**NATIONAL BUREAU OF STANDARDS REPORT**

4817-A
(Supersedes NBS Report 4817)

# Second Draft of
# Part I, Sections 1, 2
# for
# Manual on Experimental Statistics
# for Ordnance Engineers

A REPORT TO
OFFICE OF ORDNANCE RESEARCH
DEPARTMENT OF THE ARMY

NBS

**U. S. DEPARTMENT OF COMMERCE**
**NATIONAL BUREAU OF STANDARDS**

## THE NATIONAL BUREAU OF STANDARDS

### Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

### Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.25) and its Supplement ($0.75), available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.

Inquiries regarding the Bureau's reports should be addressed to the Office of Technical Information, National Bureau of Standards, Washington 25, D. C.

# NATIONAL BUREAU OF STANDARDS REPORT

**NBS PROJECT** 1103-40-5146

7 June 1957

**NBS REPORT** 4817-A

# Second Draft of Part I, Sections 1, 2 for Manual on Experimental Statistics for Ordnance Engineers

Prepared by

Statistical Engineering Laboratory

A Report

to

Office of Ordnance Research
Department of the Army



NBS

**U. S. DEPARTMENT OF COMMERCE**
**NATIONAL BUREAU OF STANDARDS**

## NOTICE

This report is a revised draft of Part I, sections 1 and 2, for the Manual on Experimental Statistics for Ordnance Engineers. It replaces and supersedes NBS Report 4817, 16 August 1956, which was a preliminary draft of the same portion of the manual. A preliminary draft of the remaining portions of the manual will be distributed separately.

The present revision has been prepared with the benefit of the numerous comments and suggestions made by various ordnance establishments and coordinated into a review report by the Office of Ordnance Research.

# PREFACE

(Proposed preface to complete manual)

This manual was prepared by the Statistical Engineering Laboratory, National Bureau of Standards, for the Office of the Chief of Ordnance, Department of the Army, under contract with the Office of Ordnance Research (D/A Project 597-01-001, Ordnance Project TB1-0006).

The Manual discusses a series of problems related to planning or analyzing experiments arising in ordnance research. Techniques appropriate to these problems are outlined in form suitable for computation, with some explanation of the general principles involved and illustrations of the interpretation of results. Worked examples are provided for each technique.

The manual is written primarily for ordnance engineers who have responsibility for planning and interpreting experiments. The statistical techniques discussed are not new. Those relative to a single class of problem are not usually all to be found in any one book. Perhaps when summarized together in a uniform notation, as in this manual, they will be used more frequently, and to better effect.

The text of the manual is primarily the work of Dr. Paul N. Somerville and Mrs. Mary G. Natrella; and was drafted under the guidance of Dr. Churchill Eisenhart, Chief, Statistical Engineering Laboratory.

## Table of Contents

# INTRODUCTION

This manual is a collection of statistical procedures useful in ordnance applications. Each section is largely independent of the other sections, and depends mainly on the understanding of a few basic statistical concepts. Every procedure, test and technique described is illustrated by means of a worked example. A list of authoritative references is included at the end of each major section. It is hoped that the manual will be useful to two types of persons; (1) the person who has had almost no contact with statistics and (2) to the person who merely wants a convenient reference where application of some specific technique is outlined clearly and concisely.

## SOME BASIC STATISTICAL CONCEPTS

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

### Population and Sample

The concepts of population and sample are basic in the use of inductive statistical methods. Any set of individuals or objects having some common observable characteristic constitutes a population (or universe). The population may refer either to the individuals measured or the measurements themselves. Examples of populations are: velocities of individual rounds of ammunition from a given lot, when fired in a standard testing device; barometric pressure at Camp X at 9 A.M., during June, July, August, 1956; all the Corporals in the Marines as of July 1, 1956; measurements of the length of an object as measured by a large number of surveyors.

An example will illustrate a third kind of population. Suppose we select 10 rounds of ammunition from a given lot, and observe their muzzle velocities when fired in a given test weapon. Let $\bar{X}$ be the average muzzle velocity of the ten rounds. If the lot is large, there are many different sets of 10 rounds which could have been obtained from the lot. For each such set of 10 rounds, there is an average muzzle velocity $\bar{X}_i$. These averages, from all possible sets of 10, themselves constitute a <u>population</u> (of averages). This kind of population is frequently called the <u>distribution</u> of $\bar{X}$.

If we were willing or able to examine an entire population, our task would be merely that of <u>describing</u> that population, using whatever numbers, figures or charts we cared to use. Since it is ordinarily inconvenient or impossible to observe every item in the population, we take a sample - i.e., a portion of the population. Our task is now to generalize from our observations on this (usually small) portion, to the population. In order to make valid generalizations from a sample (valid in the sense that we can state a probability that our generalizations are correct) we must have a particular kind of sample - i.e., a random sample. (Methods of drawing random samples are given in the following section). We may wish to know the average velocity of a given lot of .303 ammunition when used in a

standard testing device. We take a sample from the population of rounds (where the population is the rounds in the lot), and measure their velocities. We compute an average velocity (and perhaps a measure of the sample variation) and infer the average lot velocity. Either the rounds themselves or the velocities of the rounds can be taken as the individuals in the population.

Selection of the Sample

The method of choosing the sample is an important factor in determining what use can be made of it. In order for the theory of probability to be applied to statements made about the population, we must have a random sample from that population. A random sample is one in which each individual in the population has an equal chance of appearing. In practice it is not always easy to obtain a random sample. Unconscious selections and biases tend to enter. For this reason, it is often advisable to use a table of random numbers as an aid to selecting the sample. Briefly, the method is as follows. Suppose the population consists of 87 items, and we wish to select a random sample of 10. Assign to each individual a separate two digit number between 00 and 86, Now turn to a table of random numbers, and decide, before looking at the numbers, whether you will read vertically, horizontally, and where you will start. Table 0.00 is a short table of random numbers. Any rule

whatever may be used provided it is fixed in advance and is independent of the numbers occurring. Now read two digit numbers, selecting the individuals whose numbers occur, until 10 individuals are selected.

In all examples in this manual, we shall assume that we are dealing with a random sample.

Some Properties of Populations

Although it is unusual to examine populations in their entirety, the examination of a large sample (or of many small samples) from a population can give us much information about the general nature of the population. One device for studying the nature of a population is a "histogram". Suppose we have a large number of observed items, and a numerical measurement for each item. For example, we have the Rockwell hardness reading for 5000 specimens.

We now make a table showing the number of items which have certain hardness readings. (Data taken from Bowker and Goode, "Sampling Inspection by Variables," McGraw Hill Book Company, 1952).

| Hardness Reading | Frequency of Items |
|---|---|
| 55 | 1 |
| 56 | 17 |
| 57 | 135 |
| 58 | 503 |
| 59 | 1110 |
| 60 | 1470 |
| 61 | 1120 |
| 62 | 490 |
| 63 | 125 |
| 64 | 26 |
| 65 | 3 |

From this frequency table, we can make a histogram (Figure 1a). The height of the bar for any hardness range is the frequency of items in that hardness range. The bar is centered at the tabulated hardness value. If we take the sum of all the bar areas to be one square unit, then the area of an individual bar represents the proportion of the sample having hardness readings within the corresponding range.

Figure 1a

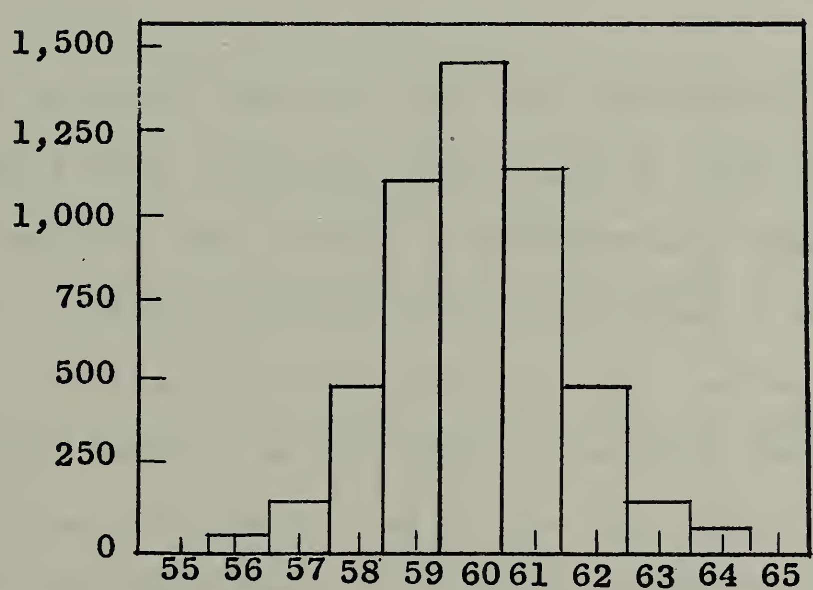
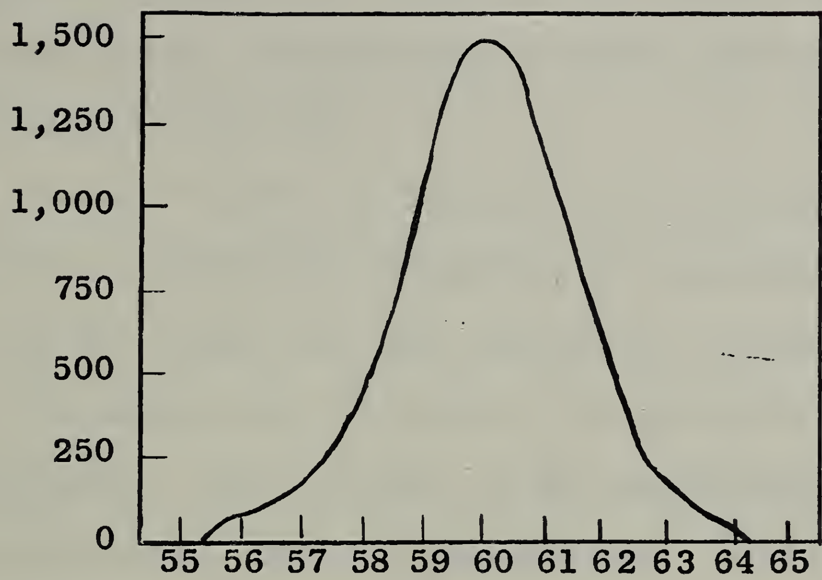
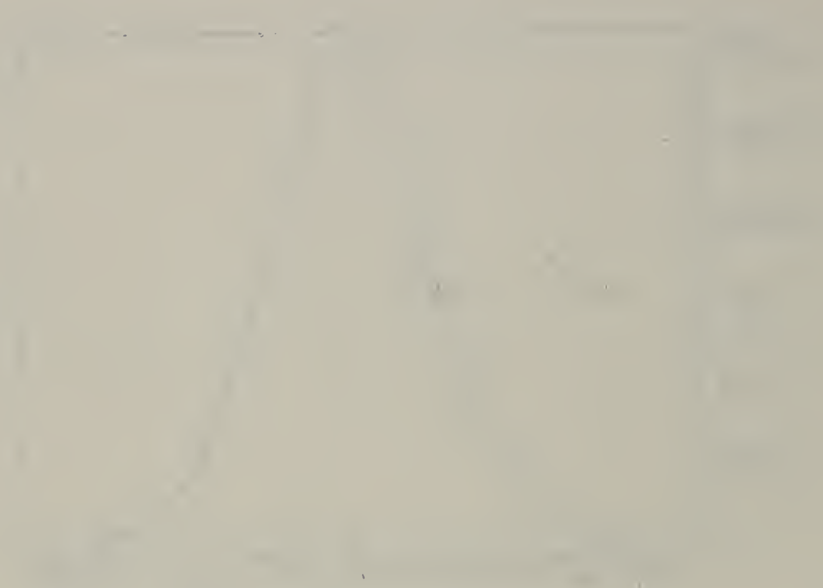


Figure 1b




If we could read our data to one decimal place and if our sample were large, we could make a finer breakdown - i.e., increase the number of groups whose frequency we tabulate. Thus, we would be able to draw a larger number of bars, smaller in width than before. As we get a larger and larger sample, we can keep increasing the number of groups (therefore bars), until the bar width become so small that we can blend their tops into a continuous curve, such as that of Figure 1b. If we were to carry out this sort of scheme on a large number of populations, we would find that many different curves would arise. Possibly the majority of them would be a class of symmetrical bell shaped curves



called "normal" or "Gaussian" distributions. An example is the curve in figure 1b. Some of the curves would not be symmetrical, and occasionally we would find some that were shaped like a J or a U.

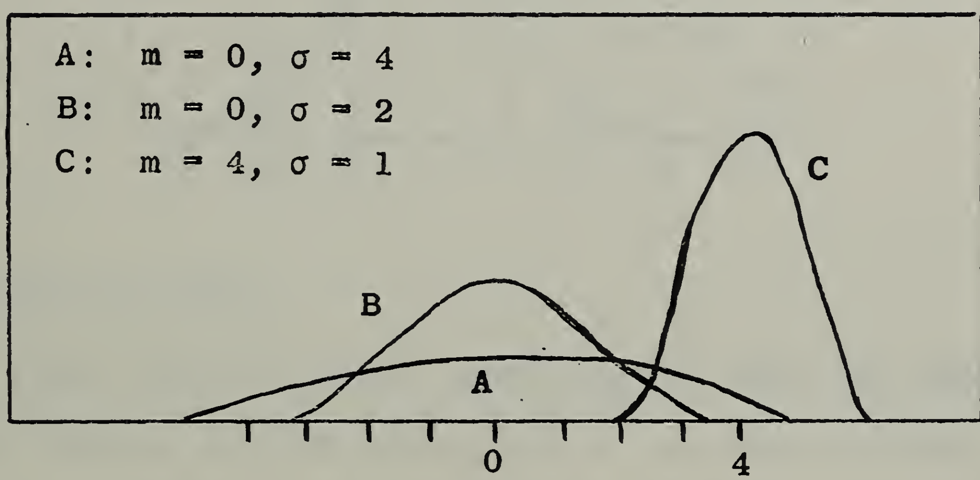
A "normal" curve is completely determined by two parameters. These two parameters are usually represented by $m$ and $\sigma$, which are the population arithmetic mean (or simply the mean) and the population standard deviation respectively. ($\sigma^2$ is known as the population variance). Since the normal curve is symmetrical, $m$ is the value for which the curve is highest. It is useful to note that $\sigma$ represents the distance between $m$ and either inflection point (the inflection point is the point at which the curve changes from concave upward to concave downward). More generally $m$ is the "centre of gravity" of the distribution, while $\sigma^2$ is the second moment about $m$.

The parameter $m$ is the location parameter, while $\sigma$ is a measure of the spread, scatter or dispersion of the population. As we have stated, if the distribution is "normal", then, knowing the values of $m$ and $\sigma$ completely determines the distribution. Three different normal curves are shown in Figure 2.

Let $X$ be any value in the population and let $z = (X-m)/\sigma$. That is, $X$ is $z$ standard deviations above the mean. Table 1 enables us to get $P$, the proportion of

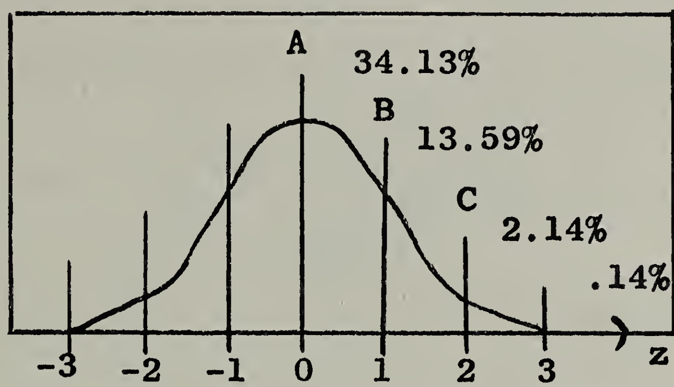
the population below any $z$. Figure 3 shows the percentages of the population in various intervals of $z$.

Figure 2, Normal Distributions




Suppose we know that the chamber pressures of a lot of ammunition form a normal population, with the average chamber pressure in p.s.i. $m = 50{,}000$, and standard deviation $\sigma = 5 \times 10^3$ (p.s.i.). Then from Figure 3, we know that if we fired the ammunition in the prescribed manner, we would expect 50 percent of the rounds to have a chamber pressure above 50,000 p.s.i., 15.9% to have pressures above 55,000 p.s.i., and 2.3% to have pressures above 60,000 p.s.i., etc.

Figure 3, Distribution of Chamber Pressures p




Estimation of m and σ

In areas where a lot of experimental work has been done, it often happens that we know m or σ or both, fairly accurately. However, in the majority of cases it will be our task to estimate them by means of a sample. Suppose we have n observations, $X_1, X_2, \ldots, X_n$ taken at random from a normal population. From the sample, what are our best estimates of m and σ? Actually, it is usual to estimate m and $\sigma^2$, taking our estimate of σ as the square root of the estimate of $\sigma^2$. The recommended estimates of m and $\sigma^2$ are: 1/

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

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

1/ The Greek symbol $\sum$ is often used as shorthand for "the sum of". For example,

$\overline{X}$ and $s^2$ are the sample mean and sample variance respectively (s is called the sample standard deviation). For computational purposes, the following formula is more convenient for $s^2$,

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

It is clear that nearly every sample will contain different individuals, and thus our estimates $\overline{X}$ and $s^2$ of m and $\sigma^2$ will differ from sample to sample. However, our estimates are such that "on the average", or "in the long run" they tend to be equal to m and $\sigma^2$ respectively. If, for example, we have a large number of random samples of size n, the average of the estimates of $\sigma^2$ will tend to be near $\sigma^2$. Furthermore, the amount of fluctuation of the estimates about $\sigma^2$ (or of the $\overline{X}$'s about m if we are estimating m) will be smaller than the fluctuation would be for any estimates other than the recommended ones.

---

1/ (Continued)

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

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

$$\sum_{i=1}^{3} (X_iY_i) = X_1Y_1 + X_2Y_2 + X_3Y_3.$$

The larger our sample size n the more faith we can put in our estimates $\overline{X}$ and $s^2$. This is perhaps not surprising, and is illustrated in figures 4a and 4b. Figure 4a shows the distribution of the sample values of $\overline{X}$ for samples of various sizes. If we define the area under any curve as being one square unit (this is a standard convention in statistics), then the area under the curve between any two $\overline{X}$ values represents the proportion of the time we will expect to get values between those two points. As the curves show, the larger our sample size the less scatter we will have.

Figure 4b shows the distribution of the sample values of $s^2$ for samples of various sizes.

Figure 4a: Distribution of sample means from samples of size n taken from a normal population with $m=0$, $\sigma=1$, $n=1,4,16$ and 25.

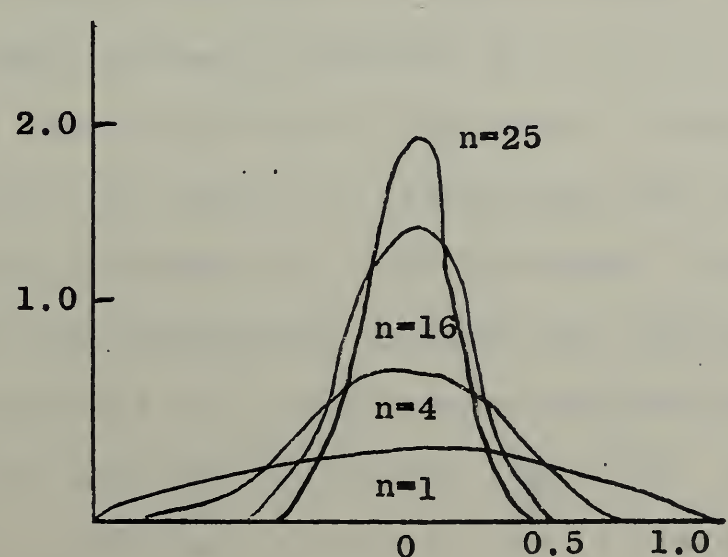
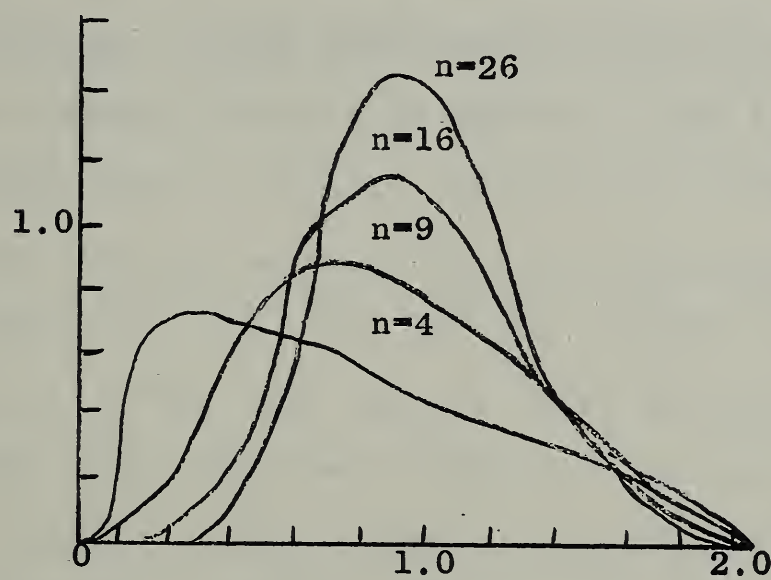


Figure 4b: Distribution of sample variances ($s^2$) from samples of size n taken from a normal population with $\sigma=1$, n=4,9,16 and 26.




## Confidence Intervals

Inasmuch as estimates of m and $\sigma$ vary from sample to sample, interval estimates may sometimes be preferred to "single-value" estimates. Provided we have a random sample from a normal population we can with a chosen degree of confidence make interval estimates of m or $\sigma$. The confidence is not associated with a particular interval, but is associated with the method of calculating the interval. The interval either brackets the true parameter value (m or $\sigma$, whichever we are estimating) or does not, and our confidence coefficient will be the proportion of samples for which our method will be expected to bracket the value. The interval is known as a confidence interval, and is always associated

with a confidence coefficient. As we would expect, larger samples tend to give narrower interval estimates.

Suppose we are given the lot of ammunition mentioned earlier, and wish to make confidence interval estimates of the average chamber pressure of rounds in the lot. Assume that the population of chamber pressures is normal, and that the true average is 50,000 p.s.i., (although this value is unknown to us). Let us take a random sample of four rounds, and from this sample, using the given procedure, calculate the upper and lower limits for our confidence interval. Consider all the possible samples of size 4 that we could have taken, and the accompanying upper and lower limits for the confidence intervals computed from each. If the limits were for a 90% confidence interval, then we should expect 90% of the intervals to cover the true value, 50,000 p.s.i.

Figure 5: Illustration showing computed confidence intervals for 100 samples of size 4 drawn at random from a normal population with $\mu = 50,000$ p.s.i., $\sigma = 5,000$ p.s.i. (Adapted from A.S.T.M. Manual on Quality Control of Materials).

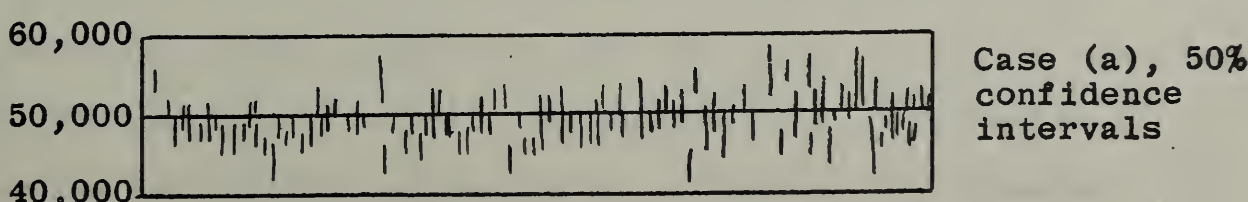
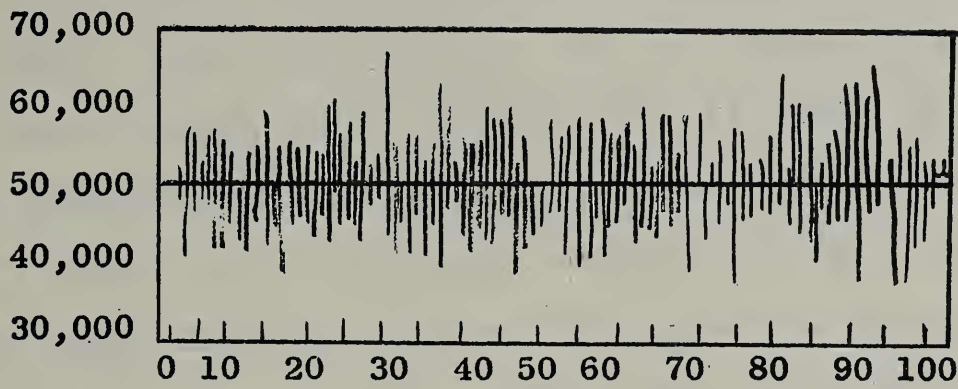


Figure 5: (Continued)




Case (b), 90% confidence intervals

In the first diagram 51 of the 100 intervals we have drawn, actually include the true mean. For 50% confidence interval estimate, we would expect in the long run that 50% of the intervals would include the true mean. Fifty one out of 100 is close, and is in fact a reasonable variation from the expected number 50.

In the second diagram 90 out of 100 of the intervals contain the true mean. This is precisely the expected number for 90% confidence intervals.

## Statistical Tolerance Limits

Sometimes what is wanted is not an estimate of the mean and variance of the population but two outer values or limits which contain nearly all of the population. For example if extremely low chamber pressures or extremely high chamber pressures might cause serious problems, we may wish to know

approximately the range of chamber pressures of a lot of ammunition. There are methods for obtaining the approximate range. More specifically, what we can do is give a lower and an upper lkmit, and say that at least P percent of the ammunition will have chamber pressures within the above limits, with a confidence coefficient of $\gamma$. If we use the prescribed method (see 1.5), then the proportion of the time that we will be making true statements will be $\gamma$. The different meanings of the terms "confidence intervals", "statistical tolerance limits" and "engineering tolerance limits" should be noted. A "confidence interval" is an interval within which we estimate a given parameter (e.g., the population mean m) to lie. "Statistical tolerance limits" for a given population are limits within which we expect a stated proportion of the population to lie. Engineering tolerance limits are specified outer limits of acceptability usually designated by a design engineer.

Using Statistics to Make Decisions

Ten rounds of a new type of shell are fired into a target, and the depth of penetration is measured for each round. The depths of penetration are 10.0, 9.8, 10.2, 10.5, 11.4, 10.8, 9.8, 12.2, 11.6, 9.9 cms. The average penetration depth of the standard comparable shell is 10.0 cm. We wish to know if the new type shells penetrate farther on the average than the standard.

If we compute the arithmetic mean of the ten shells, we find it is 10.7 cm. Our first impulse might be to state that on the average the new shell will penetrate 0.7 cm., further than the standard shell. This indeed is our best guess, but how sure can we be that this is actually the case? If we were forced to decide on the basis of the above ten shells alone whether to keep on making the standard shells or to convert our equipment to making the new shell, what would be our choice (assume for simplicity that for all other characteristics there are no differences, or that the differences are irrelevant)?

One thing that might catch our notice is the variability in the penetration depths. The standard deviation as calculated from the sample is 0.73 cm. Could it be that the new shell is on the average no better than the standard? There is variation from shell to shell, so might not our sample of ten shells have contained some of the ones which have unusually high penetrating power? If the new shell really has no more penetrating power than the standard shell, how improbable is it that we should get a sample average of shells differing from the standard by as much as our sample did? If it is highly improbable, then we should undoubtedly come to the conclusion that the new shell did indeed penetrate farther than the standard shell and we might take practical steps toward putting into production the new type of shell.

If it were not improbable (i.e., reasonably likely) that we should get a sample mean differing as much as this one then we would have no good reason to believe that the new shells penetrated farther than the standard shells.

Setting up the Decision Procedure

In our example we wish to know if the population of new shells will penetrate farther on the average than the old shells. Inasmuch as it is frequently easier to disprove than to prove, we start with what we call the null hypothesis - the hypothesis of no difference. That is, we make the hypothesis that the old shells are as good as the new shells. Then, if in our sample of new shells we get an improvement so large that it is unlikely to be due to statistical fluctuation, we reject the null hypothesis. We make Decision (i). The new shells penetrate farther on the average than the standard shells.

On the other hand we may be able to "explain" the "increase" in average penetration shown by the sample of new shells as within the realm of statistical fluctuation. In this case we make Decision (ii) - there is no reason to believe the new shells penetrate farther on the average than the standard shells.

Level of Significance

We have stated that we will reject the "null hypothesis" i.e., make decision (i), when our increase in penetration is

so large that it is improbable under the null hypothesis. We may say the increase is improbable under the null hypothesis if by sample fluctuation alone, an increase as large as the observed one would occur in at most a proportion $\alpha$ of possible samples ($\alpha$ is some small fraction decided on in advance of performing the experiment). The quantity $\alpha$ defined above is known as the "significance level." The significance level should be chosen on economical and other non-statistical grounds. Two values for $\alpha$ have been made use of in extensive tabulation of many test statistics, and it is common to choose one or the other of these. There is, however, nothing unique about them. The levels are $\alpha = 0.05$, and $\alpha = 0.01$. Using the .05 level of significance, for example, we should reject the null hypothesis (make decision (i)) no more often than .05 of the time when in fact the null hypothesis was true (i.e., the old shells were as good as the new).

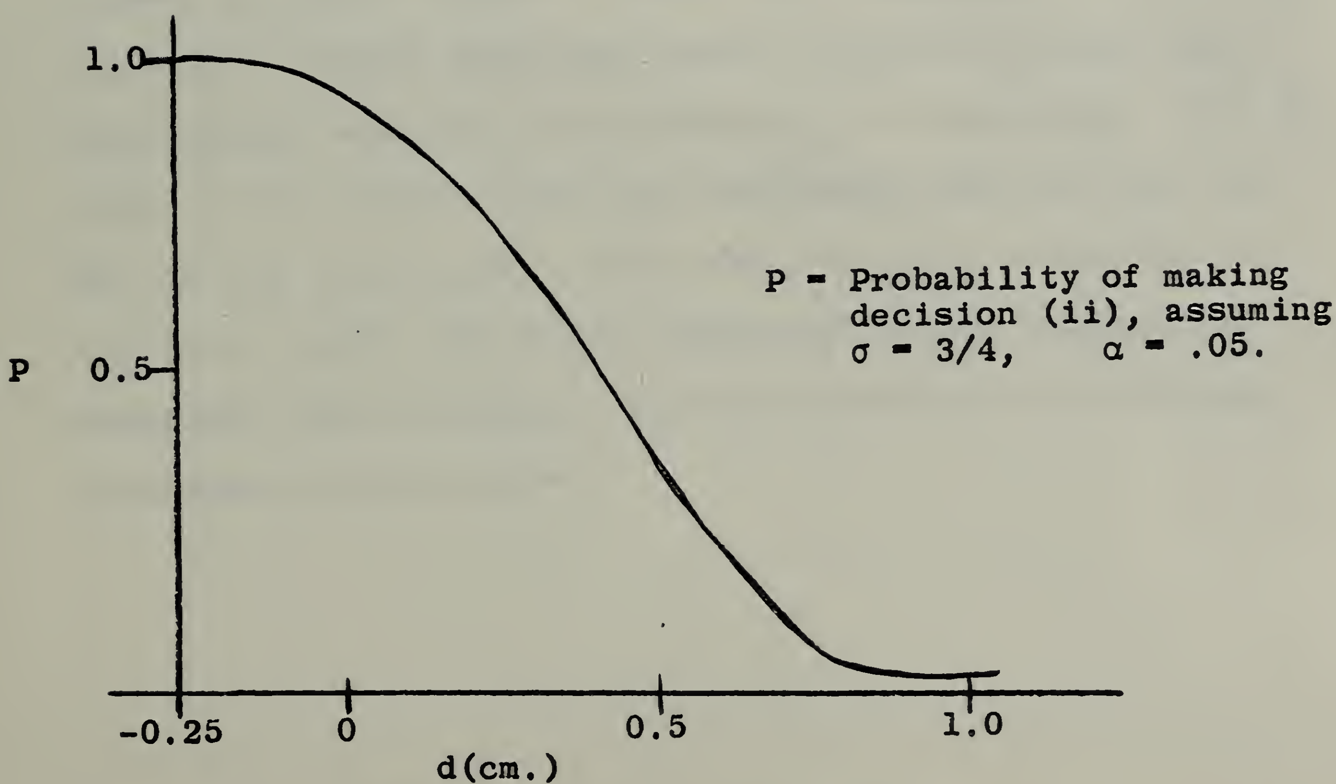
Two Kinds of Errors

Since there is fluctuation in sample means and sample standard deviations, it is obvious that we run the risk of error in our decisions. As a matter of fact we speak of errors of the first kind, and errors of the second kind. If we reject the null hypothesis when it is true, i.e., announce a difference which really does not exist, then we make an "Error of the First Kind". The "Error of the

First Kind", is equal to $\alpha$, the "Significance Level." If we fail to reject the null hypothesis when it is false, e.g., fail to find an improvement in the new shell over the old, when an improvement exists, then we make what is called an "Error of the Second Kind". Although we do not know in a given instance whether we have made an error in decision, we can know the probability of making either type of error.

For the above example of penetration depths, the graph in Figure 6 gives the probability $P$ of making decision (ii) (i.e., saying that the new shell penetrates no farther than the old) for varying values of $d$. $d$ is the true average penetration depth for the new shell minus the true average penetration depth for the old shell.

Figure 6: Operating Characteristics of Decision Procedure Of Example Given on page 16 ff.




d = Amount by which population average penetration depth for new shell exceeds that for standard shell.

# PART I

## SOME STANDARD TECHNIQUES FOR QUANTITATIVE DATA

The techniques described in Part I apply to the analysis of numerical results of experiments. The results must be expressed as actual measurements in some conventional units on a continuous scale. They do not apply to the analysis of data in the form of proportions, percentage, or counts.

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

# 1. Performance of an item.

## 1.1 Estimating average performance from a sample.

<u>Given</u>:

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

<u>Example 1.1</u>:

Ten mica washers are taken at random from a large group and their thicknesses measured as follows (inches):

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

<u>Questions</u>: The general question is "what can we say about the larger group?" - specifically,

i) What is our best guess as to the average thickness of the whole lot? (see 1.1.1)

ii) Can we give an interval which we expect, with certain confidence, to bracket the true average - i.e., a "confidence interval?" (see 1.1.2, 1.1.3, and 1.1.4)

<u>Note</u>: A common question which is quite different will be treated in 1.5; Can we give an interval within which we expect, with chosen confidence, to find a specified proportion of the individual items - i.e., can we set "statistical tolerance limits"? (see 1.5)

1.1.1 Best single estimate.

The most common, and ordinarily "the best" single estimate is simply the arithmetic mean.

**Procedure:**

Compute the arithmetic mean

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

(For some asymmetrical distributions, the arithmetic mean may not necessarily be the best single description of the over-all performance of items. In these cases, the median, mode or some percentile may be a more meaningful description of the population).

**Example:**

$$\bar{X} = \frac{1.260}{10} = .1260$$

1.1.2 Confidence interval estimate (when knowledge of the variability cannot be assumed). When we take a sample from a lot or a population, the sample average will seldom be exactly the same as the lot or population average. We do, however, hope that it is fairly close, and we might be willing to state an interval which we are confident will bracket the lot mean. If we regularly made such interval estimates, in a particular fashion, and found that over a long period of time these intervals actually did contain the true mean 99% of the time, we might say that we were operating at a 99% confidence level. Our particular kind of interval estimates might likewise be called "99% confidence intervals." Similarly if our intervals included the true average 95 % of the time, we would be operating at a 95% confidence level, and our intervals would be called 95% confidence intervals. In general, if in the long run we expect 100(1-$\alpha$)% of our intervals to contain the true value, we are operating at 100(1-$\alpha$)% confidence.

We may choose whatever confidence level we wish. Commonly used levels are 99% and 95%, which correspond to $\alpha = .01$ and $\alpha = .05$. (In later sections we speak of the "significance level" ($\alpha$)

of a test. This is the same $\alpha$ which appears here in the general expression for confidence level). If we wish to estimate the mean of a large group (population) using the results of a random sample from that group, the following procedure will allow us to make interval estimates at any chosen confi-level. (It is assumed that the large group forms a normal population.) We may make a 2-sided interval estimate, expected to bracket the mean; or make a one-sided interval estimate, to give an open interval (limited on the upper or lower side as we choose) expected to contain the mean.

1.1.2.1 Two-sided confidence interval. - This procedure gives an interval which we expect to bracket the true mean $100(1-\alpha)\%$ of the time.

| Procedure | Example |
|---|---|
| Problem: What is a $100(1-\alpha)\%$ confidence interval (2-sided) for the true mean?<br>i) Choose the desired confidence level, $1-\alpha$ | Problem: What is a 95% (2-sided) confidence interval for the true mean? (data from example 1.1)<br>i) Choose confidence level .95<br>$.95 = 1-\alpha$<br>$\alpha = .05$ |
| ii) Compute:<br>arithmetic mean $\bar{X}$<br>(see 1.1.1)<br>$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}}$ | ii)<br>$\bar{X} = .1260$ inches<br>$s = 0.00359$ inches |
| iii) Look up:<br>$t = t_{1-\alpha/2}$ for n-1 degrees of freedom in Table II. | iii)<br>$t = t_{.975}$ for 9 degrees of freedom $= 2.26$ |

| Procedure | Example |
| --- | --- |
| iv) Compute: $X_U = \bar{X} + t \frac{s}{\sqrt{n}}$ | iv) $X_U = \bar{X} + t \frac{s}{\sqrt{n}}$ $= .1260 + \frac{2.26(.00359)}{\sqrt{10}}$ $= .1286$ inches |
| $X_L = \bar{X} - t \frac{s}{\sqrt{n}}$ | $X_L = \bar{X} - t \frac{s}{\sqrt{n}}$ $= .1260 - \frac{2.26(.00359)}{\sqrt{10}}$ $= .1234$ inches |
| v) Conclude: The interval from $X_L$ to $X_U$ is a $100(1-\alpha)\%$ confidence interval for the true mean. | v) Conclude: The interval from .1234 to .1286 inches is a 95% confidence interval for the lot mean. |

1.1.2.2 One-sided confidence interval.

The example used in 1.1.2.1, can be used to make another kind of confidence interval statement. 100 $\alpha/2$ percent of the time the interval in 1.1.2.1 will be above the true mean (i.e., $X_L$ is greater than true mean). Therefore 100(1-$\alpha$/2) percent of the time, the true mean is greater than $X_L$. From the example of 1.1.2.1,

$$100\ (1 - \frac{\alpha}{2})\text{ percent} = 97.5\text{ percent.}$$

Thus, either of the two open intervals - above .1234 inches, or below .1286 inches can be called a 97.5 percent one-sided confidence interval for the population mean.

We also give the complete example for a 1-sided interval for a different choice of confidence level.

| Procedure | Example |
|---|---|
| Problem: What is a 100(1-$\alpha$)% confidence interval (one-sided) for the true mean? | Problem: What is a value, which we expect, with 99% confidence, to be exceeded by the lot mean? (data from example 1.1) |
| i) Choose the desired confidence level (1-$\alpha$) | i) (1-$\alpha$) = .99 , $\alpha$ = 0.01. |

| Procedure | Example |
|---|---|
| ii) Compute: $\overline{X}$<br>$s$ | ii) $\overline{X}$ = .1260 inches<br>$s$ = 0.00359 inches |
| iii) Look up:<br>$t = t_{1-\alpha}$ for n-1 degrees of freedom in Table II | iii)<br>$t = t_{.99}$ for 9 degrees of freedom = 2.82 |
| iv) Compute:<br>$X'_L = \overline{X} - t \frac{s}{\sqrt{n}}$<br>(or compute<br>$X'_U = \overline{X} + t \frac{s}{\sqrt{n}}$ ) | iv)<br>$X'_L = .1260 - \frac{(2.82)(.00359)}{\sqrt{10}}$<br>$X'_L = .1228$<br>(or $X'_U = .1292$) |
| v) Conclude:<br>We are 100(1-α)% confident that the lot mean is greater than $X'_L$. (or we are 100(1-α)% confident that the lot mean is less than $X'_U$) | v) Conclude:<br>We are 99% confident that the lot mean is greater than .1228 inches. (or we are 99% confident that the lot mean is less than .1292 inches). |

1.1.3 Confidence interval estimates when we have previous knowledge of the variability.

In the previous section (1.1.2) we have assumed that we had no previous information about the variability of performance of items, and were limited to using the variability estimated from the sample. Suppose that in the case of the mica washers, we had taken samples many times previously from the same process and found that, although each batch had a different average, there was always about the same amount of variation within a batch. We may then be able to assume that we know $\sigma$, the standard deviation of the lot, from this previous experience. This assumption should not be made casually, but only after real investigation of the stability of the variation among samples using control chart techniques.

The procedure for computing these confidence intervals is simple. In the example of 1.1.2, merely replace $s$ by $\sigma$ and $t_{1-\alpha/2}$ by $z_{1-\alpha/2}$, and the formulas remain the same. Values of $z_{1-\alpha/2}$ are given in Table I. Note that $z_{1-\alpha/2}$ and $t_{1-\alpha/2}$ with an infinite number of degrees of freedom have the same value.

| Procedure | Example |
| --- | --- |
| Problem: Find a $100(1-\alpha)\%$ 2-sided confidence interval for the lot mean, using known $\sigma$. | Problem: What is a 95% confidence interval (2-sided) for the lot mean? ($\sigma$ known equal to .0040 inches). (Data from example 1.1) |
| i) Choose the desired confidence level, $1-\alpha$ | i) $1-\alpha = .95$, thus $\alpha = .05$ |
| ii) Compute $\overline{X}$ | ii) $\overline{X} = .1260$ inches |
| iii) Look up: $z = z_{1-\alpha/2}$ in Table I | iii) $z = z_{1-\alpha/2} = 1.96$ |
| iv) Compute: $X_U = \overline{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ <br> $X_L = \overline{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ | iv) $X_U = .1260 + 1.96 \left(\frac{.004}{\sqrt{10}}\right)$ <br> $= .1285$ <br> $X_L = .1235$ |

### 1.1.4 Confidence intervals when normality cannot be assumed.

When the departures from normality are not great, or when the sample sizes are moderately large, confidence in the interval estimates made as described in 1.1.2 and 1.1.3, will usually be changed very little from the chosen level.

## 1.2 Estimating the variability of performance from a sample

**Given:**

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

**Example 1.2:**

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

| | |
|---|---|
| 50.7 | 69.8 |
| 54.9 | 53.4 |
| 54.3 | 66.1 |
| 44.8 | 48.1 |
| 42.2 | 35.5 |

### 1.2.1 Single estimates

#### 1.2.1.1 $s^2$ and s

In the introduction we have stated that our best estimate of $\sigma^2$, the variance of a normal population is:

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

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

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

*) It is important to carry more than the usual number of decimal places in the computation of $\Sigma X_i^2$ and $(\Sigma X_i)^2$. If we do not, the subtraction involved in the computation of $s^2$ may result in a value of $s^2$ which lacks significance. Rounding off before subtracting may even result in a negative value for $s^2$.

The estimate of $\sigma$, the population standard deviation, is

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

Example: Using the data of example 1.2, $\Sigma X_i^2 = 27987.54$, $\Sigma X_i = 519.8$ and thus $s^2 = 107.593$ seconds, $s = 10.37$ seconds.

1.2.1.2 Use of the range to estimate variability.

The "range" of n observations is defined as the difference between the highest and the lowest values. For small samples (n less than 10), the range is a reasonably efficient estimator of $\sigma$ (the standard deviation of a normal population) - not as efficient as <u>s</u>, but easier to calculate. Table 1.2.1.2 gives the factors which convert from observed range in a sample of n to an estimate of population standard deviation.

TABLE 1.2.1.2

Table of Factors for Converting the Range to an Estimate of $\sigma$, the Population Standard Deviation)

Estimate of $\sigma = b_n \times$ range (in a sample of n)

| Size of Sample n | $b_n$ |
|---|---|
| 2 | .8862 |
| 3 | .5908 |
| 4 | .4857 |
| 5 | .4299 |
| 6 | .3946 |
| 7 | .3698 |
| 8 | .3512 |
| 9 | .3367 |
| 10 | .3249 |

Note that $b_n$ is approximately equal to $1/\sqrt{n}$ for n small, say less than 10. Thus, a quick estimate of $\sigma$ can be obtained by dividing the range by $\sqrt{n}$.

1.2.2 Confidence interval estimates.

Confidence. As in 1.1.2 we say we have a confidence of 1-α in an interval estimate, if the method of constructing the interval will result in correct statements 100(1-α)% of the time; i.e., in the long run our intervals will contain the true value a proportion 1-α of the time.

1.2.2.1 Two-sided confidence interval estimates.

We are interested in an interval which beackets the true measure of variability of the normal population.

| Procedure | Example |
|---|---|
| Problem: What is a 100(1-α) percent confidence interval for σ? | Problem: What is a 95% confidence interval for σ, the variability of the burning time of the lot of powder? (Data from example 1.2) |
| i) Choose 1-α, the confidence coefficient. | i) .95 = 1-α<br>α = .05 |
| ii) Compute s,<br>$s = \sqrt{\frac{n\Sigma X_i^2 - (\Sigma X_i)^2}{n(n-1)}}$ | ii) s = 10.37 seconds |

| Procedure | Example |
| --- | --- |
| iii) Look up: $A_{\alpha/2}$, $A_{1-\alpha/2}$ for n-1 degrees of freedom in Table XXIII. | iii) For 9 degrees of freedom, $A_{.025} = .6878$ $A_{.975} = 1.826$ |
| iv) Compute: $s_L = A_{\alpha/2}\, s$, and $s_U = A_{1-\alpha/2}\, s$ | iv) $s_L = (10.37)(.6878) = 7.13$ $s_U = (10.37)(1.826) = 18.94$ |
| v) Conclude: Our two-sided interval estimate for $\sigma$ is the interval $s_L$ to $s_U$ and we are $100(1-\alpha)\%$ confident that the interval contains $\sigma$. | v) Conclude: Our two-sided interval estimate for $\sigma$ is the interval from 7.13 to 18.94 and we are 95% confident that the interval contains $\sigma$. |

1.2.2.2 One-sided confidence interval estimate.

In some cases we are not interested in a bracketing interval, but only in knowing whether the variability is large. We would then be happy with a statement such as the following:

We are $100(1-\alpha)\%$ confident that the variability as measured by $\sigma$ is less than some value C. Similarly we may be interested only in statements that the variability is greater than some number C'. Both statements are one-sided confidence interval estimates.

| Procedure | Example |
|---|---|
| **Problem:** Can we give a value C and be $100(1-\alpha)\%$ confident that $\sigma$ is less than C. | Problem: Can we give a value C, and be 95% confident that $\sigma$ is less than C? (Data from example 1.2). |
| i) Choose $1-\alpha$, the confidence coefficient | i) $1-\alpha = .95$<br>$\alpha = .05$ |
| ii) Compute s | ii) s = 10.37 seconds |
| iii) Look up: $A_{1-\alpha}$ for n-1 degrees of freedom in Table XXIII. | iii) For 9 degrees of freedom $A_{.95} = 1.645$ |

| Procedure | Example |
|---|---|
| iv) Compute: | iv) |
| $s'_U = A_{1-\alpha}\ s$ | $s'_U = (1.645)(10.37)$ |
| | $s'_U = 17.05$ seconds |

Therefore we are 95% confident that the variability as measured by $\sigma$ is <u>less</u> <u>than</u> $s'_U = 17.05$ seconds.

### 1.2.3 Estimating the Standard Deviation of an Item, Product or Process, When no Previous Data is Available.

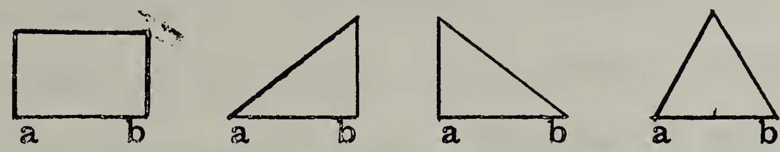
Frequently it is very desirable to have some idea of the magnitude of the variation as measured by $\sigma$, the standard deviation. In planning experiments for example, the sample size required in order to meet certain requirements is a function of $\sigma$. There is seldom a situation where one does not know something about the variance, or cannot use some existing information to get at least a very rough estimate of $\sigma$. The necessary information involves the form of the distribution and the spread of values. If the values for the individual items can be assumed to form a normal distribution, then either of the following methods can be used to get an estimate of $\sigma$.

a) Choose values $a_1$ and $b_1$ between which you expect 99.7% of all individuals to be. Estimate $\sigma$ as $\frac{|a_1 - b_1|}{6}$ or

b) Choose values $a_2$ and $b_2$ between which you expect 95% of all individuals to be. Estimate $\sigma$ as $\frac{|a_2 - b_2|}{4}$

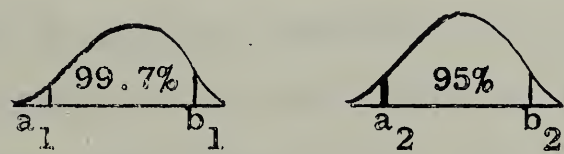
If in fact the populations are not "normal" but follow one of the forms in Figure 1.2.3, then the standard deviation may be estimated as indicated in the figure.

Distribution




Standard Deviation: $\frac{b-a}{3.5}$ $\frac{b-a}{4.2}$ $\frac{b-a}{4.2}$ $\frac{b-a}{4.9}$

Distribution: Normal




Standard Deviation: $\frac{b_1-a_1}{6}$ $\frac{b_2-a_2}{4}$

Figure 1.2.3

Reference, W. E. Deming, "Some Theory of Sampling", John Wiley and Sons, Inc., page 62.

1.3 Number of measurements required to establish the mean of a population with a stated precision.

In planning experiments, we frequently wish to know how many measurements (or how large a sample) we must take in order to be fairly certain that we know the mean $m$ of some population. Suppose we are willing to allow a margin of error $d$, and a risk $\alpha$ that our estimate of $m$ will be out by an amount $d$ or greater. Then, if we assume the population is "normal", we can ascertain the required sample size $n$ provided that we have an estimate $s$ of $\sigma$, the standard deviation of the population, or we are willing to assume some value for $\sigma$. If we do not have an estimate, or are unwilling to assume some value for $\sigma$, then we must use a two-stage sample (See 1.3.2). The latter method has the advantage that the result is independent of our estimates or guesses of $\sigma$, and will usually result in a smaller total sample size.

1.3.1 Estimation of the mean of a population using a single sample.

| Procedure | Example |
|---|---|
| Problem: We wish to know the sample size required to ascertain the mean m of a population. We are willing to take a risk $\alpha$ that our estimate is out by d or more. We have an estimate s of the population standard deviation, with $\nu$ degrees of freedom. | Problem: We wish to know the average thickness of the washers in a given lot. We are willing to take a risk of .05 of being in error in our estimate by 0.02 inches or more. From a previous sample from another lot we have an estimate of the population standard deviation of $s = .00359$ with 9 degrees of freedom. |
| i) Choose d, the allowable margin of error and $\alpha$, the risk that our estimate of m will be out by d or more. | i) $d = 0.02$ inches<br>$\alpha = .05$ |
| ii) Look up $t_{1-\alpha/2}$ for $\gamma$ degrees of freedom in Table II. | ii) $t = t_{.975}$ for 9 degrees of freedom $= 2.26$. |

| Procedure | Example |
|---|---|
| iii) Compute: $n = \frac{t^2 s^2}{d^2}$ | iii) $n = \frac{(2.26)^2(.00359)^2}{(.02)^2} = 12.3$ We will need a sample of size 13 in order to ascertain the lot mean with the required precision. |

If we know $\sigma$, or assume some value for $\sigma$, then we should replace s by $\sigma$ and $t_{1-\alpha/2}$ by $z_{1-\alpha/2}$ in the above procedure.

1.3.2 Estimation using a sample which is taken in two stages.

In many situations, we may not have a good estimate of $\sigma$, the standard deviation of the population whose mean we are trying to estimate. This is especially true when the cost of sampling is high, and rather than take a chance on using a larger sample than is really necessary, we prefer to take the sample in two stages. The method (sometimes called Stein's method) is roughly as follows: We make a guess for the value of $\sigma$, From this we can determine $n_1$ the size of our first sample. From the first sample we get an estimate s of the population standard deviation. We use this to ascertain how large the second sample should be.

| Procedure | Example |
|---|---|
| **Problem:** We wish to know the sample size required to ascertain the mean m of a population. We are willing to take a risk $\alpha$ that our estimate is out by d. | **Problem:** We have a large lot of electrical devices, and wish to determine their average life in hours under specified conditions using an accelerated test. We are willing to take a risk of .05 of our estimate being in error by 30 hours. |

| Procedure | Example |
|---|---|
| i) Choose d, the allowable margin of error, and $\alpha$, the risk that our estimate of m will be out by d or more. | i) $d = 30$ hours<br>$\alpha = .05$ |
| ii) Let $\sigma'$ be the best guess possible as to the value of $\sigma$, the standard deviation of the population (See section 1.2.3) | ii) From our knowledge of similar devices our best estimate of $\sigma$ is 200 hours. |
| iii) Look up $z_{1-\alpha/2}$ in Table I. | iii) $z_{.975} = 1.96$ |
| iv) Compute: $n' = \left(\frac{z_{1-\alpha/2}\sigma}{d}\right)^2$<br>n' is our first estimate of the total sample size that will be required. | iv) $n' = \frac{(1.96)^2(200)^2}{(30)^2} = 170.7$ |

| Procedure | Example |
| --- | --- |
| v) Choose $n_1$ the size of the first sample. This should be considerably less than $n'$. If we guessed too large a value for $\sigma$, then this will protect us against a first sample which is already larger than we need. A rough rule might be to make $n_1 \geq 30$ unless $n' < 60$, in which case we should let $n_1$ be somewhere between $.5n'$ and $.7n'$. | v) Let $n_1 = 50$ |
| vi) Observe the sample of $n_1$. Compute $s_1$, the sample standard deviation. | vi) Make life tests on 50 devices chosen at random, $s_1 = 160$ hours. |
| vii) Look up $t_{1-\alpha/2}$ for $n_1-1$ degrees of freedom. | vii) $t = t_{.975}$ for 49 degrees of freedom = 2.01. |

| Procedure | Example |
| --- | --- |
| viii) Compute: $n = \left(\frac{ts_1}{d}\right)^2$ n is the total required sample size for the first and second samples. We will then require a second sample size of $n_2 = n - n_1$. | viii) $n = \frac{(2.01)^2(160)^2}{(30)^2} = 114.9$ i.e., n = 115 $n_2 = 115 - 50$ $= 65$ We will require an additional 65 devices to be tested. |

1.4 Number of measurements required to establish the variability with given precision.

We may wish to know the size of sample required to estimate the standard deviation with certain precision. If we can express this precision as a percentage of the true (unknown) standard deviation, we can use the curves in Figure 1.4. (Figure 1.4 to be reprinted from an article by J. A. Greenwood and M. M. Sandomire, "Sample Size Required for Estimating the Standard Deviation as a Percent of its True Value," Journal of the American Statistical Association, Volume 45, No. 250, June 1950).

Problem:

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

Procedure:

If we choose P = 20%, and confidence coefficient .95, we read the curve labelled 20% at the point on the horizontal scale ("confidence coefficient") marked .95. This gives a value on the vertical scale ("degrees of freedom, n") equal to 46.

The required degrees of freedom therefore = 46. The required number of measurements in a simple series is one plus the value read from the graph = 1 + 46 = 47.

1.5 Statistical Tolerance Limits - or estimating the proportion of individual items between (above, below) given limits.

Sometimes we are more interested in the approximate range of values in a lot or population than we are in its average. We might, for example like to be able to give two values A and B between which we can be fairly certain that at least a proportion P of the population will lie, (two-sided limits), or a value A above which at least a proportion P will lie, (one-sided limit).

In the example of mica washers (see 1.1), we might want to give 2 thickness values and state (with chosen confidence) that a proportion P (at least) of the washers in the lot will have thicknesses between these 2 limits. In this case we call our confidence coefficient $\gamma$, and it refers to the proportion of the time that our method will result in correct statements.

1.5.1 Two-sided tolerance limits.

| Procedure | Example |
|---|---|
| Problem: We would like to state 2 thickness limits within which we are $100\gamma$ percent confident that 100P percent of the values lie. | Problem: We would like to state 2 thickness limits within which we are 95% confident that 90% of the values lie (data from example 1.1) |
| i) Choose P, the proportion and $\gamma$, the confidence coefficient | i) $P = .90$<br>$\gamma = .95$ |
| ii) Compute from the sample:<br>$\overline{X}$, the arithmetic mean<br>s, the standard deviation | ii) $\overline{X} = .1260$ inches<br>$s = 0.00359$ inches |
| iii) Look up K for chosen P and $\gamma$ in Table X. | iii) $K = 2.839$ |
| iv) Compute:<br>$X_U = \overline{X} + Ks$<br>$X_L = \overline{X} - Ks$ | iv)<br>$X_U = .1260+2.839(.00359)$<br>$= 0.136$ inches<br>$X_L = .1260-2.839(.00359)$<br>$= 0.116$ inches |

| Procedure | Example |
| --- | --- |
| v) Conclusion: With a confidence coefficient of $\gamma$, we may predict that a proportion P of the individuals of the population will have values between $X_L$ and $X_U$. | v) Conclusion: With 95% confidence, we may predict that 90% of the washers have thicknesses between 0.116 and 0.136 inches. |

1.5.2 One-sided tolerance limits.

Sometimes we are more interested in estimating a value above or below which a proportion P (at least) will lie.

In this case the tolerance limits will be

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

for the one-sided upper limit and

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

for the one-sided lower limit. The values for K are not the same as those of 1.5.1.

| Procedure | Example |
|---|---|
| Problem: Give a single value above which you predict with confidence $\gamma$ that a proportion P of the population will lie. | Problem: Give a single value above which you predict with 90% confidence that 97.5% of the population will lie. (Use the data of example 1.1). |
| i) Choose P the proportion and $\gamma$, the confidence coefficient. | i) P = 97.5<br>$\gamma = .90$ |
| ii) Compute:<br>$\overline{X}$, the arithmetic mean<br>s, the standard deviation | ii)<br>$\overline{X} = .1260$ inches<br>s = 0.00359 inches |

| Procedure | Example |
|---|---|
| iii) Compute: | iii) |
| $a = 1 - \frac{z_\gamma^2}{2(n-1)}$ | $a = 1 - \frac{(1.282)^2}{18} = .9085$ |
| (where z can be found in Table I). | |
| $b = z_p^2 - \frac{z_\gamma^2}{n}$ | $b = (1.960)^2 - \frac{(1.282)^2}{10} = 3.677$ |
| $K = \frac{z_p + \sqrt{z_p^2 - ab}}{a}$ | $K = \frac{1.960+\sqrt{1.960^2-(.9085)(3.677)}}{.9085}$ <br> $= 2.93$ |
| iv) $X_L = \bar{X} - Ks$ | iv) $X_L = .1260-2.93(.00359)$ <br> $= .115$ inches |

Thus we are 90% confident that 97.5% of the mica washers will have thicknesses above .115 inches.

1.5.3 Tolerance limits when the population is not normal.

The methods given in 1.5.1 and 1.5.2 are based on the assumption that the observations come from a normal population. If the population is not in fact normal, then the effect will be that the true proportion P of the population between the tolerance limits will vary from the intended P by an amount depending on the amount of departure from normality. If the departure from normality is believed to be great, then we may wish to obtain tolerance limits which do not require any assumption of normality (we assume only that the distribution has no discontinuities). The tolerance limits so obtained will be substantially longer than those assuming normality.

<u>Two-sided Tolerance Limits (no normality assumption)</u>

Table XV gives values (r,s) such that we may assert with confidence at least $\gamma$ that 100P% of a population lies between the $r^{th}$ smallest and the $s^{th}$ largest of a random sample of n from that population. For example, if we have a sample of n=60, then we can say with a confidence of at least .95 that 90% of the population will lie between the fifth largest and the fifth smallest of the sample values. That is, if we were to take many random samples of 60, and take the fifth largest and

fifth smallest of each, we should expect to find that at least 95% of the resulting intervals would contain 90% of the population.

Table XVII may be useful for sample sizes of $n \leq 100$. They give the confidence $\gamma$ with which we may assert that 100P% of the population lies between the largest and smallest values of the sample.

<u>One-sided Tolerance Limits (no normality assumption)</u>

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

## 2. Statistical Tests Concerning Averages and Dispersions.

2.1 General. - One of the most frequent uses for statistics is in testing for differences. If we wish to know whether a treatment applied to a standard round affects its muzzle velocity we conduct an experiment and make a statistical test on the results to see whether there is a difference between treated and untreated rounds. In another case, we may have two processes for manufacturing a given component: Process I is cheaper and we wish to use it unless Process II is demonstrated to be superior. We make a statistical test of experimental results to see if Process II is superior.

In a large number of cases we would be quite happy if we could, on the basis of a sample, decide between a pair of alternatives. In many cases, we should like to be able to make one of the following decisions:

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

ii) We could find no difference.

In other cases we would like to make one of these decisions:

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

ii) We do not have reason to believe the (population) average of product A is greater than that of product B.

In this section, we shall consider a number of statistical tests of differences. The result of each will limit us to making one of two decisions (as above). <u>In each case the alternative decisions are chosen before the data are observed - this is important!</u> Since we ordinarily get our information on one or both of the products by means of a sample, we may sometimes make the wrong decision. Of course, increasing the number of observations will reduce our chance of making the wrong decision.

There are two ways we can make a wrong decision. When we conclude that there is a difference, and in fact there is none, we say we make an <u>Error of the First Kind</u>. When we fail to find a difference that really exists, then we say we make an <u>Error of the Second Kind</u>.

In any particular case, we can never be absolutely sure that we have made the correct decision, but we can know the probability with which we will make either type of error, when we use a given procedure.

We usually let $\alpha$ be the probability that we will make an Error of the First Kind, and $\beta$ be the probability that we will make an Error of the Second Kind. Since the ability to detect a difference between averages will in general depend on the size of the difference ($\delta$) there will be a value of $\beta$, say $\beta(\delta)$ for each possible difference $\delta$. $\beta(\delta)$ will decrease as $\delta$ increases. $\beta$ has no meaning by itself, but is always associated with a particular difference $\delta$.

Given a particular statistical test, and any two of the three quantities $n, \alpha, \beta(\delta)$, where $n$ is the sample size (the number of observations) then the third is automatically determined.

Our procedure will be a very logical one. Suppose we wish to test whether two types of vaccum tubes have the same resistance in ohms on the average. We take samples of each type, and measure their resistances. If the sample mean of one type of tube differs sufficiently from the other sample mean, we shall say that the two kinds of tubes differ in their average resistance. Otherwise, we shall say we failed to find a difference. How large must the difference be in order that we may conclude that the two types differ or that the observed difference is "significant?"*

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

This will depend on several factors -- the amount of variability in the tubes of each type, the number of tubes of each type tested, and the risk we are willing to take of stating a difference exists when there really is none, i.e., the risk of making an error of the first kind. We might decide as follows: we would be willing to state that the true averages differ, if a difference larger than the observed difference could arise by chance less than five times in a hundred when the true averages are in fact equal. The probability of a type one error is then $\alpha = .05$, or, as we commonly say, we have a .05 "significance level." The use of a "significance level" of .05 or .01 is common, and these levels are tabulated extensively for many tests. There is nothing unique about these levels, however, and a test user may choose any value for $\alpha$ that he feels is appropriate.

<u>Operating Characteristic of a Statistical Test.</u> As we have mentioned, the ability to detect a difference will in general depend on the size of the difference ($\delta$). Let us denote by $\beta(\delta)$ the probability of failing to detect a specified difference $\delta$. If we plot $\beta(\delta)$ vs. the difference $\delta$, we have what we call an Operating Characteristic (OC) curve. (What we usually plot is not $\beta(\delta)$ vs. $\delta$, but rather $\beta(\delta)$ vs. some convenient function of $\delta$.) Figures 1.6.1 to 1.6.8 give OC curves for a

number of situations for $\alpha = .05$ and $\alpha = .01$.

An OC curve depicts the discriminatory power of a particular statistical test. For specified values of n and $\alpha$, there is a unique OC curve. The curve is useful in two ways. If we have decided upon a value for n and chosen $\alpha = .01$ or $.05$, we can use the OC curve to read $\beta(\delta)$ for various values of $\delta$. If we are at liberty to choose the sample size for our experiment and have a particular value of $\delta$ in mind, for $\alpha = .01$ or $.05$, we can choose n by looking at the OC curves.

The use of the OC curves for certain situations is described in the following sections.

It is evident that for any $\beta(\delta)$, n will increase as $\delta$ decreases. It requires larger samples to recognize smaller differences. In some cases the experiment as originally thought of will be seen to require prohibitively large sample sizes, and we must compromise between the sharp discriminatory power we think we need, and the cost of the necessary amount of testing required to achieve it.

When the experiment has already been run, and we had no choice of n, we can look at the OC curve to see just what chance we would have had of detecting a particular difference $\delta$.

To use the OC curve for either purpose, one must know the standard deviation $\sigma$, or be willing to state some range of $\sigma$. (It is generally possible at least to assign some upper bound to the variability, even without past data (See section 1.2.3).

After the experiment is run a possibly better estimate of $\sigma$ will be available and a hindsight look at the OC curve using this value will help to evaluate the experiment.

We shall outline a number of different tests in the following sections. For each test, we shall first outline the procedure to be followed for a given significance level $\alpha$ and sample size n. For most of the tests, we shall also give the OC curve which will enable us to obtain the (approximate) value of $\beta$ for any given difference. Finally, we shall give tables for determining n, the sample size when $\alpha, \delta$, and $\beta(\delta)$ have been specified.

The tests given are exact when (a) the observations for each item are taken randomly from a single population of possible observations and (b) within the population, the quality characteristic measured is normally distributed. The assumption of normality is not ordinarily crucial, particularly if the sample size is not too small.

Alternate procedures for most of the tests given in sections 2.2 to 2.6 are given in sections 2.7 and 2.8. Section 2.7 gives tests which require neither normality assumptions nor knowledge of the variability of the populations. Section 2.8 gives short-cut tests which involve less computation.

2.2 Comparison of the average of a new product with that of a standard.

GIVEN: The average performance of a standard product is known to be $m_0$. We will consider 3 different problems:

PROBLEMS:

2.2.1 To determine whether average of the new product <u>differs</u> from the standard.

2.2.2 To determine whether average of the new product <u>exceeds</u> the standard.

2.2.3 To determine whether average of the new product is <u>less</u> than the standard.

(For summary of the procedures appropriate for each of these problems see Table 2.2).

It is necessary to decide which of the three problems is appropriate before taking the observations. If this is not done and the choice of the problem is influenced by the observations, (for example 2.2.1 vs. 2.2.2), the significance level of the test, i.e., the probability of an error of the first kind, and the operating characteristics of the test may differ considerably from their normal values.

Ordinarily we will not know the amount of variation in the new product. At other times we may have previous experience which enables us to state a value of $\sigma$. We

shall outline the solutions for each of the three problems (2.2.1, 2.2.2, and 2.2.3) for both cases, i.e., where the variability is estimated from the sample, and where $\sigma$ is known from previous experience. We shall also give alternative procedures which require neither assumptions of normality nor knowledge of $\sigma$.

Symbols to be used

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

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

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

$s$ - standard deviation of n measurements on new product (used where $\sigma$ is unknown).

$\sigma$ - the known standard deviation of the new product.

Example 2.2

For a certain type of shell, specifications state that the amount of powder should average 0.735 pounds. In order to determine whether the average for present stock meets the specification, 20 shells are taken at random, and the amount of powder they contain is measured.

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

The sample standard deviation is $s = .0504$ pounds. (In illustrating the known $\sigma$ case, we assume $\sigma$ known to be 0.06 pounds).

## TABLE 2.2

**Summary Table for Problems of 2.2 - Comparison of Average of a New Product with that of a Standard**

**(For Details and Worked Example see 2.2.1, 2.2.2, 2.2.3)**

| We wish to test whether | Section Reference | Knowledge of Variation of new item | Test to be made | Operating Characteristics of the test (for $\alpha = .05$ and $\alpha = .01$) | Sample Size Required (n) | Notes |
|---|---|---|---|---|---|---|
| m differs from $m_0$ | 2.2.1.1 | $\sigma$ unknown; s = estimate of $\sigma$ from sample | $\lvert\bar{X}-m_0\rvert>u$ | See Figs. 2.2a and 2.2b* | Use Table XVIIIa. For $\alpha$=.05, add 2 to tabular value. For $\alpha$=.01 add 4 to tabular value. | $u=t_{1-\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ (t for n-1 degrees of freedom) |
| | 2.2.1.2 | known $\sigma$ | $\lvert\bar{X}-m_0\rvert>u$ | See Figs. 2.2c and 2.2d. | Use Table XVIIIa. | $u=z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ |
| m is larger than $m_0$ | 2.2.2.1 | $\sigma$ unknown; s = estimate of $\sigma$ from sample | $(\bar{X}-m_0)>u$ | See Figs. 2.2e and 2.2f* | Use Table XVIIIb. For $\alpha$=.05, add 2 to tabular value. For $\alpha$=.01 add 3 to tabular value. | $u=t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$ (t for n-1 degrees of freedom) |
| | 2.2.2.2 | $\sigma$ known | $(\bar{X}-m_0)>u$ | See Figs. 2.2g and 2.2h | Use Table XVIIIb. | $u=z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$ |
| m is smaller than $m_0$ | 2.2.3.1 | $\sigma$ unknown; s = estimate of $\sigma$ from sample | $(m_0-\bar{X})>u$ | See Figs. 2.2e and 2.2f* | Use Table XVIIIb. For $\alpha$=.05, add 2 to the tabular values. For $\alpha$=.01, add 3 to the tabular values | $u=t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$ (t for n-1 degrees of freedom) |
| | 2.2.3.2 | $\sigma$ known | $(m_0-\bar{X})>u$ | See Figs. 2.2g and 2.2h | Use Table XVIIIb. | $u=z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$ |

*) It is necessary to have some value for $\sigma$ (or two bounding values) in order to use the Operating Characteristic curve. Although $\sigma$ is unknown, in many situations it is possible to have some notion, however loose, about the magnitude of $\sigma$ and thereby to get helpful information from the OC curve. Section 1.2.3 gives assistance in "estimating" $\sigma$ from general knowledge of the process.

Problem 2.2.1.1 - Does the average of the new product differ from the standard ($\sigma$ unknown)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$, (Data from example 2.2) |
| ii) Look up $t_{1-\alpha/2}$ for n-1 degrees of freedom in Table II. | ii) $t_{.975}$ for 19 degrees of freedom = 2.09 |
| iii) Compute: $\bar{X}$, the mean; s, the standard deviation of the n measurements. | iii) $\bar{X} = .710$ pounds<br>$s = .0504$ pounds |
| iv) Compute: $u = t_{1-\alpha/2} \frac{s}{\sqrt{n}}$ | iv) $u = \frac{2.09 \times .0504}{\sqrt{20}}$<br>$u = .0236$ |
| v) If $\lvert\bar{X}-m_0\rvert > u$, decide that the average of the new type differs from that of the standard. (Otherwise, there is no reason to believe that they differ). | v) $\lvert\bar{X}-m_0\rvert = \lvert .710-.735\rvert = .025$ We conclude the average amount of powder in present stocks differs from 0.735 (the specified amount). |

| Procedure | Example |
|---|---|
| vi) Note: The interval $\overline{X} \pm u$ is a $(1-\alpha)$ confidence interval estimate of the true average of the new type. | vi) Note that $(.710 \pm .0236)$ is a 95% confidence interval estimate of true average of new product. |

<u>Operating Characteristics of the Test</u> - Figures 2.2a and 2.2b give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively, for various values of n. Although we have assumed that we do not know $\sigma$, in order to use the OC curve, we must have some value for $\sigma$. (See section 1.2.3). If we use too large a value for $\sigma$, the effect is to lower our estimate of $\beta(\delta)$.

If $\delta = |m-m_0|$ is the true absolute difference (unknown of course) between the two averages, then putting $d = \frac{|m-m_0|}{\sigma} = \frac{\delta}{\sigma}$ we can read $\beta(\delta)$, the probability of failing to detect a difference $\delta = |m-m_0|$, when we use the given test.

<u>Selection of Sample Size n</u> - If we state

$\alpha$, the significance level of the test;

$\beta$, the probability of failing to detect a difference $|m-m_0|$

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

then we may use Table XVIIIa to obtain a good approximation to the required sample size. If we use $\alpha = .01$, add 4 to the value obtained in the above table. If $\alpha = .05$, we add 2 to the tabled value. (In order to use the table, we must have a value for $\sigma$. See section 1.2.3).

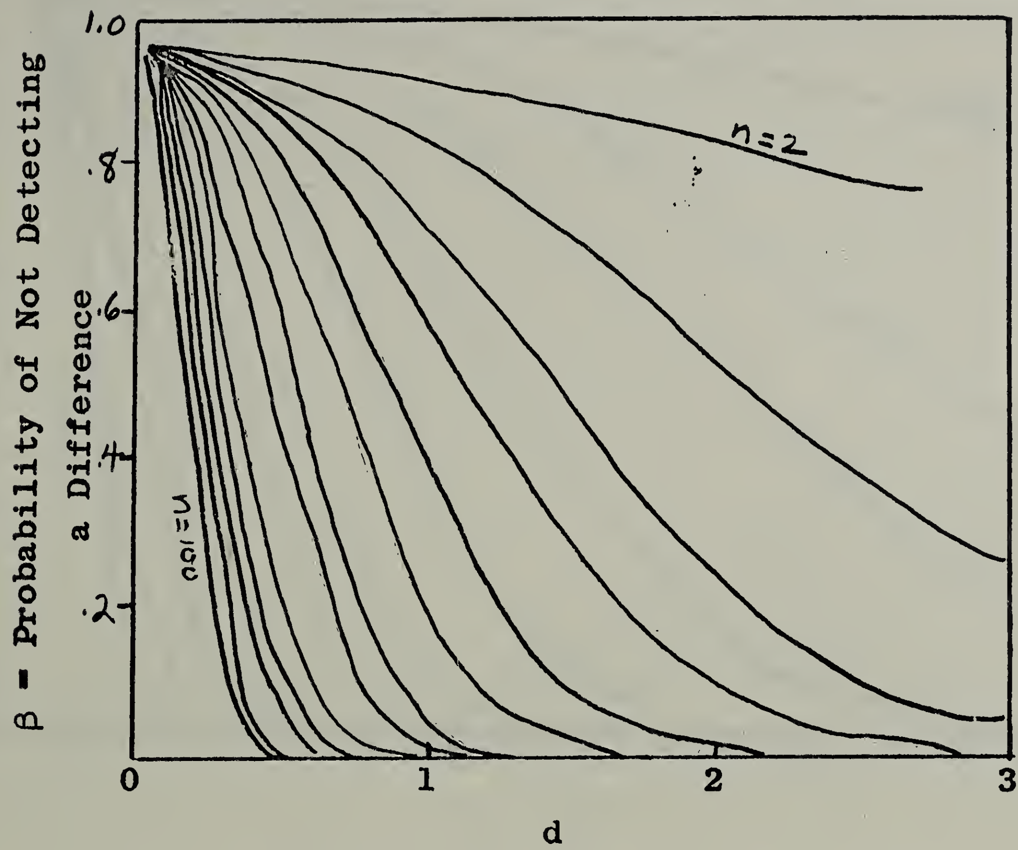



Figure 2.2a. Operating Characteristics for the test of whether:

i) The average of a new product differs from a standard ($d = \frac{|m-m_0|}{\sigma}$. See section 2.2)

ii) The averages of two products differ ($d = \frac{|m_A-m_B|}{\sqrt{\sigma_A^2+\sigma_B^2}}$ See section 2.3).

The standard deviations are assumed unknown ($\sigma_A^2=\sigma_B^2=\sigma^2$).

$\alpha = .05$

CAUTION: This is a rough tracing of curves to be reprinted from Figure 13.29 of Bowker and Lieberman, "Handbook of Industrial Statistics".

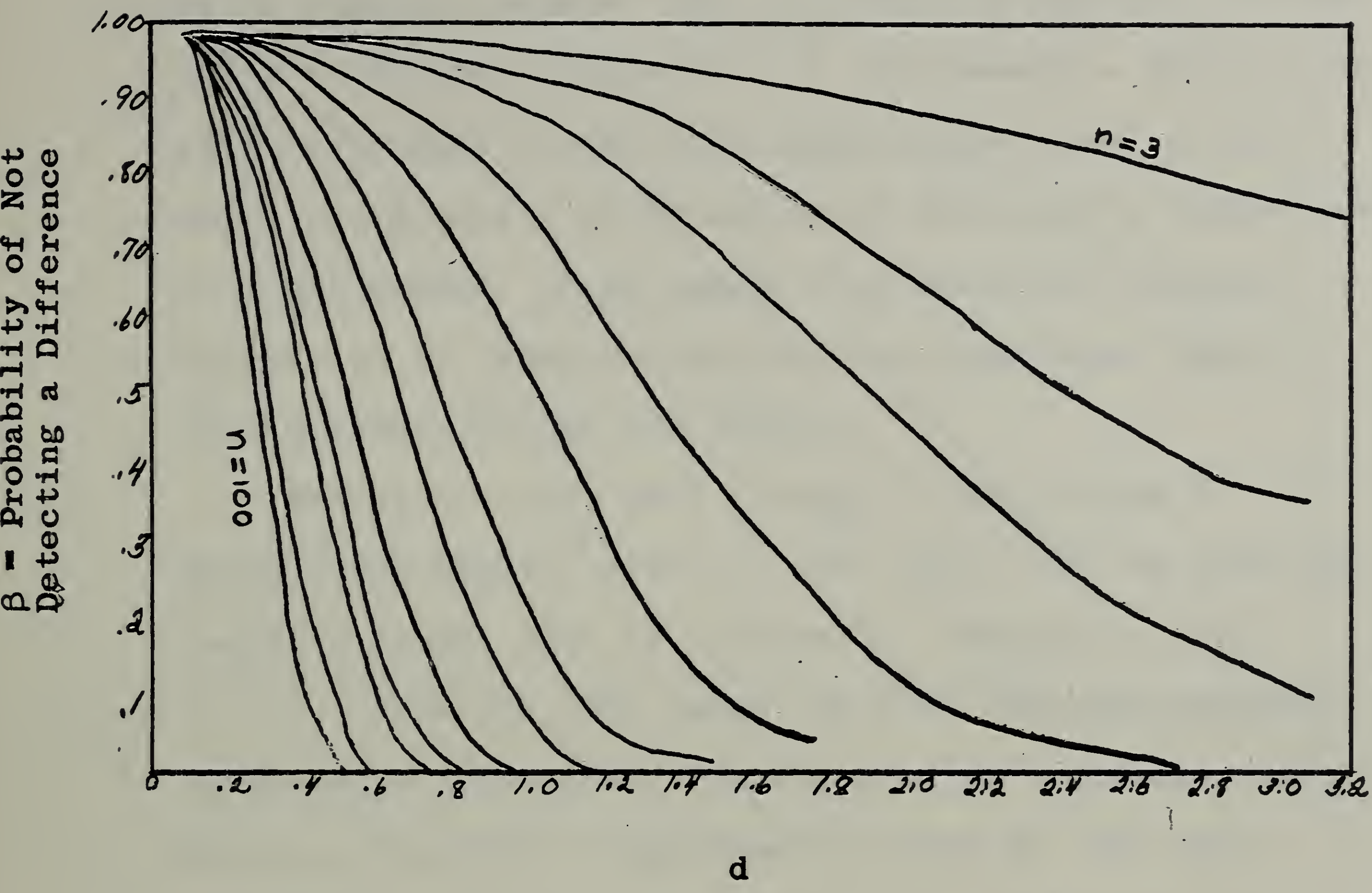



Figure 2.2b. Operating Characteristics for the Test of whether:

i) The average of a new product differs from a standard $(d = \frac{|m-m_0|}{\sigma}$. See section 2.2)

ii) The averages of two products differ $(d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2+\sigma_B^2}}$ See section 2.3)

The standard deviations are assumed unknown $(\sigma_A^2=\sigma_B^2=\sigma^2)$.

$\alpha = .01$

<u>CAUTION:</u> This is a rough tracing of curves to be reprinted from Figure 13.30 of Bowker and Lieberman, "Handbook of Industrial Statistics".

As an example, suppose that we wished to specify $\alpha = .05$, and $\beta = .50$ for a difference of .024 pounds - that is, we wish to conduct a test at a significance level of .05 which would have a 50-50 chance of detecting a difference of 0.024 pounds. What sample size should we require? Suppose it is thought from previous experience that $\sigma$ lies between .04 and .06 pounds.

Taking $\sigma = .04$, with $|m-m_0| = .024$, gives $d = .6$. Using Table XVIIIa, with $\alpha = .05$, $1-\beta = .50$, we find the required sample size as $n=11+2=13$. Taking $\sigma = .06$, $d = .4$. From the same table, we find that the required sample size is $25+2=27$. To be safe, we would use $n=27$. For $\sigma \leq .06$, with a significance level of .05, this would give a 50% chance of detecting a difference of 0.024 pounds.

Problem 2.2.1.2 - Does the average of the new product differ from the standard ($\sigma$ known)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$. (Data from example 2.2) |
| ii) Look up $z_{1-\alpha/2}$ in Table I. | ii) $z_{.975} = 1.96$ |
| iii) Compute: $\bar{X}$, the mean of the $n$ measurements. | iii) $\bar{X} = .710$ pounds<br>$\sigma$ is known to be equal to .06 pounds. |
| iv) Compute $u = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ | iv) $u = \frac{1.96(.06)}{\sqrt{20}} = .0263$ |
| v) If $\lvert\bar{X} - m_0\rvert > u$ decide that the average of the new type differs from that of the standard. (Otherwise there is no reason to believe that they differ). | v) $\lvert\bar{X} - m\rvert = \lvert .710-.735\rvert = .025$<br>We conclude that there is no reason to believe that the average amount of powder in present stocks differs from 0.735 (the specified amount). |

| Procedure | Example |
|---|---|
| vi) Note that the interval $\overline{X} \pm u$ is a $(1-\alpha)$ confidence interval estimate of the true average of the new type. | vi) Note that $(.710 \pm .0263)$ is a 95% confidence interval for the true average of the new type. |

Operating Characteristics of the Test - Figures 2.2c and 2.2d give the operating characteristics of the above test for $\alpha = .05$ and $\alpha = .01$ respectively. For any given n and $d = \frac{|m-m_0|}{\sigma}$, the value of $\beta$, the probability of failing to detect a difference of size $|m-m_0|$, can be read off directly.

Selection of Sample Size n - If we specify $\alpha$, our significance level, and $\beta$, the probability or risk we are willing to take of not detecting a difference of size $|m-m_0|$, then we can use Table XVIIIa to obtain n, the required sample size. As an example, if $\sigma$ is known to be 0.06 pounds, and we wish to have a 50-50 chance of detecting a difference of 0.036 pounds, then $d = 0.6$. From Table XVIIIa, we find that the required sample size is 11.

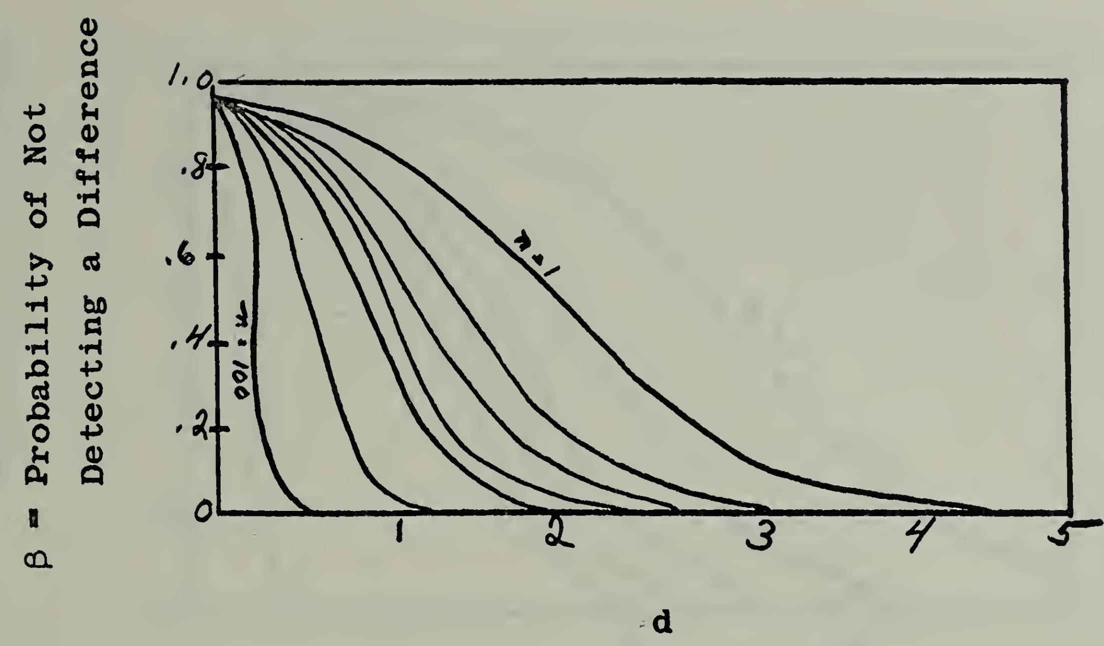



Fig. 2.2c Operating Characteristics for the test of whether

i) The average of a new product differs from a standard

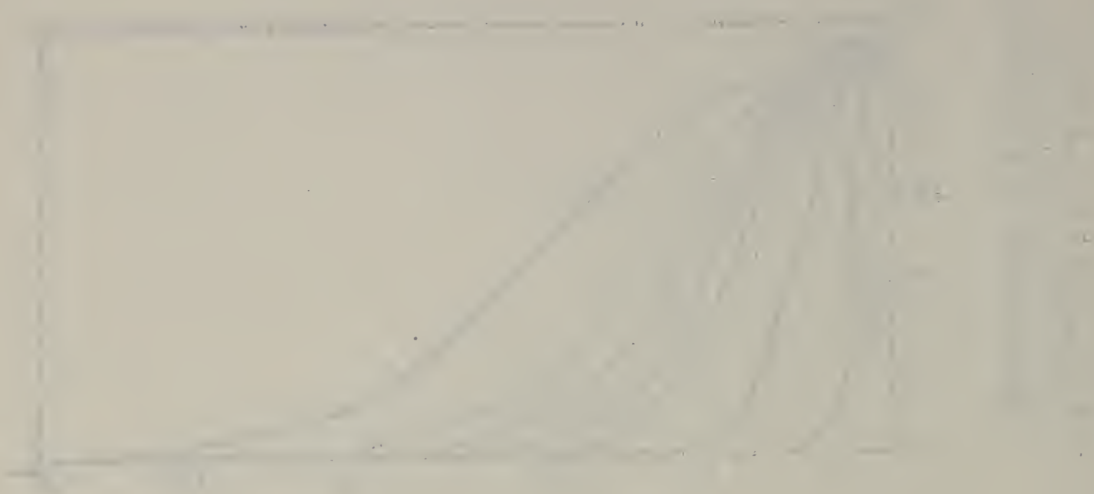
$\left(d = \frac{|m-m_0|}{\sigma}\right.$ . See section 2.2)

ii) The averages of 2 products differ $\left(d = \frac{|m_A - m_B|}{\sqrt{\sigma_A^2+\sigma_B^2}}\right.$

See Section 2.3)

The standard deviations are assumed to be known. $\alpha = .05$.

<u>CAUTION</u>: This is a rough tracing of curves to be reprinted from Figure 13.25 of Bowker and Lieberman "Handbook of Industrial Statistics."



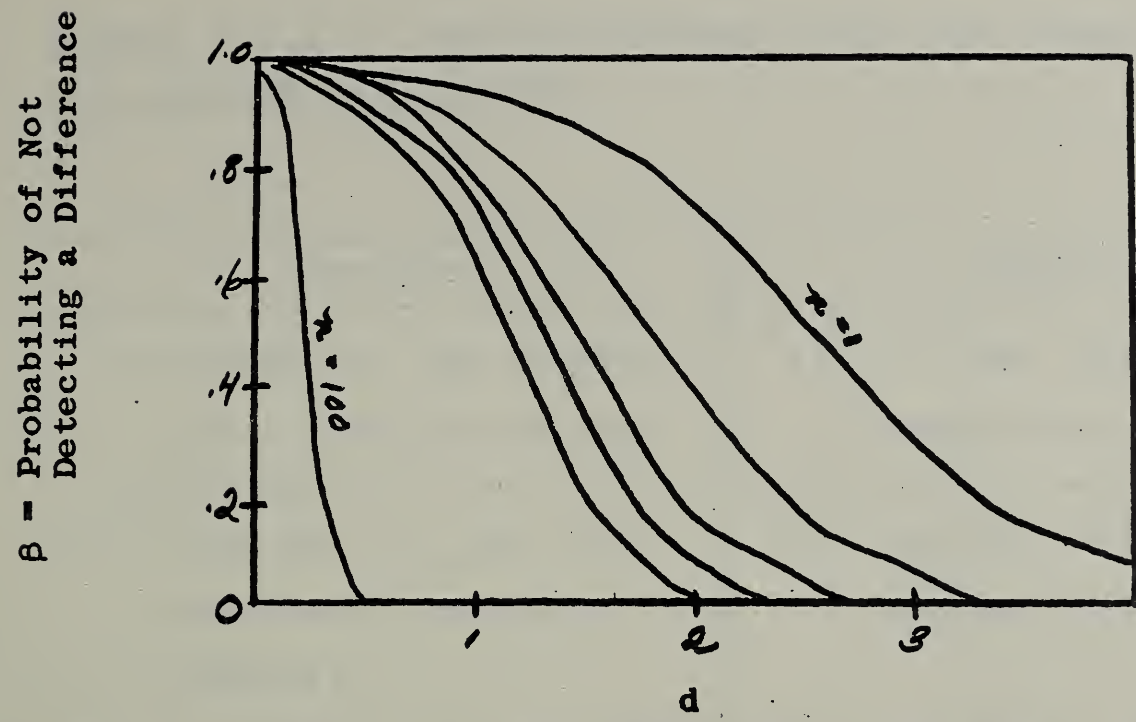



Fig. 2.2d Operating Characteristics for the test of whether:

i) The average of a new product differs from a standard $(d = \frac{|m-m_0|}{\sigma}$ . See section 2.2).

ii) The averages of two products differ $(d = \frac{|m_A-m_B|}{\sqrt{\sigma_A^2+\sigma_B^2}}$ see section 2.3).

The standard deviations are assumed to be known. $\alpha = .01$.

<u>CAUTION</u>: This is a rough tracing of curves to be reprinted from Figure 13.26 of Bowker and Lieberman "Handbook of Industrial Statistics".

Problem 2.2.2.1 - Does the average of the new product exceed the standard ($\sigma$ unknown)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$. (Data from example 2.2) |
| ii) Look up $t_{1-\alpha}$ for n-1 degrees of freedom in Table II. | ii) $t_{.95}$ for 19 degrees of freedom = 1.73. |
| iii) Compute: $\overline{X}$, the sample mean; s, the sample standard deviation | iii) $\overline{X} = .710$ pounds<br>$s = .0504$ pounds |
| iv) Compute: $u = t_{1-\alpha} \frac{s}{\sqrt{n}}$ | iv) $u = \frac{1.73(.0504)}{\sqrt{20}}$<br>$= 0.019$ |
| v) If $(\overline{X} - m_0) > + u$, decide that the average of the new type exceeds that of the standard. (Otherwise there is no reason to believe that the average of the new type exceeds that of the standard. | v) $(\overline{X} - m_0) = (.710-.735)$ $= -.025$. We conclude there is no reason to believe the average of the new product exceeds that of the standard. |

| Procedure | Example |
|---|---|
| vi) Note that the open interval from $(\bar{X} - u)$ to infinity is a one-sided confidence interval for the true mean. | vi) Note that the open interval from .691 to $\infty$ is a 95% one-sided confidence interval for true average of new product. |

Operating Characteristics of the test. See Figures 2.2e and 2.2f (Problem 2.2.2.1). Figures 2.2e and 2.2f give the operating characteristic (OC) curves of the above test for $\alpha = .05$, and $\alpha = .01$ respectively, for various values of n. Although we have assumed that we do not know $\sigma$, in order to use the OC curve, we must have some value for $\sigma$. (See section 1.2.3). If we use too large a value for $\sigma$, the effect is to lower our estimate of $\beta(\delta)$.

If $(m-m_0)$ is the true difference (unknown of course) between the two averages, then putting $d = \frac{m-m_0}{\sigma}$, we can read $\beta$, the probability of failing to detect such a difference.

Selection of Sample Size n - If we state

$\alpha$, the significance level of the test

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

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

then we may use table XVIIIb to obtain a good approximation to the required sample size. Using $\alpha = .01$, add 3 to the tabular value. Using $\alpha = .05$, add 2 to the tabular value. (In order to use the table, we must have a value for $\sigma$ - see section 1.2.3).

**CAUTION:** This is a rough tracing of curves to be reprinted from Fig. 13.31 of Bowker and Lieberman "Handbook of Industrial Statistics".

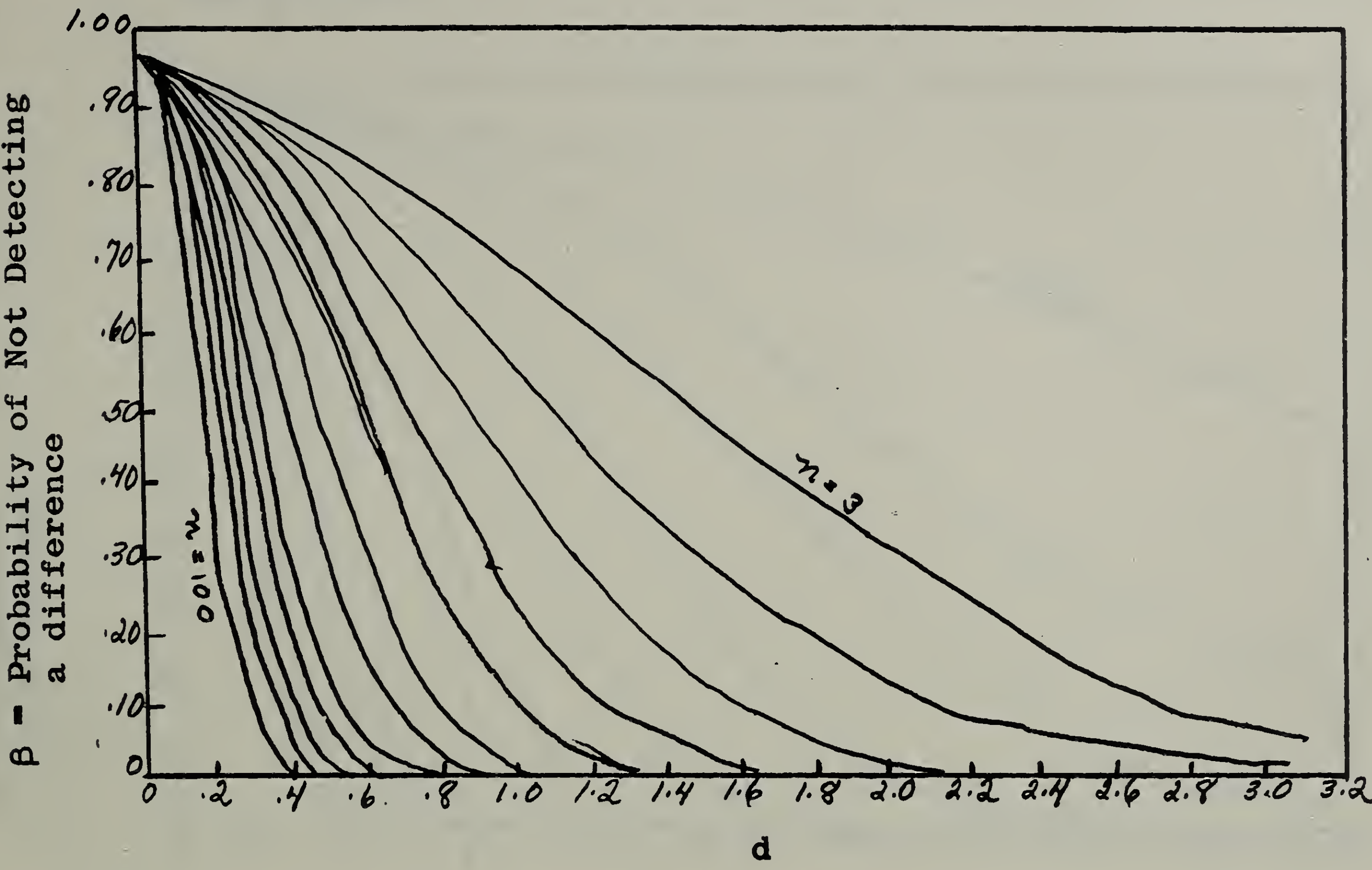



Figure 2.2e

Operating Characteristics for the Test of Whether

i) The average of a new product exceeds a standard

$$\left( d = \frac{m - m_0}{\sigma} \right)$$

ii) The average of a new product is less than a standard

$$\left( d = \frac{m_0 - m}{\sigma} \right)$$

iii) The average of product A exceeds that of product B

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

The standard deviations are assumed unknown $(\sigma_A^2 = \sigma_B^2 = \sigma^2)$ $\alpha = .05$.

**CAUTION:** This is a rough tracing of curves to be reprinted from Fig. 13.32 of Bowker and Lieberman "Handbook of Industrial Statistics."

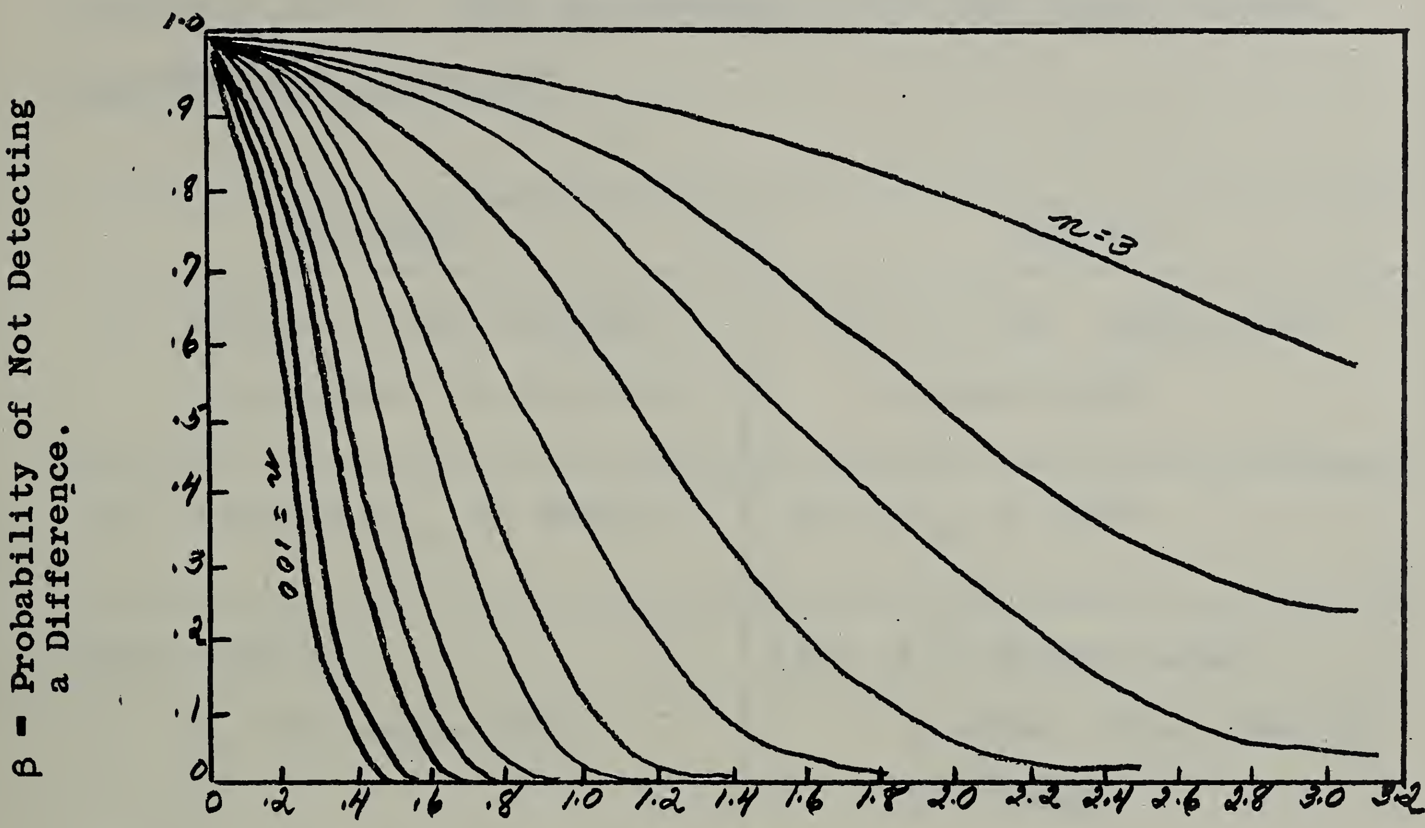



Figure 2.2f

Operating Characteristics for the Test of Whether

i) The average of a new product exceeds a standard

$$\left( d = \frac{m - m_0}{\sigma} \right)$$

ii) The average of a new product is less than a standard

$$\left( d = \frac{m_0 - m}{\sigma} \right)$$

iii) The average of product A exceeds that of product B

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

The standard deviations are assumed unknown $(\sigma_A^2 = \sigma_B^2 = \sigma^2)$ $\alpha = .01$.

Problem 2.2.2.2 - Does the average of the new product exceed the standard ($\sigma$ known)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$. (Data from example 2.2) |
| ii) Look up $z_{1-\alpha}$ in Table I. | ii) $z_{.95} = 1.64$ |
| iii) Compute: $\bar{X}$, the sample mean | iii) $\bar{X} = 0.710$ pounds ($\sigma$ known to be equal to .06 pounds) |
| iv) Compute: $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ | iv) $u = \frac{1.64(.06)}{\sqrt{20}}$ $= .022$ |
| v) If $(\bar{X} - m_0) > u$, decide that the average performance of the new type exceeds that of the standard. Otherwise there is no reason to believe that the average of the new type exceeds that of the standard. | v) $(\bar{X} - m_0) = .710-.735)$ $= -.025$, which is not larger than u. We conclude that there is no reason to believe that the average of the new product exceeds that of the standard. |

| Procedure | Example |
|---|---|
| vi) Note that the open interval from $(\bar{X} - u)$ to infinity is a one-sided confidence interval for the true mean of the new product. | vi) Note that the open interval from .688 to $\infty$ is a 95% one-sided confidence interval for the true mean of the new product. |

Operating Characteristics of the Test - Figures 2.2g and 2.2h give the operating characteristics of the above test for $\alpha = .05$ and $\alpha = .01$ respectively. For any given n and $d = \frac{m-m_0}{\sigma}$, the value of $\beta$, the probability of failing to detect a difference of $(m-m_0)$ can be read off directly.

Selection of sample size n

If we specify

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference $m-m_0$

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

then we may use Table XVIIIb to obtain the required sample size.

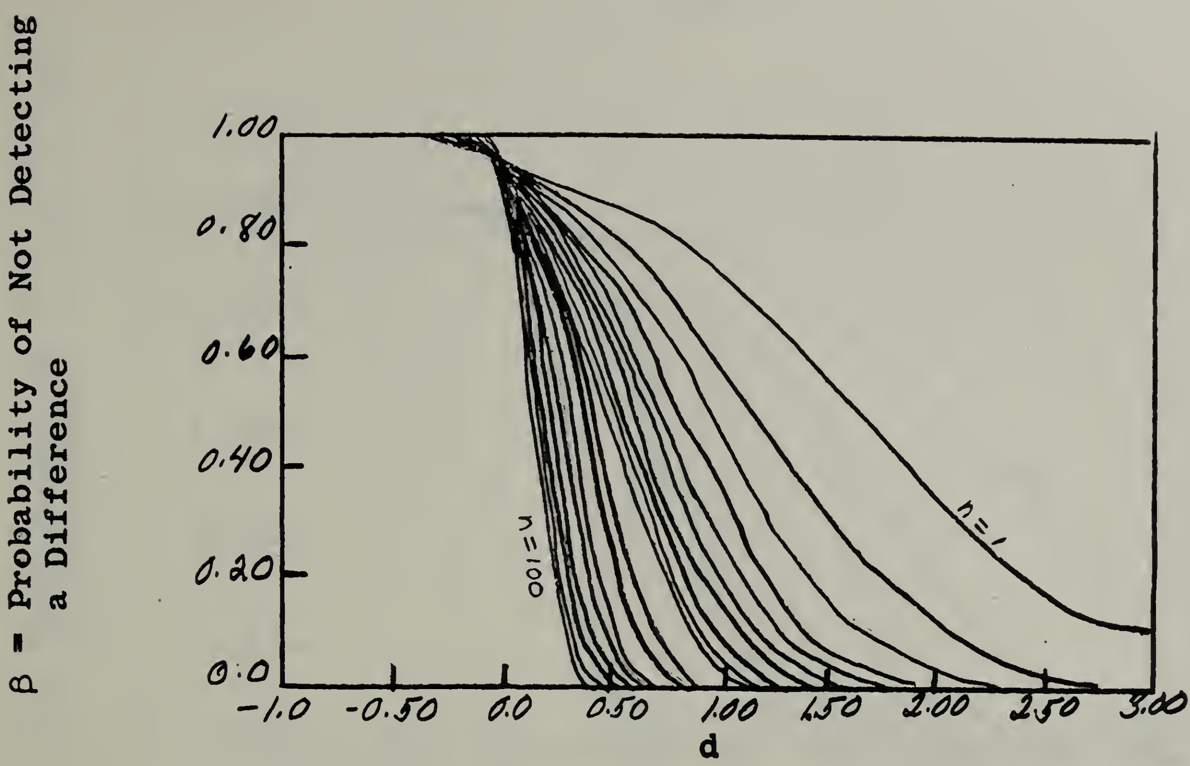



Figure 2.2g

Operating Characteristics for the Test of Whether

i) The average of a new product exceeds a standard

$$\left( d = \frac{m - m_0}{\sigma} \right)$$

ii) The average of a new product is less than a standard

$$\left( d = \frac{m_0 - m}{\sigma} \right).$$

iii) The average of a product A exceeds that of product B

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

The standard deviations are assumed to be known.

$\alpha = .05$

CAUTION: This is a rough tracing of curves to be reprinted from Fig. 13.27 of Bowker and Lieberman, Handbook of Industrial Statistics.

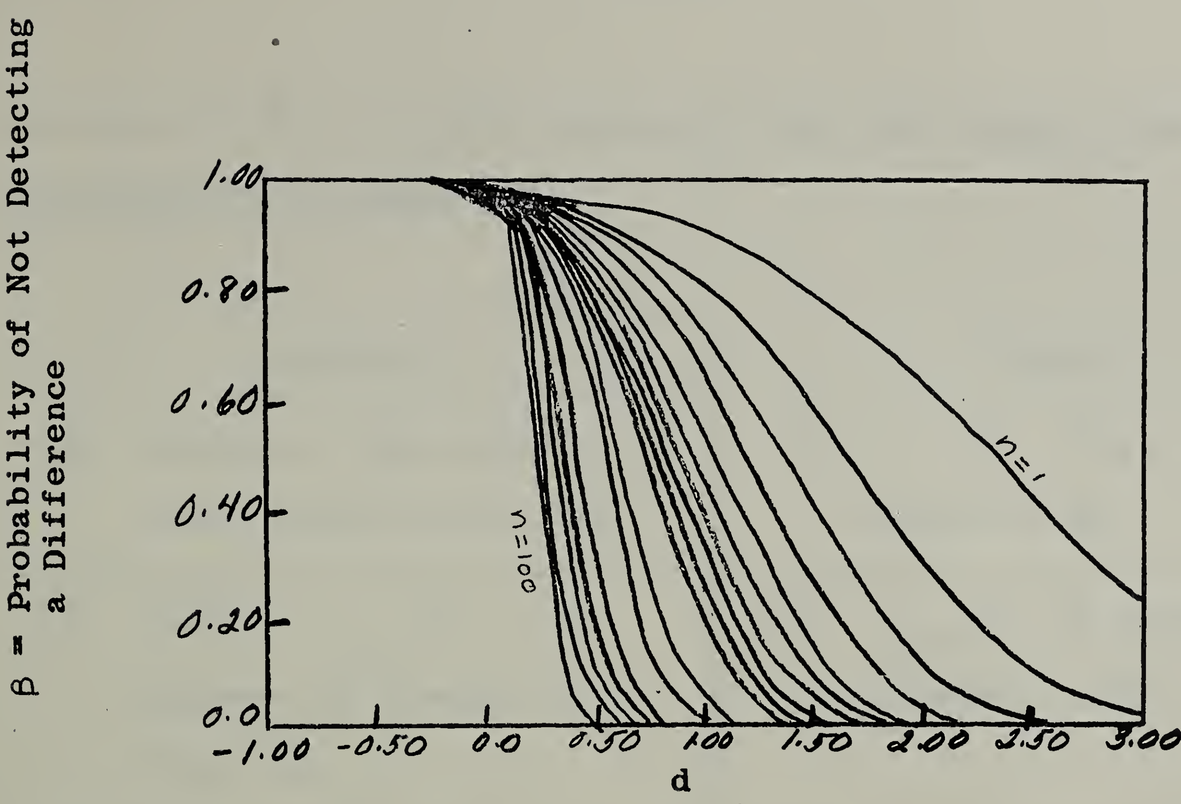



Figure 2.2h

Operating Characteristics for the Test of Whether

i) The average of a new product exceeds a standard

$$\left( d = \frac{m - m_0}{\sigma} \right)$$

ii) The average of a new product is less than a standard

$$\left( d = \frac{m_0 - m}{\sigma} \right)$$

iii) The average of product A exceeds that of product B

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

The standard deviations are assumed to be known.

$\alpha = .01$.

<u>CAUTION</u>: This is a rough tracing of curves to be reprinted from Fig. 13.28 of Bowker and Lieberman, "Handbook of Industrial Statistics.

Problem 2.2.3.1 - Is the average of the new product less than the standard ($\sigma$ unknown)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$. (Data from example 2.2) |
| ii) Look up $t_{1-\alpha}$ for n-1 degrees of freedom in Table II. | ii) $t_{.95}$ for 19 degrees of freedom = 1.73. |
| iii) Compute: $\bar{X}$, the sample mean; s, the sample standard deviation. | iii) $\bar{X} = .710$ pounds<br>$s = .0504$ pounds. |
| iv) Compute: $u = t_{1-\alpha} \frac{s}{\sqrt{n}}$ | iv) $u = \frac{1.73(.0504)}{\sqrt{20}}$<br>$= 0.019$ |
| v) If $(m_0 - \bar{X}) > u$, decide that the average of the new type is less than that of the standard. (Otherwise there is no reason to believe that the average of the new type is less than the standard). | v) $.735 - .710 = .025$ We conclude that the average of the new type is less than that of the standard. |

| Procedure | Example |
|---|---|
| vi) Note that the open interval from $-\infty$ to $(\overline{X} + u)$ is a $(1-\alpha)$ one-sided confidence interval for the true mean of the new type. | vi) Note that the open interval from $-\infty$ to .731 is a 95% one-sided oonfidence interval for the true mean of the new type. |

Operating Characteristics of the Test - Figures 2.2e and 2.2f give the operating characteristic (OC) curves of the above test for $\alpha = .05$, and $\alpha = .01$ respectively, for various values of n. Although we have assumed that we do not know $\sigma$, in order to use the OC curve, we must have some value for $\sigma$. (See Section 1.2.3). If we use too large a value for $\sigma$, the effect is to lower our estimate of $\beta(\delta)$.

If $(m_0-m)$ is the true difference (unknown of course) between the two averages, then putting $d = \frac{m_0-m}{\sigma}$ , we can read $\beta$, the probability of failing to detect such a difference.

Selection of Sample Size n - If we state

$\alpha$, the significance level of the test

$\beta$, the probability of failing to find a difference of size $(m_0 - m)$

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

then we may use Table XVIIIb to obtain a good approximation to the required sample size. Using $\alpha = .01$, add 3 to the tabular value. Using $\alpha = .05$, add 2 to the tabular value. (In order to use the table, we must have a value for $\sigma$ - see section 1.2.3).

Problem 2.2.3.2 - Is the average of the new product less than that of the standard ($\sigma$ known)?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | i) $\alpha = .05$. (Data from example 2.2) |
| ii) Look up $z_{1-\alpha}$ in Table I. | ii) $z_{.95} = 1.64$ |
| iii) Compute: $\bar{X}$, the sample mean | iii) $\bar{X} = 0.710$ pounds. ($\sigma$ known to be equal to .06 pounds) |
| iv) Compute: $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ | iv) $u = \frac{1.64(.06)}{\sqrt{20}} = 0.022$ |
| v) If $(m_0 - \bar{X}) > u$, decide that the average of the new type is less than that of the standard. Otherwise, there is no reason to believe that the average of the new type is less than that of the standard. | v) $(m_0 - \bar{X}) = (.735 - .710) = .025$, which is larger than u. We conclude that the average of the new type is less than the standard. |

| Procedure | Example |
| --- | --- |
| vi) Note that the open interval from $-\infty$ to $(\bar{X} + u)$ is a $(1-\alpha)$ one-sided confidence interval for the true mean of the new type. | vi) Note that the open interval from $-\infty$ to .732 is a 95% one-sided confidence interval for the true mean of the new type. |

<u>Operating Characteristics of the Test</u> - Figures 2.2g and 2.2h give the operating characteristics of the test for $\alpha = .05$ and $\alpha = .01$ respectively. For any given $n$ and $d = \frac{m_0 - m}{\sigma}$ the value of $\beta$, the probability of failing to detect a difference of size $(m_0 - m)$ can be read off directly.

<u>Selection of Sample Size n</u> - If we specify

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of $m_0 - m$

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

then we may use Table XVIIIb to obtain a good approximation to the required sample size. If we use $\alpha = .01$, add 4 to the tabular value. Using $\alpha = .05$, add 2 to the tabular value. (In order to use the table, we must have a value for $\sigma$ - see section 1.2.3).

## 2.3 - Comparison of the Averages of two Given Materials, Products or Processes.

We shall consider two problems.

Problem 2.3.1 - We wish to test whether the averages of two materials, products or processes differ, and we are not particularly concerned which is larger.

Problem 2.3.2 - We wish to test whether the average of material, product or process A exceeds that of material, product or process B.

It is again important to decide which problem is appropriate before making the observations. If this is not done and the choice of the problem is influenced by the observations, the significance level of the test, i.e., the probability of an error of the first kind, and the operating characteristics of the test may differ considerably from their nominal value. In the following, it is assumed that the appropriate problem has been selected and that $n_A$ and $n_B$ observations are taken from types A and B respectively.

Ordinarily one will not know $\sigma_A$, or $\sigma_B$, In some cases, it will probably be safe to assume that the variation in the performance will be approximately the same.* We shall however give the solutions for the 2 problems (2.3.1 and 2.3.2) for the following situations:

---

*) For a test to see whether $\sigma_A$ and $\sigma_B$ differ see section 2.6.

Case I - The variation in the performances of each of A and B is unknown but can be assumed to be about the same.

Case II - The variation in the performances of each of A and B is unknown, and it is not reasonable to assume the amounts of variation are the same.

Case III - The variation in the performance of each of A and B is known from previous experience and the standard deviations are $\sigma_A$ and $\sigma_B$ respectively.

Problem to be Illustrated

(Illustrative problem to be added later)

Table 2.3 - Summary Table for Problems of 2.3 - Comparison of Average Performance of Two Products
(For details and worked examples, see 2.3.1 or 2.3.2)

| We wish to test whether | Section Reference | Knowledge of Variation | Test to be made | Operating Characteristics of Test | Determination of Sample Size (n) | Notes |
|---|---|---|---|---|---|---|
| $m_A$ differs from $m_B$ | 2.3.1.1 | $\sigma_A \simeq \sigma_B$ ; both unknown | $\lvert\bar{X}_A-\bar{X}_B\rvert > u$, where $u=t_{1-\alpha/2}\, s_p \sqrt{\frac{n_A+n_B}{n_A n_B}}$ | For $\alpha$=.05 and $\alpha$=.01 see Figs. 2.2a and 2.2b* and section 2.3.1.1. | Use Table XVIIIa. For $\alpha$=.05, add 1 to the tabular value. For $\alpha$=.01, add 2 to the tabular value. | $s_p = \sqrt{\frac{(n_A-1)s_A^2+(n_B-1)s_B^2}{n_A+n_B-2}}$ |
| | 2.3.1.2 | $\sigma_A \neq \sigma_B$ ; both unknown | $\lvert\bar{X}_A-\bar{X}_B\rvert > u$, where $u = t' \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$ See notes. | | | t' is a weighted average of $t_{1-\alpha/2}$ for $n_A-1$ and $n_B-1$ degrees of freedom. |
| | 2.3.1.3 | $\sigma_A$ , $\sigma_B$ ; both known | $\lvert\bar{X}_A-\bar{X}_B\rvert > u$, where $u=z_{1-\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$ | For $\alpha$=.05 and $\alpha$=.01, see Figs. 2.2c and 2.2d. | Use Table XVIIIa. | |
| $m_A$ is greater than $m_B$ | 2.3.2.1 | $\sigma_A \simeq \sigma_B$ ; both unknown | $(\bar{X}_A-\bar{X}_B) > u$, where $u=t_{1-\alpha}\, s_p \sqrt{\frac{n_A+n_B}{n_A n_B}}$ | For $\alpha$=.05 and $\alpha$=.01 see Figs. 2.2e and 2.2f* and section 2.3.2.1. | Use Table XVIIIb. For $\alpha$=.05, add 1 to the tabular value. For $\alpha$=.01, add 2 to the tabular value. | $s_p = \sqrt{\frac{(n_A-1)s_A^2+(n_B-1)s_B^2}{n_A+n_B-2}}$ |
| | 2.3.2.2 | $\sigma_A \neq \sigma_B$ ; both unknown | $(\bar{X}_A-\bar{X}_B) > u$, where $u = t' \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$ | | | t' is a weighted average of $t_{1-\alpha}$ for $n_A-1$ and $n_B-1$ degrees of freedom. |
| | 2.3.2.3 | $\sigma_A$ , $\sigma_B$ ; both known | $(\bar{X}_A-\bar{X}_B) > u$, where $u=z_{1-\alpha}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$ | For $\alpha$=.05 and $\alpha$=.01 see Figs. 2.2g and 2.2h and section 2.3.2.3. | Use Table XVIIIb. | |

* Although common $\sigma$ is unknown, useful information may be obtained from the O.C. curve if a value (or 2 bounding values) of $\sigma$ can be assumed.

Problem 2.3.1 - Do the products A and B differ in average performance? (No particular concern over which is larger).

2.3.1.1 (Case I) - Variability of A and B unknown, but can be assumed to be about the same.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $t_{1-\alpha/2}$ for $\nu = (n_A + n_B - 2)$ degrees of freedom in Table II. | |
| iii) Compute $\bar{X}_A$, $\bar{X}_B$ and $s_A$, $s_B$ the means and standard deviations of the $n_A$ and $n_B$ measurements from A and B. | |
| iv) Compute: $s_p = \sqrt{\dfrac{(n_A-1)s_A^2+(n_B-1)s_B^2}{n_A + n_B-2}}$ | |
| v) Compute: $u = t_{1-\alpha/2}\, s_p \sqrt{\dfrac{n_A+n_B}{n_A\, n_B}}$ | |

| Procedure | Example |
| --- | --- |
| vi) If $\lvert\overline{X}_A-\overline{X}_B\rvert > u$, decide that A and B differ with regard to their average performance. Otherwise, decide that there is no reason to believe A and B differ with regard to their average performance. | |
| vii) Let $m_A$, $m_B$ be the true average performances of A and B (unknown of course). Then, it is worth noting that the interval $(\overline{X}_A-\overline{X}_B) \pm u$ is a $1-\alpha$ confidence interval estimate of $(m_A - m_B)$. | |

Operating Characteristic Curves and Determination of Sample Size.

Operating Characteristics of the Test - Figures 2.2a and 2.2b give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively, for various values of $n = n_A + n_B - 1$. Although we have assumed we do not know the standard deviation of the performances of A and B, we need to have a value for $\sigma_A = \sigma_B = \sigma$, the common standard deviation of the performance of A and B, in order to use the OC curve. This may be possible since we often know the range in which $\sigma$ lies. (See Section 1.2.3). If we use too large a value for $\sigma$, the effect is to make our estimates more conservative.

If $(m_A - m_B)$ is the true difference between the two averages, then putting

$$d = \frac{|m_A - m_B|}{\sigma} \quad \frac{1}{\sqrt{n_A + n_B - 1}} \quad \sqrt{\frac{n_A \; n_B}{n_A + n_B}}$$

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

Selection of Sample Size n

If we state

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of size $|m_A - m_B|$

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

then we may use Table XVIIIa to obtain a good approximation to the required sample size. If we use $\alpha = .01$, add 2 to the value obtained in the above table. If $\alpha = .05$, add 1 to the tabled value. (In order to use the table, we must have values for $\sigma_A$, $\sigma_B$. See Section 1.2.3).

2.3.1.2 (Case II) - Variability of A and B unknown, cannot be assumed equal.

| Procedure*# | Example |
| --- | --- |
| i) Choose $\alpha$, the significance level of the test. (Actually the procedure outlined will give a significance level of only approximately $\alpha$). | |
| ii) Compute $\overline{X}_A$, $\overline{X}_B$ and $s_A$, $s_B$ the means and standard deviations of the $n_A$ and $n_B$ measurements from A and B. | |

*) The test procedure given here is an approximation, i.e., the stated significance level is only approximately achieved. The approximation is good provided $n_A$ and $n_B$ are not too small. An exact procedure is given in "Biometrika Tables for Statisticians", Volume I, Cambridge University Press (1954), page 27 and Table 11. These tables have been extended by Tricket, Welch and James in Biometrika, Volume 43, (1956), page 203.

#) When $\sigma_A^2$ and $\sigma_B^2$ differ considerably, an alternate method of analysis involves the random pairing of observations, each pair containing a different observation from A and B. (We assume $n_A = n_B$, and thus we have $n_A = n_B$ pairs). The analysis of paired observations is given in the last part of this section and in section 2.3.2.2.

| Procedure | Example |
|---|---|
| iii) Look up $t_{1-\alpha/2}$ for $n_A-1$ degrees of freedom and for $n_B-1$ degrees of freedom. | |
| iv) Compute: $t' = \frac{w_A t_A + w_B t_B}{w_A + w_B}$ where $t_A$ is $t_{1-\alpha/2}$ for $n_A-1$ degrees of freedom, $t_B$ is $t_{1-\alpha/2}$ for $n_B-1$ degrees of freedom. $w_A = \frac{s_A^2}{n_A}$ , $w_B = \frac{s_B^2}{n_B}$ . (Note that t' is always between $t_A$ and $t_B$). | |
| v) Compute: $u = t' \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$ | |
| vi) IF $\lvert \bar{X}_A - \bar{X}_B \rvert > u$, decide that A and B differ with regard to their average performance. Otherwise, decide that there is no reason to believe A and B differ in average. | |

| Procedure | Example |
|---|---|
| vii) If $m_A$, $m_B$ are the true average performances of A and B (unknown of course), then it is worth noting that the interval $(\bar{X}_A - \bar{X}_B) \pm u$ is approximately a $1-\alpha$ confidence interval estimate of $m_A - m_B$. | |

Method of paired observations - The method of paired observations may be used in two situations:

1) We have a series of measurements on each of two products A and B, and it is believed that $\sigma_A^2$ and $\sigma_B^2$ differ considerably. If $n_A = n_B$ equals the number of observations on each, then we may pair the observations randomly, thus obtaining $n_A = n_B$ pairs, each pair containing a different observation from both A and B. The analysis shown below is valid for this situation.

2) The experiment may be planned or designed so that the observations will be taken in pairs. The usual reason for this pairing is to reduce unnecessary variation as much as possible. The pairs are usually chosen so that they are alike in as many respects as possible except in the characteristic being measured. Differences which exist are then not nearly as likely to be obscured by extraneous variation.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Compute the mean and standard deviation ($\overline{X}_d$ and $s_d$) of the $n$ differences, $X_d$. (Each $X_d$ represents an observation on A minus the paired observation on B). | |

| Procedure | Example |
|---|---|
| iii) Look up $t_{1-\alpha/2}$ for n-1 degrees of freedom in Table II. | |
| iv) Compute: $u = t_{1-\alpha/2} \frac{s_d}{\sqrt{n}}$ | |
| v) If $\|\overline{X}_d\| > u$, decide that the averages differ. (Otherwise, there is no reason to believe they differ). | |
| vi) Note: The interval $\overline{X}_d \pm u$ is a $(1-\alpha)$ confidence interval estimate of the average of A minus the average of B. | |

Operating Characteristics of the Test - Figures 2.2a and 2.2b give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively, for various values of n. Although we have assumed that we do not know $\sigma$, in order to use the OC curve, we must have some value for $\sigma$. (See section 1.2.3). If we use too large a value for $\sigma$, the effect is to lower our estimates of $\beta(\delta)$.

If $\delta = |m_A - m_B|$ is the true absolute difference (unknown of course) between the two averages, then putting $d = \frac{|m_A - m_B|}{\sigma} = \frac{\delta}{\sigma}$ we can read $\beta(\delta)$, the probability of failing to detect a difference $\delta = |m_A - m_B|$, when we use the given test.

<u>Selection of Sample Size $n = n_A = n_B$</u> - If we state

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of size $|m_A - m_B|$

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

where $\sigma'$ is the variance of the population of signed differences for the pairs, then we may use Table XVIIIa to obtain a good approximation to the required sample size. If we use $\alpha = .01$,

add 3 to the tabular value. If we use $\alpha = .05$, add 2 to the tabular value. (In order to use the table, we must have a value for $\sigma'$. See section 1.2.3).

2.3.1.3 (Case III) - The variability in performance of ea- of A and B is known from previous experience, and the standa- deviations are $\sigma_A$ and $\sigma_B$ respectively.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $z_{1-\alpha/2}$ in Table I. | |
| iii) Compute $\overline{X}_A$ and $\overline{X}_B$, the sample means of the $n_A$ and $n_B$ measurements from A and B. | |
| iv) Compute: $u = z_{1-\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$ | |
| v) If $\lvert\overline{X}_A - \overline{X}_B\rvert > u$, decide that A and B differ with regard to their average performance. Otherwise, decide that there is no reason to believe that A and B differ in average performance. | |

| Procedure | Example |
| --- | --- |
| vi) Let $m_A$, $m_B$ be the true average performances of A and B (unknown of course). Then it is worth noting that the interval $(\overline{X}_A - \overline{X}_B) \pm u$ is a $1-\alpha$ confidence interval estimate of $(m_A - m_B)$. | |

Operating Characteristics of the Test - Figures 2.2c and 2.2d give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively, for various values of $n_A$.

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

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

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

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

Selection of sample size - If we specify

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of size $m_A - m_B$

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

and wish $n_A = n_B = n$, then we may use Table XVIIIa to obtain the required sample size n.

If we wish to choose $n_A$, $n_B$ such that $n_A = cn_B$, we may compute $d = \dfrac{m_A - m_B}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$ and use the table to obtain $n_A$.

<u>2.3.2.1 (Case I)</u> - Variability of A and B unknown, but can be assumed to be about the same.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $t_{1-\alpha}$ for $\nu = n_A + n_B - 2$ degrees of freedom in Table II. | |
| iii) Compute $\bar{X}_A$, $\bar{X}_B$ and $s_A$, $s_B$ the means and standard deviations of the $n_A$ and $n_B$ measurements from products A and B respectively. | |
| iv) Compute: $s_p = \sqrt{\dfrac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2}}$ | |
| v) Compute: $u = t_{1-\alpha}\, s_p \sqrt{\dfrac{n_A + n_B}{n_A n_B}}$ | |

| Procedure | | Example |
|---|---|---|
| vi) | If $(\overline{X}_A - \overline{X}_B) > u$, decide that the average of A exceeds the average of B. Otherwise, decide there is no reason to believe that the average A exceeds average B. | |
| vii) | Let $m_A$ and $m_B$ be the true averages of A and B. Note that the interval from $\left\{(X_A - X_B) - u\right\}$ to $\infty$ is a $1-\alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$. | |

Operating Characteristics of the Test - Figures 2.2e and 2.2f give the operating characteristic (OC) curves of the test for $\alpha = .05$ and $\alpha = .01$ respectively for various values of $n = n_A + n_B - 1$. Although we have assumed that we do not know the standard deviations of A and B, in order to use the OC curve we would have to have a value of $\sigma_A = \sigma_B = \sigma$, their common standard deviation. This may be possible since we often know the range in which $\sigma$ lies (See section 1.2.3). If we use too large a value for $\sigma$, the effect is to make our estimates more conservative.

Determination of the Sample Size - If we specify

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of size $(m_A - m_B)$

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

and wish to obtain $n = n_A = n_B$, then we may use Table XVIIIb to obtain a good approximation to the required sample size n. Using $\alpha = .01$, add 2 to the tabular values. Using $\alpha = .05$, add 1 to the values given in the table. (In order to use the tables, we must have values for $\sigma_A$, $\sigma_B$. See section 1.2.3).

2.3.2.2 (Case II) - Variability of A and B unknown, cannot be assumed equal.

| Procedure* | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Compute $\overline{X}_A$, $\overline{X}_B$ and $s_A$, $s_B$, the means and standard deviations of the $n_A$ and $n_B$ measurements from A and B. | |
| iii) Look up $t_{1-\alpha/2}$ for $n_A-1$ degrees of freedom and for $n_B-1$ degrees of freedom. | |
| iv) Compute: $t' = \dfrac{w_A t_A + w_B t_B}{w_A + w_B}$ where $t_A$ is $t_{1-\alpha/2}$ for $n_A-1$ degrees of freedom $t_B$ is $t_{1-\alpha/2}$ for $n_B-1$ degrees of freedom $w_A = \dfrac{s_A^2}{n_A}$ , $w_B = \dfrac{s_B^2}{n_B}$ . (Note that t' is always between $t_A$ and $t_B$) | |

*) See footnotes of section 2.3.1.2 (Case II).

| Procedure | Example |
|---|---|
| v) Compute: $u = t' \left( \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} \right)$ | |
| vi) If $(\bar{X}_A - \bar{X}_B) > u$, decide that the average of A exceeds the average of B. Otherwise, decide that there is no reason to believe that the average of A exceeds the average of B. | |
| vii) Let $m_A$ and $m_B$ be the true averages of A and B. Note that the interval from $\left\{ (\bar{X}_A - \bar{X}_B) - u \right\}$ to $\infty$ is approximately a $1-\alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$. | |

<u>Method of Paired Observations</u> - (See Method of Paired Observations, section 2.3.1.2). Let $X_d$ represent the value of an observation on A minus the value of the paired observation on B. Then we may use the following analysis ($n = n_A = n_B$):

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Compute the mean and standard deviation ($\overline{X}_d$ and $s_d$) of the n differences, $X_d$. | |
| iii) Look up $t_{1-\alpha}$ for n-1 degrees of freedom in Table II. | |
| iv) Compute: $u = t_{1-\alpha} \dfrac{s_d}{\sqrt{n}}$ | |
| v) If $\overline{X}_d > u$, decide that the average of A exceeds that of B. (Otherwise, there is no reason to believe the average of A exceeds that of B). | |

| Procedure | Example |
|---|---|
| vi) Note that the open interval from $\overline{X}_d - u$ to $\infty$ is a one-sided confidence interval for the true difference $m_A - m_B$. | |

Operating Characteristics of the Test - Figures 2.2e and 2.2f give the operating characteristic (OC) curves of the above test for $\alpha = .05$, and $\alpha = .01$ respectively, for various values of n. Although we have assumed that we do not know $\sigma$, in order to use the OC curve, we must have some value for $\sigma$. (See Section 1.2.3). If we use too large a value for $\sigma$, the effect is to lower our estimates of $\beta(\delta)$.

If $(m_A - m_B)$ is the true difference (unknown of course) between the two averages, then putting $d = \frac{m_A - m_B}{\sigma}$, we can read $\beta$, the probability of failing to detect such a difference.

Selection of Sample Size $n = n_A = n_B$ - If we state

$\alpha$, the significance level of the test

$\beta$, the probability of failing to find a difference of size $(m_A - m_B)$

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

then we may use Table XVIIIb to obtain a good approximation to the required sample size. Using $\alpha = .01$, add 3 to the tabular value. Using $\alpha = .05$, add 2 to the tabular value. (In order to use the table, we must have a value for $\sigma$ - see section 1.2.3).

2.3.2.3 (Case III) - The variability in performance of each of A and B is known from previous experience and the standard deviations are $\sigma_A$ and $\sigma_B$ respectively.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $z_{1-\alpha}$ in Table I. | |
| iii) Compute $\bar{X}_A$ and $\bar{X}_B$, the means of the $n_A$ and $n_B$ measurements from A and B. | |
| iv) Compute: $u = z_{1-\alpha}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$ | |
| v) If $(\bar{X}_A - \bar{X}_B) > u$, decide that the average of A exceeds the average of B. Otherwise, decide that there is no reason to believe that the average of A exceeds the average of B. | |

| Procedure | Example |
|---|---|
| vii) Let $m_A$ and $m_B$ be the true averages of A and B. Note that the interval from $\left\{(\overline{X}_A - \overline{X}_B) - u\right\}$ to $\infty$ is a $1-\alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$. | |

<u>Operating Characteristics of the Test</u>. - Figures 2.2g and 2.2h give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively for various values of $n_A$.

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

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

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

If $n_A = cn_B$, we can put $\quad d^* = \dfrac{(m_A - m_B)}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$

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

<u>Selection of Sample Size</u> - If we specify

$\alpha$, the significance level of the test

$\beta$, the probability of failing to detect a difference of size $m_A - m_B$

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

and wish $n_A = n_B = n$, then we may use Table XVIIIb to obtain the required sample size n.

If we wish to choose $n_A$, $n_B$ such that $n_A = cn_B$, we may compute $\quad d = \dfrac{m_A - m_B}{\sqrt{\sigma_A^2 + c\sigma_B^2}}\quad$ and use the table to obtain $n_A$.

## 2.4 Comparison of the Averages of Several Products.

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

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

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level (the risk of concluding that the averages differ, when in fact all averages are the same). | |
| ii) Compute $s_1^2, s_2^2, \ldots, s_t^2$, the sample variance for each of the t products. | |
| iii) Compute: $s_e^2 = \frac{1}{t}(s_1^2+s_2^2+\ldots+s_t^2)$ If the $n_i$ are not all equal, the following formula will give a better result: $s_e^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+\ldots+(n_t-1)s_t^2}{(n_1+n_2+\ldots+n_t)-t}$ | |

| Procedure | Example |
| --- | --- |
| iv) Look up $q_{1-\alpha}(t,\nu)$ in Table IV where $\nu = [n_1+n_2+\ldots+n_t] - t$. | |
| v) Compute: $w = q_{1-\alpha}\, s_e/n^{1/2}$ | |
| vi) If the difference between any two sample means exceeds w, decide that the averages differ. Otherwise, decide that there is no reason to believe the averages differ. | |

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

2.5 Comparison of the variability in performance of a given material, product or process with that of a standard.

The variability of a standard material, product or process as measured by its standard deviation is known to be $\sigma$. We shall consider three problems:

2.5.1 - Does the variability of the product differ from that of the standard?

2.5.2 - Does the variability of the product exceed that of the standard?

2.5.3 - Is the variability of the product less than that of the standard?

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

The tests given are exact when (a) the observations for an item, product or process are taken randomly from a single population of possible observations and (b) within the population, the quality characteristic measured is normally distributed.

Problem 2.5.1 - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be $\sigma_0$. We wish to determine whether a given item differs in variability of performance from that of the standard. We wish, from analysis of the data, to make one of the following decisions:

i) The variability in performance of the new product differs from that of the standard.

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

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $A_{\alpha/2}$ and $A_{1-\alpha/2}$ both for n-1 degrees of freedom in Table XXIII. | |
| iii) Compute s, the sample standard deviation of the n observations. | |
| iv) Compute: $s_L = A_{\alpha/2}\, s$ and $s_U = A_{1-\alpha/2}\, s$ | |

| Procedure | Example |
|---|---|
| v) If $\sigma_0$ doesn't lie between $s_L$ and $s_U$, decide that the variability in performance of the new product differs from that of the standard. Otherwise, there is no reason to believe the new product differs with regard to variability from the standard. | |

It is worth noting that the interval from $s_L$ to $s_U$ is a $100(1-\alpha)$ percent confidence interval estimate of $\sigma$, the standard deviation of the new product. (See section 1.2.2).

Problem 2.5.2 - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be $\sigma_0$. We wish to determine whether the variability in performance of a given product exceeds that of the standard. We wish, from analysis of the data to make one of the following decisions:

i) The variability in performance of the new product exceeds that of the standard.

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

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $A_\alpha$ for n-1 degrees of freedom in Table XXIII | |
| iii) Compute s, the sample standard deviation of the n observations. | |
| iv) Compute: $s_L = A_\alpha s$. | |

| Procedure | Example |
|---|---|
| v) If $s_L$ exceeds $\sigma_0$ decide that the variability of the new product exceeds that of the standard. Otherwise, there is no reason to believe that the new product exceeds the standard with regard to variability. | |

It is worth noting that the interval above $s_L$ is a $100(1-\alpha)$ percent confidence interval estimate of $\sigma$, the standard deviation of the new product. (See section 1.2.2).

<u>Problem 2.5.3</u> - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be $\sigma_0$. We wish to determine whether the variability in performance of a given product is less than that of the standard. We wish, from analysis of the data to make one of the following decisions:

i) The variability in performance of the new product is <u>less than</u> that of the standard.

ii) There is no reason to believe the variability in performance of the new product is less than that of the standard.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $A_{1-\alpha}$ for n-1 degrees of freedom in Table XXIII. | |
| iii) Compute s, the sample standard deviation of the n observations. | |
| iv) Compute $s_U = A_{1-\alpha}s$ | |

| Procedure | Example |
|---|---|
| v) If $s_U$ is less than $\sigma_0$ decide that the variability in performance of the new product is less than that of the standard. Otherwise, there is no reason to believe the new product is less variable than the standard. | |

It is worth noting that the interval below $s_U$ is a $1-\alpha$ confidence interval estimate of $\sigma$, the standard deviation of the new product. (See section 1.2.2).

2.6 Comparison of the variability of two materials, products or processes.

We shall consider 2 problems:

2.6.1 - Does the variability of product A differ from that of product B? (We are not concerned which is larger).

2.6.2 - Does the variability of product A exceed that of product B?

It is important to decide which of the two problems is appropriate before taking the observations. If this is not done, and the choice of problem is influenced by the observations, the significance level of the test, i.e., the probability of an error of the first kind, and the operating characteristics of the test may differ considerably from their nominal values. The tests given are exact when (a) the observations for an item, product or process are taken randomly from a single population of possible observations and (b) within the population, the quality characteristic measured is normally distributed.

In the following it is assumed the appropriate problem is selected and then $n_A$, $n_B$ observations are taken from items, processes or products A, B, respectively.

Problem 2.6.1 - We wish to test whether the variability of performance of two materials, products or processes differ, and we are not particularly concerned which is larger. We wish, from analysis of the data to make one of the following decisions.

i) The two products differ with regard to their variability.

ii) There is no reason to believe the two products differ with regard to their variability.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $F_{1-\alpha/2}$ for $(n_A-1, n_B-1)$ degrees of freedom and $F_{1-\alpha/2}$ for $(n_B-1, n_A-1)$ degrees of freedom in Table III. | |
| iii) Compute $s_A^2$, $s_B^2$, the sample variances of the observations from A and B respectively. | |
| iv) Compute $F = s_A^2/s_B^2$. | |

| Procedure | Example |
|---|---|
| v) If $F > F_{1-\alpha/2}(n_A-1, n_B-1)$ or $F < \dfrac{1}{F_{1-\alpha/2}(n_B-1, n_A-1)}$ decide that the two products differ with regard to their variability. Otherwise, there is no reason to believe that they differ. | |

It is worth noting that the interval between

$$\frac{1}{F_{1-\alpha/2}(n_A-1, n_B-1)} \cdot \frac{s_A^2}{s_B^2}$$

and $F_{1-\alpha/2}(n_B-1, n_A-1) \cdot \dfrac{s_A^2}{s_B^2}$ is a 100(1-$\alpha$) percent confidence interval estimate of the ratio $\sigma_A^2/\sigma_B^2$ (See section 1.2.2).

Problem 2.6.2 - We wish to test whether the variability in performance of product A exceeds that of product B. We wish as a result of analysis of the data to make one of the following decisions:

i) The variability of product A exceeds that of product B.

ii) There is no reason to believe that the variability of A exceeds the variability of B.

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $F_{1-\alpha}$ for $n_A-1$, $n_B-1$ degrees of freedom in Table III. | |
| iii) Compute $s_A^2$, $s_B^2$, the sample variances of the observations from A and B respectively. | |
| iv) Compute $F = s_A^2/s_B^2$ . | |

| Procedure | Example |
| --- | --- |
| v) If $F > F_{1-\alpha}$, decide that the variability of product A exceeds that of B. Otherwise, there is no reason to believe that the variability of A is greater than that of B. | |

Note that the interval above $\frac{1}{F_{1-\alpha/2}(n_A-1, n_B-1)} \frac{s_A^2}{s_B^2}$ is a $1-\alpha$ confidence interval estimate of $\sigma_A^2/\sigma_B^2$ .

## 2.7 Some Tests Which are Independent of the Form of the Distribution.

This section will outline a number of test procedures in which nothing, or very little is assumed about the nature of the population distributions. In particular, the population distributions are not assumed to be "normal". These tests are often called "non-parametric" tests. If the underlying populations are indeed normal, then these tests are poorer than the ones given in 2.2 to 2.6. That is, $\beta$, the second kind of error, will always be greater (assuming $\alpha$, first kind of error and n, the sample size have the same values for both tests). For other underlying distributions however, the non-parametric tests may actually have a smaller error of the second kind. The increase in the second kind of error, in the case of non-parametric tests is surprisingly small, and is an indication that for many purposes these tests should receive more use.

Descriptions of the operating characteristic curves, and methods of obtaining sample sizes are not given for the tests in this section. However, one can obtain a rough estimate of the sample size n required to obtain a given $\alpha$, and $\beta$, by using the corresponding sections of 2.2 to 2.6, and multiplying the sample size given there by a factor of 1.1.

2.7.1 - Does the average of a new product differ from a standard $m_0$?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. Tables are given for $\alpha = .05, .02, .01$. Discard any observations which happen to be equal to $m_0$, and let n be the number of observations actually used. | |
| ii) Look up $T_{\alpha/2}(n)$ in Table XII. | |
| iii) For each observation $X_i$, compute $X_i' = X_i - m_0$. | |
| iv) Disregarding signs, rank the $X_i'$ according to their numerical value, i.e., assign the rank of 1 to the $X_i'$ which is numerically smallest, the rank of 2 to the $X_i'$ which is next smallest, etc. (In case of ties assign the average of the ranks which would have been assigned had the $X_i'$'s differed only slightly). | |
| v) To the above ranks 1,2,3, etc., prefix a + or a - according to whether the corresponding $X_i'$ was positive or negative. | |

| Procedure | Example |
|---|---|
| vi) Sum the ranks prefixed by a + sign, and the ranks prefixed by a - sign. Let T be the smaller (disregarding sign) of the two sums. | |
| vii) If $T \leq T_{\alpha/2}(n)$, decide that the average performance of the new type differs from that of the standard. (Otherwise, there is no reason to believe they differ). | |

2.7.2 Does the average of a new product exceed that of a standard $m_0$?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. Tables are given for $\alpha = .025, .01, .005$. Discard any observations which happen to be equal to $m_0$, and let n be the number of observations actually used. | |
| ii) Look up $T_\alpha(n)$ in Table XII. | |
| iii) For each observation $X_i$, compute $X_i' = X_i - m_0$. | |
| iv) Disregarding signs, rank the $X_i'$ according to their numerical value, i.e., assign the rank of 1 to the $X_i'$ which is numerically smallest, the rank of 2 to the $X_i'$ which is next smallest, etc. (In case of ties, assign the average of the ranks which would have been assigned had the $X_i'$'s differed only slightly). | |
| v) To the above ranks 1, 2, 3, etc., prefix a + or a - according to whether the $X_i'$ was positive or negative. | |

| Procedure | Example |
|---|---|
| vi) Let T be the absolute value of the sum of the ranks preceded by a negative sign. | |
| vii) If $T \leq T_{\alpha}(n)$, decide that the average performance of the new product exceeds that of the standard. (Otherwise, there is no reason to believe the average performance of the new product exceeds that of the standard). | |

2.7.3 Is the average of a new product less than that of a standard?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. Tables are given for $\alpha$ = .025, .01, and .005. Discard any observations which happen to be equal to $m_0$, and let n be the number of observations actually used. | |
| ii) Look up $T_\alpha(n)$ in Table XII. | |
| iii) For each observation $X_i$, compute $X_i' = X_i - m_0$. | |
| iv) Disregarding signs, rank the $X_i'$ according to their numerical value, i.e., assign the rank of 1 to the $X_i'$ which is numerically smallest, the rank of 2 to the $X_i'$ which is next smallest, etc. (In case of ties, assign the average of the ranks which would have been assigned had the $X_i'$'s differed only slightly). | |
| v) To the above ranks 1, 2, 3, etc., prefix a + or a - according to whether the corresponding $X_i'$ was positive or negative. | |

| Procedure | | Example |
|---|---|---|
| vi) | Let T be the sum of the ranks preceded by a + sign. | |
| vii) | If $T \leq T_{\alpha}(n)$, decide that the average performance of the new product is less than that of the standard. (Otherwise, there is no reason to believe the average performance of the new product is less than that of the standard). | |

2.7.4 Do the products A and B differ in average performance? (No particular concern over which is larger).

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. Values tabled for $n > 8$ are $\alpha = .002$, .02, .05, .10 (n is the larger of $n_A, n_B$). | |
| ii) Combine the observations from the two populations, and rank them in order of increasing size from smallest to largest. Assign the rank of 1 to the lowest score, a rank of 2 to the next lowest score, etc. (Use algebraic size, i.e., the lowest rank is assigned to the largest negative number, if there are negative numbers). In case of ties, assign to each the average of the ranks which would have been assigned had the $X_i'$s differed only slightly. | |
| iii) Compute $R_A$ = the sum of the ranks assigned to the observations from population A. | |

| Procedure | Example |
|---|---|
| iv) Compute $U_A = n_A n_B + n_A(n_A + 1)/2 - R_A$. Compute $U_B = n_A n_B - U_A$. Let U be the smaller of $U_A$, $U_B$. The remainder of the procedure depends on whether (a) $n \leq 8$, (b) $n > 8$. | |
| a) $n \leq 8$. v) Look up in Table XIII the probability corresponding to $U(n_A, n_B)$. If the probability is less than or equal to the value for α selected in (i) decide that the averages of the two materials, products or processes differ. Otherwise, decide that there is no reason to believe the averages of the two products differ. | |
| b) $n > 8$. v) Look up in Table XIV, the critical value $U_{\alpha/2}(n_1 n_2)$ corresponding to the α selected in (i). $n_2$ and $n_1$ are the larger and smaller respectively of $n_A, n_B$. | |
| vi) If $U \leq U_{\alpha/2}(n_1, n_2)$ decide that the averages of the two products or processes differ. Otherwise, decide that there is no reason to believe the averages of the two products differ. | |

2.7.5 Does the average of product A exceed that of product B?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. Values tabulated for $n > 8$ are $\alpha = .001, .01, .025, .05$. | |
| ii) Combine the observations from the two populations, and rank them in order of increasing size from smallest to largest. Assign the rank of 1 to the lowest score, a rank of 2 to the next lowest score, etc. (Use algebraic size, i.e., the lowest rank is assigned to the largest negative number if there are negative numbers). In case of ties, assign to each the average of the ranks which would have been assigned had the $X_i'$'s differed only slightly. | |
| iii) Compute $R_A$ the sum of the ranks assigned to the observations from population A. | |
| iv) Compute $U_A = n_A n_B + n_A(n_A + 1)/2 - R_A$. Compute $U_B = n_A n_B - U_A$. The remainder of the procedure depends on whether (a) $n \leq 8$, or (b) $n > 8$. | |

| Procedure | Example |
|---|---|
| .a) $n \leq 8$.<br>v) Look up in Table XIII the probability corresponding to $U(n_A,n_B)$. If the probability is less than or equal to $\alpha$, decide that observations on product A will tend to exceed those on product B*. Otherwise, decide that there is no reason to believe that observations on product A will tend to exceed those on product B. | |
| .b) $n > 8$.<br>v) Look up in Table XIV the critical value $U_\alpha(n_1,n_2)$ corresponding to the $\alpha$ selected in (i). $n_2$ and $n_1$ are the larger and smaller respectively of $n_A$, $n_B$. | |
| vi) If $U_B \leq U_\alpha(n_1,n_2)$ decide that observations on product A will tend to exceed those on product B*. Otherwise, decide that there is no reason to believe that observations on product A will tend to exceed those on product B. | |

*) If we are willing to assume that the population of observations on product A is identical to that of population B, except that their averages may differ, then we conclude that the average of product A exceeds the average of product B.

2.7.6 Do the averages of t products differ? (We have $n_1, n_2, \ldots, n_t$ observations respectively on each of the products $1, 2, \ldots, t$).

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $\chi^2_{1-\alpha}$ for t-1 degrees of freedom. | |
| iii) Let $N = n_1 + n_2 + \ldots + n_t$. Assign ranks to each observation according to its size. That is, assign 1 to the smallest, 2 to the next smallest, N to the largest, etc. | |
| iv) Compute $R_i$, the sum of the ranks of the $n_i$ observations on the $i^{th}$ product, for $i=1,2,\ldots,t$. | |
| v) Compute: $$H = \frac{12}{N(N+1)} \sum_{i=1}^{t} \frac{R_i^2}{n_i} - 3(N+1)$$ | |
| vi) If $H > \chi^2_{1-\alpha}$, conclude that the averages of the t products differ. (Otherwise, there is no reason to believe the averages differ.) | |

(Note that each of the $n_i$ should be greater than 5. If this is not so, the level of significance $\alpha$ may be considerably different from the intended value).

## 2.8 Some "Short-cut" Tests

Short-cut tests are characterized by their simplicity. The calculations are simple, and are such that they may often be done on a slide rule. Further, they are easily taught. An additional advantage in their use is that their simplicity implies fewer errors, and this may be important where time spent in checking is costly.

The main disadvantage of the "short-cut" tests is that compared to the tests given in sections 2.2 to 2.6 with the same values of $\alpha$ and n, they will in general have a larger $\beta$, - i.e., they will have a higher proportion of errors of the second kind. For the tests given in this section this increase in error will usually be rather small if the sample sizes involved are each of the order of 10 or less.

Unlike the "non-parametric" tests of 2.7, these tests require the assumption of "normality" of the underlying populations. Small departures from normality will however usually have a negligible effect on the test. That is, the values of $\alpha$ and $\beta$ will in general differ from their intended values by only a slight amount.

No description of the operating characteristics of the test are given. Neither are methods of obtaining sample sizes. A rough rule for obtaining a value for the required sample size is to use the rule given in the corresponding section of 2.2 to 2.6, multiplying the value obtained by 1.1.

2.8.1 Does the average of a new product differ from a standard $m_0$?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $\varphi_{1-\alpha/2}$ in Table XIX. | |
| iii) Compute $\overline{X}$ the mean of the n observations. | |
| iv) Compute w, the difference between the largest and smallest of the n observations. | |
| v) Compute $\varphi = (\overline{X} - m_0)/w$. | |
| vi) If $\lvert\varphi\rvert > \varphi_{1-\alpha/2}$, conclude that the average performance of the new product differs from the standard $m_0$. (Otherwise, there is no reason to believe they differ). | |

2.8.2 Does the average of a new product exceed that of a standard $m_0$?

| | Procedure | Example |
|---|---|---|
| i) | Choose $\alpha$, the significance level of the test. | |
| ii) | Look up $\varphi_{1-\alpha}$ in Table XIX. | |
| iii) | Compute $\bar{X}$, the mean of the n observations. | |
| iv) | Compute w, the difference between the largest and smallest of the n observations. | |
| v) | Compute $\varphi = (\bar{X} - m_0)/w$. | |
| vi) | If $\varphi > \varphi_{1-\alpha}$, conclude that the average of the new produce exceeds that of the standard $m_0$. (Otherwise, there is no reason to believe the average of the new product exceeds the standard). | |

2.8.3 Is the average of a new product less than that of a standard $m_0$?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $\varphi_{1-\alpha}$ in Table XIX. | |
| iii) Compute $\bar{X}$, the mean of the n observations. | |
| iv) Compute w, the difference between the largest and the smallest of the n observations. | |
| v) Compute $\varphi = (m_0 - \bar{X})/w$. | |
| vi) If $\varphi > \varphi_{1-\alpha}$, conclude that the average of the new product is less than that of a standard $m_0$. (Otherwise, there is no reason to believe the average of the new product is less than that of the standard). | |

2.8.4 Do the products A and B differ in average performance? (No particular concern over which is larger; we have a sample of $n = n_A = n_B$ observations from each of A and B)

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $\varphi'_{1-\alpha/2}$ in Table XX. | |
| iii) Compute $\bar{X}_A$, $\bar{X}_B$ the means of the samples from the two populations. | |
| iv) Compute $w_A$, $w_B$ the range, or difference between the largest and smallest values for each of A and B. | |
| v) Compute: $\varphi' = \dfrac{\bar{X}_A - \bar{X}_B}{1/2(w_A+w_B)}$ | |
| vi) If $\|\varphi'\| > \varphi'_{1-\alpha/2}$, conclude that the averages of the two products differ. (Otherwise, there is no reason to believe the averages of A and B differ). | |

2.8.5 Does the average of product A exceed that of product B?

| Procedure | Example |
| --- | --- |
| i) Choose $\alpha$, the significance level of the test. | |
| ii) Look up $\varphi'_{1-\alpha}$ in Table XX. | |
| iii) Compute $\bar{X}_A$, $\bar{X}_B$ the means of the samples from the two populations. | |
| iv) Compute $w_A$, $w_B$ the range, or difference between the largest and the smallest values for each of A and B. | |
| v) Compute: $\varphi' = \frac{\bar{X}_A - \bar{X}_B}{1/2(w_A+w_B)}$ | |
| vi) If $\varphi' > \varphi'_{1-\alpha}$, conclude that the average of A exceeds that of B. (Otherwise, there is no reason to believe the average of A exceeds that of B). | |

2.8.6 Do the averages of t products differ? (We assume we have $n = n_1 = n_2 = \ldots = n_t$ observations on each of products $1,2,\ldots,t$).

| Procedure | | Example |
|---|---|---|
| i) | Choose $\alpha$, the significance level of the test. | |
| ii) | Look up $L_\alpha$ in Table XXI corresponding to t and n. | |
| iii) | Compute $w_1, w_2, \ldots, w_t$, the range of the n observations for each of the products. | |
| iv) | Compute $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_t$ the means of the observations from each of the products. | |
| v) | Compute $w' = w_1 + w_2 + \ldots + w_t$. $w''$ = the difference between the largest and the smallest of the means $\bar{X}_i$. | |
| vi) | Compute $L = nw''/w'$. | |
| vii) | If $L > L_\alpha$, conclude that the averages of the t products differ. (Otherwise, there is no reason to believe the averages differ). | |

2.8.7 Does the variability of the performance of A and B differ? (We are not concerned with which is larger).

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $F'_{\alpha/2}(n_A, n_B)$ and $F'_{\alpha/2}(n_B, n_A)$ in Table XXII. | |
| iii) Compute $w_A$, $w_B$ the range or difference between the largest and smallest observations, for A and B respectively. | |
| iv) Compute $F' = w_A/w_B$. | |
| v) If $F' < F'_{\alpha/2}(n_A, n_B)$ or $F' > 1/F_{\alpha/2}(n_B, n_A)$, conclude that the variability in performance differs. (Otherwise, there is no reason to believe the variabilities differ). | |

2.8.8 Does the variability in performance of A exceed the variability in performance of B?

| Procedure | Example |
|---|---|
| i) Choose $\alpha$, the level of significance of the test. | |
| ii) Look up $F'_\alpha(n_B,n_A)$ in Table XXII. | |
| iii) Compute $w_A$, $w_B$ the range, or difference between the largest and smallest observations, for A and B respectively. | |
| iv) Compute $F' = w_B/w_A$. | |
| v) If $F' < F'_\alpha(n_B,n_A)$, conclude that the variability in performance of A exceeds the variability in performance of B. (Otherwise, there is no reason to believe the variability in performance of A exceeds that of B). | |

U. S. DEPARTMENT OF COMMERCE
Sinclair Weeks, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

# THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its headquarters in Washington, D. C., and its major laboratories in Boulder, Colo., is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside front cover.

## WASHINGTON, D. C.

**Electricity and Electronics.** Resistance and Reactance. Electron Devices. Electrical Instruments. Magnetic Measurements. Dielectrics. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

**Optics and Metrology.** Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Engineering Metrology.

**Heat.** Temperature Physics. Thermodynamics. Cryogenic Physics. Rheology. Engine Fuels. Free Radicals Research.

**Atomic and Radiation Physics.** Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Physics. Nuclear Physics. Radioactivity. X-rays. Betatron. Nucleonic Instrumentation. Radiological Equipment.

**Chemistry.** Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Molecular Structure and Properties of Gases. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

**Mechanics.** Sound. Mechanical Instruments. Fluid Mechanics. Engineering Mechanics. Mass and Scale. Capacity, Density, and Fluid Meters. Combustion Controls.

**Organic and Fibrous Materials.** Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

**Metallurgy.** Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

**Mineral Products.** Engineering Ceramics. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure.

**Building Technology.** Structural Engineering. Fire Protection. Air Conditioning, Heating, and Refrigeration. Floor, Roof, and Wall Coverings. Codes and Safety Standards. Heat Transfer.

**Applied Mathematics.** Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

**Data Processing Systems.** SEAC Engineering Group. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Application Engineering.

- Office of Basic Instrumentation.
- Office of Weights and Measures.

## BOULDER, COLORADO

**Cryogenic Engineering.** Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

**Radio Propagation Physics.** Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Sun-Earth Relationships. VHF Research.

**Radio Propagation Engineering.** Data Reduction Instrumentation. Modulation Systems. Navigation Systems. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Radio Systems Application Engineering. Radio Meteorology.

**Radio Standards.** High Frequency Electrical Standards. Radio Broadcast Service. High Frequency Impedance Standards. Calibration Center. Microwave Physics. Microwave Circuit Standards.

NBS