. What often occurs is that there is a greater variance between experimental units being tested than there is between the treatments!

Suppose we are comparing the corrosion resistance of a sample of a particular type of alloy, subject to a particular chemical treatment, with that of an untreated sample. Applying the treatment to one set of samples and no treatment to another set of samples will enable us to perform a t-test for comparing corrosion resistance. However, if the observed resistances of the samples were quite variable, we may not be able to tell whether a difference existed, i.e., the test would not be very sensitive. What if the samples of alloy were large enough to test both the treatment and no treatment to each of the samples?



Then we can compare the results of corrosion resistance for <u>each</u> experimental unit (samples of alloy). The test for comparing two means then becomes a single population problem by taking differences of the observed pairs. Thus, the test is called a <u>paired-t</u> test. The analysis is as follows:

$$H_0: \mu_1 - \mu_2 = \mu_d = 0$$

$$H_A: \mu_d \neq 0$$

<u>Paired-test Test</u>

<u>Treatment</u>	<u>No Treatment or Standard</u>	<u>$d_i = x_{1i} - x_{2i}$ Differences</u>	
x_{11}	x_{21}	d_1	
x_{12}	x_{22}	d_2	$E(d_1) = \mu_1 - \mu_2$
.	.	.	$Var(d_1) = Var(x_{1i} - x_{2i})$
.	.	.	$= \sigma_d^2$
.	.	.	
x_{1n}	x_{2n}	d_n	

$$\bar{x}_d = \bar{x}_1 - \bar{x}_2 \qquad s_d^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n - 1}$$

$$t_{n-1} = \frac{\bar{x}_d - \mu_d}{\sqrt{\frac{s_d^2}{n}}}$$

We note the following:

1. The degrees of freedom is n - 1 not 2(n - 1) as in the unpaired t test. Everything else being equal, a test with 2(n - 1) degrees of freedom will be more sensitive than one with n - 1 degrees of freedom. But all things are not equal here. By comparing results within each experimental unit, we reduce the variance of the observations d_i. That is, we have removed the between unit or block variation from the problem.

2. The sample size must be equal for each treatment.

3. The variables x_{1i} and x_{2i} may or may not have equal variances and in fact may be correlated. All that matters in the paired-t test is the variances of the differences, σ_d^2. In a development of this fact, the paired-t is sometimes called a correlated-t test.

The advantage of using a paired-t test for comparing means is that you block out a large source of variation in the data due to the differences among experimental units, thus developing a more sensitive test, provided that a large between unit variation exists! An important consideration, however, is that there be a natural basis for comparing treatments within experimental units. Examples of natural pairings are measurements than can be designated by such terms as before/after and with / without, or can be repeated under different conditions on the same experimental unit, such as on two halves of an ingot or by two different measuring devices. If you blocked out between unit variation where none existed, the resulting estimate of variance would not be appreciably reduced, but your t-statistic would have half the degrees of freedom (n - 1) that the more correct unpaired test would have (2(n-1)). On the other hand, if you fail to block out between unit variation that is large, the resulting pooled estimate of variance would be too high. Both situations will result in insensitive tests, i.e., tests that would accept the null hypothesis more often that they should. A rule of thumb, however, is that if you suspect a unit (block) to unit variation, use a paired-t test with as many degrees of freedom as possible. The resulting error in using, say $t_{5,0.025} = 2.571$ rather than $t_{10,0.05} = 2.228$, will usually be less than that error made in not pairing and assuming $\sigma_1^2 = \sigma_2^2$ incorrectly.

Example 6.6.

Ten sample specimens of an alloy are chosen randomly to use in testing the corrosion resistance of a particular chemical treatment. Half of each specimen is treated, the other half not.

Sample: Experimental Unit or Block	Corrosion Resistance x_1 Untreated	x_2 Treated	$x_2 - x_1$ Difference	Sample Average
1	12.1	14.7	2.6	13.40
2	10.9	14.0	3.1	12.45
3	13.1	12.9	-0.2	13.00
4	14.5	16.2	1.7	15.35
5	9.6	10.2	0.6	9.90
6	11.2	12.4	1.2	11.80
7	9.8	12.0	2.2	10.90
8	13.7	14.8	1.1	14.25
9	12.0	11.8	-0.2	11.90
10	9.1	9.7	0.6	9.40

$\bar{x}_1 = 11.60 \qquad \bar{x}_2 = 12.87 \qquad \bar{x}_d = \bar{x}_2 - \bar{x}_1 = 1.27$

H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$

H_A: $\mu_1 < \mu_2$

$$t_9 = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{10}} \qquad s_d^2 = \frac{1}{9}\Sigma(d_i - \bar{d}) = \frac{1}{9}\left[\Sigma d_i^2 - (\Sigma d_i)^2/10\right]$$

$$= \frac{27.55 - 16.129}{9}$$

$$= \frac{1.27}{0.3562} = 3.57 \qquad = \frac{11.421}{9} = 1.269$$

$$> 2.262 = t_{9,0.025} \qquad s_d = 1.126; \quad \frac{s_d^2}{10} = 0.1269$$

i.e., Reject H_0: $\mu_d = 0$. There is a difference between the treated alloy and the untreated alloy in corrosion resistance. The treated alloy yields significantly higher corrosion resistance values.

6.5 The Assumptions of the t-test

There are three underlying assumptions involved in comparing means by a t-test. There are

1) Normality
2) Equality of Variances
3) Independence

Since we are dealing with means, the central limit theorem allows us to assume that $\bar{x}$ is normally distributed. The assumption of normality of the individual observtions, then, although required for very small sample sizes ($n < 4$); is of no real practical consequence. Departures from normality must be extreme before it has any substantial effect on the result.

We have already stated that the t-test is robust with respect to inequality of variances. One way around the assumption of equal variances is to use a paired-t test. Another approach is to use equal sample sizes to minimize the chance of an erroneous conclusion.

The third assumption, independence of the observations, is very important, however. (In a paired-t test, the differences d_i must be independent.) To assure independence, randomization should be performed. That is, select the order of observations in a random fashion such as flipping a coin, or using a random number generator. We need to choose a treatment to apply to a particular response randomly. In a paired-t test, we need to randomly select the units to be tested and randomly select one treatment to be applied first (or to one side of the unit).

6.6 Comparison of k Variances

We now turn to the problem of comparing more than two populations. To compare more than two variances for possible differences there are several available methods, each with their own difficulties.

1. Hartley's Test - Max-F Test

For k populations and s_1^2, s_2^2, . . ., s_k^2 each with degrees of freedom, compute

$$F_{max} = s_{max}^2 / s_{min}^2 .$$

The critical values have been tabled for various and k at the 5% and 1% levels. However, these tables are not readily available. One source is the Handbook of Statistical Tables by Owen, [24]. An additional drawback to the maximum F-test is restriction to equal degrees of freedom for each variance estimate.

2. Cochran's Test

This test is for determining whether one of k estimates of variance is out-of-line with the others. All estimates must have equal degrees of freedom and one estimate dominate the others. The test is:

$$C = (\max s_i^2) \Big/ \sum_{i=1}^{k} s_i^2 .$$

Again, this test requires a special table which is not readily available. See Table A-17 of Dixon and Massey [8].

3. Bartlett's Test for Homogeneity of Variances

This test is more general in that it does not require that the degrees of freedom be equal for all s_i^2. However, it suffers from a sensitivity to the assumption of normality, and it is cumbersome to calculate. However, it can be programmed and requires only the use of a standard χ^2 table. For k normally distributed populations, then,

$$\chi^2_{k-1} = \frac{\nu_e \ \ell n \ s_p^2 - \sum_{i=1}^{k} \nu_i \ \ell n \ s_i^2}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu_e} \right)}$$

where $s_p^2 = \dfrac{\sum_{i=1}^{k} \nu_i s_i^2}{\nu_e}$, $\nu_e = \sum_{i=1}^{k} \nu_i$.

Example 6.7 Comparison of Four Variances

k = 4 Normal Populations

$s_1^2 = 62.5$, $\nu_1 = 4$		$\ell n \ s_1^2 = 4.13513$
$s_2^2 = 38.67$, $\nu_2 = 6$	$s_p^2 = \frac{2118}{26} = 81.46$	$\ell n \ s_2^2 = 3.65490$
$s_3^2 = 72.0$, $\nu_3 = 9$		$\ell n \ s_3^2 = 4.27667$
$s_4^2 = 141.14$, $\nu_4 = 7$		$\ell n \ s_4^2 = 4.94975$

$$\nu_e \ell n \; s_p^2 = 26 \ell n \; 81.46 = 114.41508$$

$$\nu_e = \Sigma \nu_i = 26$$

$$\chi_3^2 = \frac{\nu_e \ell n \; s_p^2 - \Sigma \nu_i \ell n \; s_i^2}{1 + \frac{1}{9}\left(\frac{1}{4} + \frac{1}{6} + \frac{1}{9} + \frac{1}{7} - \frac{1}{26}\right)} = \frac{114.41508 - 111.70820}{1.07}$$

$$= 2.53 \left[\chi^2_{3,0.025} = 9.35\right]$$

No evidence of inequality of variances.

6.7 Comparing k Means

Analysis of data from any number of normal populations can be summarized by means of an analysis of variance table, or ANOVA table for short, and a model to represent the situation. A model is always present in data analysis, but unfortunately is often not written out. Nevertheless, it is implicity there. Before attacking the problem of comparing k means, it is advantageous for us to examine the ANOVA approach to the simpler problems of testing a single mean and comparing two means.

6.7.1 The Analysis of Variance for Testing $\mu = 0$

For observing data from a single population the model is

$$x_i = \mu + \epsilon_i , \qquad (6.7.1)$$

where μ is the mean value of x and $\epsilon_i \sim N(0, \sigma^2)$, i = 1, 2, .., n. The sum of squares $\sum_{i=1}^{n} x_i^2$ can be considered as a measure of the total information in the data. This can be partitioned into two parts:

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x})^2$$
$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n \bar{x}^2 ,$$

where $n \bar{x}^2$ (or $\frac{1}{n} (\Sigma x_i)^2$) is the amount of information accounted for by the estimate $\bar{x}$ of the mean. The measure of information that remains, $\Sigma (x_i - \bar{x})^2$ is attributable to the errors ϵ_i, and when divided by n - 1 gives s^2, the estimate of variance, σ^2. This term $\Sigma (x_i - \bar{x})^2$ is called the

residual sum of squares, since $x_i - \bar{x}$ is what is left over, the residual, after estimating the mean of the population. The residuals estimate the random errors that have occurred in obtaining the observations. The sum of squares, $\Sigma(x_i - \bar{x})^2$, is in general the measure of total variability in the data and incorporates all sources of variability in the observations.

Definition

The Analysis of Variance is the partitioning of this total variability into component parts according to the model being examined, and then determining which, if any, of these components contribute significantly to the total variation in the data.

For model (6.7.1) the analysis can be summarized in the ANOVA Table as follows:

Table 6.1 ANOVA for $x_i = \mu + \epsilon_i$

Source	Sum of Squares	Degrees of Freedom	Mean Square
Total	$\sum_{i=1}^{n} x_i^2$	n	
Mean	$\frac{1}{n}(\sum x_i)^2 = n\bar{x}^2$	1	$n\bar{x}^2$
Residuals or Errors	$\sum_{i=1}^{n}(x_i - \bar{x})^2$	n - 1	s^2

where the Mean Square is the sum of squares divided by the degrees of freedom and is used to estimate a function of the variance. To test the hypothesis that $\mu = 0$, we need first to find the expected value of $n\bar{x}^2$. We know that $Var(\bar{x}) = E(\bar{x}^2) - \mu^2$ so

$$\begin{aligned} E(n\bar{x}^2) &= n[Var(\bar{x}) + \mu^2] \\ &= n[\sigma^2/n + \mu^2] \\ &= \sigma^2 + n\mu^2 . \end{aligned}$$

Thus, if H_0: $\mu = 0$ is true, we find that $n\bar{x}^2$ estimates σ^2. The residual sum of squares divided by its degrees of freedom is called the residual or error mean square and estimates σ^2, and, in fact, is the best, unbiased estimate of σ^2. There is no other source of variability in the residuals other than random error. Hence, $n\bar{x}^2$ and the residual mean square s^2 should be compatible estimates of σ^2, if $\mu = 0$. We can test this hypothesis by an

F-ratio where

$$H_0: \mu = 0, \text{ and}$$

$$H_A: \mu \neq 0,$$

$$F_{1,n-1} = \frac{n\,\bar{x}^2/1}{\Sigma(x_i-\bar{x})^2/(n-1)} = \frac{\bar{x}^2}{s^2/n}$$

That is, the test of $\mu = 0$ has been answered in terms of a test of equality of two independent estimates of σ^2 under the hypothesis that $\mu = 0$. [Note: It can be shown that $n\,\bar{x}^2$ and $\Sigma(x_i - \bar{x})^2$ are statistically independent by showing Cov ($\bar{x}$, $\Sigma(x_i - \bar{x})^2$) = 0, where Cov stands for the covariance of two random variables.] If the calculated F-value is greater that the critical $F_{1,n-1,0.05}$ value, we reject $H_0: \mu = 0$. [Note: $F_{1,n-1,0.05} = t^2_{n-1,0.025}$].

Example 6.8

The metallurgy laboratory has five observations on a zircaloy alloy produced by an arc melting process. The weight percents of zirconium in the alloys samples are 90, 91, 93, 90, and 94.

$$\Sigma x_i = 458,\ \bar{x} = 91.6,\ n = 5$$

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square
Total	$\Sigma x_i^2 = 41966$	5	
Mean	$(\Sigma x_i)^2/5 = 41952.8$	1	41952.8
Residuals or Errors	13.2	4	$s^2 = 3.3 \rightarrow \sigma^2$ (→ means estimates)

The estimate of the mean is 91.6 and of s^2 is 3.3. The F-test for significance of the mean is 41952.8/3.3 and is certainly significant. This is equivalent to the t-test

$$t_4 = \frac{\bar{x}}{s/\sqrt{5}} = \sqrt{\frac{n\,\bar{x}^2}{s^2}}\ .$$

When an F or t test is found to reject the null hypothesis, we say that the estimated parameters are significantly different from the hypothesized values. In this example, the mean μ is significantly different from zero.

6.7.2 Analysis of Variance for Comparing k Means

A question of more interest is whether or not two or more means are equal. The model associated with this problem is

$$x_{ij} = \mu_j + \epsilon_{ij} \quad , \quad j = 1, 2, \ldots, k; \; i = 1, 2, \ldots, n_j \qquad (6.7.2)$$

$$E(x_{ij}) = \mu_j \quad ; \quad \epsilon_{ij} \sim N(0, \sigma^2) \quad .$$

That is, there are n_j observations from the jth popultion and all populations are assumed to have equal variance. We may rewrite this model as

$$x_{ij} = \mu + \tau_j + \epsilon_{ij} \qquad (6.7.3)$$

where μ is the overall mean of the population and τ_j is the extra impact on the response of having come from the jth population. τ_j is usually called the treatment effect. From the definition of $\tau_j = \mu_j - \mu$, a linear constraint, $\sum n_j \tau_j = 0$, exists on the treatment.

The test of H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ is equivalent to the test of H_0: $\tau_1 = \tau_2 \ldots = \tau_k = 0$, since μ is common to all populations.

We can proceed as before and find the sum of squares due to the grand mean μ, i.e.,

$N \bar{\bar{x}}^2 = \frac{1}{N} (\sum\sum_{ji} x_{ij})^2$ where $\bar{\bar{x}} = \frac{1}{N} \sum_j^k \sum_i x_{ij}$ and $N = \sum_{j=1}^k n_j$. This is commonly referred to the <u>correction factor</u>. The remaining or <u>Corrected Sum of Squares</u> is $\sum_j \sum_i (x_{ij} - \bar{\bar{x}})^2$. This still contains a source of variability from the different treatment effects. The total sum of squares may be partitioned as follows:

Define: $\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$, then,

$$\sum_j^k \sum_i^{n_j} x_{ij}^2 = \sum_j^k \sum_i^{n_j} (x_{ij} - \bar{x}_j + \bar{x}_j - \bar{\bar{x}} + \bar{\bar{x}})^2$$

$$= \sum_j \sum_i (x_{ij} - \bar{x}_j)^2 + \sum_j n_j (\bar{x}_j - \bar{\bar{x}})^2 + N \bar{\bar{x}}^2 \quad ,$$

where the first term is the residual sum of squares, the second is the treatment sum of squares, and the last is the <u>correction factor</u> (due to the mean). The ANOVA table becomes

Table 6.2. ANOVA Table for Comparing k Means

Source	Sum of Squares	Degrees of Freedom	Mean Square	EMS
Total	$\sum_j \sum_i x_{ij}^2$	N		
Mean (Correction Factor)	$N\bar{\bar{x}}^2$	1	$N\bar{\bar{x}}^2$	$\sigma^2 + N\mu^2$
Corrected Total S_X	$\sum_j \sum_i (x_{ij} - \bar{\bar{x}})^2$	N - 1		
Treatments SST	$\sum_j^k n_j (\bar{x}_j - \bar{\bar{x}})^2$	k - 1	$\frac{\sum n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}$	$\sigma^2 + \frac{\sum n_j \tau_j^2}{k-1}$
Residuals or Errors SSE	$\sum_j^k \sum_i^{n_j} (x_{ij} - \bar{x}_j)^2$	N - k	s^2	σ^2

The treatment sum of squares represents the variability of the data attributable to differences in population means μ_j, or simply due to treatment effects τ_j. The estimate of population mean is the sample average $\bar{x}_j$; the estimate of the population effect beyond the grand mean is $\hat{\tau}_j = \bar{x}_j - \bar{\bar{x}}$. It can be shown (see Appendix D) that the expected value of the mean square for treatments, $\frac{1}{k-1} \sum_j^k n_j (\bar{x}_j - \bar{\bar{x}})^2$, is $\sigma^2 + \frac{1}{k-1} \Sigma n_j \tau_j^2$, where τ_j are the true treatment effects. The degrees of freedom here is k-1 not k since $\sum_{j=1}^k n_j(\bar{x}_j - \bar{\bar{x}}) = 0$ by definiton of $\bar{\bar{x}}$. This linear contrast restricts the freedom of the estimates by one. Of course, the residual mean square, or s^2 (pooled) estimates σ^2 unbiasedly. The treatment mean square can be considered as the estimate of between treatment variation and the residual mean square is the within treatment estimate of variance, which under the assumption H_0: $\tau_j = 0$, all j, should be the same. The F-ratio

$$F_{k-1,N-k} = \frac{MS\ (\text{treatments})}{MS\ (\text{residuals})}$$

is a test of the assumption that these estimates of variance are compatible, which in turn is a test of the hypothesis H_0: $\tau_j = 0$ for all j. If the τ's = 0, we can expect that the calculated F will be less than the critical value. If it is larger, we would reject H_0 that all treatments were equal.

In general, treatment means may be compared in the presence of other sources of variability which have been blocked out of the analysis, such as in the paired-t test. Consider the model

$$x_{ij} = \mu + \tau_j + \beta_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \qquad (6.7.4)$$

$$\sum_j \tau_j = \sum_i \beta_i = 0 \qquad j = 1, 2, \ldots k; \quad i = 1, 2, \ldots n,$$

where β_i represent block effects and in general may represent one or more block effects, each of which can be separated out individually. The ANOVA table can be summarized as follows:

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	EMS
Total	$\Sigma\Sigma x^2_{ij}$	N		
Mean (CF)	$N\bar{\bar{x}}^2$	1		
Corrected Total S_X	$\Sigma\Sigma(x_{ij} - \bar{\bar{x}})^2$	N - 1		
Treatments SST	$\sum_j n_j(\bar{x}_j - \bar{\bar{x}})^2$	k - 1	MS(Treatments)	$\sigma^2 + \frac{n}{k-1}\Sigma\tau_j^2$
Blocks SSB	$\Sigma k_i(\bar{x}_i - \bar{\bar{x}})^2$	n - 1	MS(Blocks)	$\sigma^2 + \frac{k}{n-1}\Sigma\beta_i^2$
Residuals or Errors SSE	By subtraction	(n-1)(k-1)	s^2	σ^2

The test for treatment differences is still the MS(treatments)/s^2 as above as long as the block variables have been properly blocked out, and with adjusted degrees of freedom. More details of the ANOVA table analysis will be presented at a later time when we deal with designs of experiments.

Example 6.9

Suppose the metallurgy laboratory wants to compare the mean w/o zirconium in zircaloy alloy from the arc melting process with zirconium w/o from an induction process. The data from the arc process is given in Example 6.8. Five observations from the induction process are 91, 90, 91, 89, 91 (w/o zirconium). The model for comparing the two means is

$$x_{ij} = \mu + \tau_j + \epsilon_{ij}$$

$$x_{ij} = 910, \ \bar{\bar{x}} = 91.0, \ \bar{x}_1 = 91.6, \ \bar{x}_2 = 90.4, \ n_1 = n_2 = 5$$

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square
Total	82830	10	
Mean	82810	1	
Corrected Total S_X	20	9	
Treatments	$\Sigma 5(\bar{x}_j - \bar{\bar{x}})^2 = 3.6$	1	$3.6 \rightarrow \sigma^2$ if $\tau_1 = \tau_2 = 0$
Residuals or Errors	16.4	8	$s^2 = 2.05 \rightarrow \sigma^2$

The F-test for treatment effects is

$$F_{1,8} = \frac{3.6}{2.05} = 1.76$$

The critical value of $F_{1,8,.05} = 5.32$. Therefore, we say that there is no evidence that the arc and induction processes differ.

Another way of calculating the treatment sum of squares and block sum of squares is by the use of the treatment and block sums.

$$SST = \frac{1}{n} \sum_{j=1}^{k} T_j^2 - C.F.$$

$$SSB = \frac{1}{k} \sum_{i=1}^{n} B_i^2 - C.F.$$

where T_j is the sum of all n observtions for treatment j and B_i is the sum of all k observations in block i, and C.F. is the correction factor, $N \bar{\bar{x}}^2$.

Example 6.10

Consider again the corrosion resistance data of Example 6.6. The appropriate model for this data is

$$x_{ij} = \mu + \tau_j + \beta_i + \epsilon_{ij}$$

where τ_j are the treatment effects, j = 1, 2, and β_i are the block effects, i = 1, 2, 3, ..., 10. We have

$$\sum_i x_{i1} = 116.0 \quad , \quad \sum_i x_{i2} = 128.7, \; n_1 = n_2 = 10, \; k = 2$$

$$\sum_j \sum_i x_{ij} = 244.7$$

The ANOVA table is

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square
Total	3070.33	20	
Mean	2993.9045	1	
Corrected Total S_X	76.4255	19	
Treatments	8.0645	1	8.06
Blocks	62.6505	9	6.96
Residuals	5.7105	9	$s^2 = 0.635$

The F-ratio for treatments is

$$F_{1,9} = \frac{MS(\text{Treatments})}{MS(\text{Residuals})} = \frac{8.06}{0.635} = 12.7$$

which is much larger than the critical F-value at the 0.05 level, $F_{1,9,0.05} = 5.12$. Thus, there is evidence that the treatments differ.

Although it is not immediately obvious, this F-test is the square of the paired-t test performed in Example 6.6. That is

$$F_{1,9} = 12.7 = t^2_9 = (3.57)^2 \text{ and}$$

$$F_{1,9,0.05} = 5.12 = t^2_{9,0.025} = (2.262)^2.$$

We may also test to see if the blocks had a significant effect on the response,

$$F_{9,9} = \frac{6.96}{0.635} = 10.9 > F_{9,9,0.05} = 3.18$$

The blocks do have an effect. Note that there is not a t-test to correspond to this F-test for blocks, since the degrees of freedom for blocks is greater than 1.

REVIEW

Test Statistic	Tabled Distribution	Use
1. $\dfrac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{s^2_{\bar{x}_1}+s^2_{\bar{x}_2}}}$	$t_{n_1+n_2-2}$	Test means of two unpaired groups of observations.
2. $\dfrac{\bar{d}-(\mu_2-\mu_1)}{s_{\bar{d}}}$	t_{n-1}	Test means of two groups through n paired comparisons.
3. $\sum_{i=1}^{n} \dfrac{(x_i-\bar{x})^2}{\sigma_0^2}$	χ^2_{n-1}	Test an estimated variance against a hypothesized or true variance, σ_0^2
4. s_1^2/s_2^2	F_{n_1-1,n_2-1}	Test two variances for equivalence.
5. $\dfrac{\sum_{j=1}^{k} n_j(\bar{x}_j-\bar{\bar{x}})^2/(k-1)}{\text{Residual Sum of Squares/ (error)}\ \nu_e}$	$F_{k-1,\ \nu_e}$	Test for difference between k means $H_0: \mu_1=\mu_2=\ldots=\mu_k$

where

$$d=\frac{1}{n}\sum_{i=1}^{n}(x_{1i}-x_{2i}),\ \bar{x}_j=\frac{1}{n}\sum_i x_{ij},\ \bar{x}=\frac{1}{n}\Sigma x_i,\ \bar{\bar{x}}=\frac{1}{nk}\Sigma\Sigma x_{ij},$$

$$s^2_{\bar{d}}=s^2_d/n,\ s^2_d=\frac{1}{n-1}\sum_i(d_i-\bar{d})^2,\ s^2_i=\frac{1}{n-1}\Sigma(x_{ij}-\bar{x}_i)^2,\ s^2_{\bar{x}_1-\bar{x}_2}=s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right),$$

$$s^2=\left[\sum_i(x_{1i}-\bar{x}_1)^2+\sum_i(x_{2i}-\bar{x}_2)^2\right]/(n_1+n_2-2)$$

and ν_e is the degrees of freedom for error in the analysis of variance table.

CHAPTER 7
LINEAR REGRESSION ANALYSIS

7.1 Introduction

Suppose we are interested in some response η for which it has been determined that a set of variables $X_1, X_2, \ldots, X_k$ control the response. There is more of interest here than in just determining whether the average response at one set of conditions differ from the average response at another set of conditions. It is advantageous to determine how the response changes as the levels of the independent variables X_i change. To answer this question we need to first hypothesize a model which we think may apply in some region of interest and then perform an experiment to obtain data to verify or reject the hypothesized model. This verification process is performed under a variety of names, such as curve-fitting, maximum likelihood analysis, least squares, and regression.

There are three objectives which may be approached by regression analysis. These are (1) prediction, (2) estimation and (3) model building. In the first, we are interested in fitting a model to obtain the best prediction of the response as possible. It is not of utmost importance here to have the best model in terms of accuracy of the parameters being estimated as long as the model predicts well. In the second, interest is in obtaining the best or most precise estimates of the parameters of the model under consideration. The purpose of this is usually to make precise evaluations on the significance of a term in the model. In the third objective, interest is in the process of obtaining the best explanatory model. Whether it deals with finding the best set of independent variables or finding the best function to represent the response, model building is an iterative process and requires numerous steps along the way. Of course many problems require a combination of these three objectives.

7.1.1 Models

The models suggested for describing a response may be of a theoretical nature or of an empirical nature. For example, the equation for Ohm's Law I=V/R represents the theoretical change in amperage as a function of voltage and resistance. By an empirical model we mean one suggested by the data itself. If we considered an experiment in which the resistance was held constant and plotted the change in amperage as a function of voltage we would obtain data which follows a straight line. Note, of course, that not all the observations fall on the line since random error is involved in all data taking procedures.

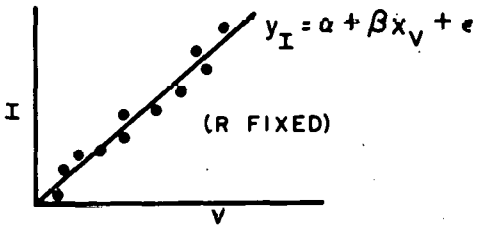


Figure 7.1: I=V/R

More typically, the exact theoretical function representing some response is unknown or too complex to use for some purposes of investigation. In these cases empirical models are used by necessity. For example, all the causative effects which control the height and weight of individuals are unknown. We could nevertheless predict one variable given knowledge of the other by fitting an empirical model to the data. That is, we can predict an

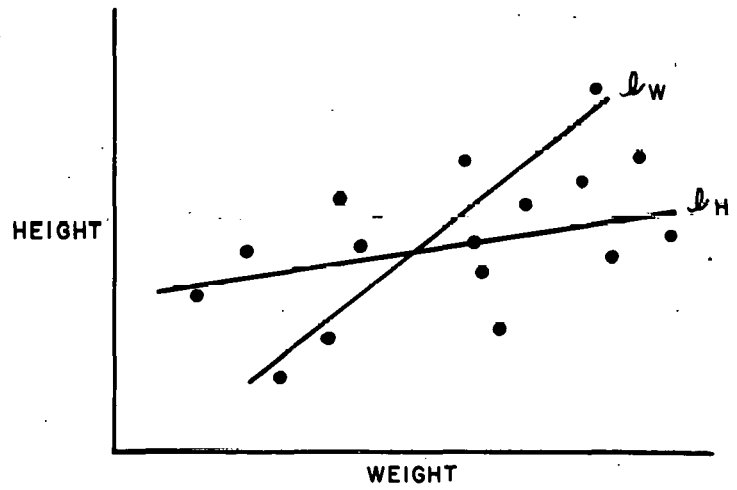


Figure 7.2: Height vs. Weight

individual's height from the fitted line ℓ_H obtained by assuming the weights to be known quantities,or predict one's weight from ℓ_W by assuming the heights to be known. In another situation a response may be quite complicated but, in a given region of interest to the experimenter, may be approximated satisfactorily by a simple model. For example, in Figure 7.3 a straight line could not possibly fit the response along the entire range of X. Divided into three segments, however, a different straight line may be a satisfactory

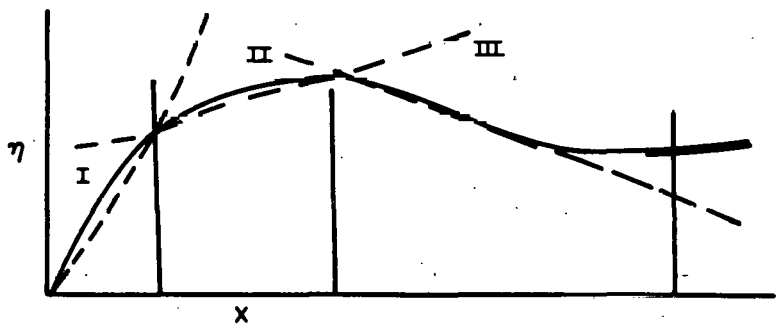


Figure 7.3: Straight line fits to Segments

representation of the response for each separate segment for some purposes of investigation. This illustration shows dramatically the dangers in extrapolating empirical models beyond the region in which data was obtained. Following any of the lines in Figure 7.3 beyond the boundary line will result in extremely large deviations from the true response curve. Part of correct analytical procedure is to provide tools to detect such departures.

In this section we shall discuss only linear models. By linear, it is meant linear in the parameters. A straight line or plane is linear not only in the parameters, β's, but in the variables, X's, as well.

Straight-line Model: $\eta = \beta_0 + \beta_1 X_1$

Planar Model: $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Models involving linear terms of variables we shall call first order models. A second order model contains quadratic terms in the X's, but is still linear in the parameters.

Quadratic or Second Order Liner Model $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$

Some non-linear models are intrinsically linear. That is, they can be converted to a linear model by a simple transformation; e.g.,

Intrinsically Linear Model

$$\eta = \alpha_0 V^{\alpha_1} f^{\alpha_2} d^{\alpha_3} A^{\alpha_4} R^{\alpha_5}$$

$$\ell n \eta = \ell n (\alpha)_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5$$

where $X_1 = \ell n(V)$, $X_2 = \ell n(f)$, $X_3 = \ell n(d)$, $X_4 = \ell n(A)$ and $X_5 = \ell n(R)$

Finally, of course, there are intrinsically non-linear models such as those which typically describe chemical rections. These models are not easily dealt with and usually require iterative procedures to obtain the best estimates of the parameters. Computer programs are available for analysis of such models.

Intrinsically Non-linear Model

$$\eta = 1 + \frac{-\theta_2 e^{-\theta_1 t} + \theta_1 e^{-\theta_2 t}}{\theta_2 - \theta_1}$$

7.1.2 The Principle of Least Squares

The analysis technique we will apply to fitting linear models is called the least squares procedure. Since all observations are subject to error, we have the model

$$y_u = \eta_u + \epsilon_u$$

where y_u is the uth observation and η_u is the value given by the hypothesized model for the settings of the independent X variables on the uth trial. The ϵ_u's are the errors involved. The least squares principle is to estimate the parameters of the model

$$\eta_u = f(\beta_1, \beta_2, \ldots \beta_k | X_1, X_2, \ldots, Xp)$$

by choosing the β's to minimize the sum of squares of the errors, i.e., choose the β's so as

$$\sum_{u=1}^{N} \epsilon_u^2 = \sum_{u=1}^{N} (y_u - \eta_u)^2$$

is a minimum.

If we further assume that these errors ϵ_u are independently and identically distributed as normal variates with mean 0 and variance σ^2, the least square estimates are also the maximum likelihood estimates (see Appendix B) of the parameters. Furthermore, if we assume that the X's are given as fixed quantities we may write the model as

$$\eta = E(y | \{\beta\}, \{X\}) .$$

The expected value of y given the X's is called the regression function. Thus, assuming any distribution for the errors, the least squares procedure is also called regression analysis whenever we consider the X's known and fixed. (The misnomer "regression" apparently came into being when an early statistical investigator studying the relative heights of fathers and sons discovered that tall fathers tended to have shorter sons and short fathers tended to have taller sons. He called this a "regressive" tendency and hence the term "regression" came into being.)

In general we shall assume the errors of observation are normally distributed and hence the curve-fitting estimation procedures of maximum-likelihood, regression and least squares are identical. We now turn to illustrating the procedure for the simple case of fitting a straight line.

7.2 Fitting a Straight Line

We suppose now that a straight line is an adequate representation of the response of interest in the region in which we are dealing. We want to (1) obtain the best straight line to fit our data, and (2) evaluate the adequacy of the straight line model. The model is

$$y_u = \beta_0 + \beta_1 X_u + \epsilon_u$$

where we assume $\epsilon_u \sim N(0, \sigma^2)$, $u = 1, 2, \ldots, N$. We want to select β_0 and β_1 such that

$$S = \sum_{u=1}^{N} \epsilon_u^2 = \sum_{u=1}^{N} (y_u - \beta_o - \beta_1 X_u)^2$$

is a minimum. That is, the deviations of the observations from the fitted or predicted line are minimum in the above sense.

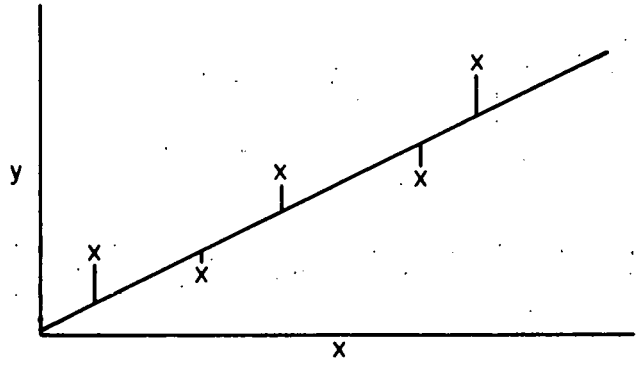


Figure 7.4. Straight Line Model

7.2.1 Analysis

We proceed to find the estimates of β_0 and β_1 by determining the normal equations by differentiating S by β_0 and β_1 and setting equal to zero.

1. $$\frac{\partial S}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \Sigma(y_u - \beta_0 - \beta_1 X_u)^2 = -2\Sigma(y - \beta_0 - \beta_1 X_u) = 0$$

$$\Sigma(\beta_0 + \beta_1 X_u) = \Sigma y_u$$

$$N\beta_0 + \beta_1 \Sigma X_u = \Sigma y_u$$

2. $$\frac{\partial S}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} (y_u - \beta_0 - \beta_1 X_u)^2 = -2\Sigma X_u (y_u - \beta_0 - \beta_1 X_u) = 0$$

$$\Sigma(\beta_0 X_u + \beta_1 X_u^2) = \Sigma X_u y_u$$

$$\beta_0 \Sigma X_u + \beta_1 \Sigma X_u^2 = \Sigma X_u y_u$$

The two equations

$$1.\quad N\beta_0 + \beta_1 \Sigma X_u = \Sigma y_u$$

$$2.\quad \beta_0 \Sigma X_u + \beta_1 \Sigma X_u^2 = \Sigma X_u y_u$$

are the normal equations of this system and may be solved simultaneously for β_0 and β_1. The resulting estimates, b_0 and b_1, are

$$1.\quad b_0 = \bar{y} - b_1 \bar{X}$$

$$2.\quad (\bar{y} - b_1 \bar{X}) \Sigma X_u - b_1 \Sigma X_u^2 = \Sigma X_u y_u$$

$$b_1 = \frac{\Sigma X_u y_u - N \bar{X} \bar{y}}{\Sigma X_u^2 - N \bar{X}^2}, \text{ where } N\bar{X} = \Sigma X_u$$

$$= \frac{\Sigma (X_u - \bar{X})(y_u - \bar{y})}{\Sigma (X_u - \bar{X})^2}$$

where

$$\Sigma X_u y_u - N\bar{X}\bar{y} = \Sigma (X_u - \bar{X})(y_u - \bar{y}) \text{ and}$$

$$\Sigma X_u^2 - N\bar{X}^2 = \Sigma (X_u - \bar{X})^2.$$

The computation can be made easier if we design the experiment so that the average of the X's is zero. Equivalently, consider the model

$$y = \alpha + \beta x_u + \epsilon_u$$

where $\beta_1 = \beta$, $\alpha = \beta_0 + \beta_1 \bar{X}$, and $x_u = X_u - \bar{X}$. Then $\Sigma x_u = 0$ $(\bar{x} = 0)$. The estimates are now

$$a = \bar{y} = b_0 + b_1 \bar{X}, \quad b_0 = y - b_1 \bar{X}$$

$$b = \frac{\Sigma_u x_u y_u}{\Sigma_u x_u^2} = b_1$$

These estimates are in fact linear unbiased estimates of the parameters and have minimum variances of all such linear unbiased estimates of α and β. The variance of b is in fact

$$\text{Var}(b) = E\left[\frac{\Sigma x_u y_u}{\Sigma x_u^2} - \beta\right]^2$$

$$= \frac{\Sigma x_u^2}{\left(\Sigma x_u^2\right)^2} \text{Var}(y_u) = \sigma^2 \frac{1}{\Sigma_u x_u^2} = \frac{\sigma^2}{\Sigma_u (X_u - \bar{X})^2}.$$

where $\sigma^2 = \text{Var}(\epsilon_u)$. For b_o

$$\text{Var}(b_o) = \sigma^2 \left(\frac{1}{N} + \frac{\bar{X}^2}{\Sigma (X_u - \bar{X})^2} \right) .$$

If we assume $\epsilon_u \sim N(0, \sigma^2)$ and using the fact that linear combinations of normal variables are distributed normally, we may obtain confidence intervals for β_o and β_1:

$$\beta_o: \quad \left[b_o \pm z\sigma \left(\frac{1}{N} + \frac{\bar{X}^2}{\Sigma (X_u - \bar{X})^2} \right)^{1/2} \right]$$

$$\text{and} \quad \beta_1: \left[b_1 \pm z\sigma \left(\frac{1}{\Sigma (X_u - \bar{X})^2} \right)^{1/2} \right]$$

However, we usually do not know σ^2 but must estimate it from the data. The estimate of the variance is obtained from the residuals

$$r_u = y_u - \hat{y}_u$$

where $y_u = b_o + b_1 X_u$ are the fitted or predicted value of the model. That is

$$\hat{\sigma}^2 = \sum_{u=1}^{N} (y_u - \hat{y}_u)^2 / (N - 2)$$

where the divisor N - 2 are the degrees of freedom for the estimate of variance in which two parameters β_o and β_1 were already estimated from the data. This is the usual unbiased estimate of variance s^2. The interval statements for β_o and β_1 are now

$$\beta_o: \left[b_o \pm t_{N-2}\, s \left(\frac{1}{N} + \frac{\bar{X}^2}{\Sigma (X_u - \bar{X})^2} \right)^{1/2} \right]$$

$$\text{and} \quad \beta_1: \left[b_1 \pm t_{N-2}\, s \left(\frac{1}{\Sigma (X_u - \bar{X})^2} \right)^{1/2} \right]$$

Statements about the significance of the parameter can be made using these estimates. If the intervals evaluated at some confidence level γ include the value zero, we say that the data does not contradict the assumption that the parameter is negligible.

Another way of examining the significance of the parameters being estimated is by the ANOVA table. The ANOVA table is constructed as follows.

ANOVA Table: Model $y_u = \beta_0 + \beta_1 X_u + \epsilon_u$

Source	Sum of Squares	Degrees of Freedom	Mean Square
Total	$\sum_{u=1}^{N} y_u^2$	N	
Regression	Σy_u^2	2	Reg SS/2
Mean	$(\Sigma y_u)^2/N$	1	
Slope (β_1)	$b_1 \sum_u (X_u - \bar{X})y_u$	1	
Residuals	$\Sigma(y_u - \hat{y}_u)^2$	N - 2	s^2=Res SS/(N-2)

The significance of the overall model can be determined by the F-test

$$F_{2,\ N-2} = \frac{\text{Reg SS}/2}{\text{Res SS}/(N-2)}$$

In other words, if the model $\beta_0 + \beta_1 X_u$ accounts for a significant proportion of the variation in the data, then this ratio will be larger than the critical value of an $F_{2,\ N-2}$ distribution at the specified confidence level. This will usually be the case since the experimenter expects something to be controlling the response besides random error, or else he would not bother to propose the model. The regression model in this case can be separated into the two component parts, mean and slope, and hence each of these parameters can be tested separately for significance, either by the t-interval approach or by the appropriate $F_{1,\ N-2}$ test suggested by the ANOVA table. For straight line models usually only the slope term is of interest since, with the exception of a line thought to pass through the original, the y-intercept is almost always significant.

Given that the straight line model is found to be significant, i.e., both the slope and intercept parameters contribute significantly to the explanation of the variation of the data, we can next calculate confidence intervals, tolerance intervals and prediction intervals for the fitted line. The variance of a fitted value is

$$\begin{aligned} \text{Var}(y_u) &= \text{Var}\,(\, b_0 + b_1 X_u) \\ &= \text{Var}(\bar{y} + b\,(X_u - \bar{X})) \\ &= \sigma^2 \left(\frac{1}{N} + \frac{(X_u - \bar{X})^2}{\Sigma (X_u - \bar{X})^2} \right) \end{aligned}$$

and is estimted by replacing σ^2 by s^2. Thus, a 95% confidence interval for the mean value of the fitted line at a point X_k is

$$\hat{y}_k \pm t_{N-2,\,0.025}\, s\left(\frac{1}{N} + \frac{(X_k - \bar{X})^2}{\Sigma(X_u - \bar{X})^2}\right)^{1/2}$$

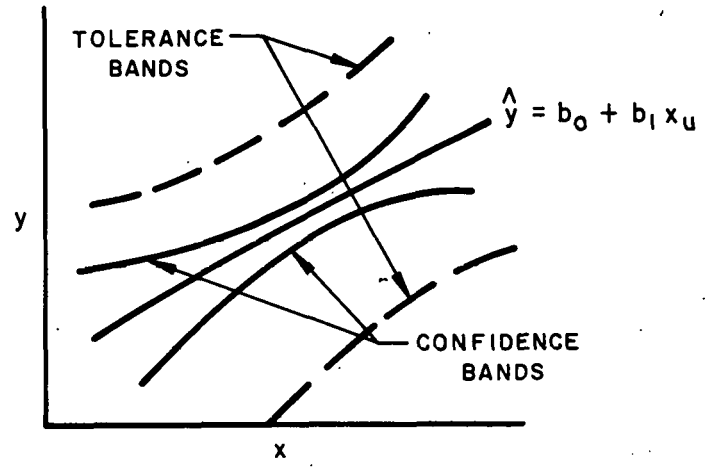


Figure 7.5. Confidence and Tolerance Bands for a Straight Line Fit

Note that the confidence bands diverge, being most narrow at $X_u = \bar{X}$.

This reflects that our knowledge of the model decreases as we move away from the center of the experimental region. The true straight line relationship (assuming it is appropriate) lies between the confidence lines. Extrapolating the fitted line much beyond the experimental region could lead to large departures from the true line.

A γ/P tolerance interval at $X = X_k$ is constructed much as before, as

$$\hat{y}_k \pm K\, s$$

only K is obtained by a complex approximation formula

$$K = z_{(1+P)/2}\left(1 + \frac{\hat{Var}(\hat{y}_k)}{2} - \frac{\hat{Var}(\hat{y}_k)^2(2z_{(1+P)/2}{}^2 - 3)}{24}\right)F^{1/2}_{\infty,\,N-2,\,1-\gamma}$$

where $\hat{Var}(\hat{y}_K)$ is an estimate of the variance of $\hat{y}_K$.

The interpretation is that at each point along the line 100P% of the population of y_k at $X = X_k$ will fall within the interval with 100γ% confidence.

These intervals are bands like the confidence bands for the mean value of the line, only much wider.

A prediction interval for a single future observation of y at $X = X_k$ can be obtained by simply recognizing that a future observation will be equal to the predicted value plus a random error,

$$y_{future} = \hat{y}_k + \epsilon$$

and its variance is

$$\text{Var}\,(y_{future}) = \text{Var}(\hat{y}_k) + \sigma^2 \quad .$$

Thus, a $100(1-\alpha)\%$ prediction interval is

$$\hat{y}_k \pm t_{N-2,\,\alpha/2}\; s \left(1 + \frac{1}{N} + \frac{(X_k - \bar{X})^2}{\Sigma (X_u - \bar{X})^2}\right)^{1/2} \quad .$$

For the mean of q future observations, a prediction interval is

$$\bar{y}_{q,\,future} - \hat{y}_k \pm t_{N-2,\alpha/2}\; s\left(\frac{1}{q} + \frac{1}{N} + \frac{(X_k - \bar{X})^2}{\Sigma (X_u - \bar{X})^2}\right)^{1/2}$$

7.2.2 Interpretation and Diagnostic Checks

We can test for the significance of the intercept or slope parameter of the fitted line and produce confidence bands, tolerane bands or prediction bands about the line provided the model under consideration and its underlying assumptions are correct. The least squares procedure provides the best fitted equation for the type of model examined, but it does not assure that the model is correct or appropriate! The question of significance of parameters β_0 and β_1 is irrelevant and the corresponding bands inappropriate if a line is a wrong model to fit to the data. There are several methods of detecting model inadequacies, but first let us identify the ways that the model may fail.

The complete representation of the model is

$$y = \eta\,(\underline{x},\, \underline{\beta}\,) + \epsilon$$

where η is the true response and is a function of a set of one or more input variables, $\underline{x}$, and a set of coefficients or parameters β. The observations are designated by y and differ from the true value because of random error. The errors are usually assumed to be independently and normally distributed with mean 0 and common variance σ^2. If this holds, then the observations themselves are $N(0, \sigma^2)$. The four assumptions involved, then, are

1. Normal distribution
2. Independence
3. Common variance
4. The function η is of the correct form

The function η may be incorrect due to wrong functions of the input variables or due to the absence of terms involving other variables not previously considered.

We shall first discuss some procedures for plotting the residuals from the analysis which have been found to be very helpful in diagnosing departures from the above assumptions and the reasons for these departures.

A. Residual Plots

1. Normality assumption

a) Histogram - A histogram of the resulting residuals may be plotted. This histogram should appear to have come from a normal distribution. There are two warnings to be made, however. For few observations, a histogram may appear quite scattered and non-normal, but may in fact be an adequate representation of a normal distribution. Secondly, although there are N residuals, there are only N-2 independent values (i.e., degrees of freedom), since two parameters were estimated from the data. This constraint on the residuals has the effect of bunching the data closer to the origin. The problem becomes much more acute as the degrees of freedom for residuals gets small compared to N. In both cases experience in the interpretation of these plots is important to avoid errors.

b) Cumulative distribution plot on probability paper

In some computer programs the cumulative distribution of the residuals are plotted on a scale (probability) so that if the assumption of normality is appropriate, a straight line may be drawn through the points. Departures from a straight line are indicative of departures from normality. Figure 7.6 illustrates typical histogram and cumulative probability plots.

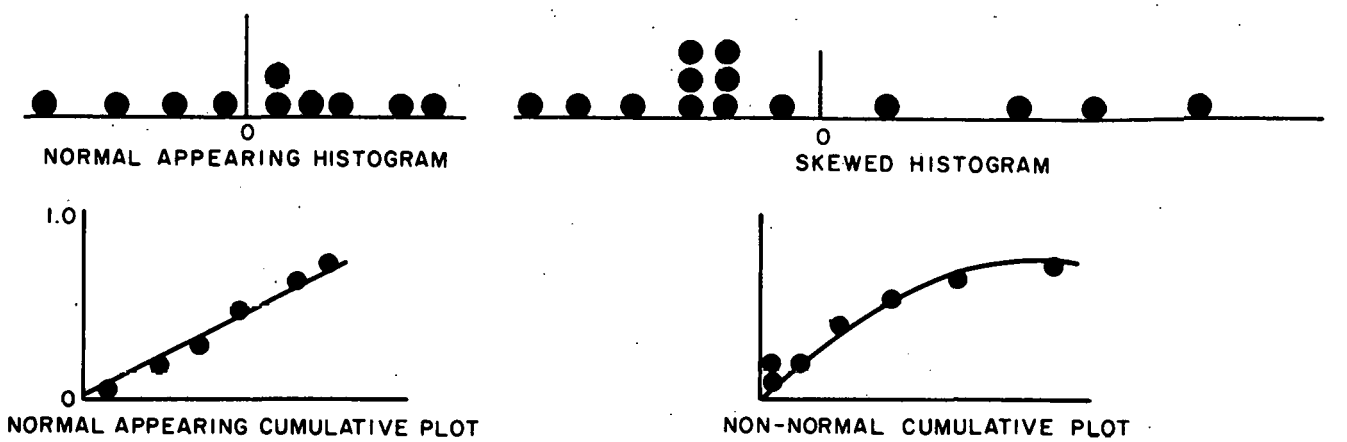


Figure 7.6. Plots for Checking Normality Assumption

2. Independence

We can check the residuals for a time dependency by plotting the residuals as a function of the order in which they were observed. Figure 7.7 illustrates such a plot. What we would like to see is that two lines horizontal to the time axis can be drawn which include all or most of the residuals. The conclusion is then that there is no apparent time trend present.

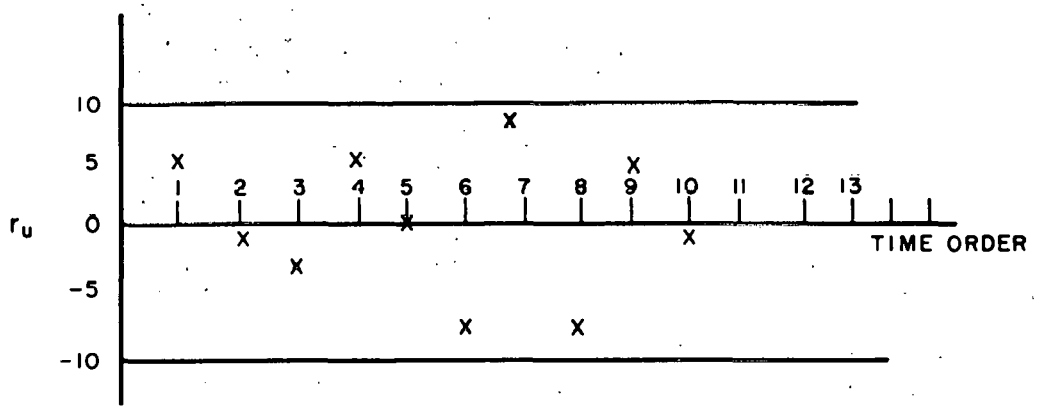


Figure 7.7. Plot of Residuals r_u vs Time Order

Figure 7.8 shows other types of bands which are drawn to include most of the data which are indicative of departure from the models. When the horizontal axis is considered to be time, then the funnel shaped band (1) is indicative that the variation or spread of the residuals is increasing with time. That is, the variance of the errors is apparently increasing as the process continues. The parallel but diagonal lines (2) indicate a linear trend in occurring in time. The curved bands (3) are indicative of a need for a quadratic term involving time.

3. Homogeneity of Variance

It may happen that the variance of the observations changes with the size of the response, thus violating the homogeneity or equality of variance assumption. This is best examined by plotting the residuals against the fitted values $\hat{y}$. The fitted values, $\hat{y}_u$, are used instead of the observations because the residuals $r_u = y_u - \hat{y}_u$ and $\hat{y}_u$ are statistically independent of each other. Again, horizontal bands are desired.

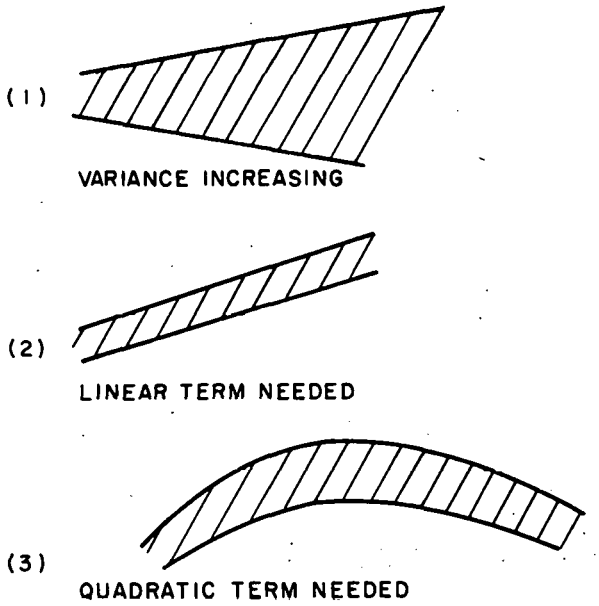


Figure 7.8. Residual Plots: Diagnostic Bands

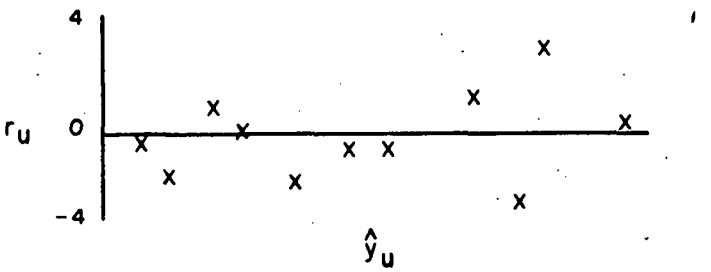


Figure 7.9. Residuals vs Fitted Values

When the horizontal axis in Figure 7.8 is considered to be $\hat{y}$, (1) indicates increasing variance and the need for a weighted least squares procedure or transformation on the observations (for transformations and weighted least squares, see Section 7.4)• Figures 7.8(2) and (3) would indicate that the model is inadequate. (2) may be caused by wrongly omitting a linear term from the model. (3) may be caused by omission of quadratic or cross-product terms.

4. Other Plots

Other residual plots may be warranted, depending on the model. If more than one x variable is in the model, a plot of r_u against each of the x's is advisable to determine omissions or departures from the model due to each variable.

Also plotting residuals within each block or treatment may provide clues to departures from assumptions that are not detected by the analysis procedure.

B. The Lack-of-Fit Test

A quantitative test may be performed to test for model inadequacy also. There are two conditions under which this lack of fit test may be performed. First, if a prior estimate of variance $\hat{\sigma}_p^2$ is available, the variance estimate s^2 resulting from the present data may be compared to it. If the model is adequate to represent the data, the F-ratio $s^2/\hat{\sigma}_p^2$ should result in an non-significant value, i.e.,

$$F_{\nu_1, \nu_2} < F_{\nu_1, \nu_2, \alpha}$$

where ν_1 are the degrees of freedom for residuals from the analysis and ν_2 are the degrees of freedom in $\hat{\sigma}_p^2$. The drawback to this test is that a sufficiently reliable prior estimate of σ^2 is rarely available. In addition, care must be taken to ensure that the data is obtained under the same conditions as was the estimate $\hat{\sigma}_p^2$. This is often extremely difficult to assure.

The second condition under which a lack-of-fit test may be performed is that replicate observations be made in the current experiment. If observations are repeated under the same conditions and same x settings, then the only explanation for differences in the observations is pure random error. The sum of squares of these repeat points about their average may be used to estimate the variance σ_{pe}^2 of this pure error.

Let y_{ij} be the j_{th} of n_i observations at the ith setting or condition. Then the estimate of variance from these n_i observations is

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1}$$

Pooling the estimates s_i^2 over all sets of repeat points, we obtain

$$s^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{\sum_{i=1}^{k} (n_i - 1)}$$

$$= \text{SS p.e.} / \nu_e$$

where $\nu_e = \sum_{i=1}^{k} (n_i - 1) = \Sigma n_i - k.$

The residual sum of squares from the ANOVA table, however, contains the information about the repeated points in addition to other unexplained variation. The residuals can be broken up, then, into two independent parts:

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) - (\hat{y}_i - \bar{y}_i) \quad .$$

The first term we have already identified as the pure error estimate. The latter term, then, represents the discrepancy between the averages of repeated points and the fitted model itself. The sum of squares of the lack of fit term estimates $\sigma^2_{p.e.}$ plus the size of the departure from the model. The sum of squares for lack of fit and the degrees of freedom are obtained by subtraction.

$$SS_{lof} = SS_{residuals} - SS_{p.e.}$$

$$\nu_{lof} = N - 2 - \nu_{p.e.}$$

Thus $s^2_{lof} = \frac{SS_{lof}}{\nu_{lof}}$

and $F_{\nu_{lof}, \nu_{p.e.}} = \frac{s^2_{lof}}{s^2_{p.c.}}$

tests the assumption that the model is adequate to fit the data. If the calculated F-ratio is larger than $F_{\nu_{lof}, \nu_{p.e.},}$, then we may take this to mean that the data is testifying to the fact that the model does not adequately fit the data. Since rejecting a hypothesized model is usually a serious consequence, some statisticians suggest that the model not be declared inadequate unless the F-ratio is twice the critical value. This procedure is arbitrary but nevertheless points out the feeling that we should not be too anxious to reject a hypothesized model unless a good alternative is available.

C. Shapiro-Wilk Statistic for Non-Normality

For $n \leq 50$ observations a test for non-normality of the distribution of the residuals can be made. This test, known as the Shapiro-Wilk Test [33], compares a distribution-free estimate of the variance of the residuals against the usual sum of squared deviations from the mean. Specifically, let

$$b = \sum_{i=1}^{k} a_{n-i+1}(r_{(n-i+1)} - r_{(i)}), \quad k = \begin{cases} \frac{n}{2}, & n \text{ even} \\ \frac{n-1}{2}, & n \text{ odd} \end{cases}.$$

Each difference $r_{(n-i+1)} - r_{(i)}$ is a range estimate of the standard deviation σ, where $r_{(j)}$ is the jth ordered residual and the a's are constants (see [33]). The test, then, is to compare b^2 against $S^2 = \Sigma(r_i - \bar{r})^2$, $(s^2 = S^2/(n-2))$. If the ratio

$$W = \frac{b^2}{S^2}$$

is quite small, the residuals are showing evidence of non-normality. Only values which are significant at a small error level (e.g., $\alpha = 0.01$) are considered important.

D. Coefficient of Determination

Another measuring tool to use to evaluate the usefulness of a model is the coefficient of determination, also known as the squared correlation coefficient. The parameters of a model, the y-intercept and slope under consideration here, may be judged to be significant even if the variation is high. A question of interest is how much of the variability is accounted for by the model? The coefficient of determination, R^2, measures the proportion of total variation about the average $\bar{y}$ that is explained by the model. Functionally,

$$R^2 = \frac{\sum_u (\hat{y}_u - \bar{y})^2}{\sum_u (y_u - \bar{y})^2}.$$

In general the larger R^2 the better. However, R^2 can be made artifically large by adding parameters to the model which are not really significant. In fact, if we fit N parameters to N distinct data points, R^2 will be 1, obviously an artificial result. However, when no or few replicated observations appear in the data and the number of parameters is small compared to the number of distinct data points, R^2 serves as a useful tool for evaluating the model. It is especially useful when trying to evaluate the worth of an extra parameter under consideration for inclusion in the model. This will be discussed further in the section about model building.

All of these diagnostic tools should be used together whenever possible. The lack-of-fit test indicates an inadequate model but does not tell you why it is inadequate. The residual plots are very useful for this purpose. In fact, the residual plots are useful even when a lack-of-fit test is not available.

7.2.3 Summary

Evaluation of a Straight Line Model

I. Objective - To determine the straight line which best fits the data for the purpose of predicting the response in a specified region of interest and to examine this fitted model for inadequacies.

II. Assumptions

A. Model: $y_u = \beta_0 + \beta_1 X_u + \epsilon_u$, $u = 1, 2, \ldots, n$,

where

y_u = the observed value of the dependent variable on the uth run,

β_0 = the true y-intercept of the line,

β_1 = the true slope of the line,

X_u = the value of the independent variable on the uth run,

and ϵ_u = the random error associated with observing y. The errors are assumed to be independent with a common mean of zero and a common variance σ^2.

Alternative Model:

$$y_u = \alpha + \beta (X_u - \bar{X}) + \epsilon_u$$

where

$$\alpha = \beta_0 + \beta_1 \bar{X}, \quad \bar{X} = \sum_{u=1}^{n} X_u/n$$

$$\beta_1 = \beta$$

B. Procedure - Least Squares Principle

$$\text{Minimize } S = \sum_{u=1}^{n} (y_u - \alpha - \beta (X_u - \bar{X}))^2$$

III. Computation

A. Estimate of Slope Parameter

$$b = b_1 = \sum_u (X_u - \bar{X})(y_u - \bar{y}) / \sum_u (X_u - \bar{X})^2$$

B. Estimate of y-intercept

$$a = \bar{y}$$

$$b_0 = a - b_1 \bar{X}$$

C. Fitted or Predicted Values

$$\hat{y}_u = a + b (X_u - \bar{X}) = b_0 + b_1 X_u$$

D. Estimate of Variance σ^2

$$s^2 = \sum_u (y_u - \hat{y}_u)^2/(n - 2)$$

E. The Analysis of Variance Table

1. Total Sum of Squares: $\sum_u y_u^2$

This is a measure of the total information contained in the data.

2. Regression Sum of Squares: $\sum_u \hat{y}_u^2$

This is a measure of the amount of information in the data explained by the model. For a straight line model, the regression sum of squares can be separated into two parts.

a. Sum of squares due to the presence of a mean: SS(a)

b. Sum of squares due to the slope: SS(b)

3. Residual Sum of Squares: $\sum_u (y_u - \hat{y}_u)^2$

This is a measure of the variation in the data which is unexplained by the model; i.e., it is the amount of variation attributable to chance and, perhaps, an inadequate model.

IV. Evaluation

A. F-test for the Slope Parameter

Assuming now that the errors ϵ_u are normally distributed, the statistical significance of the slope parameter of the fitted straight line can be tested. If the calculated value is significantly large, this is evidence that the slope is non-zero.

$$F = \frac{SS(b)}{s^2}$$, with 1 and n-2 degrees of freedom .

B. Coefficient of Determination (Squared Correlation Coefficient)

$$R^2 = \frac{\sum_u (\hat{y}_u - \bar{y})^2}{\sum_u (y_u - \bar{y})^2} .$$

This value measures the proportion of total variation about the mean of y that is explained by the model.

C. Confidence Interval for the Slope Parameter

The 100 $(1 - \alpha)$% confidence interval for β_1 is obtained from

$$b_1 \pm t_{n-2,\ 1-\alpha/2}\ s\ (b_1)$$

where $s(b)_1 = s/\left[\sum (X_u - \bar{X})^2\right]^{1/2}$.

If the confidence interval does not include the point $\beta_1 = 0$, then β_1 is said to be significantly different from zero.

D. Confidence Interval for the Predicted Mean Value of y at $X = X_k$

A 100 $(1 - \alpha)$% confidence band about the predicted line can be obtained from

$$\hat{y}_k \pm t_{n-2,\ 1-\alpha/2}\ s(\hat{y}_k)$$

where

$$s(\hat{y}_k) = s\left\{\frac{1}{n} + \frac{(X_k - \bar{X})^2}{(X_u - \bar{X})^2}\right\}^{1/2}$$

E. Tolerance Interval for the Population of y's at $X = X_k$

A tolerance band within which at least 100 P% of the population of y's at $X = X_k$ will lie can be constructed with
100$(1 - \alpha)$% confidence from

$$\hat{y}_k \pm K\ s$$

where K is a constant depending P, α, n, $\bar{X}$ and X_k.

V. Checks on the Assumptions

A. Test for Model Inadequacy

1. If a prior estimate $\hat{\sigma}^2_p$ of the variance σ^2 is available, compare s^2 against $\hat{\sigma}^2_p$.

2. If repeat observations are available, the residual sum of squares can be divided into two parts: (1) the pure error or replicates sum of squares measuring the variation within replicates and (2) the remaining portion of the residual sum of squares, called the lack-of-fit or linearity sum of squares, measures the variability about the fitted line only.

Compute:

i. $SSp.e. = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$,

where $\bar{y}_i$ is the average of n_i observations at X_i,

ii. $\nu_{p.e} = \sum_{i=1}^{k} n_i - k$,

iii. $s^2_{p.e.} = SS_{p.e.} / \nu_{p.e.}$

iv. $SS_{lof} = SS_{residuals} - SS_{p.e.}$

v. $\nu_{lof} = N - 2 - \nu_{p.e.}$

vi. $s^2_{lof} = SS_{lof} / \nu_{lof}$.

Test:

$$F_{lof} = \frac{s^2_{lof}}{s^2_{p.e.}} ,$$

with ν_{lof} and $\nu_{p.e.}$ degrees of freedom.

If the straight line model is inadequate, an excessive amount of variation about the fitted line will result,

i.e., $s_{lof}{}^2$ will be large compared to $s^2_{p.e.}$,

resulting in a significantly large F_{lof} value.

B. Shapiro-Wilk

The Shapiro-Wilk Statistic W for $n \leq 50$ observations indicates non-normality of the residuals (and hence the observations) if it is significant at a low Type I error level, e.g., $\alpha = 0.01$ or 0.05.

C. Residual Plots

The residuals $r_u = y_u - \hat{y}_u$ can be plotted in several ways to check on the assumptions of

i. Normal distribution

ii. Equality of Variance for all ϵ_u

iii. Independence of the Errors

Some typical plots are:

i. Normal probability plot

ii. r_u versus time sequence

iii. r_u versus $\hat{y}_u$, the fitted values

iv. r_u versus X_u

7.3 An Example

Suppose that a new assay gage is to be evaluated for determining the end-of-life fissile loading of LWBR rods. Let

$$y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij}.$$

where y_{ij} is the gage response, and X_i is the fissile loading of the ith rod obtained by destructive analysis.

The hypothetical data given in Table 7.2 show that 24 observations are provided for 13 rods. Seventeen gage responses represent repeated measurements from 6 rods, while seven other responses represent single observations of a rod. The data were analyzed utilizing the computations provided in Section 7.2.3. The analysis is summarized by use of a worksheet.

Table 7.2
Data for Assay Gage Example

EOL Fissile Loading by Destructive Analysis X	Assay Gage Response Y
8.00 gms f_U	50.49, 51.56
8.30	55.69, 56.44
8.50	57.17
8.55	57.80, 56.93
8.65	58.30
8.75	59.49
8.80	59.37, 60.60, 60.30 59.74, 59.88, 60.00
8.85	59.21, 59.92, 59.92
9.00	61.19
9.05	62.70, 62.85
9.15	59.88
9.20	63.23
9.45	61.28

Worksheet for Simple Linear Regression

Ind. Var. X w/o Total Uranium

Dep. Var. Y Assay Gage Values

$\Sigma X = 209.850$

$\Sigma Y = 1413.94$

$\bar{X} = 8.74375$ No. of observations $N = 24$ $\bar{Y} = 58.9142$

(1) $\Sigma XY = 12386.074$

(2) $(\Sigma X)(\Sigma Y)/N = 12363.13788$

(3) $S_{xy} = (1) - (2) = 22.93612$

(4) $\Sigma X^2 = 1837.6975$

(5) $(\Sigma X)^2/N = 1834.875938$

(6) $S_{xx} = (4) - (5) = 2.821562$

(7) $\Sigma Y^2 = 83521.6274$

(8) $(\Sigma Y)^2/N = 83301.09683$

(9) $S_{yy} = (7) - (8) = 220.53057$

(10) $b_1 = \frac{S_{xy}}{S_{xx}} = \frac{(3)}{(6)} = 8.1289$

(11) $\bar{Y} = 58.9142$

(12) $b_1\bar{X} = 71.0774$

(13) $b_o = \bar{Y} - b_1\bar{X} = (11) - (12) = -12.1629$

(14) $\frac{(S_{xy})^2}{S_{xx}} = \frac{(3)(3)}{(6)} = 186.4448$

(15) $(N-2)\, s_y^2 = (9) - (14) = 34.0858$

(16) $s_y^2 = (15) \div (N-2) = 1.5494$

(17) $s_y = \sqrt{(16)} = 1.245$

(18) $s_{b_1}^2 = \frac{s_y^2}{S_{xx}} = \frac{(16)}{(6)} = 0.5491$

(19) $s_{b_o}^2 = s_y^2 \left(\frac{1}{N} + \frac{\bar{X}^2}{S_{xx}}\right) = 42.0472$

$\hat{Y} = b_o + b_1 X = -12.163 + 8.129X$

$s_{b_o} = 6.484$

$s_{b_1} = 0.741$

$s_y = 1.245$

for $\nu = N - 2$, $t_{22,0.025} = 2.074$

for $X = X_k$, $\hat{Y}_k = b_o + b_1 X_k$

Conf. interval on η (line) for $X = X_k$:

$$\hat{Y}_k \pm t_{\nu,\alpha/2}\; s_y \left[\frac{1}{N} + \frac{(X_k - \bar{X})^2}{S_{xx}}\right]^{1/2}$$

Pred. interval on next observation for $X = X_k$:

$$\hat{Y}_k \pm t_{\nu,\alpha/2}\; s_y \left[1 + \frac{1}{N} + \frac{(X_k - \bar{X})^2}{S_{xx}}\right]^{1/2}$$

$R^2 = \frac{(14)}{(9)} = 0.85$

Note: $S_{xx} = \Sigma(X-\bar{X})^2$ $S_{yy} = \Sigma(Y-\bar{Y})^2$

$S = \Sigma(X-\bar{X})(Y-\bar{Y})$

The fitted model is

$$\hat{y} = -12.163 + 8.129\ X$$

and the ANOVA table is

Table 7.3
ANOVA Table

Source	SS	df	MS	F-Ratio
Intercept (Mean y)	83301.10	1		
Slope, β_1	186.44	1	186.44	120.34
Residuals	34.09	22	1.55	
Lack of Fit	31.58	11	2.87	12.6
Pure Error	2.51	11	0.23	

$R^2 = 0.85$

Using the residual mean square as our estimate of σ^2, a 95% confidence interval for β_1 is $0.66 < \beta_1 < 0.97$. Thus, β_1 is certainly significantly non-zero. The F-ratio of 120.34 testifies to this conclusion also. A 95% confidence band for selected X values is listed below in Table 7.4 along with 95/99 tolerance bands at these points.

Table 7.4
Confidence and Tolerance for New Assay Gage

		95% Confidence		95/99 Tolerance	
X	y	lower	upper	lower	upper
8.000	52.87	51.61	54.13	48.18	57.55
8.362	55.82	55.03	56.60	51.35	60.28
8.725	58.76	58.23	59.29	54.39	63.13
9.087	61.71	60.96	62.45	57.26	66.16
9.450	64.66	63.45	65.86	59.99	69.32

The center of the region of experimentation is $\bar{X} = 8.74375$. Note that the confidence and tolerance intervals become wider as X moves away from $\bar{X}$. Our information about a system is best at the center of the region of experimentation and gets progressively worse as we move away from the center.

Of course, the above intervals and tests of significance are of no value if the model is inadequate. We have several ways to study this question. First note that the lack-of-fit test gives an F-ratio of 12.60.

The critical F for 11 degrees of freedom in both numerator and denominator is 2.82 for $\alpha = 0.05$. This is strong evidence of lack of fit. Furthermore, the cumulative distribution plot of residuals on a probability scale shows marked departures from a straight line. The bow in the middle of the data may be indicative of a need for a quadratic term in the model. (See Figure 7.10). A plot of residuals against the fitted values, or equivalently for a straight line model against the X values, reinforces this conclusion. Figure 7.11 shows large negative residuals at each end of the scale for X and predominantly positive residuals in the middle. A tentative conclusion is that a X^2 term would be beneficial in the model. Of course, other information about the processes under study may dictate other conclusions. Nevertheless, the data has witnessed to the apparent inadequacy of the straight line model and the residual plots suggest a remedy in terms of a quadratic model.

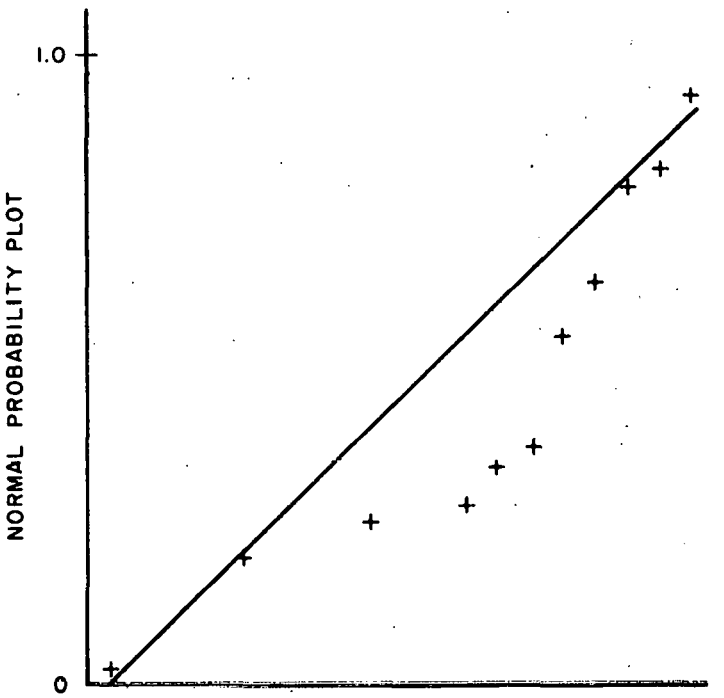


Figure 7.10. Cumulative Distribution of Residuals

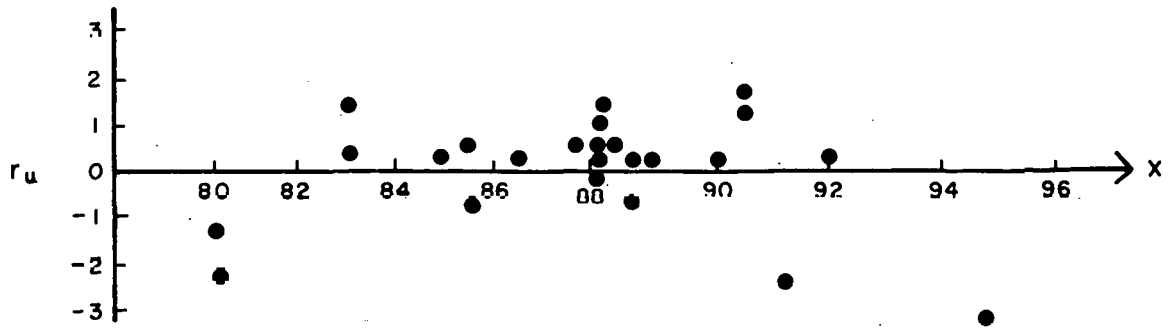


Figure 7.11. Residuals versus X

7.4 Transformation on the Dependent Variable

In some instances a transformation of the response variable is required to meet the assumptions of normality and equality of variance. That is, if it is known that the distribution of observations is skewed or if the variance increases with increasing response,then a properly chosen function of y will result in the distribution of the transformed variable being more nearly normal and the variance more nearly constant over the range of experimentation.

Briefly, the most common transformations are:

1) ℓog y or ℓn y: If $Var(y_j)$ is proportional to the square of the mean, η_j^2, a common log transformation, ℓog, or a natural log, ℓn, is applicable; i.e.,

   i.e., $Var(y_j) \alpha \eta_j^2$, use $y' = \ell og\ y$ or $y' = \ell n\ y$.

2) Square Root, $\sqrt{y}$: If the variance of y_j is proportional to the mean η_j, the square root of y should be used; i.e.,

   $Var(y_j) \alpha \eta_j$, use $y' = \sqrt{y}$.

3) Inverse If the variance of y_j increases faster than proportionally with the mean η_j, the reciprocal or inverse function is applicable;

   i.e., $Var(y_j) > \eta_j$, use $y' = 1/y$.

4) Arcsin: If y is the number of observed successes out of n events (i.e., y/n is the proportion of successes) and y/n is near 0 or 1, then an appropriate transformation to reduce non-normality is $z = \arcsin(\sqrt{y/n})$.

In general, if the plot of y versus x is curved, some sort of transformation is required to straighten the line. However, if the line turns over, usually a quadratic function is required to fit the data rather than a log or $\sqrt{}$ or inverse function.

Another procedure to handle heterogeneity of variances is the weighted least squares approach. Suppose $\mathrm{Var}(y_j) \neq \sigma^2$ but is a function of y or x. Let $w_j = 1/\mathrm{Var}(y_j)$. Then the least squares estimation of the parameters of the model, η_j, is to minimize

$$S = \sum_{j=1} w_j(y_j - \eta_j)^2 .$$

7.5 Multiple Linear Regression

Consider the linear model

$$y_u = \beta_0 + \beta_1 X_{1_u} + \beta_2 X_{2_u} \cdots + \beta_k X_{k_u} + \epsilon_u$$

where $\epsilon_u \sim N(0, \sigma^2)$ for all u. The variables X_{ju}, j = 1, 2, ..., k may represent different factors under investigation or functions of one or more of the factors in the experiment. The basic tools for fitting this general linear model have already been discussed. The general approach is to account for as much of the total variation in the data which can be accounted for by the estimation of the coefficients of the variables in the model and the constant term, β_0. The regression sum of squares $\Sigma \hat{y}_u^2$, is the measure of the variability associated with the model, and the residual sum of squares represents what is unexplained.

ANOVA Table

Source	SS	df	MS
Total	Σy_u^2	N	
Regression	$\Sigma \hat{y}^2_u$	k + 1	MS Reg.
Residuals	$\Sigma(y_u - \hat{y}_u)^2$	N-k-1	MS Res.

The F-ratio, MS Reg/MS Res, can be compared to an $F_{k+1,\,N-k-1}$ value to determine the significance of the model. The residuals may still be partitioned into pure error and lack of fit components, if replicate points are available.

In general the estimate of the parameters and the corresponding sum of squares are not easily attained by the usual procedures given in Section 7.2. A terse matrix approach to the least squares analysis will be given later. First, however, it is useful to note that the analysis of a multivariable model can be performed quite easily under the appropriate conditions.

We restrict our discussion here to k different factors X_j, and do not allow higher order terms. The model is the same as given above. If the design, that is the settings of the X factors, is orthogonal, the analysis is a simple extension of the straight line problem. By orthogonal we mean that

$$\sum_u X_{iu} X_{ju} = 0 \text{ for all } i \neq j.$$

Under this condition all the estimates in the model are independent and

$$b_j = \hat{\beta}_j = \frac{\sum_{u=1}^{N} (X_{ju}-\bar{X})(y_u-\bar{y})}{\sum_u (X_{ju}-\bar{X})^2}$$

$$b_o = \hat{\beta}_o = \bar{y} - \sum_j b_j \bar{X}_j$$

The regression sum of squares can be best calculated by

$$\text{Reg SS} = \Sigma \hat{y}_u^2 = b_o \sum_u y_u + \sum_{ju} \sum b_j X_{ju} y_u \ .$$

The residual sum of squares is the difference between the total sum of squares and the regression sum of squares. The estimate of σ^2 is then

$$s^2 = \frac{\Sigma y_u^2 - \Sigma \hat{y}_u^2}{N-k-1} \ .$$

Because of the independence of the estimates, however, each term b_j can be examined for significance independently by

$$F_{1,N-k-1} = \frac{b_j \sum_{ju} X_{ju} y_u}{s^2}$$

or by

$$t_{N-k-1} = \frac{b_j}{s \sqrt{\sum_u (X_{ju} - \bar{X})^2}}$$

In fact these are equivalent since $t^2_{N-k-1} = F_{1,N-k-1}$. This may be seen by noting that $b_j = \Sigma(X_{ju} - \bar{X})(y_u - \bar{y}) / \sum_u (X_{ju} - \bar{X})^2$. Thus, separate confidence intervals may be computed for each parameter. This cannot be done if the design is not orthogonal since the estimates are not then independent of each other.

There is one further principle which we will present next, that enables us to break out the effect of each factor from the others, even if it is a function of other X terms. This principle is known as the extra sum of squares principle.

7.5.1 The Extra-Sum-of-Squares Principle

Suppose we have two factors in a linear model of some response,

$$y_u = \beta_0 + \beta_1 X_{1u} + \beta_2 X_{2u} + \epsilon_u .$$

If the settings of X_1 and X_2 were not chosen to be orthogonal, we cannot estimate and evaluate β_1 and β_2 independently. Suppose, however, we first fit the model

$$y_u = \beta_0 + \beta_1 X_{1u} + \epsilon_u .$$

The resulting sum of squares for β_0 and β_1 is

$$S_1 = SS(b_0, b_1) = b_0 \Sigma y_u + b_1 \Sigma X_{1u} y_u .$$

We now fit the data to the larger model. We should find that we obtain a new value for b_0 and b_1 as well as a value for b_2. If we denote the new estimates by b_j', we can compute the regression sum of squares for the model,

$$S_2 = SS(b_0', b_1', b_2') = b_0' \Sigma y_u + b_1' \Sigma X_{1u} y_u + b_2' \Sigma X_{2u} y_u .$$

Now, the extra sum of squares due to including the term $\beta_2 X_2$ in the model is

$$SS(b_2' \mid b_0', b_1') = S_2 - S_1 .$$

That is, if β_0 and β_1 are already in the model, the additional information obtained by including $\beta_2 X_2$ in the model is measured by $S_2 - S_1$. If we added a third term, $\beta_3 X_3$ to the model, the extra sum-of-squares due to $\beta_3 X_3$ would be

$$SS(b_3' \mid b_0', b_1', b_2') = SS(b_0'', b_1'', b_2'', b_3'') - SS(b_0', b_1', b_2') .$$

We should note one important thing, here. The extra-sum of-squares for a given term is dependent on what has gone before it into the model. Thus, the extra sum of squares for $\beta_3 X_3$ will be different if we go from

$\beta_0 + \beta_1 X_1$ to $\beta_0 + \beta_1 X_1 + \beta_3 X_3$ than if we went from $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ to $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. The only circumstance in which it will not be different is if X_3 is orthogonal to X_1 and X_2. The procedure is also general with respect to addition terms which are functions of previous terms. Thus,

$$SS(b_{11} \mid b_0, b_1) = SS(b_0, b_1, b_{11}) - SS(b_0, b_1),$$

where b_{11} is the coefficient of X_1^2.

The principle of computing the extra-sum-of-squares is very important in the development of models, as we shall see in discussing model building in Section 7.6. The significance of the extra-sum-of-squares can be tested, then, by comparing it to the residual variance estimate of the fuller model, i.e.,

$$F_{1,\nu e} = \frac{SS(b_{11} \mid b_0, b_1)}{s^2(b_0, b_1, b_{11})}$$

*7.5.2 Matrix Approach to the Least Squares Analysis of the Multiple Regression Model

We present here a brief description of the analysis of a multiple regression model using matrix notation.

Let $\underline{y}$ be the column vector of N observations, $\underline{X}$ be the N by p matrix of X settings and $\underline{\beta}$ be the column vector of p parameters to be estimated in the model. That is,

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix}; \quad \underline{X} = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & & X_{k2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & X_{1N} & X_{2N} & & X_{kN} \end{pmatrix}, \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \\ \cdot \end{pmatrix}$$

where $p = k + 1$. The model may be written

$$\underline{y} = \underline{X}\ \underline{\beta} + \underline{\epsilon}\ ,$$

where $\underline{\epsilon}$ is a column vector of N errors. The least squares principle says to minimize the sum of squares

$$\Sigma \epsilon_u^2 = \underline{\epsilon}'\underline{\epsilon} = (\underline{y} - \underline{X}\,\underline{\beta})'\ (\underline{y} - \underline{X}\ \underline{\beta}) = \sum_u (y_u - \sum_j \beta_j X_{ju})^2$$

where ' means the transpose of the vector or matrix. The least squares estimates are given by

$$\underline{b} = \hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\ \underline{X}'\underline{y}\ ,$$

where $\underline{X}'\underline{X}$ is a non-singular matrix and hence invertible. The variance of the parameter estimates is

$$\text{Var}\ (\underline{b}) = (\underline{X}'\underline{X})^{-1}\sigma^2$$

and is estimated by replacing σ^2 by s^2. Note that if the X's are orthogonal, $\underline{X}'\ \underline{X}$ is a diagonal matrix and hence so is $(\underline{X}'\ \underline{X})^{-1}$. Then all the estimates are independent of each other and separate confidence intervals may be constructed. In general this is not the case. Constructing separate confidence intervals may lead to erroneous results. As an example, a joint confidence interval may be elliptical. The corresponding separate t-confidence intervals form a rectangle.

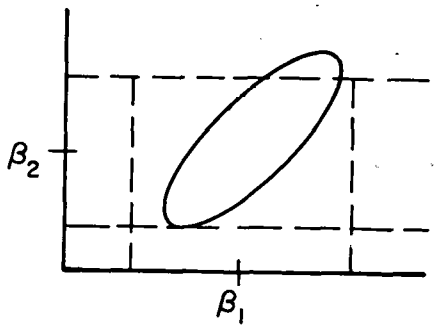


Figure 7.12. Joint Confidence Region vs. Separate Confidence Intervals

The variance of fitted values may be obtained from

$$\text{Var}(\hat{\underline{y}}) = \underline{X}\ (\underline{X}'\ \underline{X})^{-1}\ \underline{X}'\sigma^2.$$

For one point, not necessarily one of the design points,

$$\text{Var}\ (\hat{y}(\underline{x}_0)\ ) = \underline{x}_0'\ (\underline{X}'\ \underline{X})^{-1}\ \underline{x}_0\,\sigma^2.$$

Again, we estimate the variance by replacing σ^2 by s^2. We can then compute confidence bands about the fitted model,

$$\hat{y}_k \pm t_{\nu_e,\, \alpha/2}\, s\sqrt{\underline{x}'_0\, (\underline{X}'\ \underline{X})^{-1} \underline{x}_0}$$

The regression sum of squares can be compiled from

$$\text{SS Reg} = \hat{\underline{y}}'\ \hat{\underline{y}} = \underline{b}'\ \underline{x}'\ \underline{x}\ \underline{b}$$

but is subject to less roundoff error if computed as

$$\text{SS Reg} = \underline{b}'\ \underline{X}'\ \underline{y} \quad = \quad \underline{y}'\ \underline{X}(\underline{X}\ '\ \underline{X})^{-1}\ \underline{X}'\ \underline{y} \quad .$$

The residual sum of squares is, as always, obtained by subtraction.

ANOVA Table: Multiple Regression

Source	SS	df	MS
Total	$\underline{y}'\underline{y}$	N	
Regression	$\hat{\underline{y}}'\hat{\underline{y}} = \underline{b}\,\underline{X}'\underline{y}$	k + 1	MS Reg
Residuals	$\underline{y}'\underline{y} - \hat{\underline{y}}'\hat{\underline{y}}$	N-k-1	s^2 = MS Res

*7.6 Model Building

In this section we will summarize an iterative procedure for finding the best empirical model to fit the data. The process of finding the best-fitting model, whether empirical or theoretical, is called model building. The general procedure is to start with a given model and to add or subtract terms from the model until the best set of variables is included in the model. In general, there are four procedures and we evaluate the results by examining the coefficient of determination R^2 and the residual estimate of variance s^2. We seek that model which gives the highest R^2 and the lowest s^2, while keeping the number of parameters to be fitted as small as possible.

The four basic procedures are (1) all possible regressions, (2) backwards elimination, (3) forward selection, and (4) stepwise regression.

7.6.1 All Possible Regressions

If there are k candidates for entry into a model, we can perform all possible regressions and compare the results. There will be k models using one variable, k(k - 1)/2! using two terms, k(k - 1) (k - 2)/3! using four terms, etc. For k = 5, there are $5 + \frac{5\cdot4}{2!} + \frac{5\cdot4\cdot3}{3!} + \frac{5\cdot4\cdot3\cdot2}{4!} + \frac{5!}{5!} = 31$ possible regressions. As an extra variable is added to the model, R^2 will increase. However, s^2 may also increase. The experimenter must make a choice from the available information about the best model.

An analysis of thoria pellet grain size with respect to 5 thoria powder characteristics is provided below to illustrate model building procedures. The powder characteristics are

X_1 = Maximum particle size

X_2 = Average particle size

X_3 = Porosity

X_4 = Surface area

X_5 = Bulk density.

Only linear (first order) terms will be considered in this example. The results of all regressions on pellet grain size are:

Number	Terms in Model	R^2	s^2
1	X_1	62.1	3.65
2	X_2	46.7	5.13
3	X_3	37.3	6.05
4	X_4	48.6	4.95
5	X_5	23.6	7.37
6	X_1, X_2	68.8	3.19
7	X_1, X_3	63.9	3.69
8	X_1, X_4	85.4	1.49
9	X_1, X_5	63.6	3.73
10	X_2, X_3	54.7	4.64
11	X_2, X_4	52.7	4.89
12	X_2, X_5	58.7	4.22
13	X_3, X_4	57.2	4.38
14	X_3, X_5	43.2	5.82
15	X_4, X_5	48.6	5.26
16	X_1, X_2, X_3	69.3	3.35
17	X_1, X_2, X_4	91.6	0.92
18	X_1, X_2, X_5	74.9	2.74
19	X_1, X_3, X_4	86.2	1.51
20	X_1, X_3, X_5	64.9	3.84
21	X_1, X_4, X_5	94.0	0.65
22	X_2, X_3, X_4	58.9	4.48
23	X_2, X_3, X_5	63.6	3.97
24	X_2, X_4, X_5	63.5	3.98
25	X_3, X_4, X_5	57.5	4.64
26	X_1, X_2, X_3, X_4	93.6	0.75
27	X_1, X_2, X_3, X_5	75.1	2.91
28	X_1, X_2, X_4, X_5	94.1	0.69
29	X_1, X_3, X_4, X_5	96.0	0.47
30	X_2, X_3, X_4, X_5	67.3	3.83
31	X_1, X_2, X_3, X_4, X_5	96.2	0.48

The best first order model; i.e., the model which produces the highest R^2, and smallest s^2 for the fewest number of parameters possible, is

$$y_u = b_0 + b_1X_{1u} + b_3X_{3u} + b_4X_{4u} + b_5X_{5u} \quad ,$$

for which $R^2 = 96.0$ and $s^2 = 0.47$. This four variable model is preferable to the complete five variable model since it requires fitting one less parameter (and hence requires one fewer observation) at the cost of 0.2 in R^2 and actually has a smaller estimated residual variance. Similarly, a good argument can be made in favor of the three variable model involving X_1, X_4 and X_5 since it still has a large value of R^2 (94.0 vs. 96.0) and a small variance ($s^2 = 0.65$ vs. $s^2 = 0.47$). The final choice between these two models may be based on examination of residuals through lack of fit tests and residual plots.

7.6.2 Backwards Elimination

When k is large, all regressions become cumbersome and expensive to do. A better approach which chooses which new variables to add or delete is required. One such procedure begins with all terms in the model and eliminates them one at a time until a decision is reached. If the analysis is performed in the proper way, the extra sum of squares (see Section 7.5.1) for each term is available for comparison; i.e., compare SS $(b_4 \mid b_0, b_1, b_2, b_3)$ with s^2 for b_0, b_1, b_2, b_3, b_4. We then simply throw out the variables which have the insignificant extra-sum-of-squares when the other terms are included in the model.

Using the backwards elimination procedure, we would first throw out X_2 from the complete model since the F-value for its extra-sum-of-squares is definitely insignificant (see Appendix E, pages 215,216). The next candidate for elimination is X_3. The F-test for the extra-sum-of-squares for X_3, given X_1, X_4, and X_5 are in the model is 6.86 which is significant at the 1% level. Thus, the best model obtained by backwards elimination is the same as found by all regressions.

7.6.3 Forward Selection

In the forward selection procedure, we begin with the smallest model. The first step is to find the variable with the greatest absolute correlation with the response y, i.e.,

$$r_{X_i,y} = \frac{\Sigma(X_{iu} - \bar{X}_i)(y_u - \bar{y})}{\left[\Sigma(X_{iu} - \bar{X}_i)^2\Sigma(y_u - \bar{y})^2\right]^{1/2}}$$

The X variable with the highest $r_{X,y}$ is the first term to enter the model, X_1 in the example. The values for $r_{X,y}$ and the correlation matrix between X's are given below.

	$r_{X_i,y}$	X_1	X_2	X_3	X_4	X_5
X_1	-0.788	-	-0.606	-0.643	-0.300	-0.482
X_2	0.684	-	-	0.547	0.808	0.915
X_3	0.610	-	-	-	0.515	0.437
X_4	0.697	-	-	-	-	0.721
X_5	0.485	-	-	-	-	-

Variable X_1 is the first to enter the model since it has the highest absolute correlation with y. The second and each subsequent variable to enter the model is that variable which has the greatest absolute correlation to the residuals from the previous fit. Thus, X_4 enters the model second because it has the highest absolute correlation of the remaining variables to the residuals, $y-b_0-b_1X_1$. Similarly, X_5 enters the model next since it has the highest absolute correlation to the residuals, $y-b_0-b_1X_1-b_4X_4$. Next comes X_3. The variable X_2 does not add significant information and so is not accepted into the model.

Combinations and modifications of these forward and backward procedures can be proposed. One such modification is the stepwise procedure.

7.6.4 Stepwise Regression

The stepwise regression procedure is a modified forward selection procedure in which at each step, the previously selected variable is re-examined for deletion from the model. The procedure begins just as above in the forward selection procedure, first selecting X_1 and then X_4. The next step, however, was to re-evaluate X_1 given X_4 was in the model. The variable X_1 could be dropped if the extra sum of squares for X_1 given X_4 was found to be negligible. However, X_1 was maintained in the model and X_5 was added next. Both X_1 and X_4 were kept. Finally X_3 was added to complete the model. This, of course, is the same model as was obtained by the other procedures. As a final step for all procedures, the residuals should be plotted to examine for departures from assumptions undetected by a lack of fit test.

REFERENCES

1. Anderson, R.L., and Bancroft, T.A., Statistical Theory in Research, McGraw-Hill, New York, 1952.

2. Bennett, Carl A., and Franklin, Norman L., Statistical Analysis in Chemistry and the Chemical Industry, John Wiley, New York, 1954.

3. Brownlee, K.A., Statistical Theory and Methodology in Science and Engineering, 2nd Edition, John Wiley, New York, 1965.

4. Chew, Victor, Experimental Designs in Industry, John Wiley, New York, 1958.

5. Cochran, W.G., and Cox, G.M., Experimental Designs, 2nd Edition, John Wiley, New York, 1957.

6. Cox, D.R., Planning of Experiments, John Wiley, New York, 1958.

7. Davies, Owen L.,The Design and Analysis of Industrial Experiments, 2nd Edition, Hafner, New York, 1960.

8. Dixon, W.J., and Massey, F.J., Introduction to Statistical Analysis, 2nd Edition, McGraw-Hill, New York, 1957.

9. Dodge, H.F., and Romig, H.T., Sampling Inspection Tables, John Wiley and Sons, Inc., New York, 1954.

10. Draper, N.R., and Smith, H., Applied Regression Analysis, John Wiley and Sons, New York, 1966.

11. Duncan, Acheson J., Quality Control and Industrial Statistics, 3rd Edition, Richard D. Irwin, Inc., Homewood, Illinois, 1965.

12. Eisenhart, C., Hastay, M.W., and Wallis, W.A., Techniques of Statistical Analysis, McGraw-Hill, New York, 1947.

13. Federer, Walter T.,Experimental Design, Theory and Application, MacMillan, New York, 1955.

14. Fisher, R.A., The Design of Experiments, Hafner, New York, 1960.

15. Graybill, Franklin, An Introduction to Linear Statistical Models, McGraw-Hill, New York 1961.

16. Hald, A., Statistical Theory with Engineering Applications, John Wiley, New York, 1952.

17. Hicks, Charles, Fundamental Concepts in the Design of Experiments, Holt, Rinehart and Winston, New York, 1964.

18. Hooke, Robert, Introduction to Scientific Inference, Holden-Day, San Francisco, 1963.

19. Johnson, N.L., and Leone, F., Statistics and Experimental Designs, Volumes I and II, Wiley, New York, 1964.

20. United States Department of Defense, Military Standard, Sampling Procedures and Tables for Inspection by Attributes (MIL-STD-105D), Washington, D.C., U.S. Government Printing Office, 1963.

21. United States Department of Defense, Military Standard, Sampling Procedures and Tables for Inspection by Variables for Percent Defective (MIL-STD-414), Washington, D.C., U.S. Government Printing Office, 1957.

22. Kempthorne, Oscar, The Design and Analysis of Experiments, John Wiley, New York, 1952.

23. Mandel, John, The Statistical Analysis of Experimental Data, Interscience Publishers, New York, 1964.

24. Owen, D.B., Handbook of Statistical Tables, Addison-Wesley, 1962.

25. Miller, I. and Freund, J.E., Probability and Statistics for Engineers, Prentice-Hall, Engelwood Cliffs, N.J., 1965.

26. Natrella, Mary, Experimental Statistics, NBS Handbook 91, U.S. Government Printing Office, Washington, D.C., 1963.

27. Pearson and Hartley, Biometrika Tables for Statisticians, Cambridge University Press, Cambridge, 1954.

28. Scheffe, Henry, The Analysis of Variance, John Wiley, New York, 1959.

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

30. Wine, R. Lowell, Statistics for Scientists and Engineers, Prentice-Hall, Englewood Cliffs, N.J., 1964.

31. Tanur, Judith M., Frederick Mosteller, William H. Kruskal, Richard F. Link, Richard S. Pieters, and Gerald R. Rising, Statistics: A Guide to the Unknown, Holden-Day, Inc., San Francisco, 1972.

32. Weissberg, Alfred, and Beatty, Glenn H., "Tables of Tolerance-Limit Factors for Normal Distributions", Technometrics 2, 1960, pp. 483-500.

33. Shapiro, S.S., and Wilk, M.B., "An Analysis of Variance Test for Normality (complete sample)", Biometrika, Vol. 52, 1965, pp. 591-611.

34. Beggs, W.J. and Ginsburg, H., "Design In-Information Out," Experimental Mechanics, Vol. 12, No. 6, June 1972, pp. 289-296.

ACKNOWLEDGMENTS

This text has been several years in development with much of the original writing begun in 1971. Several years and courses have resulted in many additions, deletions, and modifications. I wish to thank many people for their help, contributions and encouragement toward the completion of this text:

Dr. Clyde E. Vogeley, under whose supervision I began this project, Jane Toben, Tets Shimamoto, Ozzie Willner, and especially Herb Ginsburg, my colleagues who made many helpful suggestions, Dr. Norman Draper, my major professor at the University of Wisconsin from whom I received much of my original training, especially in design of experiments, the many students who have pointed out typographical and logical errors and have prompted clearer explanations, and a special thanks to Connie Simboli who typed the original text, to Helen Dishong for typing the final draft, and to Jake Koffler for his many editorial comments and for seeing this project through to its completion.

APPENDIX A. UNCERTAINTY ANALYSIS

1. Introduction

Whenever a response y can be related to a group of contributing factors or variables, $\{x_i\}$, which are subject to error, the true response value for a given setting of the x's is also subject to error. The description and analysis of the induced variability in y is sometimes called uncertainty analysis. In other words, the uncertainty of the value of the x's leads to an uncertainty in the resulting value of y. If certain information about the uncertainties in the x's is available, the uncertainty of y can be determined or at least closely approximated. The best type of information on the x's is the probability distribution function, or at least knowledge of the mean and variance of the unknown distribution is essential. (Note: If a probability distribution can be described by a mathematical function, f(x), the mean μ of the distribution of x is the center of gravity, $\mu = \int xf(x)dx$, and the variance is the moment of inertia, $\sigma^2 = \int(x-\mu)^2 f(x)dx$. The mean locates the distribution and the standard deviation (square-root of the variance) measures the spread of the distribution.)

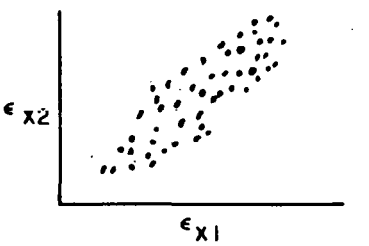
To determine the uncertainty of a response y, then three characteristics must be known or assumed:

(a) The Model - The mathematical description $y = f(x_1, x_2, \ldots, x_k)$ of the functional relationship between y and the x's. The model may represent an approximation to the true relationship,

(b) The Distribution of each x - i.e., knowledge of the probability function of the uncertainties in the x's or at least the means and variances,

(c) The Interdependencies among the x's - i.e., knowledge as to whether uncertainties in the x's act together (correlated) or separately (independent).

Independent

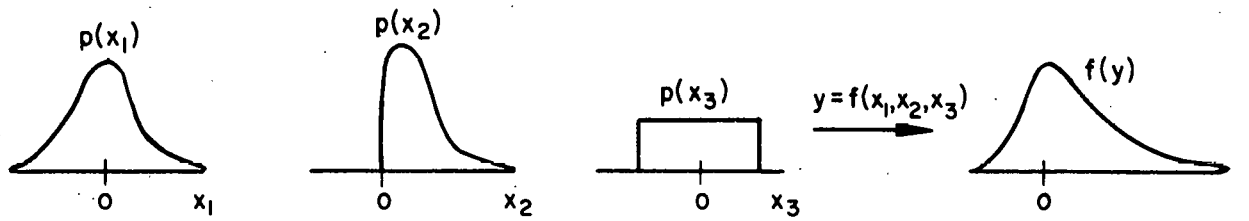


Positive Correlation

These features will be discussed repeatedly in the remainder of this Appendix. The mathematical basis for analysis of uncertainties in x's as they propagate through a model to y will be discussed. Keep in mind that the objective is to determine the size of an uncertainty (error) and the frequency (probability) of their occurrence. This is illustrated in the figure below.

Distribution of Uncertainty in

$y = f(x_1, x_2, x_3)$



2. Simple Propagation of Error

Suppose $y = a + bx$, a, b are constants and x has a mean value μ about which the true value of x varies in a random (unpredictable) fashion. Assume the variability of x about μ is given by its variance σ^2. Then, denoting the mean by E( ), (read as "expected value of") and variance by Var,

$$\text{Mean of } y = E(y) = a + b\mu \qquad \text{(A1)}$$

$$\text{Variance of } y = \text{Var}(y) = b^2 \sigma^2 \qquad \text{(A2)}$$

This is found to be the case as follows:

$$\text{Mean of } y = E(y) = E(a+bx)$$

$$= a + b\,E(x)$$

$$= a + b \int x\, f(x)\, dx$$

$$= a + b\mu$$

where $f(x)$ is the probability distribution function for x.

The variance of a constant is zero, and by definition of variance $\text{Var}(bx) = b^2\int(x-\mu)^2 f(x)dx = b^2\int (x-\mu)^2 f(x)\ dx$. The variance of y becomes

$$\text{Variance of } y = \text{Var}(y) = \text{Var}(a + bx)$$

$$\text{Var}(y) = b^2\sigma^2$$

Since the model contains only one random variable, x, the distribution function for y is the same as the distribution function for x except the mean is $a+b\mu$ and the variance $b^2\sigma^2$. Hence, all questions about the probable value of y can be answered by knowing the distribution of x. For example, if x is a normal distribution with mean 0 and variance 1, and a=5, b=2, then y is a normal distribution with mean 5 and variance 4. Thus, approximately 95% of the values for y can be expected to be between $5\pm2\sigma$ ($1<y<9$) and 99.7% of the values of y can be expected to lie between $5\pm3\sigma$ ($-1<y<11$).

3. Linear Combination of Several x's.

Suppose the model for y can be written as a linear combination of a set of x's,

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k. \qquad \text{(A3)}$$

In particular, if $a_0=0$, $a_i=1$, $i=1,2,\ldots,k$, then y is simply the sum of k values of x.

Let μ_i be the mean of x_i and σ_i^2 be the variance. Then the mean of y_i is given by

$$E(y) = E(a_0 + \sum_{i=1}^{k} a_ix_i) = a_0 + \sum_{i=1}^{k} a_iE(f_i) = a_0 + \sum_{i=1}^{k} a_i\mu_i\ . \qquad \text{(A4)}$$

Furthermore, if the x's are independent of each other (i.e., a change in one x does not force a change in any other x), then the variance of y is

$$\text{Var}(y) = \text{Var}(a_0 + \sum_{i=1}^{k} a_iy_i) = \sum_{i=1}^{k} a_i^2\sigma_i^2\ , \text{ x's independent} \qquad \text{(A5)}$$

For $a_0=0$, $a_i=1$, this says that <u>the mean of a sum is the sum of the means, regardless of correlation,</u> and <u>the variance of a sum is the sum of variances for independent variables.</u>

To illustrate the correlated case, consider first only 2 variables x_1 and x_2 for which the correlation coefficient ρ must lie between -1 and +1. (The covariance is given by $\rho\sigma_1\sigma_2$). Then

$$\text{Var}(x_1 + x_2) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \tag{A6}$$

Expanding to a general linear combination,

$$\text{Var}(a_1x_1 + a_2x_2) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2 \tag{A6a}$$

For complete generality,

$$\text{Var}\left(\sum_{i=1}^{k} a_i x_i\right) = \sum_{i=1}^{k} a_i^2\sigma_i^2 + 2\sum\sum_{i<i} a_i a_{i'}\rho_{ii'}\sigma_i\sigma_{i'} \tag{A6b}$$

or for $\sigma_i = \sigma$, $a_i = 1$, $i=1, 2, \ldots, k$,

$$\text{Var}\left(\sum_{i=1}^{k} x_i\right) = k\sigma^2 + 2k\frac{(k-1)}{2}\rho\sigma^2. \tag{A6c}$$

Note that ρ may be positive or negative. Thus, the variance of y may be increased or decreased by the presence of correlation, depending also on the signs of the a_i's. Another interesting note can be seen from (A6c). Since the variance of y can never be negative (by definition), this implies that the most negative value of ρ is not -1, but in (A6c) $\rho > -1/(k-1)$. Of course, ρ may be +1, which yields the result that the standard deviation of y is the sum of the standard deviations:

$$\text{Var}(y) = \sum_{i=1}^{k}\sigma_i^2 + \sum\sum_{i=i}(1)\,\sigma_i\sigma_{i'} = \left(\sum_{i=1}^{k}\sigma_i\right)^2$$

$$\sigma_y = \sum_{i=1}^{k}\sigma_i$$

4. Variance of Simple Product

Consider now the model

$$y = x_1x_2 \tag{A7}$$

The best way to illustrate the variance of y here is to write y as a Taylor Series expansion about the means of x_1 and x_2, i.e.,

$$y = f(\mu_1,\mu_2) + \left(\frac{\partial f}{\partial x_1}\right)_{\mu_1,\mu_2}(x_1-\mu_1) + \left(\frac{\partial f}{\partial x_2}\right)_{\mu_1,\mu_2}(x_2-\mu_2) \tag{A8}$$

$$+ \left(\frac{\partial^2 f}{\partial x_1\,\partial x_2}\right)_{\mu_1,\mu_2}\frac{2(x_1-\mu)(x_2-\mu_2)}{2!}$$

$$= \mu_1\mu_2 + \mu_2(x_1-\mu_1) + \mu_2(x_2-\mu_2) + (x_1-\mu_1)(x_2-\mu_2)$$

Thus, an exact representation of the mean of y for x_1 and x_2 independent (i.e., $E[(x_1 - \mu_1)(x_2 - \mu_2)] = 0$), is $\mu_1 \mu_2$.

Then, since $Var(y) = E(y - \mu_1 \mu_2)^2$,

$$Var(y) = E[\mu_2 (x_1 - \mu_1) + \mu_1 (x_2 - \mu_2) + (x_1 - \mu_1)(x_2 - \mu_2)]^2 \qquad (A9)$$
$$= \mu_2^2 \sigma_1^2 + \mu_1^2 \sigma_2^2 + \sigma_1^2 \sigma_2^2$$

(x_1, x_2 independent)

Since nuclear engineers and scientists often deal with relative errors, the relative variance of y can be obtained as follows:

$$Var(y)_{\%} = \frac{(100)^2 Var(y)}{(\mu_1 \mu_2)^2} = (100)^2 \left[\frac{\mu_2^2 \sigma_1^2}{\mu_1^2 \mu_2^2} + \frac{\mu_1^2 \sigma_2^2}{\mu_1^2 \mu_2^2} + \frac{\sigma_1^2 \sigma_2^2}{\mu_1^2 \mu_2^2} \right] \qquad (A10)$$
$$= (\sigma\%)_1^2 + (\sigma\%)_2^2 + \frac{(\sigma\%)_1^2 (\sigma\%)_2^2}{(100)^2}$$

and since the last term is often quite small, it is usually dropped, and

$$Var(y)_{\%} = (\sigma\%)_1^2 + (\sigma\%)_2^2 \qquad (A10a)$$

which is the well-known result that relative errors of a product add.

If x_1 and x_2 are correlated, the cross product term in (A8) is usually dropped and the result is

$$Var(y) \doteq \mu_2^2 \sigma_1^2 + \mu_1^2 \sigma_2^2 + 2\mu_1 \mu_2 \rho \sigma_1 \sigma_2 \qquad (A11)$$

and

$$Var(y)_{\%} \doteq (\sigma\%)_1^2 + (\sigma\%)_2^2 + 2\rho (\sigma\%)_1 (\sigma\%)_2 \qquad (A11a)$$

5. Variance of Complex Models Using Taylor Series Expansions

Exact variances for a linear combination of input variables and for products have been obtained above. For most other cases, of a more complex model, only approximate variances can usually be obtained. The Taylor Series expansion as used in the previous section is the tool used to obtain the approximation. The technique will be illustrated for a simple quotient $y = x_1/x_2$ and for a more complex model.

a. Simple Quotient

Let

$$y = x_1/x_2, \quad x_2 \neq 0 \tag{A12}$$

Then

$$y = f(x_1, x_2)\Big|_{\mu,\mu_2} + \left(\frac{\partial f}{\partial x_1}\right)_{\mu_1,\mu_2}(x_1-\mu_1) + \left(\frac{\partial f}{\partial x_2}\right)_{\mu_1,\mu_2}(x_2-\mu_2)$$

$$+ \left(\frac{\partial^2 f}{\partial x_1^2}\right)_{\mu_1,\mu_2}\frac{(x_1-\mu_1)^2}{2!} + \left(\frac{\partial^2 f}{\partial x_2^2}\right)_{\mu_1,\mu_2}\frac{(x_2-\mu_2)}{2!} \tag{A13}$$

$$+ \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2_{\mu_1,\mu_2} 2\frac{(x_1-\mu_1)(x_2-\mu_2)}{2} + \cdots$$

$$= \mu_1/\mu_2 + (x_1-\mu_1)/2 + (-\mu_1/\mu_2)(x_2-\mu_2)$$
$$+ (o)(x_1-\mu_1)^2/2 + 2(\mu_1/\mu_2^3)(x_2-\mu_2)^2/2 + (-1/\mu_2^2)(x_1-\mu_1)(x_2-\mu_2)$$
$$+ \cdots$$

It can be seen that an infinite number of derivatives with respect to x_2 exist. Thus, the Taylor Series is usually truncated after the linear terms. The mean of y is then an approximation,

$$E(y) \doteq \mu_1/\mu_2 \quad , \mu_2 \neq 0 \tag{A14}$$

and

$$\mathrm{Var}(y) \doteq \mathrm{Var}\left[\frac{1}{\mu_2}(x_1-\mu_1) - \frac{\mu_1}{\mu_2^2}(x_2-\mu_2)\right]$$

$$\doteq \frac{1}{\mu_2^2}\sigma_1^2 + \frac{\mu_1^2}{\mu_2^4}\sigma_2^2 \quad (x_1, x_2 \text{ independent*}) \tag{A15}$$

$$\doteq \frac{1}{\mu_2^2}\sigma_1^2 + \frac{\mu_1^2}{\mu_2^4}\sigma_2^2 - \frac{2\mu_1}{\mu_2^2}\rho\sigma_1\sigma_2, \quad (x_1, x_2 \text{ correlated}) \tag{A15a}$$

*In some cases, it is advisable to include cross product term $(1/\mu_2^4)\sigma_1^2\sigma_2^2$

The relative variance of y is given as

$$\mathrm{Var}(y)_{\%} = \begin{cases} (\sigma\%)_1^2 + (\sigma\%)_2^2 \quad , & (x_1,\ x_2,\ \text{independent}) \\ (\sigma\%)_1^2 + (\sigma\%)_2^2 - 2\rho(\sigma\%)_1\ (\sigma\%)_2, & (x_1,\ x_2\ \text{correlated}) \end{cases} \tag{A16}$$

as expected.

b. Complex Model

Suppose $y = \dfrac{x_1 e^{-ax_2}}{1-bx_3}$

$$y \doteq \frac{\mu_1 e^{-a\mu_2}}{1-b\mu_3} + \left(\frac{e^{-a\mu_2}}{1-b\mu_3}\right)(x_1-\mu_1) + \left(-\frac{a\mu_1 e^{-a\mu_2}}{1-b\mu_3}\right)(x_2-\mu_2) + \left(-\frac{b\mu_1 e^{-a\mu_2}}{(1-b\mu_3)^2}\right)(x_3-\mu_3) \tag{A17}$$

(truncated after linear terms). Then,

$$E(y) \doteq \frac{\mu_1 e^{-a\mu_2}}{1-b\mu_3} \tag{A18}$$

$$\mathrm{Var}(y) \doteq \left(\frac{e^{-a\mu_2}}{1-b\mu_3}\right)^2 \sigma_1^2 + \left(\frac{a\mu_1 e^{-a\mu_2}}{(1-b\mu_3)}\right)^2 \sigma_2^2 + \left(\frac{b\mu_1 e^{-a\mu_2}}{(1-b\mu_3)^2}\right)^2 \sigma_3^2 \tag{A19}$$

if x_1, x_2, x_3 are independent.

If x_1, x_2 and x_3 are correlated, the following covariance terms need to be added to (A19),

$$\frac{-2a\mu_1 e^{-2a\mu_2}}{(1-b\mu_3)^2}\rho_{12}\sigma_1\sigma_2 \quad \frac{-2b\mu_1 e^{-2a\mu_2}}{(1-b\mu_3)^3}\rho_{13}\sigma_1\sigma_2 \qquad \text{(A19a)}$$

$$+\frac{2ab\mu_1^2 e^{-a\mu_2}}{(1-b\mu_3)^3}\rho_{23}\sigma_2\sigma_3$$

6. The Distribution of y

Thus far, the discussion has concentrated on the variance of y for a given model for independent and correlated x's. In order to make statements about the frequency of given values of y, it is necessary to know the distribution of y. In some cases, the distribution of y can be inferred directly from the model and the distribution of the x's. More often, however, the distribution of y can only be approximated.

For a linear model of the type given in Equation (A3), the exact distribution of y can be determined if the distributions of the x's are of the proper form. In particular, if all the x's are normally distributed with means μ_i and variances σ_i^2, then $y = \sum_{i=1}^{k} a_i x_i$ is also normally distributed with mean $\sum_{i=1}^{k} a_i \mu_i$ and variance $\sum_{i=1}^{k} a_i^2 \sigma_i^2$ (+ covariance term if x's are correlated). Symbolically, if all x_i are distributed as $N(\mu_i, \sigma_i^2)$, then

$$y = \sum_{i=1}^{k} a_i x_i \text{ is distributed as } N\left(\sum_{i=1}^{k} a_i \mu_i, \sum_{i=1}^{k} a_i^2 \sigma_i^2\right).$$

In general, however, if the x's are not all normally distributed, the distribution of a linear combination cannot be inferred exactly. (The only other distribution that sums to the same distribution is the gamma, and then only for a simple sum).

Theoretically, given the distributions of the x's and a model, the statistician can, through his knowledge of mathematics and distribution theory, derive the exact distribution of y. In practice, however, this is usually impossible except for the simplest cases.

Fortunately, however, there is a very powerful theorem that enables a useful approximation to be made in many practical situations. The theorem is known as the Central Limit Theorem and says simply that a linear combination of several independent variables with the same distribution tends to have a normal distribution with the mean equal to the sum of means and a variance equal to the sum of variances. How closely the normal distribution approximates the true distribution depends on the number of variables involved and the individual distributions, but the theory, nevertheless, allows a reasonable approximation to be made in most circumstances. More generally, if the distributions of the independent variables are not identical, but no one variable dominates the others, the theorem still holds. Furthermore, since any functional form of a model can be approximated by a Taylor Series expansion truncated after the linear terms, the theorem can be seen to be applicable to any model, provided no one term dominates.

8. Conclusion

It has been stated that to be able to say anything about the uncertainty of a response y, which is a function of one or more input variables, x, it is necessary to know or assume three important facts: (1) the form of the model, (2) the distribution, or at least the mean and variance of the distribution, of the x's, and (3) the correlation structure of the x's. By use of the Taylor Series expansion, exact or approximate expressions for the variance of the response y can be obtained. Finally, under certain conditions, the distribution of y can be inferred, or by the use of the Central Limit Theorem, an approximate normal distribution can be applied to make probability statements about y.

The key to this uncertainty analysis is the assumptions. Like "Garbage In-Garbage Out," if the assumptions of an uncertainty analysis are grossly incorrect, any conclusions based on these assumptions must be viewed with caution.

APPENDIX B. ESTIMATION THEORY

1. Maximum Likelihood Principle

Having determined which distribution functions describe variables of interest, it usually remains to determine the values of the parameters of these distributions. We may not know, for example, where the distribution of data is centered, or exactly how spread out it is. Hence, we must estimate these values by collecting data. Just what the best guess or estimate of a particular parameter is, and how precise we think it is, is the subject matter of estimation theory. We must gather our information, i.e., the data, in such fashion that will give use the best value according to some criterion. There are many criteria which may be used, but the most common approach is known as the maximum likelihood procedure. The maximum likelihood procedure, as its name implies, simply finds that value of the parameter that makes the event of obtaining a given set of data most probable. To illustrate, consider an urn full of red and blue balls. Let x be recorded as 1 if a red ball is drawn out of the urn, and 0 if a blue ball is drawn. The distribution of obtaining a red ball in a single draw from the urn is a binomial with n = 1 and p being the probability of drawing a red ball,

$$p(x) = p^x (1-p)^{1-x} \quad , x = \begin{cases} 1, \text{ if red} \\ 0, \text{ if not red} \end{cases}$$

Suppose that one person claims that three-fourths of the balls are red, while another claims that the red and blue balls are equally distributed. The object is to determine which person is more probably correct. To answer the question involves taking data. Suppose one ball is drawn from the urn, the color recorded and the ball replaced until four balls have been drawn. Each draw has exactly the same distribution as every other and each result is independent of the others. Thus, if the results were red, red, blue, red, the probability of obtaining this result for a given p would be

$$\begin{aligned} \Pr\ (x_1=1, x_2=1, x_3=0, x_4=1) &= \Pr(x_1=1)\Pr(x_2=1)\Pr(x_3=0)\Pr(x_4=1) \\ &= p(x_1)p(x_2)p(x_3)p(x_4) \\ &= p^1(1-p)^0\ p^1(1-p)^0\ p^0(1-p)^1\ p^1(1-p)^0 \\ &= p^3(1-p). \end{aligned}$$

Now which value of p, 1/2 or 3/4, will result in the larger probability?

For p = 1/2, $p^3\ (1-p) = 1/16$

for p = 3/4, $p^3\ (1-p) = 27/256 = (27/16) \times 1/16$

Obviously, four draws from the urn resulting in 3 reds and one blue is more probable if the proportion of reds were $p = 3/4$ than if $p = 1/2$.

What is the most probable value of p, given the results obtained? It is that value of p that maximizes the joint probability function

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3)p(x_4)$$
$$= p^3(1-p).$$

When this function is written as a function of the unknown parameter p, given the observed x values, it is called the likelihood function, $L(p)$. Since the range (i.e., possible outcomes) of x_i does not depend on p, we may maximize $L(p)$ by taking its derivative with respect to p, equating to zero and solving for p. Thus,

$$\frac{d\,L(p)}{dp} = \frac{d}{dp}\,(p^3(1-p)) = 3p^2 - 4p^3 = 0$$

Solving for p, we obtain

$$p^2\,(3-4p) = 0$$
$$p^2 = 0$$
$$p = 3/4$$

It is quickly obvious that $p = 0$ does not maximize $L(p)$, thus we have the unique maximum likelihood estimate of p, given the results (red, red, blue, red), to be 3/4. Of course, given a different set of data, e.g., red, red, blue, blue, we would have found a different maximum likelihood estimate.

The maximum likelihood principle then is to maximize the likelihood function in the light of the data on hand. To distinguish between the joint probability density function $f(x_1, x_2 \ldots, x_n \mid p)$, of n independent random variables, given the parameter value, and the likelihood function of the parameter given the observed values of the random variables, we write the likelihood function more formally as $L(p \mid x_1, x_2, \ldots, x_n)$.

If the range of the x_i do not depend on the parameter, the maximization procedure is to take the derivative of the likelihood function, or equivalently the derivative of the natural log of the likelihood function, set to zero and solve.

We have already seen how to obtain the maximum likelihood estimate of p for a binomial distribution for $n = 1$. For $n > 1$ the result generalizes in a straight-forward manner and we find that the maximum likelihood estimate of p for a binomial is $\bar{x}$. The maximum likelihood estimates of the parameters in Poisson, exponential, and normal distributions are given below.

a. Poisson Distribution

$$p(x|\lambda) = e^{-\lambda} \lambda^x / x! \quad , x = 0, 1, 2, \ldots\ldots$$

Let $x_1, x_2, \ldots, x_n$ be n independent observations from this Poisson distribution. Then

$$L(\lambda| x_1, x_2, \ldots, x) = \sum_{i=1}^{n} e^{-\lambda} \lambda^{x_i}/x_i!$$

$$= e^{-n\lambda} \lambda^{\Sigma x_i} / \pi x_i!$$

where π means the product of $x_1 x_2 \ldots x_n$. Then

$$\ell n\, L(\lambda) = -n + \Sigma x_1 \ell n \lambda - \sum_{1=1}^{n} \ell n\, x_1!$$

$$\frac{d\, \ell n\, L(\lambda)}{d\lambda} = -n + \frac{\Sigma x_i}{\lambda} = 0$$

Solving for λ gives the maximum likelihood estimate.

$$\hat{\lambda}_{ML} = \sum_{i=1}^{n} x_i/n = \bar{x} \quad .$$

b. Exponential Distribution

$$f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda} \quad , x > 0, \lambda > 0.$$

Let $x_1, x_2, \ldots, x_n$ be n independent observations from $f(x|\lambda)$. Then

$$L(\lambda| x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{-x_i/\lambda}$$

$$= \frac{e^{-\Sigma x_i/\lambda}}{\lambda^n}$$

$$\ell n\, L(\lambda) = -n\, \ell n \lambda - \Sigma x_i/\lambda$$

$$\frac{d\, \ell n\, L(\lambda)}{d\lambda} = \frac{-n}{\lambda} + \frac{\Sigma x_i}{\lambda^2} = 0$$

which gives $\hat{\lambda}_{ML} = \bar{x}$

c. Normal Distribution

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{1}{2\sigma^2}(x-\mu)^2\right] \quad , \quad \begin{matrix} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma^2 > 0. \end{matrix}$$

Let $x_1, x_2, \ldots, x_n$ be n independent observations from $f(x|\mu,\sigma^2)$. Then

$$L(\mu,\sigma^2 \mid x_1, x_2,\ldots,x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2\right].$$

$$\ell n\, L(\mu,\sigma^2) = -\frac{n}{2}\ell n\, 2\pi - \frac{n}{2}\ell n\sigma^2 - \frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2$$

$$\frac{\partial \ell n\ L}{\partial \mu} = \frac{-2}{2\sigma^2}\Sigma(x_i-\mu)$$

$$\frac{\partial \ell n\ L}{\partial \sigma^2} = \frac{-n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\Sigma(x_i-\mu)^2$$

Setting both equations to zero and solving for μ and σ^2, we obtain,

$$\hat{\mu}_{ML} = \bar{x}$$

$$\hat{\sigma}^2_{ML} = \frac{\Sigma(x_i-\bar{x})^2}{n}$$

Note: The estimate $\hat{\sigma}^2_{ML} = \frac{1}{n}\Sigma(x_i-\bar{x})^2$ is a <u>biased</u> estimate of the variance σ^2 of a normal distribution. That is, the expected value or mean of the distribution of $\hat{\sigma}^2_{ML}$ is not σ^2 but $(n-1)\sigma^2/n$. As a result, the unbiased estimate of σ^2, $s^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$, is preferred in most cases and will be used where required in the next sections.

2. Other Estimation Criteria

a. Unbiased Estimator:

Let $\hat{\theta}$ be an estimate of θ. If $E(\hat{\theta}) = \theta$, then $\hat{\theta}$ is an unbiased estimate of θ. For example, $E(\bar{x}) = \mu$, hence $\bar{x}$ is unbiased for μ.

b. Minimum Variance Estimator:

Let $\hat{\theta}_1$ be one of a class of estimators $\{\theta_i\}$. If $Var(\hat{\theta}_1)$ is less than or equal to $Var(\hat{\theta}_i)$, $i \neq 1$, then $\hat{\theta}_1$ is the minimum variance estimator of that class of estimators, i.e., $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_i)$, all i. For example,

$$\bar{x} = \frac{1}{n}x_i + \frac{1}{n}x_2 + \ldots + \frac{1}{n}x_n$$

is a linear, unbiased estimate of $\mu = E(x_i)$. The Gauss-Markov Theorem of least squares tells us that $\bar{x}$ is also the minimum variance linear unbiased estimator of μ.

c. Minimum Mean Square Error Estimator:

Let $\hat{\theta}_1$ be an estimtor of θ. If $\hat{\theta}_1$ gives the minimum value of $E(\hat{\theta}_i - \theta)^2$ among a class of estimators $\{\hat{\theta}_i\}$, it is the minimum mean square error estimator. Note, if $E(\hat{\theta}_1) = \theta$, i.e., if $\hat{\theta}_1$ is unbiased, then the mean square error is the same as the variance. However, $\hat{\theta}_1$ need not be unbiased. If $E(\hat{\theta}_1) \neq \theta$, then

$$E(\hat{\theta}_1 - \theta)^2 = E[(\hat{\theta}_1 - E(\hat{\theta}_1)]^2 + [E(\hat{\theta}_1) - \theta]^2$$
$$= \text{Var}(\hat{\theta}_1) = (\text{Bias})^2$$

APPENDIX C. THE CENTRAL LIMIT THEOREM

Theorem:

Let $x_1, x_2, \ldots, x_n$ be independently and identically distributed as $f(x)$ with mean μ and variance σ^2. Then x has in the limit a normal distribution with mean μ and variance σ^2/n for n sufficiently large.

Proof:

Every probability distribution function with at least a finite mean and variance has an associated moment generating function.

$$M_x(t) = E\ (e^{tx})$$

$$= E(1 + tx + (tx)^2/2! + (tx)^3/3! + \ldots.)$$

$$= 1 + \mu'_1\, t + \mu'_2\, t^2/2! + \mu'_3\, t^3/3! + \ldots.$$

where μ'_r are the raw moments and can be obtained by

$$\mu'_r = \frac{d^r M_x(t)}{d\ t^r} \quad .$$

The central moments for the distribution of x can be obtained as follows:

$$M_{x-u}\ (t) = E[e^{t(x-u)}]$$

$$= e^{-ut} M_x(t)$$

Expanding e^{-ut} and $M_x(t)$, multiplying the two series together and collecting terms, and making use of the relationship between central and raw moments given in Section 2.3.2, we have

$$M_{x-u}(t) = 1 + \mu_2\, t^2/2! + \mu_3 t^3/3!$$

$$\text{where } \mu_r = \frac{d^r M_{x-u}(t)}{d\ t^r} \quad .$$

Now consider $w = \sum_i (x_i - \mu)$ the mean of which is 0. Then since x_i are independent,

$$M_w(t) = E[e^{t \Sigma (x_i - \mu)}]$$

$$= \int \ldots \int e^{t \Sigma (x_i - \mu)}\ f(x_1) \ldots f(x_n)\ dx_1 \ldots dx_n$$

$$= \int e^{t(x_1 - \mu)}\ f(x_1) dx_1 \int e^{t(x_2 - \mu)}\ f(x_2) dx_2 \ldots \int e^{t(x_n - \mu)}\ f(x_n) dx_n$$

$$= \prod_{i=1}^{n} E\ [e^{t(x_i - \mu)}]$$

$$= [M_x(t)]^n$$

$$= [1 + \mu_2\, t^2/2! + R_2(t)]^n$$

where $R_2(t)$ is a remainder term that converges.

Recall that the variance of n independent variables with common variance σ^2 is $n\sigma^2$. Then we can standardize w by dividing by its standard deviation $\sqrt{n\sigma}$. Thus,

$$z = \frac{w}{\sqrt{n\sigma}}$$

and

$$M_z(t) = M_{\frac{w}{\sqrt{n\sigma}}}(t) = E\left[e^{tw/\sqrt{n\sigma}}\right]$$

which can also be written as

$$M_z(t) = E[e^{w(t/\sqrt{n\sigma})}]$$

$$= M_w(t/\sqrt{n\sigma})$$

$$= \left[1 + \mu_2\left(\frac{t}{\sqrt{n\sigma}}\right)^2/2! + R_2\left(\frac{t}{\sqrt{n\sigma}}\right)\right]^n$$

Since $R_2(t)$ converges, so does $R_2(t/\sqrt{n\sigma})$, so that

$$\text{Lim } M_z(t) = e^{\mu_2 t^2/2\sigma^2}$$

$$= e^{t^2/2}$$

since $\mu_2 = \sigma^2$. But this is the central moment generating function of a normal distribution with mean 0 and variance 1. This can be seen as follows: For a N(0, 1) variable z,

$$M_z(t) = E[e^{tz}] = \int e^{tz}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

$$= \frac{1}{\sqrt{2\pi}}\int e^{-(z^2-2tz+t^2-t^2)/2}dz$$

$$= e^{t^2/2}\frac{1}{\sqrt{2\pi}}\int e^{-1/2(z-t)^2}dz$$

$$= e^{t^2/2}$$

Since the function inside the integral is a normal distribution with mean t and variance 1 and thus integrates to 1.

Now since z is a N(0, 1) variable, it follows that $w = \sqrt{n}\sigma z$ follows a $N(0, n\sigma^2)$ distribution and $\bar{x} = \mu + (\sigma/\sqrt{n})z$ follows a $N(\mu, \sigma^2/n)$ distribution as n becomes large.

APPENDIX D. CALCULATION OF EXPECTED MEAN SQUARES

Consider the case of comparing k different treatments on a set of N homogeneous experimental units. The j^{th} treatment (j = 1, 2, ..., k) is assigned to n randomly chosen units, and N = nk. Let the i^{th} observation (i = 1,2, ..., n) on the j^{th} treatment be represented by x_{ij}. The mathematical model is $x_{ij} = \mu + \tau_j + \epsilon_{ij}$, where μ and τ_j are unknown true constants or parameters (subject to the contraint that $\Sigma_j \ \tau_j = 0$) and the ϵ_{ij} are random variables independent of each other and the τ's, and have a mean of zero and common variance σ^2 for all i, j. The calculational procedures for the ANOVA for this fixed-effects model are given in Section 6.7.2.

Substituting into the model the least squares estimates, we have $x_{ij} = \bar{\bar{x}} + (\bar{x}_j - \bar{\bar{x}}) + (x_{ij} - \bar{x}_j)$. Note that this is an identity for all values of x_{ij}. Transposing, we may write $(x_{ij} - \bar{\bar{x}}) = (\bar{x}_j - \bar{\bar{x}}) + (x_{ij} - \bar{x}_j)$. Thus, the deviation of the response of the i^{th} observation on the j^{th} treatment from the grand mean is equal to the deviation of the j^{th} treatment mean from the grand mean plus the deviation of x_{ij} from the jth treatment mean. Now, squaring both sides of the identity and summing over all N observations, we have

$$\sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{\bar{x}})^2 = \Sigma_j \Sigma_i (\bar{x}_j - \bar{\bar{x}})^2 + 2\sum_j \Sigma_i (\bar{x}_j - \bar{\bar{x}})(x_{ij} - \bar{x}_j) + \sum_j \Sigma_i (x_{ij} - \bar{x}_j)^2$$

But $\Sigma_j \Sigma_i (\bar{x}_j - \bar{\bar{x}})(x_{ij} - \bar{x}_j) = \Sigma_j (\bar{x}_j - \bar{\bar{x}}) \sum_i (x_{ij} - \bar{x}_j) = 0$

Thus, $$\Sigma\Sigma_i (x_{ij} - \bar{\bar{x}}) = \sum_j \sum_i (\bar{x}_j - \bar{\bar{x}})^2 + \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$$

$$= \sum_j n (\bar{x}_j - \bar{\bar{x}})^2 + \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$$

Let $T_j = \sum_{i=1} x_{ij}$ and C.F. - $(\sum_j \sum_i x_{ij})^2/N$.

Then the sum of squares for treatments SST can be written

$$SST = \sum_j n (\bar{x}_j - \bar{\bar{x}})^2 - \frac{\sum T_j^2}{n} - C.F.$$

Substituting the model $x_{ij} = \mu + \tau_j + \epsilon_{ij}$ into T_j^2 we obtain

$$T_j^2 = (n\mu + n\tau_j + \epsilon_{1j} + \ldots + \epsilon_{nj})^2$$

Squaring and taking expectations, recalling that μ and τ_j are constants,

$\sum_j \tau_j = 0$, $E(\epsilon_{ij}) = 0$, and $E(\epsilon_{i'j}\ \epsilon_{ij}) = \begin{cases} \sigma^2, & i=i' \\ 0, & i \neq i' \end{cases}$, we have

$$E(T_j^2) = E(n^2\mu^2) + E(n^2\tau_j^2) + \sum_{i=1}^{n} E(\epsilon_{ij}^2) + E(2n^2\mu\tau_j)$$

$$= n^2\mu^2 + n^2\tau_j^2 + n\sigma^2 + 2n^2\mu\tau_j.$$

Then

$$E\left[\frac{\sum_{j=1}^{k} T_j^2}{n}\right] = \frac{1}{n}\sum_{j=1}^{k} E(T_j^2)$$

$$= \frac{1}{n}\left[kn^2\mu^2 + n^2\sum_{j=1}^{k}\tau_j^2 + kn\sigma^2 + 2n^2k\mu\sum_{j=1}^{k}\tau_j\right]$$

$$= kn\mu^2 + n\sum_j \tau_j^2 + k\sigma^2 .$$

Substituting the model into C.F. gives

$$CF = \frac{1}{nk}\left(nk\mu + n\sum_j \tau_j + \sum_j \sum_i \epsilon_{ij}\right)^2.$$

Squaring and taking expectations term by term,

$$E(C.F.) = \frac{1}{nk}\left[n^2k^2\mu^2 + nk\sigma^2\right] = nk\mu^2 + \sigma^2 .$$

Thus,

$$E(SST) = E(\Sigma T_j^2/n) - E(C.F.)$$

$$= kn\mu^2 + n\Sigma\tau_j^2 + k\sigma^2 - [nk\mu^2 + \sigma^2],$$

$$= (k-1)\sigma^2 + n \sum_j \tau_j^2 .$$

Define the mean square by treatments by $MST = \frac{SST}{k-1}$.

$$\text{Then, } E(MST) = \sigma^2 + \frac{n}{k-1}\Sigma\tau_j^2$$

The error sum of squares (SSE) or, alternatively, the sum of squares within treatments (SSW), can be written as

$$SSW = SSE = \sum_{j=1}^{k} (x_{1j}^2 + x_{2j}^2 \cdots + x_{nj}^2 - \frac{T_j^2}{n}),$$

where

$$x_{ij}^2 = (\mu + \tau_j + \epsilon_{ij})^2 = \mu^2 + \tau_j^2 + \epsilon_{ij}^2 + 2\mu\tau_j + 2\mu\epsilon_{ij} + 2\tau_j\epsilon_{ij}.$$

Then

$$E(x_{ij}^2) = \mu^2 + \tau_j^2 + \sigma^2 + 2\mu\tau_j.$$

Thus,

$$E(SSE) = E[\Sigma\Sigma x_{ij}^2] - E[\sum_j T_j^2/n]$$

$$= nk\mu^2 + n\Sigma\tau_j^2 + nk\sigma^2 - [kn\mu^2 + n\tau_j^2 + k\sigma^2]$$

$$= k(n-1)\sigma^2$$

The mean square for within treatments is $MSW = s_e^2 = \frac{SSE}{k(n-1)}$.

Thus, $E(MSE) = E(s_e^2) = \sigma^2$. Hence, the within treatment or residual or error

mean square calculated in the ANOVA is an unbiased estimate of σ^2, the

variance of ϵ_{ij}. The between treatment mean square has an

expectation of $\sigma^2 + \frac{n}{k-1}\sum_j \tau_j^2$.

Appendix E - Model Building (see Section 7.6)

Listing of Input Data

```
*    Fuel Powder Characteristics
*    X1 = Maximum Particle Size
*    X2 = Average Particle Size
*    X3 = Porosity
*    X4 = Surface Area
*    X5 = Bulk Density
*
*    Y  = Center Grain Size (ASTM No.)
*
=    Grain Size of Fuel Pellets
```

X_1	X_2	X_3	X_4	X_5	y
46.0	1.57	0.7600	6.33	1.45	4.5
46.0	1.57	0.7600	6.33	1.45	3.5
44.5	1.415	0.7725	6.25	1.40	5.5
44.5	1.415	0.7725	6.25	1.40	5.5
44.5	1.415	0.7725	6.25	1.40	5.5
47.0	1.52	0.7735	6.51	1.44	3.5
47.0	1.52	0.7735	6.51	1.44	3.5
44.5	1.58	0.7725	6.34	1.48	3.5
44.5	1.50	0.7725	6.34	1.48	3.5
46.0	1.86	0.7750	7.08	1.51	8.5
46.0	1.86	0.7750	7.08	1.51	7.0
37.0	1.925	0.7800	6.74	1.52	11.0
37.0	1.925	0.7800	6.74	1.52	11.5
41.0	1.765	0.7790	6.80	1.48	9.5
41.0	1.765	0.7790	6.80	1.48	9.5
41.0	1.765	0.7790	6.80	1.48	10.0
44.5	1.705	0.7745	6.38	1.48	3.5
44.5	1.705	0.7745	6.38	1.48	1.5
42.5	1.690	0.7770	6.32	1.46	5.5

A selection of the 31 possible first order regression models for the 5 independent variables listed above are given in the next pages. The results needed to follow through the steps of backwards elimination and forward selection are included. The final model and the residuals are provided. The results presented can be obtained from most standard least squares analysis computer programs or can be obtained by hand calculations by following the procedures discussed in Chapter 7.

GRAIN SIZE OF FUEL PELLETS

RESPONSE VARIABLE Y(1) (MEAN = 6.10526)

CASE 1 REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	40.8545			
(1)	1	-0.796424	.150823	-5.28**	43.6316

(*, ** DENOTE SIGNIFICANT AT ALPHA = 0.05, 0.01, RESPECTIVELY)

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	1	101.7536	101.7536	27.88
RESIDUAL	17	62.03584	3.649167	(S.D. = 1.91)
LACK OF FIT	4	32.34239	8.085598	3.54
REPLICATES	13	29.69345	2.284112	
TOTAL (CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 62.1 PERCENT

CASE 2 REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	-13.8779			
(2)	2	12.0648	3.12495	3.86**	1.65632

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY)

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	1	76.51939	76.51939	14.91
RESIDUAL	17	87.27009	5.133535	(S.D. = 2.27)
LACK OF FIT	8	83.35342	10.41918	23.94
REPLICATES	9	3.916667	0.4351852	
TOTAL (CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 46.7 PERCENT

CASE 3

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	-248.963			
(3)	3	329.624	103.751	3.18**	0.773816

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F-RATIO
REGRESSION	1	61.01942	61.01942	10.09
RESIDUAL	17	102.7700	6.045297	(S.D. = 2.46)
LACK OF FIT	6	94.05338	15.67556	19.78
REPLICATES	11	8.716667	0.7924242	
TOTAL (CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 37.3 PERCENT

CASE 4

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	-43.4005			
(1)	1	7.57152	1.88912	4.01**	6.53842

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F-RATIO
REGRESSION	1	79.56276	79.56276	16.06
RESIDUAL	17	84.22672	4.954513	(S.D. = 2.23)
LACK OF FIT	7	80.31005	11.47286	29.29
REPLICATES	10	3.916667	0.3916667	
TOTAL (CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 48.6 PERCENT

CASE 5

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	-50.3337			
(5)	5	38.4903	16.8172	2.29*	1.46632

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	1	38.58146	38.58146	5.24
RESIDUAL	17	125.2080	7.365177	(S.D. = 2.71)
LACK OF FIT	5	44.10087	8.820174	1.30
REPLICATES	12	81.10714	6.758929	
TOTAL (CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 23.6 PERCENT

CASE 6

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	-1.78431			
(1)	1	-0.643029	0.101078	-6.36**	43.6316
(4)	4	5.49765	1.08671	5.06**	6.53842

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	2	139.9256	69.96279	46.91
RESIDUAL	16	23.86389	1.491493	(S.D. = 1.22)
LACK OF FIT	6	19.94723	3.324538	8.49
REPLICATES	10	3.916667	0.3916667	
TOTAL(CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 85.4 PERCENT

CASE 7

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	37.1473			
(1)	1	-0.779617	7.294200E-02	-10.69**	43.6316
(4)	4	8.67526	0.990771	8.76**	6.53842
(5)	5	- 36.6556	7.87631	-4.65**	1.46632

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	3	154.0249	51.34163	78.87
RESIDUAL	15	9.764582	0.6509721	(S.D. = 0.81)
LACK OF FIT	5	5.847915	1.169583	2.99
REPLICATES	10	3.916667	0.3916667	
TOTAL(CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 94.0 PERCENT

CASE 8

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	122.228			
(1)	1	-0.943688	9.534583E-02	-9.90**	43.6316
(2)	2	-2.66410	3.53233	-0.75	1.65632
(3)	3	-115.534	43.6176	-2.65*	0.773816
(4)	4	10.3221	1.22214	8.45**	6.53842
(5)	5	-33.1608	11.1524	-2.97*	1.46632

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	5	157.5117	31.50234	65.23
RESIDUAL	13	6.277795	0.4829073	(S.D.=0.69)
LACK OF FIT	4	2.361128	0.5902821	1.36
REPLICATES	9	3.916667	0.4351852	
TOTAL(CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 96.2 PERCENT

FINAL MODEL: $y = 126.631 - 0.904X_1 - 111.7X_3 + 9.74X_4 - 39.77X_5$

CASE 9

REGRESSION COEFFICIENTS

NO.	I	B(I)	STD. DEV. OF B	T-VALUE	MEAN OF X
(0)	0	126.631			
(1)	1	-0.903525	7.786194E-02	-11.60**	43.6316
(3)	3	-111.743	42.6547	-2.62*	0.773816
(4)	4	9.74044	0.933323	10.44**	6.53842
(5)	5	-39.7744	6.78380	-5.86**	1.46632

(*, ** DENOTE SIGNIFICANCE AT ALPHA = 0.05, 0.01, RESPECTIVELY))

ANALYSIS OF VARIANCE TABLE

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F=RATIO
REGRESSION	4	157.2370	39.30925	83.99
RESIDUAL	14	6.552484	0.4680346	(S.D.= 0.68)
LACK OF FIT	4	2.635818	0.6589545	1.68
REPLICATES	10	3.916667	0.3916667	
TOTAL(CORRECTED)	18	163.7895		

SUM OF SQUARES ACCOUNTED FOR BY REGRESSION = 96.2 PERCENT

Correlation of Parameter Estimates

I=	J= 1	3	4	5
1	1.00	0.61	-0.32	0.42
3		1.00	-0.44	0.18
4			1.00	- 0.69
5				1.00

DISTRIBUTION OF DEVIATIONS FROM REGRESSION

CLASS UPPER LIMIT	FREQUENCY	C.D.F.	C.D.F.(%)
1.25000E+00	1	19	100.0
1.10000E+00	0	18	94.7
9.50000E-01	0	18	94.7
8.00000E-01	0	18	94.7
6.50000E-01	3	18	94.7
5.00000E-01	1	15	78.9
3.50000E-01	5	14	73.7
2.00000E-01	1	9	47.4
5.00000E-02	0	8	42.1
-1.00000E-01	1	8	42.1
-2.50000E-01	3	7	36.8
-4.00000E-01	2	4	21.1
-5.50000E-01	1	2	10.5
-7.00000E-01	0	1	5.3
-8.50000E-01	0	1	5.3
-1.00000E+00	0	1	5.3
-1.15000E+00	0	1	5.3
-1.30000E+00	0	1	5.3
-1.45000E+00	1	1	5.3

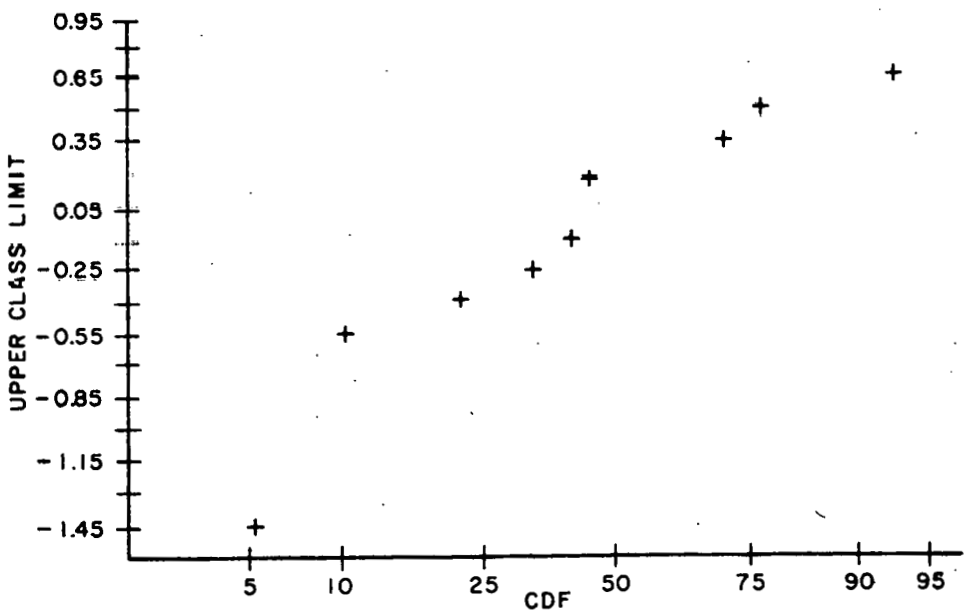
SKEWNESS = G1 = -0.86, T-TEST OF G1 = -1.64

KURTOSIS = G2= 2.13, T-TEST OF G2 = 2.10

SHAPIRO-WILK W STATISTIC = 0.926.

WHICH IS SIGNIFICANT AT THE ALPHA = 0.50 LEVEL

NORMAL PROBABILITY PLOT



PREDICTION AND DEVIATIONS

OBSERVED	PREDICTED	DEVIATION
4.500	4.128	0.3720
5.500	5.296	0.2040
3.500	3.867	-0.3670
8.500	7.371	1.129
11.50	11.23	0.2658
10.000	9.907	9.2727E-02
5.500	4.896	0.6045
3.500	4.128	-0.6280
5.500	5.296	0.2040
3.500	2.991	0.5093
7.000	7.371	-0.3707
9.500	9.907	-0.4073
3.500	3.157	0.3432
5.500	5.296	0.2040
3.500	3.867	-0.3670
3.500	2.991	0.5093
11.00	11.23	-0.2342
9.500	9.907	-0.4073
1.500	3.157	-1.657

Blank

LIST OF STATISTICAL TABLES

Table 1
Binomial Distribution Function*

$$Pr(k \leq x) = B(x; n, p) = \sum_{k=0}^{x} \binom{n}{k} p^k (1 - p)^{n-k}$$

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6562
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0570	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980

*Reprinted by permission from Irwin Miller and John E. Freund Probability and Statistics for Engineers, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, Copyright 1965.

Table I

BINOMIAL DISTRIBUTION FUNCTION (*Continued*)

n	x	p 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0232	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009

Table I

BINOMIAL DISTRIBUTION FUNCTION (*Continued*)

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
14	2	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9996	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245

Table I

BINOMIAL DISTRIBUTION FUNCTION *(Continued)*

n	x	p 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
17	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	0.9868	0.8850	0.6841	0.4551	0.2630	0.1332	0.0591	0.0230	0.0077	0.0022
	4	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table I

BINOMIAL DISTRIBUTION FUNCTION (*Continued*)

n	x	p 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9420
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II
Poisson Distribution Function*

$$\Pr(k \leq x) = F(x; \lambda) = \sum_{k=0}^{x} e^{-\lambda} \frac{\lambda^k}{k!}$$

λ \ x	0	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	0.999	1.000						
0.20	0.819	0.982	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.994	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.55	0.577	0.894	0.982	0.998	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.65	0.522	0.861	0.972	0.996	0.999	1.000				
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.75	0.472	0.827	0.959	0.993	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.85	0.427	0.791	0.945	0.989	0.998	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
0.95	0.387	0.754	0.929	0.984	0.997	1.000				
1.00	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.998	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.833	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.995	0.999	1.000	

*Reproduced with permission from E.C. Molina, Poisson's Exponential Binomial Limit, D. Van Nostrand Company, Inc., Princeton, New Jersey, 1947.

Table II

POISSON DISTRIBUTION FUNCTION (*Continued*)

λ \ x	0	1	2	3	4	5	6	7	8	9
2.2	0.111	0.355	0.623	0.819	0.928	0.975	0.993	0.998	1.000	
2.4	0.091	0.308	0.570	0.779	0.904	0.964	0.988	0.997	0.999	1.000
2.6	0.074	0.267	0.518	0.736	0.877	0.951	0.983	0.995	0.999	1.000
2.8	0.061	0.231	0.469	0.692	0.848	0.935	0.976	0.992	0.998	0.999
3.0	0.050	0.199	0.423	0.647	0.815	0.916	0.966	0.988	0.996	0.999
3.2	0.041	0.171	0.380	0.603	0.781	0.895	0.955	0.983	0.994	0.998
3.4	0.033	0.147	0.340	0.558	0.744	0.871	0.942	0.977	0.992	0.997
3.6	0.027	0.126	0.303	0.515	0.706	0.844	0.927	0.969	0.988	0.996
3.8	0.022	0.107	0.269	0.473	0.668	0.816	0.909	0.960	0.984	0.994
4.0	0.018	0.092	0.238	0.433	0.629	0.785	0.889	0.949	0.979	0.992
4.2	0.015	0.078	0.210	0.395	0.590	0.753	0.867	0.936	0.972	0.989
4.4	0.012	0.066	0.185	0.359	0.551	0.720	0.844	0.921	0.964	0.985
4.6	0.010	0.056	0.163	0.326	0.513	0.686	0.818	0.905	0.955	0.980
4.8	0.008	0.048	0.143	0.294	0.476	0.651	0.791	0.887	0.944	0.975
5.0	0.007	0.040	0.125	0.265	0.440	0.616	0.762	0.867	0.932	0.968
5.2	0.006	0.034	0.109	0.238	0.406	0.581	0.732	0.845	0.918	0.960
5.4	0.005	0.029	0.095	0.213	0.373	0.546	0.702	0.822	0.903	0.951
5.6	0.004	0.024	0.082	0.191	0.342	0.512	0.670	0.797	0.886	0.941
5.8	0.003	0.021	0.072	0.170	0.313	0.478	0.638	0.771	0.867	0.929
6.0	0.002	0.017	0.062	0.151	0.285	0.446	0.606	0.744	0.847	0.916

λ \ x	10	11	12	13	14	15	16
2.8	1.000						
3.0	1.000						
3.2	1.000						
3.4	0.999	1.000					
3.6	0.999	1.000					
3.8	0.998	0.999	1.000				
4.0	0.997	0.999	1.000				
4.2	0.996	0.999	1.000				
4.4	0.994	0.998	0.999	1.000			
4.6	0.992	0.997	0.999	1.000			
4.8	0.990	0.996	0.999	1.000			
5.0	0.986	0.995	0.998	0.999	1.000		
5.2	0.982	0.993	0.997	0.999	1.000		
5.4	0.977	0.990	0.996	0.999	1.000		
5.6	0.972	0.988	0.995	0.998	0.999	1.000	
5.8	0.965	0.984	0.993	0.997	0.999	1.000	
6.0	0.957	0.980	0.991	0.996	0.999	0.999	1.000

Table II

POISSON DISTRIBUTION FUNCTION (*Continued*)

λ \ x	0	1	2	3	4	5	6	7	8	9
6.2	0.002	0.015	0.054	0.134	0.259	0.414	0.574	0.716	0.826	0.902
6.4	0.002	0.012	0.046	0.119	0.235	0.384	0.542	0.687	0.803	0.886
6.6	0.001	0.010	0.040	0.105	0.213	0.355	0.511	0.658	0.780	0.869
6.8	0.001	0.009	0.034	0.093	0.192	0.327	0.480	0.628	0.755	0.850
7.0	0.001	0.007	0.030	0.082	0.173	0.301	0.450	0.599	0.729	0.830
7.2	0.001	0.006	0.025	0.072	0.156	0.276	0.420	0.569	0.703	0.810
7.4	0.001	0.005	0.022	0.063	0.140	0.253	0.392	0.539	0.676	0.788
7.6	0.001	0.004	0.019	0.055	0.125	0.231	0.365	0.510	0.648	0.765
7.8	0.000	0.004	0.016	0.048	0.112	0.210	0.338	0.481	0.620	0.741
8.0	0.000	0.003	0.014	0.042	0.100	0.191	0.313	0.453	0.593	0.717
8.5	0.000	0.002	0.009	0.030	0.074	0.150	0.256	0.386	0.523	0.653
9.0	0.000	0.001	0.006	0.021	0.055	0.116	0.207	0.324	0.456	0.587
9.5	0.000	0.001	0.004	0.015	0.040	0.089	0.165	0.269	0.392	0.522
10.0	0.000	0.000	0.003	0.010	0.029	0.067	0.130	0.220	0.333	0.458

λ \ x	10	11	12	13	14	15	16	17	18	19
6.2	0.949	0.975	0.989	0.995	0.998	0.999	1.000			
6.4	0.939	0.969	0.986	0.994	0.997	0.999	1.000			
6.6	0.927	0.963	0.982	0.992	0.997	0.999	0.999	1.000		
6.8	0.915	0.955	0.978	0.990	0.996	0.998	0.999	1.000		
7.0	0.901	0.947	0.973	0.987	0.994	0.998	0.999	1.000		
7.2	0.887	0.937	0.967	0.984	0.993	0.997	0.999	0.999	1.000	
7.4	0.871	0.926	0.961	0.980	0.991	0.996	0.998	0.999	1.000	
7.6	0.854	0.915	0.954	0.976	0.989	0.995	0.998	0.999	1.000	
7.8	0.835	0.902	0.945	0.971	0.986	0.993	0.997	0.999	1.000	
8.0	0.816	0.888	0.936	0.966	0.983	0.992	0.996	0.998	0.999	1.000
8.5	0.763	0.849	0.909	0.949	0.973	0.986	0.993	0.997	0.999	0.999
9.0	0.706	0.803	0.875	0.926	0.959	0.978	0.989	0.995	0.998	0.999
9.5	0.645	0.752	0.836	0.898	0.940	0.967	0.982	0.991	0.996	0.998
10.0	0.583	0.697	0.792	0.864	0.917	0.951	0.973	0.986	0.993	0.997

λ \ x	20	21	22
8.5	1.000		
9.0	1.000		
9.5	0.999	1.000	
10.0	0.998	0.999	1.000

Table II

POISSON DISTRIBUTION FUNCTION (*Continued*)

λ \ x	0	1	2	3	4	5	6	7	8	9
10.5	0.000	0.000	0.002	0.007	0.021	0.050	0.102	0.179	0.279	0.397
11.0	0.000	0.000	0.001	0.005	0.015	0.033	0.079	0.143	0.232	0.341
11.5	0.000	0.000	0.001	0.003	0.011	0.028	0.060	0.114	0.191	0.289
12.0	0.000	0.000	0.001	0.002	0.008	0.020	0.046	0.090	0.155	0.242
12.5	0.000	0.000	0.000	0.002	0.005	0.015	0.035	0.070	0.125	0.201
13.0	0.000	0.000	0.000	0.001	0.004	0.011	0.026	0.054	0.100	0.166
13.5	0.000	0.000	0.000	0.001	0.003	0.008	0.019	0.041	0.079	0.135
14.0	0.000	0.000	0.000	0.000	0.002	0.006	0.014	0.032	0.062	0.109
14.5	0.000	0.000	0.000	0.000	0.001	0.004	0.010	0.024	0.048	0.088
15.0	0.000	0.000	0.000	0.000	0.001	0.003	0.008	0.018	0.037	0.070

λ \ x	10	11	12	13	14	15	16	17	18	19
10.5	0.521	0.639	0.742	0.825	0.888	0.932	0.960	0.978	0.988	0.994
11.0	0.460	0.579	0.689	0.781	0.854	0.907	0.944	0.968	0.982	0.991
11.5	0.402	0.520	0.633	0.733	0.815	0.878	0.924	0.954	0.974	0.986
12.0	0.347	0.462	0.576	0.682	0.772	0.844	0.899	0.937	0.963	0.979
12.5	0.297	0.406	0.519	0.628	0.725	0.806	0.869	0.916	0.948	0.969
13.0	0.252	0.353	0.463	0.573	0.675	0.764	0.835	0.890	0.930	0.957
13.5	0.211	0.304	0.409	0.518	0.623	0.718	0.798	0.861	0.908	0.942
14.0	0.176	0.260	0.358	0.464	0.570	0.669	0.756	0.827	0.883	0.923
14.5	0.145	0.220	0.311	0.413	0.518	0.619	0.711	0.790	0.853	0.901
15.0	0.118	0.185	0.268	0.363	0.466	0.568	0.664	0.749	0.819	0.875

λ \ x	20	21	22	23	24	25	26	27	28	29
10.5	0.997	0.999	0.999	1.000						
11.0	0.995	0.998	0.999	1.000						
11.5	0.992	0.996	0.998	0.999	1.000					
12.0	0.988	0.994	0.997	0.999	0.999	1.000				
12.5	0.983	0.991	0.995	0.998	0.999	0.999	1.000			
13.0	0.975	0.986	0.992	0.996	0.998	0.999	1.000			
13.5	0.965	0.980	0.989	0.994	0.997	0.998	0.999	1.000		
14.0	0.952	0.971	0.983	0.991	0.995	0.997	0.999	0.999	1.000	
14.5	0.936	0.960	0.976	0.986	0.992	0.996	0.998	0.999	0.999	1.000
15.0	0.917	0.947	0.967	0.981	0.989	0.994	0.997	0.998	0.999	1.000

Table II

POISSON DISTRIBUTION FUNCTION (*Continued*)

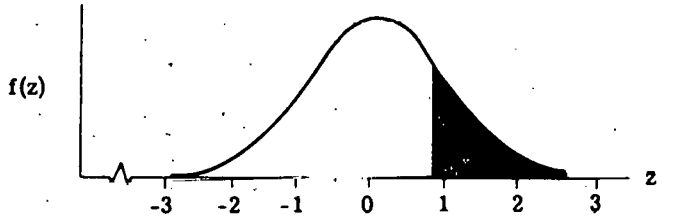
λ \ x	4	5	6	7	8	9	10	11	12	13
16	0.000	0.001	0.004	0.010	0.022	0.043	0.077	0.127	0.193	0.275
17	0.000	0.001	0.002	0.005	0.013	0.026	0.049	0.085	0.135	0.201
18	0.000	0.000	0.001	0.003	0.007	0.015	0.030	0.055	0.092	0.143
19	0.000	0.000	0.001	0.002	0.004	0.009	0.018	0.035	0.061	0.098
20	0.000	0.000	0.000	0.001	0.002	0.005	0.011	0.021	0.039	0.066
21	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.013	0.025	0.043
22	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.008	0.015	0.028
23	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.009	0.017
24	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.011
25	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.006

λ \ x	14	15	16	17	18	19	20	21	22	23
16	0.368	0.467	0.566	0.659	0.742	0.812	0.868	0.911	0.942	0.963
17	0.281	0.371	0.468	0.564	0.655	0.736	0.805	0.861	0.905	0.937
18	0.208	0.287	0.375	0.469	0.562	0.651	0.731	0.799	0.855	0.899
19	0.150	0.215	0.292	0.378	0.469	0.561	0.647	0.725	0.793	0.849
20	0.105	0.157	0.221	0.297	0.381	0.470	0.559	0.644	0.721	0.787
21	0.072	0.111	0.163	0.227	0.302	0.384	0.471	0.558	0.640	0.716
22	0.048	0.077	0.117	0.169	0.232	0.306	0.387	0.472	0.556	0.637
23	0.031	0.052	0.082	0.123	0.175	0.238	0.310	0.389	0.472	0.555
24	0.020	0.034	0.056	0.087	0.128	0.180	0.243	0.314	0.392	0.473
25	0.012	0.022	0.038	0.060	0.092	0.134	0.185	0.247	0.318	0.394

λ \ x	24	25	26	27	28	29	30	31	32	33
16	0.978	0.987	0.993	0.996	0.998	0.999	0.999	1.000		
17	0.959	0.975	0.985	0.991	0.995	0.997	0.999	0.999	1.000	
18	0.932	0.955	0.972	0.983	0.990	0.994	0.997	0.998	0.999	1.000
19	0.893	0.927	0.951	0.969	0.980	0.988	0.993	0.996	0.998	0.999
20	0.843	0.888	0.922	0.948	0.966	0.978	0.987	0.992	0.995	0.997
21	0.782	0.838	0.883	0.917	0.944	0.963	0.976	0.985	0.991	0.994
22	0.712	0.777	0.832	0.877	0.913	0.940	0.959	0.973	0.983	0.989
23	0.635	0.708	0.772	0.827	0.873	0.908	0.936	0.956	0.971	0.981
24	0.554	0.632	0.704	0.768	0.823	0.868	0.904	0.932	0.953	0.969
25	0.473	0.553	0.629	0.700	0.763	0.818	0.863	0.900	0.929	0.950

λ \ x	34	35	36	37	38	39	40	41	42	43
19	0.999	1.000								
20	0.999	0.999	1.000							
21	0.997	0.998	0.999	0.999	1.000					
22	0.994	0.996	0.998	0.999	0.999	1.000				
23	0.988	0.993	0.996	0.997	0.999	0.999	1.000			
24	0.979	0.987	0.992	0.995	0.997	0.998	0.999	0.999	1.000	
25	0.966	0.978	0.985	0.991	0.994	0.997	0.998	0.999	0.999	1.000

Table III
Normal Distribution*



Area under the Normal distribution to the right of z

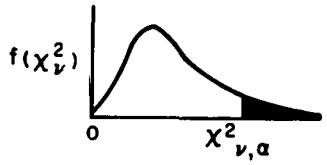
EXAMPLE: Prob($z \geq 0.84$) = 0.2005

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100

*Reproduced with permission of the Biometrika trustees from E.S. Pearson and H.O. Hartley, Biometrika Tables for Statisticians, Volume 1, Cambridge University Press, New York, 1956.

TABLE IV

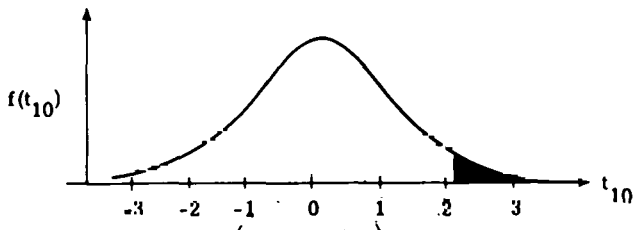
Chi-Square Distribution $Pr(\chi^2_\nu > \chi^2_{\nu,\alpha}) = \alpha$



ν	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	26.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.526	16.047	17.708	42.557	45.722	49.588	52.336	29
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67	30
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77	40
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95	60

*Adapted with permission of the Biometrika trustees from E. S. Pearson and H. A. Hartley, Biometrika Tables for Statisticians, Volume 1, Cambridge Univ. Press, New York, 1956.

Table V
t-Distribution*



EXAMPLE: Prob $\left(t_{10} \geq 2.228\right) = 0.025$

Percentage Points of the t distribution

$Q = 1 - P(t|\nu)$

ν	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001
1	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307
60	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232
120	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160
∞	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090

$Q = 1 - P(t|\nu)$ is the upper-tail area of the distribution for ν degrees of freedom, appropriate for use in a single-tailed test. For a two-tailed test, 2Q must be used.

*Adapted with permission of Biometrika trustees from E.S. Pearson and H.O. Harley, Biometrika Tables for Statisticians. Volume 1, Cambridge Univ. Press, New York, 1956.

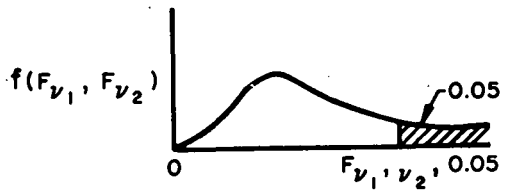
Table VI
The F-Distribution*

$$Pr(F_{\nu_1, \nu_2} > F_{\nu_1, \nu_2, \alpha}) = \alpha$$

ν_1 is degrees of freedom for numerator

ν_2 is degrees of freedom for denominator of ratio of two variance estimates

(a) $\alpha = 0.05$



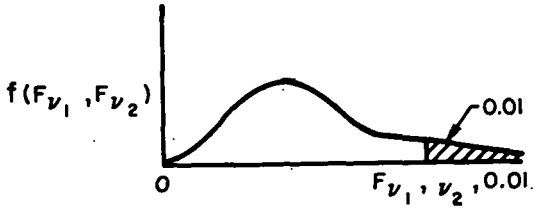
ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4 12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

*Adapted with permission of Biometrika trustees from E.S. Pearson, and H.O. Hartley, Biometrika Tables for Statisticians, Volume I, Cambridge Univ. Press, New York, 1956.

Table VI

$$\Pr(F_{\nu_1,\nu_2} > F_{\nu_1,\nu_2,\alpha}) = \alpha$$

(b) $\alpha = 0.01$



$P = 0.99$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Table VII(a) Factors for Two-Sided Tolerance Limits for Normal Distributions*

Factors K such that the probability is γ that at least a proportion P of the distribution will be included between $\bar{X} \pm Ks$, where $\bar{X}$ and s are estimates of the mean and the standard deviation computed from a sample size of n.

Example For $\gamma = 0.90$, P = 0.95, n = 20, then K = 2.564

	$\gamma = 0.75$					$\gamma = 0.90$				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	4.498	6.301	7.414	9.531	11.920	11.407	15.978	18.800	24.167	30.227
3	2.501	3.538	4.187	5.431	6.844	4.132	5.847	6.919	8.974	11.309
4	2.035	2.892	3.431	4.471	5.657	2.932	4.166	4.943	6.440	8.149
5	1.825	2.599	3.088	4.033	5.117	2.454	3.494	4.152	5.423	6.879
6	1.704	2.429	2.889	3.779	4.802	2.196	3.131	3.723	4.870	6.188
7	1.624	2.318	2.757	3.611	4.593	2.034	2.902	3.452	4.521	5.750
8	1.568	2.238	2.663	3.491	4.444	1.921	2.743	3.264	4.278	5.446
9	1.525	2.178	2.593	3.400	4.330	1.839	2.626	3.125	4.098	5.220
10	1.492	2.131	2.537	3.328	4.241	1.775	2.535	3.018	3.959	5.046
11	1.465	2.093	2.493	3.271	4.169	1.724	2.463	2.933	3.849	4.906
12	1.443	2.062	2.456	3.223	4.110	1.683	2.404	2.863	3.758	4.792
13	1.425	2.036	2.424	3.183	4.059	1.648	2.355	2.805	3.682	4.697
14	1.409	2.013	2.398	3.148	4.016	1.619	2.314	2.756	3.618	4.615
15	1.395	1.994	2.375	3.118	3.979	1.594	2.278	2.713	3.562	4.545
16	1.383	1.977	2.355	3.092	3.946	1.572	2.246	2.676	3.514	4.484
17	1.372	1.962	2.337	3.069	3.917	1.552	2.219	2.643	3.471	4.430
18	1.363	1.948	2.321	3.048	3.891	1.535	2.194	2.614	3.433	4.382
19	1.355	1.936	2.307	3.030	3.867	1.520	2.172	2.588	3.399	4.339
20	1.347	1.925	2.294	3.013	3.846	1.506	2.152	2.564	3.368	4.300
21	1.340	1.915	2.282	2.998	3.827	1.493	2.135	2.543	3.340	4.264
22	1.334	1.906	2.271	2.984	3.809	1.482	2.118	2.524	3.315	4.232
23	1.328	1.898	2.261	2.971	3.793	1.471	2.103	2.506	3.292	4.203
24	1.322	1.891	2.252	2.959	3.778	1.462	2.089	2.489	3.270	4.176
25	1.317	1.883	2.244	2.948	3.764	1.453	2.077	2.474	3.251	4.151
26	1.313	1.877	2.236	2.938	3.751	1.444	2.065	2.460	3.232	4.127
27	1.309	1.871	2.229	2.929	3.740	1.437	2.054	2.447	3.215	4.106

*Adapted from Techniques of Statistical Analysis by C. Eisenhart, M.W. Hastay, and W.A. Wallis, Copyright 1947. Used with permission of McGraw-Hill Book Company.

Table VII(a) Continued

n \ P	$\gamma = 0.95$					$\gamma = 0.99$				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	22.858	32.019	37.674	48.430	60.573	114.363	160.193	188.491	242.300	303.054
3	5.922	8.380	9.916	12.861	16.208	13.378	18.930	22.401	29.055	36.616
4	3.779	5.369	6.370	8.299	10.502	6.614	9.398	11.150	14.527	18.383
5	3.002	4.275	5.079	6.634	8.415	4.643	6.612	7.855	10.260	13.015
6	2.604	3.712	4.414	5.775	7.337	3.743	5.337	6.345	8.301	10.548
7	2.361	3.369	4.007	5.248	6.676	3.233	4.613	5.488	7.187	9.142
8	2.197	3.136	3.732	4.891	6.226	2.905	4.147	4.936	6.468	8.234
9	2.078	2.967	3.532	4.631	5.899	2.677	3.822	4.550	5.966	7.600
10	1.987	2.839	3.379	4.433	5.649	2.508	3.582	4.265	5.594	7.129
11	1.916	2.737	3.259	4.277	5.452	2.378	3.397	4.045	5.308	6.766
12	1.858	2.655	3.162	4.150	5.291	2.274	3.250	3.870	5.079	6.477
13	1.810	2.587	3.081	4.044	5.158	2.190	3.130	3.727	4.893	6.240
14	1.770	2.529	3.012	3.955	5.045	2.120	3.029	3.608	4.737	6.043
15	1.735	2.480	2.954	3.878	4.949	2.060	2.945	3.507	4.605	5.876
16	1.705	2.437	2.903	3.812	4.865	2.009	2.872	3.421	4.492	5.732
17	1.679	2.400	2.858	3.754	4.791	1.965	2.808	3.345	4.393	5.607
18	1.655	2.366	2.819	3.702	4.725	1.926	2.753	3.279	4.307	5.497
19	1.635	2.337	2.784	3.656	4.667	1.891	2.703	3.221	4.230	5.399
20	1.616	2.310	2.752	3.615	4.614	1.860	2.659	3.168	4.161	5.312
21	1.599	2.286	2.723	3.577	4.567	1.833	2.620	3.121	4.100	5.234
22	1.584	2.264	2.697	3.543	4.523	1.808	2.584	3.078	4.044	5.163
23	1.570	2.244	2.673	3.512	4.484	1.785	2.551	3.040	3.993	5.098
24	1.557	2.225	2.651	3.483	4.447	1.764	2.522	3.004	3.947	5.039
25	1.545	2.208	2.631	3.457	4.413	1.745	2.494	2.972	3.904	4.985
26	1.534	2.193	2.612	3.432	4.382	1.727	2.469	2.941	3.865	4.935
27	1.523	2.178	2.595	3.409	4.353	1.711	2.446	2.914	3.828	4.888

FACTORS FOR TWO-SIDED TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

	$\gamma = 0.75$					$\gamma = 0.90$				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
30	1.297	1.855	2.210	2.904	3.708	1.417	2.025	2.413	3.170	4.049
35	1.283	1.834	2.185	2.871	3.667	1.390	1.988	2.368	3.112	3.974
40	1.271	1.818	2.166	2.846	3.635	1.370	1.959	2.334	3.066	3.917
45	1.262	1.805	2.150	2.826	3.609	1.354	1.935	2.306	3.030	3.871
50	1.255	1.794	2.138	2.809	3.588	1.340	1.916	2.284	3.001	3.833
55	1.249	1.785	2.127	2.795	3.571	1.329	1.901	2.265	2.976	3.801
60	1.243	1.778	2.118	2.784	3.556	1.320	1.887	2.248	2.955	3.774
65	1.239	1.771	2.110	2.773	3.543	1.312	1.875	2.235	2.937	3.751
70	1.235	1.765	2.104	2.764	3.531	1.304	1.865	2.222	2.920	3.730
75	1.231	1.760	2.098	2.757	3.521	1.298	1.856	2.211	2.906	3.712
80	1.228	1.756	2.092	2.749	3.512	1.292	1.848	2.202	2.894	3.696
85	1.225	1.752	2.087	2.743	3.504	1.287	1.841	2.193	2.882	3.682
90	1.223	1.748	2.083	2.737	3.497	1.283	1.834	2.185	2.872	3.669
95	1.220	1.745	2.079	2.732	3.490	1.278	1.828	2.178	2.863	3.657
100	1.218	1.742	2.075	2.727	3.484	1.275	1.822	2.172	2.854	3.646
110	1.214	1.736	2.069	2.719	3.473	1.268	1.813	2.160	2.839	3.626
120	1.211	1.732	2.063	2.712	3.464	1.262	1.804	2.150	2.826	3.610
130	1.208	1.728	2.059	2.705	3.456	1.257	1.797	2.141	2.814	3.595
140	1.206	1.724	2.054	2.700	3.449	1.252	1.791	2.134	2.804	3.582
150	1.204	1.721	2.051	2.695	3.443	1.248	1.785	2.127	2.795	3.571
160	1.202	1.718	2.047	2.691	3.437	1.245	1.780	2.121	2.787	3.561
170	1.200	1.716	2.044	2.687	3.432	1.242	1.775	2.116	2.780	3.552
180	1.198	1.713	2.042	2.683	3.427	1.239	1.771	2.111	2.774	3.543
190	1.197	1.711	2.039	2.680	3.423	1.236	1.767	2.106	2.768	3.536
200	1.195	1.709	2.037	2.677	3.419	1.234	1.764	2.102	2.762	3.529
250	1.190	1.702	2.028	2.665	3.404	1.224	1.750	2.085	2.740	3.501
300	1.186	1.696	2.021	2.656	3.393	1.217	1.740	2.073	2.725	3.481
400	1.181	1.688	2.012	2.644	3.378	1.207	1.726	2.057	2.703	3.453
500	1.177	1.683	2.006	2.636	3.368	1.201	1.717	2.046	2.689	3.434
600	1.175	1.680	2.002	2.631	3.360	1.196	1.710	2.038	2.678	3.421
700	1.173	1.677	1.998	2.626	3.355	1.192	1.705	2.032	2.670	3.411
800	1.171	1.675	1.996	2.623	3.350	1.189	1.701	2.027	2.663	3.402
900	1.170	1.673	1.993	2.620	3.347	1.187	1.697	2.023	2.658	3.396
1000	1.169	1.671	1.992	2.617	3.344	1.185	1.695	2.019	2.654	3.390
∞	1.150	1.645	1.960	2.576	3.291	1.150	1.645	1.960	2.576	3.291

Table VII(a) Continued

FACTORS FOR TWO-SIDED TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

	$\gamma = 0.95$					$\gamma = 0.99$				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
30	1.497	2.140	2.549	3.350	4.278	1.668	2.385	2.841	3.733	4.768
35	1.462	2.090	2.490	3.272	4.179	1.613	2.306	2.748	3.611	4.611
40	1.435	2.052	2.445	3.213	4.104	1.571	2.247	2.677	3.518	4.493
45	1.414	2.021	2.408	3.165	4.042	1.539	2.200	2.621	3.444	4.399
50	1.396	1.996	2.379	3.126	3.993	1.512	2.162	2.576	3.385	4.323
55	1.382	1.976	2.354	3.094	3.951	1.490	2.130	2.538	3.335	4.260
60	1.369	1.958	2.333	3.066	3.916	1.471	2.103	2.506	3.293	4.206
65	1.359	1.943	2.315	3.042	3.886	1.455	2.080	2.478	3.257	4.160
70	1.349	1.929	2.299	3.021	3.859	1.440	2.060	2.454	3.225	4.120
75	1.341	1.917	2.285	3.002	3.835	1.428	2.042	2.433	3.197	4.084
80	1.334	1.907	2.272	2.986	3.814	1.417	2.026	2.414	3.173	4.053
85	1.327	1.897	2.261	2.971	3.795	1.407	2.012	2.397	3.150	4.024
90	1.321	1.889	2.251	2.958	3.778	1.398	1.999	2.382	3.130	3.999
95	1.315	1.881	2.241	2.945	3.763	1.390	1.987	2.368	3.112	3.976
100	1.311	1.874	2.233	2.934	3.748	1.383	1.977	2.355	3.096	3.954
110	1.302	1.861	2.218	2.915	3.723	1.369	1.958	2.333	3.066	3.917
120	1.294	1.850	2.205	2.898	3.702	1.358	1.942	2.314	3.041	3.885
130	1.288	1.841	2.194	2.883	3.683	1.349	1.928	2.298	3.019	3.857
140	1.282	1.833	2.184	2.870	3.666	1.340	1.916	2.283	3.000	3.833
150	1.277	1.825	2.175	2.859	3.652	1.332	1.905	2.270	2.983	3.811
160	1.272	1.819	2.167	2.848	3.638	1.326	1.896	2.259	2.968	3.792
170	1.268	1.813	2.160	2.839	3.527	1.320	1.887	2.248	2.955	3.774
180	1.264	1.808	2.154	2.831	3.616	1.314	1.879	2.239	2.942	3.759
190	1.261	1.803	2.148	2.823	3.606	1.309	1.872	2.230	2.931	3.744
200	1.258	1.798	2.143	2.816	3.597	1.304	1.865	2.222	2.921	3.731
250	1.245	1.780	2.121	2.788	3.561	1.286	1.839	2.191	2.880	3.678
300	1.236	1.767	2.106	2.767	3.535	1.273	1.820	2.169	2.850	3.641
400	1.223	1.749	2.084	2.739	3.499	1.255	1.794	2.138	2.809	3.589
500	1.215	1.737	2.070	2.721	3.475	1.243	1.777	2.117	2.783	3.555
600	1.209	1.729	2.060	2.707	3.458	1.234	1.764	2.102	2.763	3.530
700	1.204	1.722	2.052	2.697	3.445	1.227	1.755	2.091	2.748	3.511
800	1.201	1.717	2.046	2.688	3.434	1.222	1.747	2.082	2.736	3.495
900	1.198	1.712	2.040	2.682	3.426	1.218	1.741	2.075	2.726	3.483
1000	1.195	1.709	2.036	2.676	3.418	1.214	1.736	2.068	2.718	3.472
∞	1.150	1.645	1.960	2.576	3.291	1.150	1.645	1.960	2.576	3.291

Table VII (b) Factors for One-Sided Tolerance Limits for Normal Distributions*

Factors K such that the probability is γ that at least a proportion P of the distribution will be less than $\bar{X}$ + Ks (or greater than $\bar{X}$ - Ks) where $\bar{X}$ and s are estimates of the mean and the standard deviation computed from a sample size of n.

EXAMPLE: For γ = 0.90, P = 0.95, n = 20, then K = 2.208

	γ = 0.75					γ = 0.90				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
3	1.464	2.501	3.152	4.396	5.805	2.602	4.258	5.310	7.340	9.651
4	1.256	2.134	2.680	3.726	4.910	1.972	3.187	3.957	5.437	7.128
5	1.152	1.961	2.463	3.421	4.507	1.698	2.742	3.400	4.666	6.112
6	1.087	1.860	2.336	3.243	4.273	1.540	2.494	3.091	4.242	5.556
7	1.043	1.791	2.250	3.126	4.118	1.435	2.333	2.894	3.972	5.201
8	1.010	1.740	2.190	3.042	4.008	1.360	2.219	2.755	3.783	4.955
9	0.984	1.702	2.141	2.977	3.924	1.302	2.133	2.649	3.641	4.772
10	0.964	1.671	2.103	2.927	3.858	1.257	2.065	2.568	3.532	4.629
11	0.947	1.646	2.073	2.885	3.804	1.219	2.012	2.503	3.444	4.515
12	0.933	1.624	2.048	2.851	3.760	1.188	1.966	2.448	3.371	4.420
13	0.919	1.606	2.026	2.822	3.722	1.162	1.928	2.403	3.310	4.341
14	0.909	1.591	2.007	2.796	3.690	1.139	1.895	2.363	3.257	4.274
15	0.899	1.577	1.991	2.776	3.661	1.119	1.866	2.329	3.212	4.215
16	0.891	1.566	1.977	2.756	3.637	1.101	1.842	2.299	3.172	4.164
17	0.883	1.554	1.964	2.739	3.615	1.085	1.820	2.272	3.136	4.118
18	0.876	1.544	1.951	2.723	3.595	1.071	1.800	2.249	3.106	4.078
19	0.870	1.536	1.942	2.710	3.577	1.058	1.781	2.228	3.078	4.041
20	0.865	1.528	1.933	2.697	3.561	1.046	1.765	2.208	3.052	4.009
21	0.859	1.520	1.923	2.686	3.545	1.035	1.750	2.190	3.028	3.979
22	0.854	1.514	1.916	2.675	3.532	1.025	1.736	2.174	3.007	3.952
23	0.849	1.508	1.907	2.665	3.520	1.016	1.724	2.159	2.987	3.927
24	0.845	1.502	1.901	2.656	3.509	1.007	1.712	2.145	2.969	3.904
25	0.842	1.496	1.895	2.647	3.497	0.999	1.702	2.132	2.952	3.882
30	0.825	1.475	1.869	2.613	3.454	0.966	1.657	2.080	2.884	3.794
35	0.812	1.458	1.849	2.588	3.421	0.942	1.623	2.041	2.833	3.730
40	0.803	1.445	1.834	2.568	3.395	0.923	1.598	2.010	2.793	3.679
45	0.795	1.435	1.821	2.552	3.375	0.908	1.577	1.986	2.762	3.638
50	0.788	1.426	1.811	2.538	3.358	0.894	1.560	1.965	2.735	3.604

*Adapted with permission from G.J. Lieberman, "Table for One-Sided Statistical Tolerance Limits", Industrial Quality Control, Volume XIV, 10, 1958.

Table VII (b)

FACTORS FOR ONE-SIDED TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

	$\gamma = 0.95$					$\gamma = 0.99$				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
3	3.804	6.158	7.655	10.552	13.857	—	—	—	—	—
4	2.619	4.163	5.145	7.042	9.215	—	—	—	—	—
5	2.149	3.407	4.202	5.741	7.501	—	—	—	—	—
6	1.895	3.006	3.707	5.062	6.612	2.849	4.408	5.409	7.334	9.540
7	1.732	2.755	3.399	4.641	6.061	2.490	3.856	4.730	6.411	8.348
8	1.617	2.582	3.188	4.353	5.686	2.252	3.496	4.287	5.811	7.566
9	1.532	2.454	3.031	4.143	5.414	2.085	3.242	3.971	5.389	7.014
10	1.465	2.355	2.911	3.981	5.203	1.954	3.048	3.739	5.075	6.603
11	1.411	2.275	2.815	3.852	5.036	1.854	2.897	3.557	4.828	6.284
12	1.366	2.210	2.736	3.747	4.900	1.771	2.773	3.410	4.633	6.032
13	1.329	2.155	2.670	3.659	4.787	1.702	2.677	3.290	4.472	5.826
14	1.296	2.108	2.614	3.585	4.690	1.645	2.592	3.189	4.336	5.651
15	1.268	2.068	2.566	3.520	4.607	1.596	2.521	3.102	4.224	5.507
16	1.242	2.032	2.523	3.463	4.534	1.553	2.458	3.028	4.124	5.374
17	1.220	2.001	2.486	3.415	4.471	1.514	2.405	2.962	4.038	5.268
18	1.200	1.974	2.453	3.370	4.415	1.481	2.357	2.906	3.961	5.167
19	1.183	1.949	2.423	3.331	4.364	1.450	2.315	2.855	3.893	5.078
20	1.167	1.926	2.396	3.295	4.319	1.424	2.275	2.807	3.832	5.003
21	1.152	1.905	2.371	3.262	4.276	1.397	2.241	2.768	3.776	4.932
22	1.138	1.887	2.350	3.233	4.238	1.376	2.208	2.729	3.727	4.866
23	1.126	1.869	2.329	3.206	4.204	1.355	2.179	2.693	3.680	4.806
24	1.114	1.853	2.309	3.181	4.171	1.336	2.154	2.663	3.638	4.755
25	1.103	1.838	2.292	3.158	4.143	1.319	2.129	2.632	3.601	4.706
30	1.059	1.778	2.220	3.064	4.022	1.249	2.029	2.516	3.446	4.508
35	1.025	1.732	2.166	2.994	3.934	1.195	1.957	2.431	3.334	4.364
40	0.999	1.697	2.126	2.941	3.866	1.154	1.902	2.365	3.250	4.255
45	0.978	1.669	2.092	2.897	3.811	1.122	1.857	2.313	3.181	4.168
50	0.961	1.646	2.065	2.863	3.766	1.096	1.821	2.296	3.124	4.096

TABLE VIII

Two-Sided Prediction Intervals for k Additional Observations from a Normal Distribution*

$$\bar{x} \pm r(k, n, \gamma)s$$

$\bar{x}$ is average of n observations, s is estimated standard deviation

(a) γ = .90 Prediction Interval

n=Size of Previous Sample	k = Number of Additional Observations										
	1	2	3	4	5	6	8	10	12	15	30
4	2.63	3.33	3.74	4.03	4.26	4.43	4.71	4.93	5.09	5.30	5.55
5	2.34	2.91	3.25	3.48	3.67	3.81	4.04	4.22	4.36	4.53	4.74
6	2.18	2.69	2.98	3.19	3.35	3.48	3.68	3.84	3.96	4.11	4.30
7	2.08	2.55	2.82	3.01	3.15	3.27	3.46	3.60	3.71	3.85	4.02
8	2.01	2.45	2.71	2.88	3.02	3.13	3.30	3.44	3.54	3.67	3.83
9	1.96	2.38	2.63	2.79	2.92	3.03	3.19	3.32	3.42	3.54	3.70
10	1.92	2.33	2.56	2.73	2.85	2.95	3.11	3.23	3.33	3.44	3.59
11	1.89	2.29	2.52	2.67	2.79	2.89	3.04	3.16	3.25	3.36	3.51
12	1.87	2.26	2.48	2.63	2.75	2.84	2.99	3.10	3.19	3.30	3.44
15	1.82	2.19	2.39	2.54	2.65	2.74	2.87	2.98	3.06	3.17	3.30
20	1.77	2.12	2.32	2.45	2.56	2.64	2.77	2.87	2.95	3.04	3.16
25	1.74	2.09	2.27	2.40	2.50	2.58	2.71	2.80	2.88	2.97	3.08
30	1.73	2.06	2.25	2.37	2.47	2.55	2.67	2.76	2.83	2.92	3.03
40	1.71	2.03	2.21	2.34	2.43	2.50	2.62	2.71	2.78	2.86	2.97
60	1.69	2.00	2.18	2.30	2.39	2.46	2.57	2.66	2.72	2.81	2.91
∞	1.64	1.95	2.11	2.23	2.31	2.38	2.48	2.56	2.62	2.70	2.79

*Adapted by permission from C. J. Hahn, "Factors for Calculating Two-Sided Prediction Intervals for Samples from a Normal Distribution," Journal of the American Statistical Association, Vol. 64, No. 327, 1969, and "Calculating Prediction Intervals for Samples from a Normal Distribution," Journal of the American Statistical Association, Vol. 65, No. 332, 1970.

TABLE VIII (cont'd)

(b) $\gamma = 0.95$ Prediction Interval*

n = size of previous sample	k = Number of Additional Observations													
	1	2	3	4	5	6	7	8	9	10	11	12	15	20
4	3.56	4.41	4.92	5.29	5.56	5.79	5.98	6.14	6.28	6.41	6.52	6.62	6.88	7.21
5	3.04	3.70	4.09	4.36	4.58	4.75	4.90	5.02	5.13	5.23	5.32	5.39	5.60	5.85
6	2.78	3.34	3.66	3.90	4.09	4.23	4.35	4.45	4.55	4.63	4.70	4.47	4.93	5.15
7	2.62	3.11	3.41	3.61	3.78	3.91	4.01	4.10	4.20	4.26	4.33	4.40	4.54	4.73
8	2.51	2.97	3.24	3.43	3.58	3.70	3.80	3.88	3.96	4.02	4.09	4.14	4.28	4.46
9	2.43	2.86	3.12	3.29	3.43	3.55	3.64	3.72	3.79	3.86	3.91	3.96	4.09	4.25
10	2.37	2.78	3.03	3.20	3.33	3.44	3.52	3.59	3.67	3.72	3.77	3.83	3.94	4.10
11	2.33	2.73	2.96	3.12	3.24	3.34	3.43	3.50	3.56	3.62	3.67	3.72	3.83	3.99
12	2.29	2.68	2.90	3.05	3.17	3.28	3.35	3.43	3.49	3.54	3.59	3.63	3.75	3.89
13	2.06	2.64	2.85	3.01	3.12	3.22	3.29	3.36	3.42	3.47	3.52	3.56	3.68	3.82
14	2.24	2.61	2.81	2.97	3.08	3.16	3.24	3.31	3.37	3.42	3.47	3.51	3.61	3.75
15	2.22	2.58	2.78	2.93	3.04	3.12	3.20	3.26	3.32	3.37	3.41	3.46	3.56	3.70
20	2.15	2.49	2.68	2.81	2.91	2.99	3.06	3.12	3.17	3.22	3.26	3.30	3.39	3.51
25	2.10	2.43	2.62	2.74	2.84	2.91	2.98	3.04	3.09	3.13	3.17	3.21	3.30	3.41
31	2.07	2.39	2.57	2.69	2.79	2.86	2.92	2.98	3.03	3.06	3.10	3.14	3.22	3.34
41	2.05	2.35	2.52	2.64	2.73	2.80	2.86	2.91	2.96	3.01	3.03	3.07	3.15	3.26
61	2.02	2.32	2.48	2.60	2.68	2.75	2.81	2.85	2.90	2.94	2.97	3.01	3.08	3.18
121	1.99	2.28	2.44	2.54	2.63	2.69	2.75	2.79	2.84	2.88	2.91	2.94	3.01	3.10
	1.96	2.24	2.39	2.50	2.58	2.64	2.69	2.74	2.78	2.82	2.84	2.88	2.95	3.04

(c) $\gamma = 0.99$ Prediction Interval*

4	6.53	7.94	8.80	9.41	9.88	10.27	10.59	10.87	11.12	11.33	11.52	11.70	12.14	12.70
5	5.04	5.97	6.54	6.94	7.25	7.51	7.72	7.91	8.07	8.22	8.35	8.47	8.77	9.15
6	4.36	5.08	5.51	5.82	6.07	6.27	6.44	6.58	6.71	6.83	6.93	7.02	7.26	7.57
7	3.96	4.56	4.93	5.19	5.39	5.56	5.69	5.82	5.93	6.02	6.10	6.10	6.38	6.64
8	3.71	4.24	4.56	4.78	4.96	5.10	5.23	5.34	5.43	5.51	5.58	5.65	5.83	6.05
9	3.54	4.02	4.30	4.51	4.66	4.80	4.90	5.00	5.08	5.16	5.23	5.29	5.45	5.65
10	3.41	3.85	4.11	4.31	4.45	4.57	4.67	4.76	4.83	4.90	4.96	5.02	5.16	5.35
11	3.31	3.73	3.99	4.14	4.28	4.39	4.49	4.57	4.64	4.71	4.76	4.82	4.95	5.12
12	3.23	3.63	3.86	4.03	4.15	4.26	4.35	4.43	4.49	4.55	4.60	4.65	4.78	4.95
13	3.17	3.55	3.77	3.93	4.05	4.15	4.23	4.31	4.37	4.43	4.48	4.52	4.65	4.80
14	3.12	3.48	3.70	3.86	3.96	4.06	4.15	4.21	4.27	4.33	4.38	4.42	4.53	4.69
15	3.07	3.43	3.64	3.78	3.90	3.98	4.07	4.13	4.19	4.24	4.29	4.34	4.44	4.58
20	2.93	3.26	3.44	3.56	3.66	3.74	3.81	3.87	3.92	3.97	4.01	4.05	4.14	4.27
25	2.85	3.15	3.32	3.44	3.53	3.61	3.67	3.73	3.70	3.81	3.86	3.89	3.98	4.09
31	2.79	3.08	3.24	3.35	3.44	3.51	3.57	3.62	3.66	3.71	3.74	3.78	3.86	3.97
41	2.74	3.01	3.16	3.27	3.35	3.42	3.48	3.52	3.56	3.60	3.64	3.66	3.74	3.84
61	2.68	2.94	3.08	3.19	3.26	3.33	3.38	3.42	3.46	3.50	3.52	3.56	3.63	3.72
121	2.63	2.88	3.01	3.11	3.17	3.24	3.28	3.33	3.36	3.40	3.43	3.45	3.52	3.61
	2.58	2.80	2.94	3.03	3.10	3.15	3.19	3.24	3.27	3.31	3.33	3.35	3.41	3.50

TABLE VIII (d)

One-Sided Prediction Intervals for k Additional Observations from a Normal Distribution*

$\bar{x} + r'(k, n, \gamma)$ or $\bar{x} - r'(k, n, \gamma)$

Size of previous sample (n)	Number of additional observations (k)										
	1	2	3	4	5	6	8	10	12	15	20
	γ = 0.90 prediction interval										
4	1.83	2.48	2.87	3.15	3.36	3.54	3.81	4.02	4.18	4.39	4.64
5	1.68	2.24	2.57	2.80	2.98	3.12	3.34	3.52	3.66	3.82	4.04
6	1.59	2.11	2.40	2.61	2.76	2.89	3.09	3.25	3.37	3.52	3.71
7	1.54	2.02	2.29	2.48	2.63	2.75	2.93	3.08	3.19	3.33	3.50
8	1.50	1.96	2.22	2.40	2.54	2.65	2.83	2.96	3.07	3.20	3.36
9	1.47	1.92	2.17	2.34	2.47	2.58	2.75	2.87	2.98	3.10	3.26
10	1.45	1.88	2.13	2.29	2.42	2.53	2.69	2.81	2.91	3.02	3.18
11	1.43	1.86	2.09	2.26	2.38	2.48	2.64	2.76	2.85	2.97	3.11
12	1.42	1.84	2.07	2.23	2.35	2.45	2.60	2.71	2.81	2.92	3.06
15	1.39	1.79	2.01	2.16	2.28	2.37	2.52	2.62	2.71	2.82	2.95
20	1.36	1.75	1.96	2.11	2.21	2.30	2.44	2.54	2.62	2.72	2.85
25	1.34	1.72	1.93	2.07	2.18	2.26	2.39	2.49	2.57	2.67	2.79
30	1.33	1.71	1.91	2.05	2.15	2.24	2.36	2.46	2.54	2.63	2.75
40	1.32	1.69	1.89	2.02	2.12	2.20	2.33	2.42	2.50	2.59	2.70
60	1.31	1.67	1.86	1.99	2.09	2.17	2.29	2.38	2.46	2.54	2.65
	1.28	1.63	1.82	1.94	2.04	2.11	2.22	2.31	2.38	2.46	2.56
	γ = 0.95 Prediction interval										
4	2.63	3.40	3.87	4.21	4.47	4.69	5.02	5.28	5.49	5.74	6.06
5	2.34	2.95	3.32	3.58	3.79	3.95	4.22	4.42	4.58	4.78	5.03
6	2.18	2.72	3.03	3.26	3.43	3.58	3.80	3.97	4.11	4.28	4.49
7	2.08	2.57	2.86	3.06	3.22	3.34	3.55	3.70	3.82	3.98	4.17
8	2.01	2.47	2.74	2.93	3.07	3.19	3.37	3.52	3.63	3.77	3.95
9	1.96	2.40	2.65	2.83	2.97	3.08	3.25	3.38	3.49	3.62	3.79
10	1.92	2.35	2.59	2.76	2.89	2.99	3.16	3.28	3.39	3.51	3.67
11	1.89	2.30	2.54	2.70	2.82	2.93	3.08	3.21	3.30	3.42	3.58
12	1.87	2.27	2.50	2.65	2.78	2.87	3.03	3.14	3.24	3.35	3.50
15	1.82	2.20	2.41	2.56	2.67	2.76	2.90	3.01	3.10	3.21	3.34
20	1.77	2.13	2.33	2.47	2.57	2.66	2.79	2.89	2.97	3.07	3.19
25	1.74	2.09	2.29	2.42	2.52	2.60	2.73	2.82	2.90	2.99	3.11
30	1.73	2.07	2.26	2.39	2.48	2.56	2.68	2.78	2.85	2.94	3.05
40	1.71	2.04	2.22	2.35	2.44	2.52	2.63	2.72	2.79	2.88	2.99
60	1.69	2.01	2.19	2.31	2.40	2.47	2.58	2.67	2.74	2.82	2.92
	1.64	1.95	2.12	2.23	2.32	2.39	2.49	2.57	2.63	2.71	2.80
	γ = 0.99 Prediction interval										
4	5.07	6.30	7.07	7.63	8.07	8.43	9.00	9.43	9.79	10.22	10.76
5	4.10	4.94	5.46	5.83	6.13	6.37	6.75	7.04	7.28	7.57	7.95
6	3.63	4.30	4.70	4.99	5.22	5.41	5.71	5.94	6.12	6.35	6.64
7	3.36	3.93	4.27	4.51	4.70	4.86	5.11	5.30	5.46	5.65	5.90
8	3.18	3.69	3.99	4.20	4.37	4.51	4.73	4.90	5.04	5.21	5.43
9	3.05	3.52	3.79	3.99	4.14	4.27	4.46	4.62	4.74	4.90	5.09
10	2.96	3.39	3.65	3.83	3.97	4.09	4.27	4.41	4.53	4.67	4.85
11	2.89	3.30	3.54	3.71	3.84	3.95	4.12	4.25	4.36	4.50	4.67
12	2.83	3.22	3.45	3.61	3.74	3.84	4.00	4.13	4.23	4.36	4.52
15	2.71	3.07	3.27	3.42	3.53	3.62	3.77	3.88	3.97	4.08	4.22
20	2.60	2.93	3.11	3.24	3.34	3.42	3.55	3.65	3.73	3.83	3.95
25	2.54	2.85	3.02	3.14	3.24	3.31	3.43	3.52	3.60	3.69	3.80
30	2.50	2.80	2.97	3.08	3.17	3.24	3.36	3.44	3.51	3.60	3.71
40	2.46	2.74	2.90	3.01	3.09	3.16	3.27	3.35	3.41	3.49	3.59
60	2.41	2.68	2.83	2.94	3.02	3.08	3.18	3.26	3.32	3.39	3.49
	2.33	2.58	2.71	2.81	2.88	2.93	3.02	3.09	3.14	3.21	3.29

Table IX

NUMBER OF OBSERVATIONS FOR t-TEST OF MEAN

The entries in this table show the numbers of observations needed in a t-test of the significance of a mean in order to control the probabilities of errors of the first and second kinds at α and β respectively.

		LEVEL OF t-TEST																				
SINGLE-SIDED TEST		$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
DOUBLE-SIDED TEST		$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$		0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
	0.05																					0.05
	0.10																					0.10
	0.15																				122	0.15
	0.20										139					99					70	0.20
	0.25					110					90				128	64			139	101	45	0.25
	0.30				134	78				115	63			119	90	45		122	97	71	32	0.30
	0.35			125	99	58			109	85	47		109	88	67	34		90	72	52	24	0.35
	0.40		115	97	77	45		101	85	66	37	117	84	68	51	26	101	70	55	40	19	0.40
	0.45		92	77	62	37	110	81	68	53	30	93	67	54	41	21	80	55	44	33	15	0.45
VALUE OF	0.50	100	75	63	51	30	90	66	55	43	25	76	54	44	34	18	65	45	36	27	13	0.50
$D = \frac{\delta}{\sigma}$	0.55	83	63	53	42	26	75	55	46	36	21	63	45	37	28	15	54	38	30	22	11	0.55
	0.60	71	53	45	36	22	63	47	39	31	18	53	38	32	24	13	46	32	26	19	9	0.60
	0.65	61	46	39	31	20	55	41	34	27	16	46	33	27	21	12	39	28	22	17	8	0.65
	0.70	53	40	34	28	17	47	35	30	24	14	40	29	24	19	10	34	24	19	15	8	0.70
	0.75	47	36	30	25	16	42	31	27	21	13	35	26	21	16	9	30	21	17	13	7	0.75
	0.80	41	32	27	22	14	37	28	24	19	12	31	22	19	15	9	27	19	15	12	6	0.80
	0.85	37	29	24	20	13	33	25	21	17	11	28	21	17	13	8	24	17	14	11	6	0.85
	0.90	34	26	22	18	12	29	23	19	16	10	25	19	16	12	7	21	15	13	10	5	0.90
	0.95	31	24	20	17	11	27	21	18	14	9	23	17	14	11	7	19	14	11	9	5	0.95
	1.00	28	22	19	16	10	25	19	16	13	9	21	16	13	10	6	18	13	11	8	5	1.00

*Reproduced with permission from O.L. Davies, Design and Analysis of Industrial Experiments, Oliver and Boyd, Edinburgh, 1956.

Number of Observations for t-Test of Mean

The entries in this table show the numbers of observations needed in a t-test of the significance of a mean in order to control the probabilities of errors of the first and second kinds at α and β respectively.

		Level of t-test																				
Single-sided test		$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
Double-sided test		$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$		0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
	1.1	24	19	16	14	9	21	16	14	12	8	18	13	11	9	6	15	11	9	7		1.1
	1.2	21	16	14	12	8	18	14	12	10	7	15	12	10	8	5	13	10	8	6		1.2
	1.3	18	15	13	11	8	16	13	11	9	6	14	10	9	7		11	8	7	6		1.3
	1.4	16	13	12	10	7	14	11	10	9	6	12	9	8	7		10	8	7	5		1.4
	1.5	15	12	11	9	7	13	10	9	8	6	11	8	7	6		9	7	6			1.5
	1.6	13	11	10	8	6	12	10	9	7	5	10	8	7	6		8	6	6			1.6
	1.7	12	10	9	8	6	11	9	8	7		9	7	6	5		8	6	5			1.7
	1.8	12	10	9	8	6	10	8	7	7		8	7	6			7	6				1.8
Value of	1.9	11	9	8	7	6	10	8	7	6		8	6	6			7	5				1.9
$D = \frac{\delta}{\sigma}$	2.0	10	8	8	7	5	9	7	7	6		7	6	5			6					2.0
	2.1	10	8	7	7		8	7	6	6		7	6				6					2.1
	2.2	9	8	7	6		8	7	6	5		7	6				6					2.2
	2.3	9	7	7	6		8	6	6			6	5				5					2.3
	2.4	8	7	7	6		7	6	6			6										2.4
	2.5	8	7	6	6		7	6	6			6										2.5
	3.0	7	6	6	5		6	5	5			5										3.0
	3.5	6	5	5			5															3.5
	4.0	6																				4.0

Table X

Number of Observations for t-Test of Difference between Two Means

The entries in this table show the number of observations needed in a t-test of the significance of the difference between two means in order to control the probabilities of the errors of the first and second kinds at α and β respectively.

		LEVEL OF t-TEST																				
SINGLE-SIDED TEST		$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
DOUBLE-SIDED TEST		$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$		0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
	0.05																					0.05
	0.10																					0.10
	0.15																					0.15
	0.20																				137	0.20
	0.25															124					88	0.25
	0.30										123					87					61	0.30
	0.35					110					90					64				102	45	0.35
	0.40					85					70				100	50			108	78	35	0.40
	0.45				118	68				101	55			105	79	39		108	86	62	28	0.45
VALUE OF	0.50				96	55			106	82	45		105	86	64	32		88	70	51	23	0.50
$D = \frac{\delta}{\sigma}$	0.55			101	79	46		106	88	68	38		87	71	53	27	112	73	58	42	19	0.55
	0.60		101	85	67	39		90	74	58	32	104	74	60	45	23	89	61	49	36	16	0.60
	0.65		87	73	57	34	104	77	64	49	27	88	63	51	39	20	76	52	42	30	14	0.65
	0.70	100	75	63	50	29	90	66	55	43	24	76	55	44	34	17	66	45	36	26	12	0.70
	0.75	88	56	55	44	26	79	58	48	38	21	67	48	39	29	15	57	40	32	23	11	0.75
	0.80	77	58	49	39	23	70	51	43	33	19	59	42	34	26	14	50	35	28	21	10	0.80
	0.85	69	51	43	35	21	62	46	38	30	17	52	37	31	23	12	45	31	25	18	9	0.85
	0.90	62	46	39	31	19	55	41	34	27	15	47	34	27	21	11	40	28	22	16	8	0.90
	0.95	55	42	35	28	17	50	37	31	24	14	42	30	25	19	10	36	25	20	15	7	0.95
	1.00	50	38	32	26	15	45	33	28	22	13	38	27	23	17	9	33	23	18	14	7	1.00

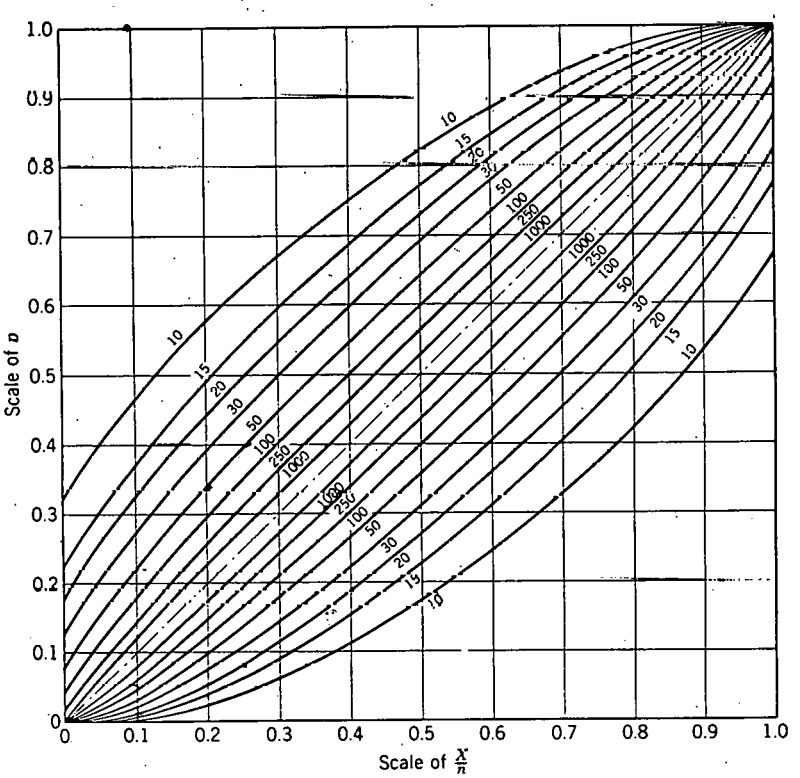
Number of Observations for t-Test of Difference between Two Means

The entries in this table show the number of observations needed in a t-test of the significance of the difference between two means in order to control the probabilities of the errors of the first and second kinds at α and β respectively.

		LEVEL OF t-TEST																				
SINGLE-SIDED TEST		$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
DOUBLE-SIDED TEST		$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$		0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
	1.1	42	32	27	22	13	38	28	23	19	11	32	23	19	14	8	27	19	15	12	6	1.1
	1.2	36	27	23	18	11	32	24	20	16	9	27	20	16	12	7	23	16	13	10	5	1.2
	1.3	31	23	20	16	10	28	21	17	14	8	23	17	14	11	6	20	14	11	9	5	1.3
	1.4	27	20	17	14	9	24	18	15	12	8	20	15	12	10	6	17	12	10	8	4	1.4
	1.5	24	18	15	13	8	21	16	14	11	7	18	13	11	9	5	15	11	9	7	4	1.5
	1.6	21	16	14	11	7	19	14	12	10	6	16	12	10	8	5	14	10	8	6	4	1.6
	1.7	19	15	13	10	7	17	13	11	9	6	14	11	9	7	4	12	9	7	6	3	1.7
	1.8	17	13	11	10	6	15	12	10	8	5	13	10	8	6	4	11	8	7	5		1.8
VALUE OF	1.9	16	12	11	9	6	14	11	9	8	5	12	9	7	6	4	10	7	6	5		1.9
$D = \frac{\delta}{\sigma}$	2.0	14	11	10	8	6	13	10	9	7	5	11	8	7	6	4	9	7	6	4		2.0
	2.1	13	10	9	8	5	12	9	8	7	5	10	8	6	5	3	8	6	5	4		2.1
	2.2	12	10	8	7	5	11	9	7	6	4	9	7	6	5		8	6	5	4		2.2
	2.3	11	9	8	7	5	10	8	7	6	4	9	7	6	5		7	5	5	4		2.3
	2.4	11	9	8	6	5	10	8	7	6	4	8	6	5	4		7	5	4	4		2.4
	2.5	10	8	7	6	4	9	7	6	5	4	8	6	5	4		6	5	4	3		2.5
	3.0	8	6	6	5	4	7	6	5	4	3	6	5	4	4		5	4	3			3.0
	3.5	6	5	5	4	3	6	5	4	4		5	4	4	3		4	3				3.5
	4.0	6	5	4	4		5	4	4	3		4	4	3			4					4.0

Table XI (a)

CONFIDENCE BELTS FOR PROPORTIONS
(Confidence coefficient 0.95)



*Reproduced with permission of Biometrika trustees from C.J. Clopper and E.S. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," Biometrika 26, 1934.

Table XI (b)

CONFIDENCE BELTS FOR PROPORTIONS
(Confidence coefficient 0.99)

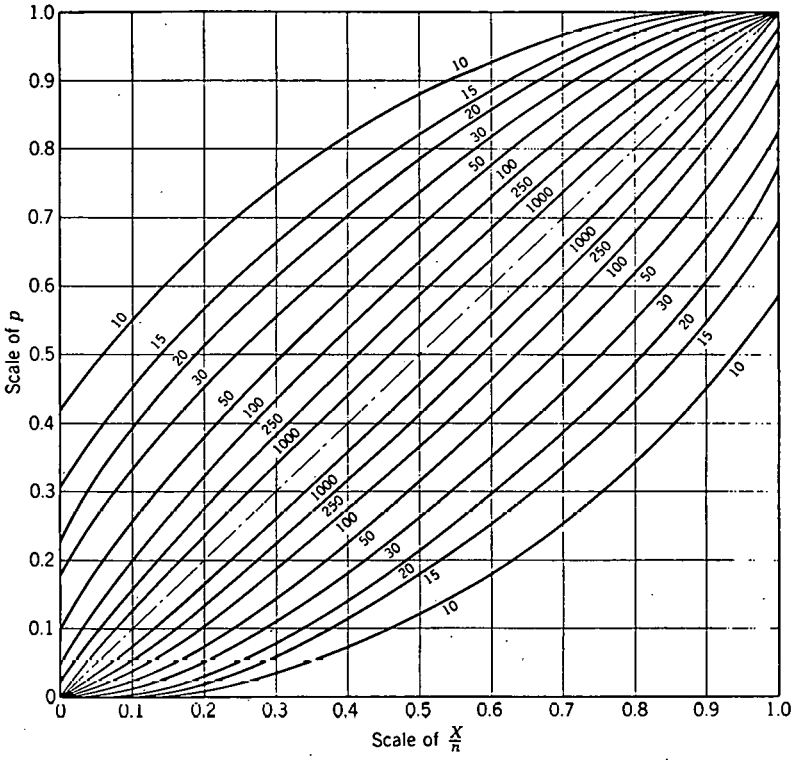


TABLE XII

CONFIDENCE LIMITS FOR THE EXPECTED VALUE OF A POISSON DISTRIBUTION*

TOTAL OBSERVED COUNT $x_0 = \Sigma x_i$	SIGNIFICANCE LEVEL				TOTAL OBSERVED COUNT $x_0 = \Sigma x_i$	SIGNIFICANCE LEVEL			
	$\alpha = 0.01$		$\alpha = 0.05$			$\alpha = 0.01$		$\alpha = 0.05$	
	Lower Limit	*Upper Limit*	*Lower Limit*	*Upper Limit*		*Lower Limit*	*Upper Limit*	*Lower Limit*	*Upper Limit*
0	0.0	5.3	0.0	3.7					
1	0.0	7.4	0.1	5.6	26	14.7	42.2	17.0	38.0
2	0.1	9.3	0.2	7.2	27	15.4	43.5	17.8	39.2
3	0.3	11.0	0.6	8.8	28	16.2	44.8	18.6	40.4
4	0.6	12.6	1.0	10.2	29	17.0	46.0	19.4	41.6
5	1.0	14.1	1.6	11.7	30	17.7	47.2	20.2	42.8
6	1.5	15.6	2.2	13.1	31	18.5	48.4	21.0	44.0
7	2.0	17.1	2.8	14.4	32	19.3	49.6	21.8	45.1
8	2.5	18.5	3.4	15.8	33	20.0	50.8	22.7	46.3
9	3.1	20.0	4.0	17.1	34	20.8	52.1	23.5	47.5
10	3.7	21.3	4.7	18.4	35	21.6	53.3	24.3	48.7
11	4.3	22.6	5.4	19.7	36	22.4	54.5	25.1	49.8
12	4.9	24.0	6.2	21.0	37	23.2	55.7	26.0	51.0
13	5.5	25.4	6.9	22.3	38	24.0	56.9	26.8	52.2
14	6.2	26.7	7.7	23.5	39	24.8	58.1	27.7	53.3
15	6.8	28.1	8.4	24.8	40	25.6	59.3	28.6	54.5
16	7.5	29.4	9.4	26.0	41	26.4	60.5	29.4	55.6
17	8.2	30.7	9.9	27.2	42	27.2	61.7	30.3	56.8
18	8.9	32.0	10.7	28.4	43	28.0	62.9	31.1	57.9
19	9.6	33.3	11.5	29.6	44	28.8	64.1	32.0	59.0
20	10.3	34.6	12.2	30.8	45	29.6	65.3	32.8	60.2
21	11.0	35.9	13.0	32.0	46	30.4	66.5	33.6	61.3
22	11.8	37.2	13.8	33.2	47	31.2	67.7	34.5	62.5
23	12.5	38.4	14.6	34.4	48	32.0	68.9	35.3	63.6
24	13.2	39.7	15.4	35.6	49	32.8	70.1	36.1	64.8
25	14.0	41.0	16.2	36.8	50	33.6	71.3	37.0	65.9

NOTE. There is at least $1 - \alpha$ confidence that $n\theta$ will be between the limits when Σx_i is the total number of occurrences of an event in n independent observations on a Poisson variable with expected value θ.

*Reproduced with permission from W. E. Ricker, "The Concept of Confidence on Fiducial Limits Applied to the Poisson Frequency Distribution," The Journal of the American Statistical Association, 32, 1937.

Table XIII (a)

Sample Sizes for Two-Sided Distribution-Free Tolerance Limits*

n is sample size required to assure with $100\gamma\%$ confidence that at least 100 P % of the population will lie between the smallest and largest observations.

EXAMPLE $\gamma = 0.95$, P = 0.90, then n = 46

$$np^{n-1} - (n-1)p^n \leq 1 - \gamma$$

	P						
γ	0.500	0.700	0.750	0.800	0.850	0.900	0.950
0.500	3	6	7	9	11	17	34
0.700	5	8	10	12	16	24	49
0.750	5	9	10	13	18	27	53
0.800	5	9	11	14	19	29	59
0.850	6	10	13	16	22	33	67
0.900	7	12	15	18	25	38	77
0.950	8	14	18	22	30	46	93
0.975	9	17	20	26	35	54	110
0.980	9	17	21	27	37	56	115
0.990	11	20	24	31	42	64	130
0.995	12	22	27	34	47	72	146
0.999	14	27	33	42	58	89	181
0.9995	15	29	36	46	63	96	196
0.9999	18	34	42	54	73	113	230

	P						
γ	0.975	0.980	0.990	0.995	0.999	0.9995	0.9999
0.500	67	84	168	336	1679	3357	16783
0.700	97	122	244	488	2439	4878	24392
0.750	107	134	269	538	2692	5385	26926
0.800	119	149	299	598	2994	5988	29943
0.850	134	168	337	674	3372	6744	33724
0.900	155	194	388	777	3889	7778	38896
0.950	188	236	473	947	4742	9486	47437
0.975	221	277	555	1113	5570	11141	55715
0.980	231	290	581	1165	5832	11666	58337
0.990	263	330	662	1325	6636	13274	66381
0.995	294	369	740	1483	7427	14858	74299
0.999	366	458	920	1843	9230	18463	92330
0.9995	396	496	996	1996	9995	19993	99983
0.9999	465	583	1171	2346	11751	23508	117559

*Reproduced with permission from D.B. Owen, Handbook of Statistical Tables, Addison-Wesley Publishing Company, Inc., 1962.

Table XIII (b)

Proportion of Population Contained in A Two-Sided Distribution-Free Tolerance Interval*

P is proportion of population which lies between the smallest and largest of n observations with $100\gamma\%$ confidence.

EXAMPLE For $\gamma = 0.90$, $n = 15$, then $P = 0.7644$

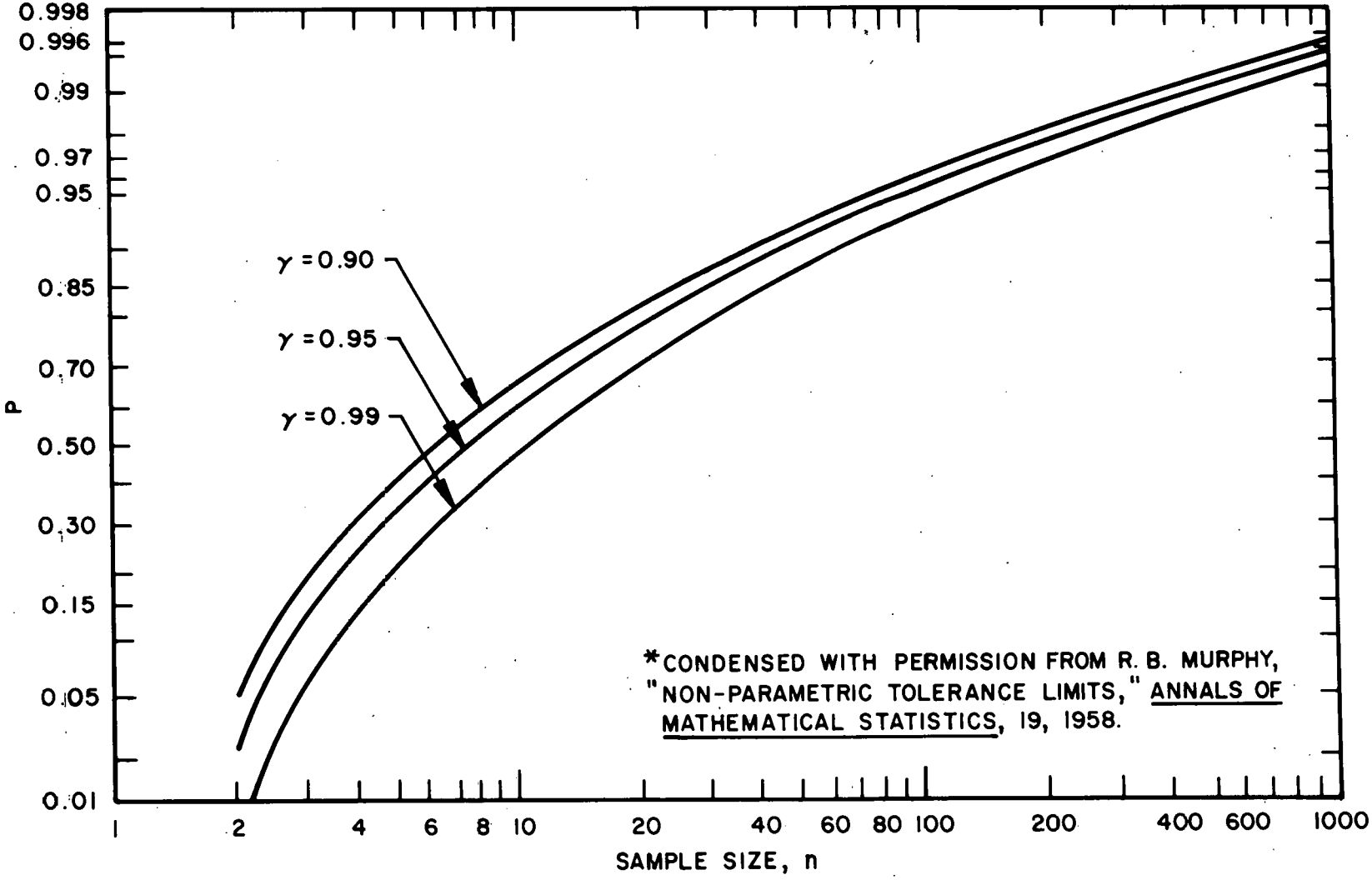
Sample size n	$\gamma = 0.75$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.99$	$\gamma = 0.999$
2	0.1340	0.0513	0.0253	0.00501	0.00050
3	0.3263	0.1958	0.1353	0.0589	0.01838
4	0.4563	0.3205	0.2486	0.1409	0.06410
5	0.5458	0.4161	0.3426	0.2221	0.12196
6	0.6105	0.4897	0.4182	0.2943	0.1812
7	0.6593	0.5474	0.4793	0.3566	0.2375
8	0.6973	0.5938	0.5293	0.4101	0.2888
9	0.7277	0.6316	0.5709	0.4560	0.3350
10	0.7526	0.6631	0.6058	0.4956	0.3763
15	0.8303	0.7644	0.7206	0.6321	0.5283
20	0.8710	0.8190	0.7839	0.7112	0.6216
25	0.8959	0.8531	0.8239	0.7625	0.6854
30	0.9127	0.8764	0.8514	0.7984	0.7308
35	0.9249	0.8933	0.8715	0.8249	0.7598
40	0.9341	0.9062	0.8868	0.8452	0.7911
45	0.9413	0.9163	0.8989	0.8614	0.8123
50	0.9471	0.9244	0.9086	0.8744	0.8299
60	0.9558	0.9367	0.9234	0.8944	0.8557
70	0.9620	0.9456	0.9340	0.9089	0.8749
80	0.9667	0.9522	0.9421	0.9199	0.8898
90	0.9704	0.9575	0.9484	0.9285	0.9016
100	0.9733	0.9617	0.9534	0.9354	0.9114
150	0.9821	0.9743	0.9688	0.9566	0.9398
200	0.9866	0.9807	0.9765	0.9673	0.9546
250	0.9893	0.9845	0.9812	0.9737	0.9625
300	0.9911	0.9871	0.9843	0.9781	0.9697
350	0.9923	0.9889	0.9865	0.9812	0.9739
400	0.9933	0.9903	0.9882	0.9835	0.9772
450	0.9940	0.9914	0.9895	0.9853	0.9797
500	0.9946	0.9922	0.9905	0.9868	0.9817
600	0.9955	0.9935	0.9921	0.9890	0.9847
700	0.9962	0.9945	0.9932	0.9906	0.0969
800	0.9966	0.9951	0.9941	0.9917	0.9885
900	0.9970	0.9957	0.9947	0.9926	0.9898
1000	0.9973	0.9961	0.9953	0.9934	0.9908

*Adapted with permission from R.B. Murphy, Non-Parametric Tolerance Limits", Annals of Mathematical Statistics, 19, 1948.

GRAPHS OF P SUCH THAT AT LEAST A PROPORTION P OF THE POPULATION IS BETWEEN THE MINIMUM AND MAXIMUM OBSERVATION WITH CONFIDENCE LEVELS*

$\gamma = 0.90, 0.95, 0.99$

TWO-SIDED DISTRIBUTION-FREE TOLERANCE INTERVALS



*CONDENSED WITH PERMISSION FROM R. B. MURPHY, "NON-PARAMETRIC TOLERANCE LIMITS," ANNALS OF MATHEMATICAL STATISTICS, 19, 1958.

FIGURE XIII (C)

Table XIII (a)

Confidence Associated with a Two-Sided Distribution-Free Tolerance Interval*

Confidence γ with which we may assert that 100P percent of the population lies between the largest and smallest of a random sample of n from that population (continuous distribution assumed)

EXAMPLE For P = 0.95, n = 11, then γ = 0.10

n	P = .75	P = .90	P = .95	P = .99
3	.16	.03	.01	.00
4	.26	.05	.01	.00
5	.37	.08	.02	.00
6	.47	.11	.03	.00
7	.56	.15	.04	.00
8	.63	.19	.06	.00
9	.70	.23	.07	.00
10	.76	.26	.09	.00
11	.80	.30	.10	.01
12	.84	.34	.12	.01
13	.87	.38	.14	.01
14	.90	.42	.15	.01
15	.92	.45	.17	.01
16	.94	.49	.19	.01
17	.95	.52	.21	.01
18	.96	.55	.23	.01
19	.97	.58	.25	.02
20	.98	.61	.26	.02
25	.99	.73	.36	.03
30	1.00−	.82	.45	.04
40	—	.92	.60	.06
50	—	.97	.72	.09
60	—	.99	.81	.12
70	—	.99	.87	.16
80	—	1.00−	.91	.19
90	—	—	.94	.23
100	—	—	.96	.26

*Adapted with permission from P.N. Somerville, "Tables for Obtaining Non-Parametric Tolerance Limits," Annals of Mathematical Statistics, 29, 1958.

Table XIII (e)

Sample Sizes for One-Sided Distribution-Free Tolerance Limits*

n is sample size required to assure that the $100\gamma\%$ confidence that at least 100 P % of the population sampled will lie above the smallest observation (or below the largest).

EXAMPLE For $\gamma = 0.90$, $P = 0.95$, then $n = 45$.

$$p^n \leq 1 - \gamma$$

	P						
γ	0.500	0.700	0.750	0.800	0.850	0.900	0.950
0.500	1	2	3	4	5	7	14
0.700	2	4	5	6	8	12	24
0.750	3	4	5	7	9	14	28
0.800	3	5	6	8	10	16	32
0.850	3	6	7	9	12	19	37
0.900	4	7	9	11	15	22	45
0.950	5	9	11	14	19	29	59
0.975	6	11	13	17	23	36	72
0.980	6	11	14	18	25	38	77
0.990	7	13	17	21	29	44	90
0.995	8	15	19	24	33	51	104
0.999	10	20	25	31	43	66	135
0.9995	11	22	27	35	47	73	149
0.9999	14	26	33	42	57	88	180

	P						
γ	0.975	0.980	0.990	0.995	0.999	0.9995	0.9999
0.500	28	35	69	139	693	1386	6932
0.700	48	60	120	241	1204	2408	12040
0.750	55	69	138	277	1386	2772	13863
0.800	64	80	161	322	1609	3219	16094
0.850	75	94	189	379	1897	3794	18971
0.900	91	114	230	460	2302	4605	23025
0.950	119	149	299	598	2995	5990	29956
0.975	146	183	368	736	3688	7376	36887
0.980	155	194	390	781	3911	7823	39119
0.990	182	228	459	919	4603	9209	46050
0.995	210	263	528	1058	5296	10594	52981
0.999	273	342	688	1379	6905	13813	69075
0.9995	301	377	757	1517	7598	15199	76006
0.9999	364	456	917	1838	9206	18417	92099

*Reproduced with permission from D.B. Owen, Handbook of Statistical Tables, Addison-Wesley Publishing Company, Inc., 1962.

Table XIII (f)

Proportion of Population Contained in a One-Sided Distribution-Free Tolerance Interval*

P is proportion of population that lies below the maximum observation (or above the minimum observation) for sample size n and confidence level γ.

EXAMPLE $\gamma = 0.90$, $n = 15$, then $P = 0.8577$

Sample size n	$\gamma = 0.80$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.99$	$\gamma = 0.999$
1	0.2000	0.1000	0.0500	0.0100	0.0010
2	0.4472	0.3162	0.2236	0.1000	0.0316
3	0.5848	0.4641	0.3684	0.2155	0.1000
4	0.6687	0.5623	0.4729	0.3162	0.1778
5	0.7248	0.6310	0.5493	0.3981	0.2512
6	0.7647	0.6813	0.6069	0.4641	0.3162
7	0.7945	0.7196	0.6517	0.5178	0.3726
8	0.8177	0.7499	0.6876	0.5623	0.4217
9	0.8362	0.7743	0.7169	0.5995	0.4642
10	0.8514	0.7943	0.7411	0.6310	0.5012
15	0.8983	0.8577	0.8190	0.7356	0.6310
20	0.9227	0.8912	0.8609	0.7943	0.7079
25	0.9377	0.9120	0.8871	0.8317	0.7586
30	0.9478	0.9261	0.9050	0.8577	0.7944
35	0.9551	0.9364	0.9180	0.8767	0.8209
40	0.9606	0.9441	0.9278	0.8912	0.8414
45	0.9649	0.9501	0.9356	0.9027	0.8577
50	0.9683	0.9550	0.9418	0.9120	0.8709
60	0.9735	0.9624	0.9513	0.9261	0.8912
70	0.9772	0.9676	0.9581	0.9363	0.9060
80	0.9801	0.9716	0.9633	0.9440	0.9173
90	0.9823	0.9747	0.9673	0.9501	0.9262
100	0.9840	0.9772	0.9705	0.9549	0.9332
150	0.9893	0.9848	0.9802	0.9698	0.9549
200	0.9920	0.9885	0.9852	0.9772	0.9661
250	0.9936	0.9908	0.9881	0.9817	0.9728
300	0.9946	0.9924	0.9901	0.9848	0.9773
350	0.9955	0.9934	0.9915	0.9869	0.9805
400	0.9960	0.9943	0.9925	0.9886	0.9828
450	0.9964	0.9949	0.9934	0.9899	0.9848
500	0.9968	0.9954	0.9941	0.9908	0.9863
600	0.9973	0.9962	0.9951	0.9924	0.9886
700	0.9977	0.9967	0.9958	0.9934	0.9901
800	0.9980	0.9971	0.9963	0.9943	0.9914
900	0.9982	0.9974	0.9967	0.9949	0.9923
1000	0.9984	0.9977	0.9970	0.9954	0.9931

*Adapted with permission from R.B. Murphy, "Non-Parametric Tolerance Limits," Annals of Mathematical Statistics, 19, 1948.

GRAPHS OF P SUCH THAT AT LEAST A PROPORTION P OF THE POPULATION IS ABOVE THE MINIMUM (OR BELOW THE MAXIMUM) OBSERVATION WITH CONFIDENCE LEVELS*
γ=0.90, 0.95, 0.99

ONE-SIDED DISTRIBUTION-FREE TOLERANCE INTERVALS

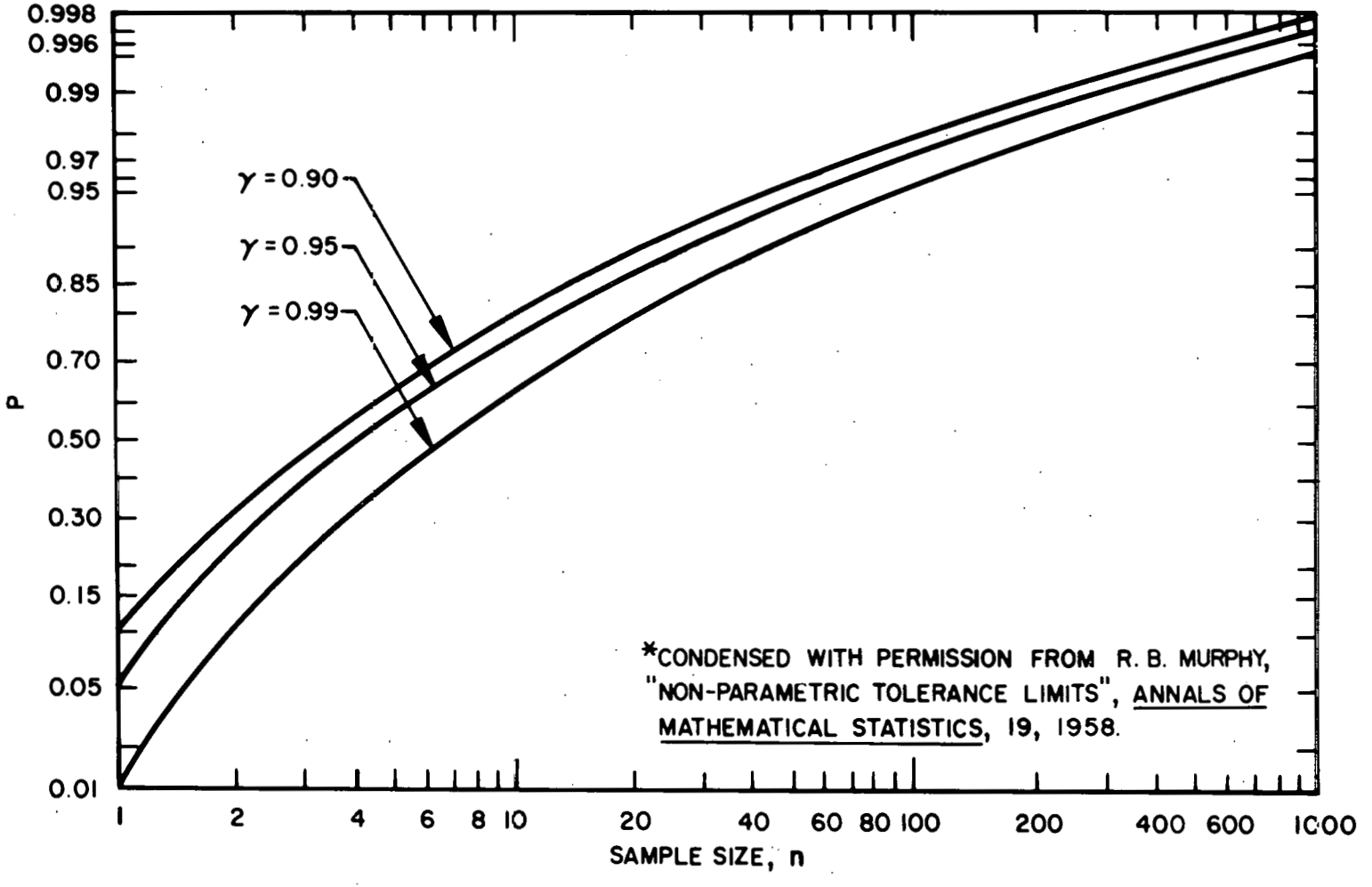


FIGURE XIII (g)

Table XIV

Factors for Computing Control Limits*

Number of Observations in Sample, n	$\bar{X}$ CHART		R CHART		
	Factors for Control Limits		Factor for Central Line	Factors for Control Limits	
	A	A_2	d_2	D_3	D_4
2	2.121	1.880	1.128	0	3.267
3	1.732	1.023	1.693	0	2.575
4	1.500	0.729	2.059	0	2.282
5	1.342	0.577	2.326	0	2.115
6	1.225	0.483	2.534	0	2.004
7	1.134	0.419	2.704	0.076	1.924
8	1.061	0.373	2.847	0.136	1.864
9	1.000	0.337	2.970	0.184	1.816
10	0.949	0.308	3.078	0.223	1.777
11	0.905	0.285	3.173	0.256	1.744
12	0.866	0.266	3.258	0.284	1.716
13	0.832	0.249	3.336	0.308	1.692
14	0.802	0.235	3.407	0.329	1.671
15	0.775	0.223	3.472	0.348	1.652
16	0.750	0.212	3.532	0.364	1.636
17	0.728	0.203	3.588	0.379	1.621
18	0.707	0.194	3.640	0.392	1.608
19	0.688	0.187	3.689	0.404	1.596
20	0.671	0.180	3.735	0.414	1.586
21	0.655	0.173	3.778	0.425	1.575
22	0.640	0.167	3.819	0.434	1.566
23	0.626	0.162	3.858	0.443	1.557
24	0.612	0.157	3.895	0.452	1.548
25	0.600	0.153	3.931	0.459	1.541

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

TABLE XV

CRITICAL VALUES FOR TESTING OUTLIERS*
(x_1 is the extreme value)

STATISTIC	NUMBER OF MEANS, n	CRITICAL VALUES† $\alpha = 0.05$	$\alpha = 0.01$
$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	3	0.941	0.988
	4	0.765	0.889
	5	0.642	0.780
	6	0.560	0.698
	7	0.507	0.637
$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	8	0.554	0.683
	9	0.512	0.635
	10	0.477	0.597
$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	11	0.576	0.679
	12	0.546	0.642
	13	0.521	0.615
$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	14	0.546	0.641
	15	0.525	0.616
	16	0.507	0.595
	17	0.490	0.577
	18	0.475	0.561
	19	0.462	0.547
	20	0.450	0.535
	21	0.440	0.524
	22	0.430	0.514
	23	0.421	0.505
	24	0.413	0.497
	25	0.406	0.489
	26	0.399	0.486
	27	0.393	0.475
	28	0.387	0.469
	29	0.381	0.463
	30	0.376	0.457

*Reproduced with permission frim W. J. Dixon and F. J. Massey, Introduction to Statistical Analysis, McGraw-Hill Book Company, 1951.

Table XVI

*Table of Critical Values for T (One-sided Test) When Standard Deviation is Calculated from the Same Sample**

Number of Observations n	5% Significance Level	2.5% Significance Level	1% Significance Level
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94
23	2.62	2.78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01
30	2.75	2.91	
35	2.82	2.98	
40	2.87	3.04	
45	2.92	3.09	
50	2.96	3.13	
60	3.03	3.20	
70	3.09	3.26	
80	3.14	3.31	
90	3.18	3.35	
100	3.21	3.38	

$$T_n = \frac{x_n - \bar{x}}{s} \qquad s = \left\{\frac{\sum(x_i - \bar{x})^2}{n-1}\right\}^{\frac{1}{2}} = \left\{\frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}\right\}^{\frac{1}{2}}$$

$$T_1 = \frac{\bar{x} - x_1}{s} \qquad x_1 \leq x_2 \leq \cdots \leq x_n$$

NOTE: For $n > 25$, the values of T are approximated. All values have been adjusted for division by $n - 1$ instead of n in calculating s.

*Reproduced with permission from F.E. Grubbs, "Procedures for Detecting Outlying Observations in Samples", Technometrics, Volume 2, 1969.

TABLE XVII

Critical Values for T When Standard Deviation s_ν is Independent of Present Sample

$$T' = \frac{x_n - \bar{x}}{s_\nu} \text{ or } \frac{\bar{x} - x_1}{s_\nu}$$

n	3	4	5	6	7	8	9	10	12
$\nu = df$	α = 1% points								
10	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.04
11	2.72	3.02	3.24	3.39	3.52	3.63	3.72	3.79	3.93
12	2.67	2.96	3.17	3.32	3.45	3.55	3.64	3.71	3.84
13	2.63	2.92	3.12	3.27	3.38	3.48	3.57	3.64	3.76
14	2.60	2.88	3.07	3.22	3.33	3.43	3.51	3.58	3.70
15	2.57	2.84	3.03	3.17	3.29	3.38	3.46	3.53	3.65
16	2.54	2.81	3.00	3.14	3.25	3.34	3.42	3.49	3.60
17	2.52	2.79	2.97	3.11	3.22	3.31	3.38	3.45	3.56
18	2.50	2.77	2.95	3.08	3.19	3.28	3.35	3.42	3.53
19	2.49	2.75	2.93	3.06	3.16	3.25	3.33	3.39	3.50
20	2.47	2.73	2.91	3.04	3.14	3.23	3.30	3.37	3.47
24	2.42	2.68	2.84	2.97	3.07	3.16	3.23	3.29	3.38
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15	3.21	3.30
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08	3.13	3.22
60	2.29	2.52	2.68	2.79	2.88	2.95	3.01	3.06	3.15
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95	3.00	3.08
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88	2.93	3.01
	α = 5% points								
10	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
11	1.98	2.24	2.42	2.56	2.67	2.76	2.84	2.91	3.03
12	1.96	2.21	2.39	2.52	2.63	2.72	2.80	2.87	2.98
13	1.94	2.19	2.36	2.50	2.60	2.69	2.76	2.83	2.94
14	1.93	2.17	2.34	2.47	2.57	2.66	2.74	2.80	2.91
15	1.91	2.15	2.32	2.45	2.55	2.64	2.71	2.77	2.88
16	1.90	2.14	2.31	2.43	2.53	2.62	2.69	2.75	2.86
17	1.89	2.13	2.29	2.42	2.52	2.60	2.67	2.73	2.84
18	1.88	2.11	2.28	2.40	2.50	2.58	2.65	2.71	2.82
19	1.87	2.11	2.27	2.39	2.49	2.57	2.64	2.70	2.80
20	1.87	2.10	2.26	2.38	2.47	2.56	2.63	2.68	2.78
24	1.84	2.07	2.23	2.34	2.44	2.52	2.58	2.64	2.74
30	1.82	2.04	2.20	2.31	2.40	2.48	2.54	2.60	2.69
40	1.80	2.02	2.17	2.28	2.37	2.44	2.50	2.56	2.65
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47	2.52	2.61
120	1.76	1.96	2.11	2.22	2.30	2.37	2.43	2.48	2.57
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.52

*Reproduced from H. A. David, "Revised Upper Percentage Points of the Extreme Studentized Deviate from the Sample Mean", Biometrika, 43, 1956, by permission of the Biometrika trustees.

*U.S. GOVERNMENT PRINTING OFFICE: 1981/703-014/155