NATIONAL BUREAU OF STANDARDS REPORT

4817

PROGRESS REPORT FOR THE PERIOD
ENDING 30 JUNE 1956

On

MANUAL ON EXPERIMENTAL STATISTICS FOR
ORDNANCE ENGINEERS

(NBS Project 1103-40-5146)

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

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 field laboratories in Boulder, Colorado, 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 reports and publications, appears on the inside back cover of this report.

WASHINGTON, D. C.

Electricity and Electronics. Resistance and Reactance. Electron Tubes. Electrical Instruments. Magnetic Measurements. Process Technology. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

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

Heat and Power. Temperature Measurements. Thermodynamics. Cryogenic Physics. Engines and Lubrication. Engine Fuels.

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

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. 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. Organic Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

Mineral Products. Ceramic Engineering. Porcelain and Pottery. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure.

Building Technology. Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

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

Data Processing Systems. Components and Techniques. Digital Circuitry. Digital Systems. Analogue Systems. Applications 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.

Radio Propagation Engineering. Frequency Utilization Research. Tropospheric Propagation Research.

Radio Standards. High Frequency Standards Branch: High Frequency Electrical Standards. Radio Broadcast Service. High Frequency Impedance Standards. Microwave Standards Branch: Extreme High Frequency and Noise. Microwave Frequency and Spectroscopy. Microwave Circuit Standards.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT
1103-40-5146/55-93

NBS REPORT
4817

16 August 1956

MANUAL ON EXPERIMENTAL STATISTICS
FOR ORDNANCE ENGINEERS

PROGRESS REPORT FOR THE PERIOD
ENDING 30 JUNE 1956

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

PROGRESS REPORT FOR THE PERIOD
ENDING 30 JUNE 1956

This project has as its aim the preparation of a manual on "Experimental Statistics" for inclusion in the Army Ordnance Engineering Handbook.

As of June 30, 1956, work on the manual had been underway for one year.

The pages to follow present an advanced draft of the first part of the proposed manual; Introduction, chapter on Some Basic Statistical Concepts, and Part I (Some Standard Statistical Techniques for Quantitative Data). This material is being submitted at this time in the interest of receiving comments on the proposed form, format, content, etc., of the manual at this intermediate stage.

Note: The complex system of numbering the problems (e.g., section 1.6.2.2.1) has proved useful during the writing of the various sections of the manual, but will be replaced in the final version by a simpler system.

MANUAL ON EXPERIMENTAL STATISTICS

FOR ORDNANCE ENGINEERS

Prepared by

THE STATISTICAL ENGINEERING LABORATORY

NATIONAL BUREAU OF STANDARDS

PREFACE

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

The use of statistical methods has increased greatly in the past quarter of a century. Since the initial development of these methods occurred largely in agriculture and the biological sciences, it is natural that these fields were the first to make extensive use of the methods. With the introduction of techniques for statistical quality control of production and acceptance sampling, in the 1920's industry began to take notice of statistics, However, it was World War II which brought statistics to their notice. This was largely due to the influence of the military. They adopted scientifically designed sampling inspection procedures, putting indirect pressure on the suppliers of military material to adopt quality control procedures to lower the rejection of their output by the service procurement officers. Also, they set up an educational program giving intensive courses in Statistical Quality Control. Mention should be made of the wartime research in statistical quality control and statistics and especially of the Statistical Research Group, Columbia University, working under the Applied Mathematics Panel, the National Defense Research Committee and the Office of Scientific Research and Development.

Control charts and sampling plans formed the opening wedge for the use of statistics in industry. Today, many other statistical tools are used in laboratories and research departments in military and industrial establishments, for

example, analysis of variance, regression analysis and the design of experiments.

This manual is a collection of statistical procedures useful in ordnance applications. Each section is independent of all the other sections, and depends only on the first few pages of the introduction. 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 both to the person who has had almost no contact with statistics and to the person who merely wants a convenient reference where application of some specific technique is outlined clearly and concisely.

SOME BASIC STATISTICAL CONCEPTS

Everything which deals even remotely with the collection, analysis, interpretation and presentation of numerical data may be classified as belonging to the domain of statistics. We may divide statistical methods into two classes -- descriptive and inductive. Descriptive statistical methods are those which are used to summarize or describe data and are the kind we see used everyday in the newspapers and magazines. Inductive statistical methods are those which attempt to make generalizations, predictions, or estimates from given data, and apply it to a larger mass of similar data. 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). Any subset of a population is a sample from that population. 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.

Usually, we wish to know something about a population. Since it is ordinarily inconvenient or impossible to observe every item in the population, we take a sample, observe the items in the sample, and from them make inferences about the entire population. 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 random sample from the population of rounds (all 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 the sample. If some individual in the population is more likely to be chosen in the sample than others, then the sample is said to be biased. In practice it is found that subjective methods of selection, due to unconscious or conscious preferences, usually leads to biased samples. To avoid possibilities of bias, and to protect against unwarranted assumptions about the individuals in the population, a non-subjective method of choosing a sample should be employed. Section 00.00 of the Appendix describes a method of obtaining random samples by means of a table of random numbers.

Properties of Populations

Although we seldom examine the entire population, we obtain much information from populations in general by observing samples. Large samples do in fact tell us a great deal about populations. Below is a histogram representing the distribution of 5,000 Rockwell hardness readings. 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 a given range of edge-widths. If the sample is large, then we can draw the histogram by increasing the number of bars, each of which will be smaller in width. Imagine the

Figure 1a

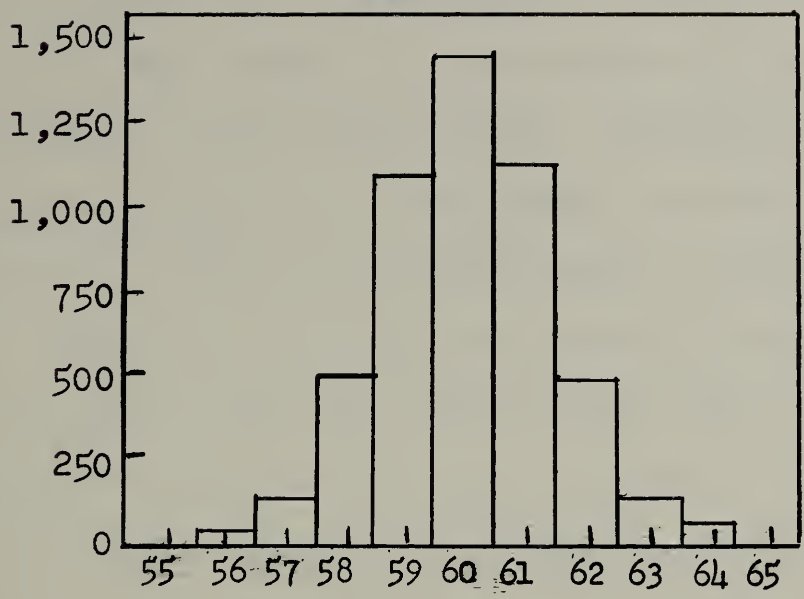
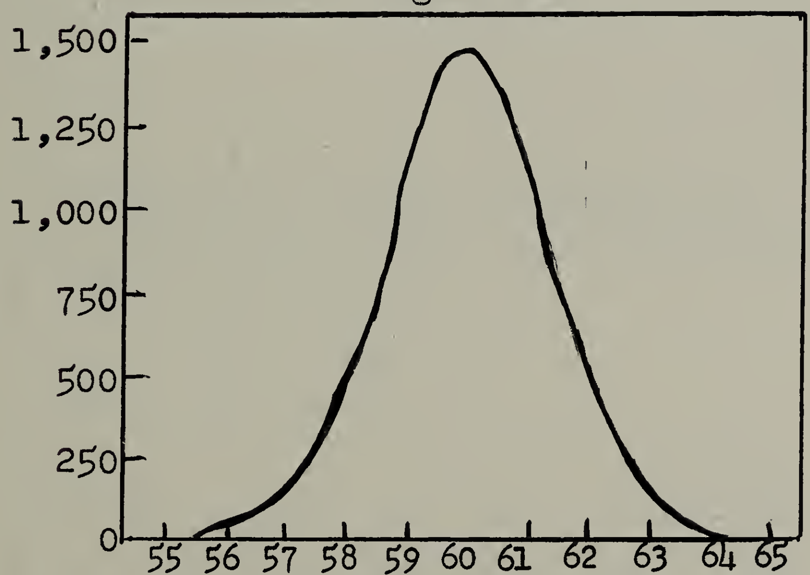


Figure 1b

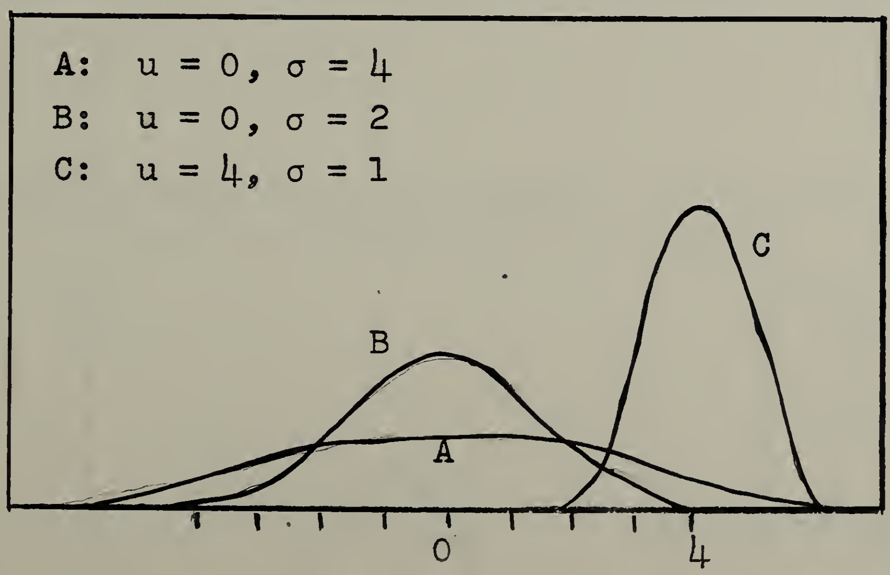


sample size continuing to increase while the number of bars also continues to increase and their width becomes so small that the tops of the bars blend into a continuous curve, such

as that of Figure 1b. If we were to carry out this sort of a scheme on a large number of populations, we would find that many different curves would arise. Possibly, the majority of them would be symmetrical bell shaped curves, such as that in Figure 1b. These are called normal or "Gaussian" distributions. We would also find some which were not symmetrical, and occasionally some that were shaped like a J or a U.

"Normal" curves can be represented by a two parameter family of curves. These two parameters are usually represented by μ and σ, and are called the arithmetic mean (or simply the mean), and the standard deviation. On our normal curve, μ is the value for which the curve is highest, and σ represents the distance between the inflection point and μ.

Figure 2, Normal Distributions

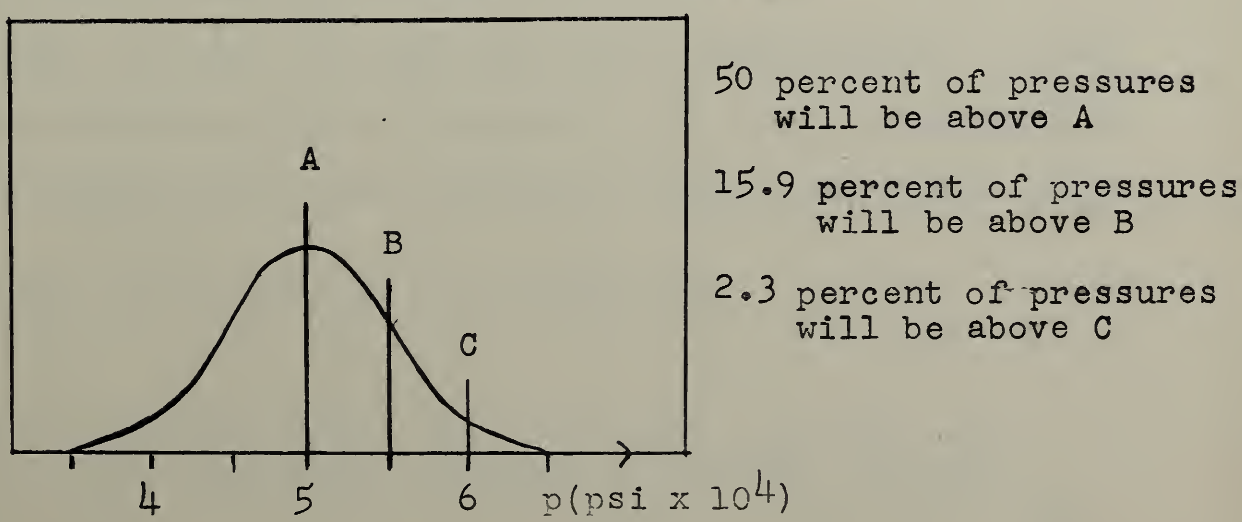


The parameter u is also the arithmetic mean and a measure of central tendency of the population. The parameter σ is a measure of the spread, scatter or dispersion of the population.

From the above, we may state that if the population is indeed normal, then we know everything about it if we know the two parameters u and σ.

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 = u = 50,000, and standard deviation $\sigma = 5x10^3$ (p.s.i.). Then by means of tables of the normal distribution it is very easy to show 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 percent to have pressures above 55,000 p.s.i., and 2.3 percent to have pressures above 60,000 p.s.i.

Figure 3, Distribution of Chamber Pressures p.



In areas where a lot of experimental work has been done, it often happens that we do know μ 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 that we have n observations $x_1, x_2, \ldots, x_n$ taken at random from a normal population. What then are our best estimates of μ and σ? They are:

for μ, $$\bar{x} = \sum_{i=1}^{n} x_i$$

for σ, $$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

For computational purposes, the following formula is more convenient for s^2,

$$s^2 = \frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}$$

Although $\bar{x}$ and s^2 are the "best" estimates of μ and σ^2 for populations which are normal, it should be obvious that different samples will have different values for $\bar{x}$ and s^2.*

How much then can we rely on our estimates?

* For some populations which are non-normal, and for some purposes, there are better measures of central tendency than μ. For a discussion of this point, see Section 00.00 of the Appendix.

Figure 4a Distribution of means of samples from normal distributions, various sample sizes

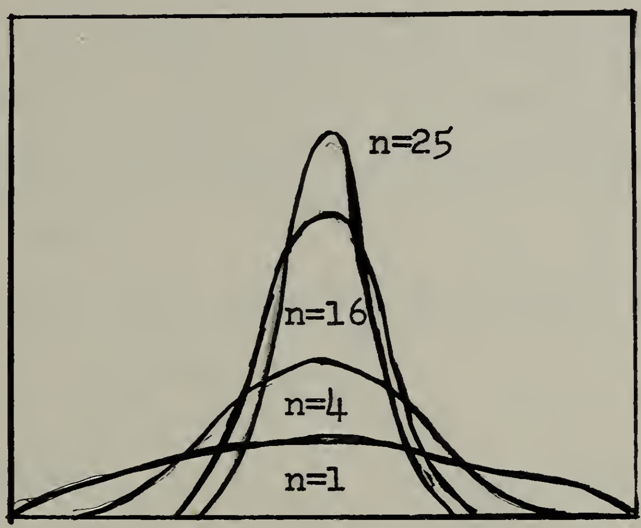


Figure 4b Distribution of variances of samples from normal distributions, various sample sizes

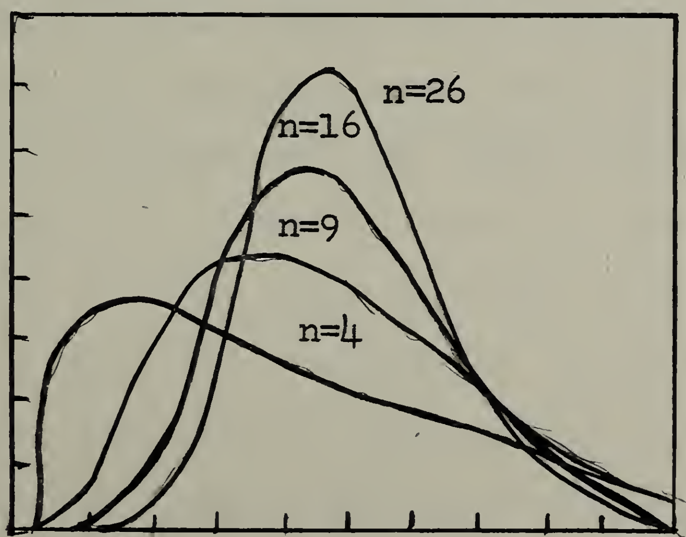


Figure 4a shows the distribution of the sample values of $\bar{x}$ for samples of various sizes. If we define the area

under any curve as being one square unit (this is a standard convention in statistics), then the area under the curve between any two $\bar{x}$ 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.

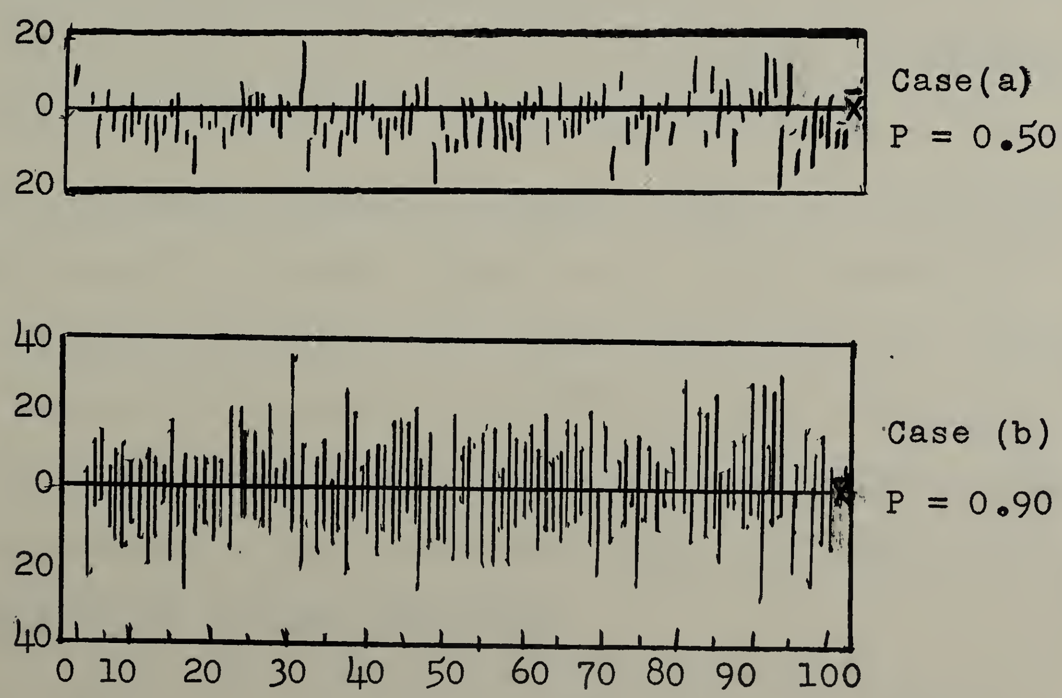
Confidence Intervals

$\bar{x}$ and s, our sample estimates of u and σ thus are seen to vary from sample to sample. Increasing the sample size decreases the amount of variation, but does not eliminate it. Even though we cannot state the exact value of the parameter, we can with a chosen degree of confidence state an interval within which the parameter lies. We can state an interval and a probability that the interval range contains the true value of the parameter, provided that our sample has been randomly choosen. The probability will of course be closely related to the interval that we give. This interval is an estimate of the parameter and is known as a "confidence interval". The procedure for obtaining the intervals is given in the section 0.00. However, it may be worthwhile giving an illustration.

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. Suppose

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 point for our confidence interval. Consider all the possible samplesof size 4 that we could take, and the accompanying upper and lower points for the confidence intervals computed from each. If the points were for a 95 percent confidence interval, then we should expect 95 percent of the intervals to cover the true value, 50,000 p.s.i.

Figure 5 Confidence intervals for samples of size 5 drawn at random from a normal population with μ = 50,000 p.s.i, σ = 5,000 p.s.i. (Samples drawn with aid of tables of the RAND Corporation).



It should be noted that an increase in the confidence will always result in an increase in the width of the confidence interval.

Tolerance Limits

Sometimes what is wanted is not an inference as to the mean and variance of the population but two 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, then we may wish to know approximately the range of chamber velocities 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 limit, and say that at least P percent of the ammunition will have chamber pressures within the above limits, with a confidence coefficient of α. If we use the method (See Problem 0.00), then the proportion of the time that we will be making true statements will be α.

The difference between tolerance limits and confidence intervals should be noted. They are two quite different things. Tolerance limits for a given population are limits between which we estimate a stated proportion of the population to lie. A confidence interval is an interval within which we estimate a given parameter (e.g. the mean) will lie.

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 from 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. Their standard deviation 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 shells differing from the standard by as much as our sample did?

If it is highly improbable, than 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 was not improbable that we should get a sample differing as much as ours then we should decide that there was no good reason to believe that the new shells penetrated farther than the standard shells.

Setting up the Decision Procedure

In our example, if the new shells penetrated farther than the old shells, we wished to know it. Inasmuch as it is frequently easier to disprove than to prove, we start with what we call the null hypothesis. 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 hypothesis. We make Decision (i) - The new shells penetrate further on the average than the standard shells.

If, on the other hand the "increase" in penetration of the new shells over the old shells is not large enough to be considered improbable, we make Decision (ii) - There is no reason to believe the new shells penetrate further 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 fluctuation for sample alone, a larger increase would have occurred only a small proportion α of the time. (α is some small number decided on in advance of performing the experiment). The quantity α defined above is known as the "significance level." The significance level should be chosen on economical and other non-statistical grounds. Two values for α 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 sacred in their use. These values are 0.05 and 0.01. Using the .05 level of significance, for example, we should reject the null hypothesis and make decision (i) only if the sample increase in penetration for the new shell was so large that it would occur less frequently than .05 of the time in repeated trials, when the new shells were in fact no better.

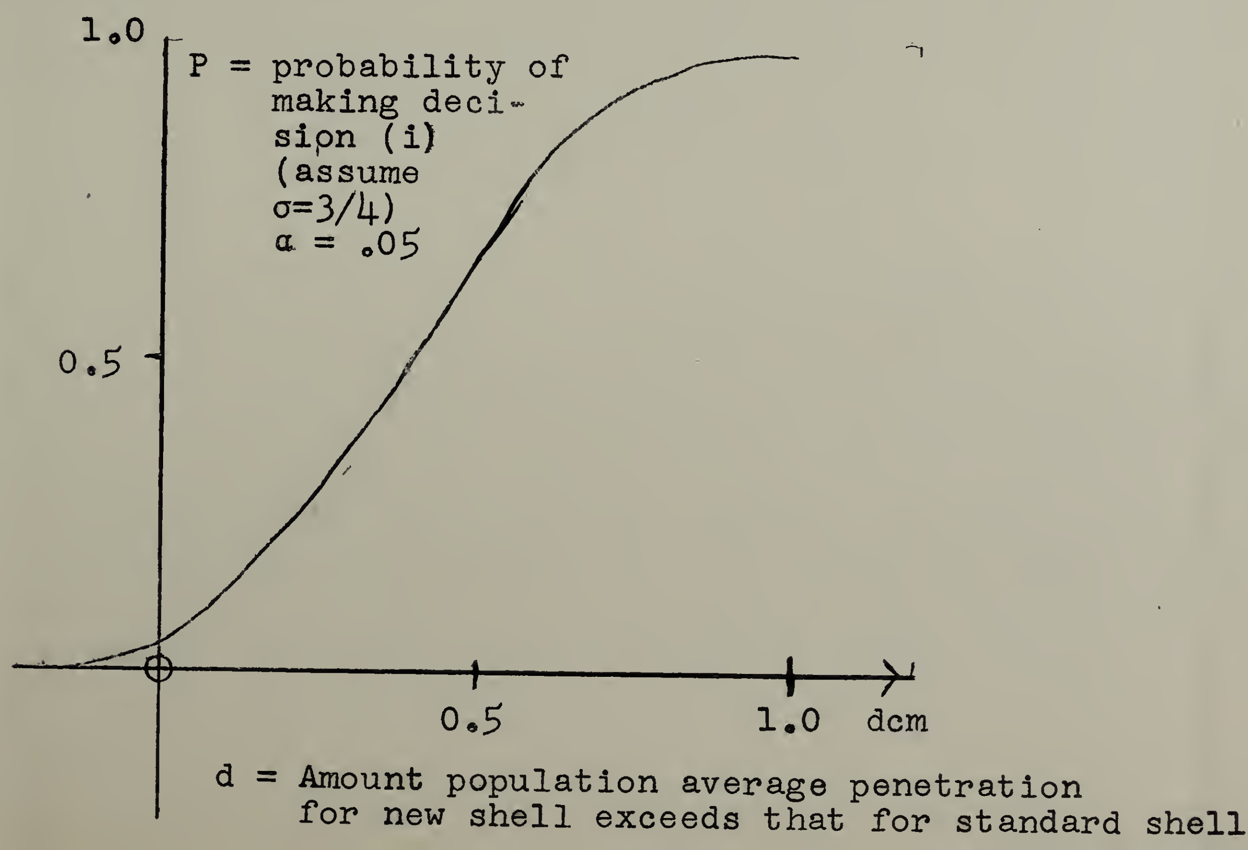
Two Kinds of Errors

Since we know that 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., find 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 α, 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." Of course we do not know in a given instance whether we have made an error in decision. We can, however, know the probability with which we will make either type of error.

For the above example, the probabilities of error for possible true values of the difference are given in Figure 6.

Figure 6 Operating Characteristics of Decision Procedure of Example



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 (dimensional units ... etc.). 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 technique. 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.

Given:

n independent measurements $x_1, x_2, \ldots, x_n$ selected at random from a much larger group.

Example:

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

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

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

(2) 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)

Note: 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

$$\bar{x} = \frac{1}{n}\left(\sum_{x=1}^{n} x_i\right)$$

Example:

$$\bar{x} = \frac{1.260}{10} = .126$$

(For some assymetrical distributions, the arithmetic mean may not necessarily be the best single description of the over-all performance of items. For a discussion of this point, see Section 00.00, Appendix

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 would expect to 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 percent of the time, we might say that we were operating at a 99 percent confidence level. Our particular kind of interval estimates might likewise be called "99 percent confidence intervals."

Similarly if our intervals included the true average 95 percent of the time, we would be operating at a 95 percent confidence level, and our intervals would be called 95 percent confidence intervals. In general, if in the long run we expect 100 $(1-\alpha)$ percent of our intervals to contain the true value, we are operating at 100 $(1-\alpha)$ percent confidence.

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

level" (α) of a test. This is the same α 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 confidence level. (It is assumed that the large group forms a normal population, and that each observation or individual is quite independent of the others in the sample). 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 $(1-\alpha)$ percent of the time; 100 $(\frac{\alpha}{2})$ percent of the time the interval will be above the true mean, and 100 $(\frac{\alpha}{2})$ percent of the time it will be below the true mean.

Procedure	Example
Problem: What is a 100 $(1-\alpha)$ percent confidence interval (2 sided) for the true mean? i) Choose the desired confidence level, $1 - \alpha$	Problem: What is a 95 percent (2 sided) confidence interval for the true mean? i) Choose confidence level .95 $.95 = 1 - \alpha$ $\alpha = .05$
ii) Compute: arithmetic mean $\bar{x}$ (see 1.1.1) $s = \sqrt{\frac{n\Sigma x^2 - (\Sigma x)^2}{n\,(n-1)}}$	ii) $\bar{x}$ = .126 inches s = 0.00359 inches
iii) Look up: $t = t_{1-\alpha/2}$ for n-1 degrees of freedom in Table I	iii) $t = t_{.975,9} = 2.26$

Procedure	Example
iv) Compute: $x_U = \bar{x} + t \frac{s}{\sqrt{n}}$ $x_L = \bar{x} - t \frac{s}{\sqrt{n}}$	iv) $x_U = \bar{x} + t \frac{s}{\sqrt{n}}$ $x_U = .126 + \frac{2.26(.00359)}{\sqrt{10}} = .1341$ inches $x_L = \bar{x} - t \frac{s}{\sqrt{n}}$ $x_L = .126 - \frac{2.26(.00359)}{\sqrt{10}} = .1179$ inches
v) Conclude: The interval from x_L to x_U is a 100 $(1-\alpha)$ percent confidence interval for the true mean.	v) Conclude: The interval from .1179 to .1341 inches is a 95 percent confidence interval for the lot mean.

1.1.2.2 One-sided confidence intervals.

The example used in 1.1.2.1, can be used to make another kind of confidence interval statement. As we said 100 $(\frac{\alpha}{2})$ percent of the time the interval will be above the true mean (i.e. x_L is greater than true mean). Therefore 100 $(1 - \frac{\alpha}{2})$ percent of the time, the true mean is greater than x_L.

From the example of 1.1.2.1,

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

Thus, the either of the two open intervals - above .1179 inches, or below .1341 inches can be called a <u>97.5 percent one-sided</u> confidence interval for the average.

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 $(1-\alpha)$ percent confidence interval (one-sided) for the true mean?	Problem: What is a value, which we expect, with 99 percent confidence, to be exceeded by the lot mean?
i) Choose the desired confidence level $(1 - \alpha)$	i) $(1 - \alpha) = .99$ $\alpha = .1$

Procedure	Example
ii) Compute $\bar{x}$, s	ii) $\bar{x}$ = .126 inches s = 0.00359 inches
iii) Look up: $t = t_{1-\alpha}$ for n-1 degrees of freedom in Table II	iii) $t = t_{.99}$ for 9 degrees of freedom = 3.25
iv) Compute: $x_L' = \bar{x} - t \frac{s}{\sqrt{n}}$	iv) $X_L' = .126 - \frac{(3.25)(.00359)}{\sqrt{10}}$ $X_L' = .1223$
v) Conclude: We are 100 (1-α) percent confident that the lot mean is greater than x_L'.	v) Conclude: We are 99 percent confident that the lot mean is greater than .1223 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 <u>know</u> σ, 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 techniques of sections 1.2 and/or 1. (Control chart procedures).

The procedure for computing these confidence intervals is simple. In the example of 1.1.2, merely replace s by σ 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.

Procedure	Example
Problem: Find a 100 $(1-\alpha)$ percent 2-sided confidence interval for the lot mean, using known σ.	Problem: What is a 95 percent confidence interval (2-sided) for the lot mean? (σ known equal to .0040 inches).

Procedure	Example
i) Choose the desired confidence level, $1-\alpha$	i) $1-\alpha = .95$, thus $\alpha = .05$
ii) Compute $\bar{x}$	ii) $\bar{x} = .126$ inches
iii) Look up: $z = z_{1-\alpha/2}$ in Table I	iii) $z = z_{1-\alpha/2} = 1.96$
iv) Compute: $x_U = \bar{x} + z\left(\frac{\sigma}{\sqrt{n}}\right)$ $x_L = \bar{x} - z\left(\frac{\sigma}{\sqrt{n}}\right)$	iv) $x_U = .126 + 1.96 \left(\frac{.004}{\sqrt{10}}\right) = .128$ $x_L = .124$

1.1.4 Confidence intervals when normality cannot be assumed.

1.1.4.1 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 be changed very little from the chosen level. For other cases, a method of making interval estimates is given below which is independent of the population distribution.

1.1.4.2 Interval estimates which are independent of the population distribution.

Let $x_1 \leq x_2 \leq \ldots \leq x_n$ be the observed values arranged in order of increasing magnitude. Let M be the population median.

Consider any interval formed by taking the i^{th} largest and the i^{th} smallest observation. The probability that such an interval contains M, the true population median, is $1-2I_{.5}(n-i+1,i)$ where I is the incomplete beta function tabulated in Table 16 of [1].

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:

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 Preface (see pp. 4 and 5) we have stated that our best estimate of σ^2, the variance of a normal population is:

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

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

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

The estimate of σ, the population standard deviation, is

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

NOTE: For small samples (say n less than 10), a fair approximation to $\underline{s}$ is $\frac{R}{\sqrt{n}}$, where R = difference between highest and lowest values in the sample.

1.2.1.2 Use of the range to estimate variability.

The range of n observations is the difference between the highest and the lowest values observed in the sample. For small samples (n less than 10), the range is a reasonably efficient estimator of σ (the standard deviation of a normal population) - not as efficient as s, but easier to calculate. Table 1.2.1.2 Col. 2, gives the factors which convert from observed range in a sample of n to an estimate of population standard deviation.

Estimate of $\sigma = \frac{1}{d_n}$ x range in a sample of n

Table 1.2.1.2

Size of Sample n	d_n	$\frac{1}{d_n}$	[See note] $\sqrt{n}$
2	1.1284	.8862	1.414
3	1.6926	.5908	1.732
4	2.0588	.4857	2.000
5	2.3259	.4299	2.236
6	2.5344	.3946	2.449
7	2.7044	.3698	2.646
8	2.8472	.3512	2.828
9	2.9700	.3367	3.000
10	3.0775	.3249	3.162

NOTE: The last column ($\sqrt{n}$) is included in order to show a quick and rough method of estimating σ from small samples. Note how closely $\sqrt{n}$ approximates d_n for small n. For samples of n = 10 or less, therefore, one may estimate σ by taking the range and dividing by $\sqrt{n}$.

1.2.1.3 The calculation of standard deviation - division by n or (n-1)?

There exists some confusion over the calculation of standard deviation, particularly over the proper choice between the two relationships given below:

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

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

Either of these formulas give a measure of variability of test results. Either may be used to compare the variability of different samples or processes, provided that: (a) the same formula is used in calculating all the measures to be compared and (b) all calculations are based on the same sample size n. When n is not the same, values of s_{n-1}^2 calculated from (2) are still comparable, but not values of s^2 calculated from (1). Furthermore, when the n's are different, values of the standard deviation or s obtained by taking the square root of the right-hand side of either (1) or (2) are no longer comparable, the magnitude of the resulting systematic error depending on the actual values of n involved. Whichever divisor (n or n-1) is used, the estimate

of σ obtained is dependent on n, i.e., the average of a very large number of estimates calculated in the same way will not quite be equal to σ. Table 1.2.1.3 shows how the number of observations n affects the goodness of estimation. Note that s_{n-1} calculated from (2) is substantially less likely to underestimate the real σ in the case of small samples.

TABLE 1.2.1.3

TWO COMMON ESTIMATORS OF σ - HOW THEY RELATE TO σ

	$s_n = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$		$s_{n-1} = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$	
Size of Sample n	Average value of s_n (in σ units)	Probability that σ is underestimated by s_n	Average value of s_{n-1} (in σ units)	Probability that σ is underestimated by s_{n-1}
2	.564	.843	.798	.683
3	.724	.777	.886	.632
4	.798	.739	.921	.608
5	.841	.713	.940	.594
6	.869	.694	.951	.584
7	.888	.679	.959	.577
8	.903	.667	.965	.571
9	.914	.658	.969	.567
10	.923	.650	.973	.563
12	.936	.637	.978	.557
15	.949	.622	.982	.550
20	.962	.605	.987	.543
24		.596	.989	.539
30	.975	.586	.991	.535

In addition, all of the tables that have been developed to facilitate the application of modern statistical methods presuppose that s_{n-1} is calculated from (2). (The noteworthy exception is in quality control chart work, and there the necessary adjustments have been made in calculating the factors given for central lines and control limits). In this Handbook we shall always use s_{n-1}.

Inasmuch as our estimate is subject to fluctuation about the true value, we may also wish to have an interval estimate (described in 1.2.2).

1.2.2 Confidence Interval estimates.

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

1.2.2.1 Two sided confidence interval estimates.

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

Problem.

What is a 95 percent confidence interval for σ, the variability of the burning time of the lot of powder?

Procedure	Example
i) Choose α, the significance level	i) $.95 = 1-\alpha$ $\alpha = .05$
ii) Compute s, $s = \sqrt{\frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n-1)}}$	ii) $s = 10.37$ seconds
iii) Look up: $\chi^2_{\alpha/2}$, $\chi^2_{1-\alpha/2}$ for n-1 degrees of freedom in Table V .	iii) For 9 degrees of freedom, $\chi^2_{\alpha/2} = \chi^2_{.025} = 2.70$ $\chi^2_{1-\alpha/2} = \chi^2_{.975} = 19.02$

Procedure	Example
iv) Compute: $s_L = s\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}}}$, and $s_U = s\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}}$	iv) $s_L = 10.37\sqrt{\frac{9}{19.02}} = 7.13$ $s_U = 10.37\sqrt{\frac{9}{2.70}} = 18.94$
v) Conclude: Our two-sided interval estimate for σ is the interval s_L to s_U and we have confidence 100 (1-α) percent that the interval contains σ.	v) Conclude: Our two-sided interval estimate for σ is the interval from 7.13 to 18.94 and we are 95 percent confident that the interval contains σ.

For given degrees of freedom (n-1) and significance level α, part of the formula, i.e.,

$$\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}} \quad \text{and} \quad \sqrt{\frac{n-1}{\chi^2_{\alpha/2}}}$$

can be tabulated in advance, and merely multiplied by s in a particular problem. Such a table has been made for certain values of $n-1 = \nu$ and α. (Table 1.2.2).

Table 1.2.2
Tables for Computing Confidence Limits for σ

Degrees of Freedom ν	$\sqrt{\frac{\nu}{\chi^2_{.05}}}$	$\sqrt{\frac{\nu}{\chi^2_{.95}}}$	$\sqrt{\frac{\nu}{\chi^2_{.025}}}$	$\sqrt{\frac{\nu}{\chi^2_{.975}}}$	$\sqrt{\frac{\nu}{\chi^2_{.005}}}$	$\sqrt{\frac{\nu}{\chi^2_{.995}}}$
1	.5103	15.947	.4461	31.910	.3562	159.576
2	.5778	4.415	.5207	6.285	.4344	14.124
3	.6196	2.920	.5665	3.729	.4834	6.467
4	.6493	2.372	.5992	2.874	.5188	4.396
5	.6721	2.089	.6242	2.453	.5464	3.485
6	.6903	1.915	.6444	2.202	.5688	2.980
7	.7054	1.797	.6612	2.035	.5875	2.660
8	.7183	1.711	.6754	1.916	.6037	2.439
9	.7293	1.645	.6878	1.826	.6177	2.278
10	.7391	1.593	.6987	1.755	.6301	2.154
11	.7477	1.551	.7084	1.698	.6412	2.056
12	.7554	1.515	.7171	1.651	.6512	1.976
13	.7624	1.485	.7250	1.611	.6603	1.909
14	.7688	1.460	.7321	1.577	.6686	1.854
15	.7747	1.437	.7387	1.548	.6762	1.806
20	.7979	1.358	.7650	1.444	.7071	1.640
25	.8149	1.308	.7843	1.380	.7299	1.542
30	.8279	1.274	.7991	1.337	.7477	1.475
40	.8470	1.228	.8210	1.279	.7740	1.390
50	.8606	1.199	.8367	1.243	.7931	1.337
60	.8710	1.179	.8487	1.217	.8078	1.299
70	.8793	1.163	.8583	1.198	.8196	1.272
80	.8861	1.151	.8662	1.183	.8293	1.250
90	.8919	1.141	.8728	1.171	.8376	1.233
100	.8968	1.133	.8785	1.161	.8446	1.219

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 statememt such as the following:

We are 100 $(1-\alpha)$ percent confident that the variability as measured by σ is less than some value A.

Similarly we may be interested only in statements that the variability is greater than some number B. Both statements are one-sided confidence interval estimates.

Problem: Can we give a value A; and have 95 percent confidence that σ is less than A?

Procedure	Example
i) Compute s,	i) $s = 10.37$ seconds
ii) Look up: χ^2_α for n-1 degrees of freedom	ii) For 9 degrees of freedom $\chi^2_{.05} = 3.33$
iii) Compute: $s'_U = s\sqrt{\frac{(n-1)}{\chi^2_\alpha}}$	iii) $s'_U = 10.37\sqrt{\frac{9}{3.33}}$ $s'_U = 17.05$ seconds

Therefore we are 95 percent confident that the variability as measured by σ is <u>less 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 σ, the standard deviation. In planning experiments for example, the sample size required in order to meet certain requirements is a function of σ.

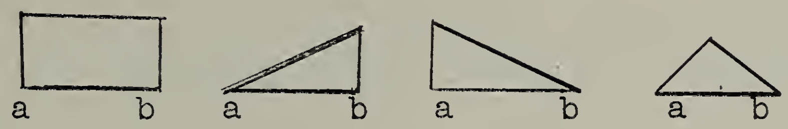
There is seldom a situation where one does not know something about the variance, or cannot use some existing information to get at least a very rough estimate of σ. The necessary information involves the form of the distribution and the spread of values. If the values for the individual items can be assumed to form a normal distribution, then either of the following methods can be used to get an estimate of σ.

a) Choose values a_1 and b_1 between which you expect 99.9 percent of all individuals to be. Estimate σ as $\frac{|a_1 - b_1|}{6}$ or

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

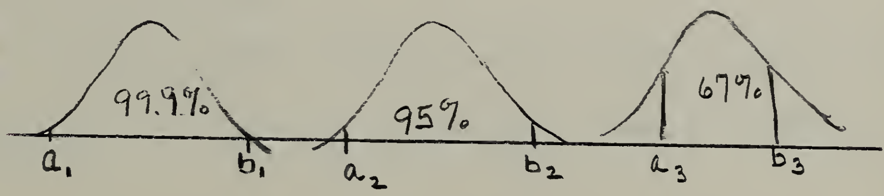
If in fact the populations are not "normal" but follow one of the forms in Figure 00.00, then the standard deviation may be estimated as indicated in the figure.

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:	$\frac{b_1-a_1}{6}$	$\frac{b_2-a_2}{4}$	$\frac{b_3-a_3}{2}$

See Deming "Some Theory of Sampling", John Wiley and Sons, Inc. page 62.

1.3 Number of measurements required to establish the mean with certain precision.

In planning experiments we may wish to ask the question: How many measurements must be taken in order to be fairly certain that our estimate of the mean (i.e. $\bar{x}$ = sample mean) does not differ from the true mean (μ) by more than a specified amount <u>d</u>?

We must choose the following:

(1) α, the significance level = some small probability of making an incorrect statement, say $\alpha = .05$ or .01.

(2) d, the precision of our estimate which is of importance ($d = \bar{x} - \mu$).

and we must have an estimate of the variability (s).

<u>Example:</u>

Using the data of 1.1 to provide an estimate of the variability, how many washers would we have to measure from a new lot in order to say (with 95 percent chance of being right) that the mean of this sample did not differ from the true lot mean by more than   inches?

Procedure	Example
i) Choose α, the significance level.	i) $\alpha = .05$

Procedure	Example
ii) Choose <u>d</u> the amount of difference from the true mean that is of practical significance.	ii) $d =$
iii) Look up t_α for n-1 degrees of freedom in Table II.	iii) $t_{.95}$ for 9.f. =
iv) Compute: s^2 the sample variance	iv) $s^2 =$
v) Compute: $n = \frac{t^2 s^2}{d^2}$	

Stein's Method of Two-stage sampling.

The method described above, of course, depends upon how good an estimate of variability(s) we have. Suppose we used an <u>s</u> calculated from a previous lot (or lots) of washers to calculate the sample size required from a new lot. In other cases, we may only "guess" at s, based on experience with similar material. The correctness of our sample size calculation will depend on whether or not the assumed s is typical also of the new lot.

To avoid this difficulty, a method has been developed which uses the information about s from the lot (or population) being sampled, and which gives a chosen degree of precision, regardless of the correctness of initial guesses or estimates of s.

The sample is taken in 2 parts. The first part, of size n_1, supplies an estimate of variability s, and a preliminary estimate of $\bar{x}$. A formula is given showing how to calculate the additional number of observations needed to have a specified confidence.

Procedure	Example
i) Choose α, the significance level	i) $\alpha = .05$
ii) Choose d	ii) d =
iii) Select a value for n_1, the size of the initial sample	iii) Select n_1 =

(continued next page)

Procedure	Example
iii) continued - (For this, it is helpful to know how large the final sample would have to be if the assumed value of s happened to be correct $n = \frac{t^2 s^2}{d^2}$)	
iv) Take the first sample and compute s^2	iv) s^2 =
v) Look up t_1 for n-1 degrees of freedom and $\alpha = .05$	v) $t_{.05}$ for degrees of freedom =
vi) Compute $\ell = \frac{t_1 s}{\sqrt{n}}$	vi) l =
vii) If $\ell < d$, sample of n_1 was sufficient If $\ell > d$, compute $n = \frac{t_1^2 s^2}{d^2}$ (Round n to next higher integer) n = total sample size $= n_1 + n_2$ Take n_2 additional observations	

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 . (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, Vol. 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 within P percent of its true value, with prescribed confidence?

Procedure:

If we choose P = 20 percent, and confidence coefficient .95, we read the curve labelled 20 percent 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 <u>range</u> 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 <u>P</u> of the population will lie, (two-sided limits), or a value <u>A</u> above which at least a proportion <u>P</u> 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 γ, and it refers to the proportion of the time that our method will result in correct statements.

1.5.1 Two-sided tolerance limits.

Problem: We would like to state 2 thickness limits within which we are 95 percent confident that 90 percent of the values lie.

Procedure	Example
i) Choose P, the proportion and γ, the confidence coefficient	i) $P = .90$ $\gamma = .95$
ii) Compute from the sample: $\bar{x}$, the arithmetic mean s, the standard deviation	ii) $\bar{x} = .126$ inches $s = 0.00359$ inches
iii) Look up K for chosen P and γ in Table IX.	iii) $K = 2.839$
iv) Compute: $X_U = \bar{X} + Ks$ $X_L = \bar{X} + Ks$	iv) $X_U = .126+2.839(.00359)=0.136$ in $X_L = .126-2.839(.00359)=0.116$ in
v) Conclusion: With a confidence coefficient of γ, 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 percent confidence, we may predict that 90 percent 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

$$X_U = \bar{X} + K s$$

for the one-sided upper limit and

$$X_L = \bar{X} - K s$$

for the one-sided lower limit.

<u>Problem</u>: Give a single value above which you predict with confidence γ that a proportion P of the population will lie.

Procedure	Example
i) Choose P the proportion and γ, the confidence coefficient.	i) P = 97.5 σ = .90
ii) Compute: $\bar{X}$, the arithmetic mean s,	ii) $\bar{X}$ = .126 inches s = 0.00359 inches

Procedure	Example
iii) Compute: $a = \frac{1 - z^2_{1-\gamma}}{2(n-1)}$ (where z can be found in Table I). $b = z^2_P - \frac{z^2_{1-\gamma}}{n}$ $K = \frac{z_P + \sqrt{z^2_P - ab}}{a}$	iii) $a = 1 - \frac{(1.282)^2}{18} = .9085$ $b = (1.96)^2 - \frac{(1.282)^2}{10} = 3.677$ $K = \frac{1.96 + \sqrt{(1.96)^2 - ab}}{a} = 2.93$
iv) $X_L = \overline{X} - Ks$	iv) $X_L = .126 - 2.93(.00359) = .115$ in

Thus we are 90 percent confident that 97.5 percent 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. The difference will decrease with increasing sample size, and for most purposes the normality assumption will probably not cause serious error. Occasionally we may wish to obtain tolerance limits for populations which are considerably different from normal. For these populations, provided only that their distributions are continuous, we may state that the probability of including a fraction P or more of the population between the i^{th} largest and the i^{th} smallest observation is $I_{1-P}(2i, n-2i+1)$, where I is tabulated in [4] and [5].

TESTS

1.6 Statistical Tests Concerning Averages and Dispersions.

1.6.1 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. 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, from analysis of the data, 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 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 average of product A is greater than that of product B.

ii) We do not have reason to believe the 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). In each case the alternative decisions are chosen before the data are observed - this is important! Since we ordinarily get our information on one or both of the products by means of a sample, every decision will be subject to error. Of course, other things being equal, the more observations we have, the smaller will be our chance of error.

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

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 α be the probability that we will make an Error of the First Kind, and β 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 (δ) there will be a value of β, say $\beta(\delta)$ for each possible difference δ. $\beta(\delta)$ will decrease as δ increases. β has no meaning by itself, but is always associated with a particular difference δ.

Given a particular statistical test, and any two of the three quantities n, α, β, 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 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. Just how large must the difference be in order that we may conclude that the two types differ or that the observed difference is "significant?"* This will depend on several factors--the amount of variability in the tubes of each type, the number of tubes of each type 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 that observed 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

---

* Or more accurately "statistically significant." For the distinction between statistically significant and "practially important," see

levels are tabulated extensively for many tests. There is nothing sacred about these levels, however, and a test user may choose any value for α that he feels is appropriate.

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

An OC curve depicts the discriminatory power of a particular statistical test. In the simplest cases we shall discuss (see Figure , for example), if we choose α there will be a whole family of OC curves depending on n. If we choose both n and α. there will be a unique OC curve. The curve can be useful in 2 ways:

1) If our n is already settled upon, we can use the OC curve to read $\beta(\delta)$ for various values of δ or (2), if we are at liberty to choose the sample size for our experiment and have a particular value of δ in mind, we can choose n in rational fashion by looking at the OC curves. What must we do to determine sample size?

(a) First choose α, the significance level. This leads us to a family of OC curves.

(b) The next choice involves δ, the true difference between the two averages. (Of course it is impossible to know the value of δ, but we can choose a δ of the size that it would be practically important to recognize and, perhaps, embarrassing to miss if it did exist.

(c) From the family of OC curves choose one which gives a suitably small value of $\beta(\delta)$ for the chosen δ (conventionally $\beta(\delta) = .05$ or .01). This gives us the required n.

It is evident that for any $\beta(\delta)$, n will increase as δ decreases. It requires larger samples to recognize smaller differences. In some cases the experiment as originally thought of will be seen to require prohibitively large sample sizes, 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 δ.

To use the OC curve for either purpose, one must know the variance σ, or be willing to state some range of σ. (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 σ 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 α and sample size n. For most of the tests, we shall also give the OC curve and a formula which will give the (approximate) value of β for any given difference. Finally, we shall give a formula for determining n, the sample size when α, δ and $\beta(\delta)$ have been specified.

The tests given are exact when (a) the observations for each item are taken randomly from a single population of possible observations and (b) within the population, the quality characteristic measured is normally distributed. The assumption of normality is not ordinarily crucial. In cases where non-normality is thought to be serious, the methods of section may be used.

1.6.2 Comparison of the average of a new product with that of a standard - single measured characteristic.

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

PROBLEMS: 1.6.2.1 To determine whether average of the new product <u>differs</u> from the standard.

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

1.6.2.3 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 1.6.2).

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 1.6.2.2 vs 1.6.2.3), the significance level of the test, i.e. the probability of an error of the first kind, and the operating characteristics of the test may differ considerably from their nominal values.

Ordinarily 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 σ. We shall outline the solutions for each of the 3 problems (1.6.2.1, 1.6.2.2, and 1.6.2.3) for both cases - i.e. where the variability is estimated from the sample, and where σ is known from previous experience.

Symbols to be used

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

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

$\bar{x}$ = average of sample of n measurements on new product,

s = standard deviation of n measurements on new product (used where σ is unknown).

σ = the known standard deviation of the new product.

Problem to be Illustrated in 1.6.2.

For a certain type of shell, specifications state that the amount of powder should average 0.735 lbs. 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, $\bar{x}$ = .710 lbs.

The sample standard deviation s = .0504 lbs.

(In illustrating the known σ case, we assume σ known to be 0.06 lbs.)

Table 1.6.2 - Summary Table for Problems of 1.6.2 - Comparison of Average of a New Product with that of a Standard

(For Details and Worked Example see 1.6.2.1, 1.6.2.2, or 1.6.2.3)

We wish to test whether	Section Reference	Knowledge of Variation of new item	Test to be made	Operating Characteristic of the test (for α=.05 and α=.01)	Sample Size Required (n)	Notes
m differs from m_0	1.6.2.1.1	σ unknown; s = estimate of σ from sample	$\lvert\bar{x}-m_0\rvert > u$	See Figs. 1.3.1 and 1.3.2 *	Figs. 1.3.1 and 1.3.2 can be read in reverse *	$u=t_{1-\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
	1.6.2.1.2	known σ	$\lvert\bar{x}-m_0\rvert > u$	See Figs. 1.3.5 and 1.3.6	For specified α and β, read Figs 1.35 and 1.36 in reverse. For more accurate n, see 1.6.2.1.2	$u=z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$
m is <u>larger</u> than m_0	1.6.2.2.1	σ unknown; s = estimate of σ from sample	$(\bar{x}-m_0) > u$	See Figs. 1.3.3 and 1.3.4 *	Figs 1.3.3 and 1.3.4 can be read in reverse*	$u=t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$
	1.6.2.2.2	σ known	$(x-m_0) > u$	See Figs 1.3.7 and 1.3.8	For specified α and β, Read Figs 1.3.7 and 1.3.8 in reverse For more accurate n, see 1.6.2.2.2	$u=z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$

(Table continued on following page)

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

Table 1.6.2 (Continued)

Summary Table for Comparison of Average of a New Product with that of a Standard

We wish to Test whether	Section Reference	Knowledge of Variation of new item	Test to be made	Operating Characteristic of the test (for α=.05 and α=.01)	Sample Size Required (n)	Notes
m is <u>smaller</u> than m_0	1.6.2.3.1	σ unknown; s = estimate of σ from sample	$(m_0-\bar{x}) > u$	See Figs 1.3.3 and 1.3.4*	Figs 1.3.3 and 1.3.4 can be read in reverse*	$u=t_{1-\alpha}\left(\frac{s}{\sqrt{n}}\right)$
	1.6.2.3.2	σ known	$(m_0-\bar{x}) > u$	See Figs 1.3.7 and 1.3.8	For specified α and β read Figs 1.3.7 and 1.3.8 in reverse. For more accurate n, see 1.6.2.3.2	$u=z_{1-\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)$

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

Problem 1.6.2.1.1 - Does the average of the new product differ from the standard (σ unknown)?

Procedure	Example (for problem see p.)
i) Choose α, the significance level of the test.	i) Choose $\alpha = .05$, for example
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) $\bar{x} = .710$ lbs. $s = .0504$ lbs.
iv) Compute $u = t_{1-\alpha/2} \frac{s}{\sqrt{n}}$	iv) $u = \frac{2.09 \times .0504}{\sqrt{20}}$ $u = .0236$
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) $\lvert\bar{x}-m\rvert = \lvert .710 - .735\rvert = .025$ We conclude the average amount of powder in present stocks differs from 0.735 (the specified amount).

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

Problem 1.6.2.1.1 - Operating Characteristics and Determination of Sample Size for Unknown σ.

Operating Characteristics of the Test - Figures 1.3.1 and 1.3.2 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 σ, in order to use the OC curve, we must have some value for σ. (See section 1.2.3). If we use too large a value for σ, the effect is to lower our estimates of $\beta(\delta)$.

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

Selection of Sample Size, n - If we specify α, our significance level, and β, the risk we are willing to take of not detecting a difference of size $|m - m_0|$, then we can use the above OC curves in reverse to read off n, the required sample size.

A more accurate value for n may be obtained from the following formula:

$$n = \frac{b + \sqrt{b^2 - 4a}}{2a}$$

where $a = \left[\frac{\Delta}{z_{1-\alpha/2} + z_{1-\beta}}\right]^2$, $b = 1+a \left(1 + \frac{z^2_{1-\alpha/2}}{2}\right)$

For example, suppose that we wished to specify $a = .05$, and $\beta = .50$ for a difference of .030 lbs. - that is, we wish to conduct a test with a significance level of .05, and one which would have a 50-50 chance of detecting a difference of 0.030 lbs. What sample size should we require? Suppose it is thought from previous experience that σ lies between .04 and .06 lbs.

Taking $\sigma = .04$, with $|m - m_0| = .030$, gives $\Delta = .75$. Since $\alpha = .05$ and $\beta = .50$, $z_{1-\alpha/2} = 1.96$, $z_{.50} = 0$. Hence, $a = .1465$, $b = 1.4278$ and $n = 11(10.4)$. Taking $\sigma = .06$, we find $\Delta = .50$, $a = .06508$, $b = 1.1900$ and $n = 18(17.4)$. To be safe, we would use $n = 18$; and for $\sigma \leq .06$, with a significance level of .05, this would give a 50 percent chance of detecting a difference of 0.030 lbs.

Problem 1.6.2.1.2 - Does the average of the new product differ from the standard (σ known)?

Procedure	Example
i) Choose α, the significance level of the test.	i) Choose $\alpha = .05$, for example
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) $\bar{x} = .710$ lbs σ known to be equal to .06 lbs.
iv) Compute $u = z_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$	iv) $u = \dfrac{1.96(.06)}{20} = \dfrac{.1176}{4.47} =$ $u = .0263$
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) $\lvert\bar{x} - m\rvert = \lvert .710 - .735\rvert = .025$ 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 $\bar{x} \pm w$ is a $(1-\alpha)$ confidence interval estimate of the true average of the new type	vi) Note that $(.710 \pm .0263)$ is a 95 percent confidence interval for the true average of the new type.

Problem 1.6.2.1.2 - Operating Characteristics and Determination of Sample Size for Known σ.

Operating Characteristics of the Test - Figures 1.3.5 and 1.3.6 give the operating characteristics of the above test for $\alpha = .05$ and $\alpha = .01$ respectively. For any given n and $\Delta = \frac{|m - m_0|}{\sigma}$, the value of β, 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 α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of $|m - m_0|$, then we can use the above OC curves in reverse to read off n, the required sample size.

A more accurate method for obtaining n is from the formula

$$n = \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\Delta} \right)^2$$

Problem 1.6.2.2.1 - Does the average of the new product exceed the standard (σ unknown)?

Procedure	Example
i) Choose α, the significance level of the test.	i) Choose $\alpha = .05$, for example.
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) $\bar{x} = .710$ lbs. $s = .0504$ lbs.
iv) Compute: $u = t_{1-\alpha} \frac{s}{\sqrt{n}}$	iv) $u = \frac{1.73(.0504)}{\sqrt{20}} = \frac{.0872}{4.47}$ $u = .0195$
v) If $(\bar{x} - m_0) > + u$, decide that the average of the new type exceeds that of the standard. (Otherwise there is no reason to believe that they differ.)	v) $(\bar{x} - m_0) = (.710-.735) = -.025$. We conclude there is no reason to believe the new product differs from the standard.

Procedure	Example
vi) Note that the open interval from ($\overline{x}$ - w) to infinity is a one-sided confidence interval for the true mean .	vi) Note that the open interval from .690 to ∞ is a 95 percent one-sided confidence interval for true average of new product.

Problem 1.6.2.3.1 - Is the average of the new product less than the standard (σ unknown)?

	Procedure		Example
i-iv)	Complete Steps i thru iv as in Problem 1.6.2.2.1	i-iv)	u = .0195
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 they differ).	v)	.735 - .710 = .025 We conclude that the average of the new type is less than that of the standard.
vi)	Note that the open interval from $-\infty$ to $(\bar{x} + w)$ 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 .730 is a 95 percent one-sided confidence interval for the true mean of the new type.

Problem 1.6.2.2.1 - Operating Characteristic Curves and Determination of Sample Size for unknown σ.

Operating Characteristics of the Test - Figures 1.3.3 and 1.3.4 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 σ, in order to use the OC curve, we must have some value for σ. (See Section 1.2.3). If we use too large a value for σ, 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 $\Delta = \frac{m-m_0}{\sigma}$, we can read β, the probability of failing to detect such a difference.

Selection of Sample Size n - If we specify α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of $(m - m_0)$, then we can use the above OC curves in reverse to read off n, the required sample size.

A more accurate method for obtaining n is from the following formula:

$$n = \frac{b + \sqrt{b^2 - 4a}}{2a}$$

where

$$a = \left[\frac{\Delta}{z_{1-\alpha} + z_{1-\beta}}\right]^2, \quad b = 1 + a\left(1 + \frac{z^2_{1-\alpha}}{2}\right),$$

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

Problem 1.6.2.3.1 - Operating Characteristic Curves and Determination of sample size for Unknown σ.

Operating Characteristics of the Test - Same as Problem 1.6.2.2.1, except replace $(m - m_0)$ by $(m_0 - m)$.

Selection of Sample Size n - Same as Problem 1.6.2.2.1 except replace $(m - m_0)$ by $(m_0 - m)$.

Problem 1.6.2.2.2 - Does the average of the new product exceed the standard (σ known)?

Test Procedure

Procedure	Example
i) Choose α, the significance level of the test.	i) Choose $\alpha = .05$, for example.
ii) Look up $z_{1-\alpha}$ in Table I.	ii) $z_{.95} = 1.64$
iii) Compute: $\bar{x}$, the sample mean	iii) $\bar{x} = .710$ lbs. (σ known to be equal to .06 lbs).
iv) Compute: $u = z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$	iv) $u = \frac{1.64(.06)}{\sqrt{20}} = \frac{.0984}{4.47}$ $u = .022$
v) If $(\bar{x} - m_0) > u$, decide that the average performance 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 they differ.

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 ∞ is a 95 percent one-sided confidence interval for the true mean of the new product.

Problem 1.6.2.3.2 - Is the average of the new product less than that of the standard (σ known)?

Test Procedure

i-iv) Complete steps i - iv as in Problem 1.6.2.2.2	i-iv) $u = .022$
v) If $(m_0 - \bar{x}) > u$, decide 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.
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 percent one-sided confidence interval for the true mean of the new type.

Problem 1.6.2.2.2 - Operating Characteristics and Determination of Sample Size for Known σ.

Operating Characteristics of the Test - Figures 1.3.7 and 1.3.8 give the operating characteristics of the above test for $\alpha = .05$ and $\alpha = .0;$ respectively. For any given n and $\Delta = \frac{m - m_0}{\sigma}$, the value of β, 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 α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of $(m - m_0)$, then we can use the above OC curves in reverse to read of n, the required sample size.

An exact method for obtaining n is from the formula

$$n = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\Delta} \right)^2$$

Problem 1.6.2.3.2 - Operating Characteristic Curves and Determination of Sample Size for Known σ.

Operating Characteristics of the Test and Selection of Sample Size n. Proceed as in Problem 1.6.2.2.2 using $(m_0 - m)$ instead of $(m - m_0)$.

1.6.3 - Comparison of the Averages of two Given Materials, Products or Processes - Single Measured Characteristic.

We shall consider two problems.

Problem 1.6.3.1 - We wish to test whether the averages of two materials, products or processes differ, and we are not particularly concerned which is larger.

Problem 1.6.3.2 - 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 σ_A, or σ_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 (1.6.3.1 and 1.6.3.2) for the following situations:

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 σ_A and σ_B respectively.

Problem to be Illustrated

(Illustrative problem to be added later)

Table 1.6.3 - Summary Table for Problems of 1.6.3 - Comparison of Average Performance of Two Products

(For details and worked examples, see 1.6.3.1 or 1.6.3.2)

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	1.6.3.1.1	$\sigma_A \approx \sigma_B$; 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 α=.05 and α=.01 see Figs 1.3.1 and 1.3.2 ** and section 1.6.3.1.1	If we wish to have $n_A = n_B$, we can get a good approximation from sec. 1.6.3.1.1	$s_P = \sqrt{\frac{(n_A-1)s^2_A + (n_B-1)s^2_B}{n_A + n_B - 2}}$
	1.6.3.1.2	$\sigma_A \neq \sigma_B$; both unknown	$\lvert\bar{x}_A - \bar{x}_B\rvert > u$, where $u = t^*_{1-\alpha/2}\sqrt{\frac{s^2_A}{n_A} + \frac{s^2_B}{n_B}}$ See notes			$t^*_{1-\alpha/2}$ is t for f' degrees of freedom; formula for f' given in section 1.6.3.1.2
	1.6.3.1.3	σ_A, σ_B; both known	$\lvert\bar{x}_A - \bar{x}_B\rvert > u$, where $u = z_{1-\alpha/2}\sqrt{\frac{\sigma^2_A}{n_A} + \frac{\sigma^2_B}{n_B}}$	For α=.05 and α=.01, see Figs 1.3.5 and 1.3.6	See section 1.6.3.1.3	
m_A is greater than m_B	1.6.3.2.1	$\sigma_A \approx \sigma_B$; 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 α=.05 and α=.01 see Figs 1.3.3 and 1.3.4 ** and section 1.6.3.2.1	If we wish to have $n_A = n_B$, we can get a good approximation from sec. 1.6.3.2.1	$s_P = \sqrt{\frac{(n_A-1)s^2_A + (n_B-1)s^2_B}{n_A + n_B - 2}}$
	1.6.3.2.2	$\sigma_A \neq \sigma_B$; both unknown	$(\bar{x}_A - \bar{x}_B) > u$, where $u = t^*_{1-\alpha}\sqrt{\frac{s^2_A}{n_A} + \frac{s^2_B}{n_B}}$ See notes			$t^*_{1-\alpha}$ is t for f' degrees of freedom; formula for f' given in section 1.6.3.2.2
	1.6.3.2.3	σ_A, σ_B; both known	$(\bar{x}_A - \bar{x}_B) > u$, where $u = z_{1-\alpha}\sqrt{\frac{\sigma^2_A}{n_A} + \frac{\sigma^2_B}{n_B}}$	For α=.05 and α=.01 see Figs 1.3.7 and 1.3.8 and section 1.6.3.2.3	For $n_A = n_B$ or $n_A = Cn_B$, see formulas of 1.6.3.2.3	

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

Problem 1.6.3.1 - Do the products A and B differ in average performance? (No particular concern over which is larger).

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

Test Procedure	Example
i) Choose α, 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 $s_P = \sqrt{\frac{(n_A-1)s_A^2+(n_B-1)s_B^2}{n_A + n_B - 2}}$	
v) Compute $u = t_{1-\alpha/2}\, s_P \sqrt{\frac{n_A + n_B}{n_A n_B}}$	

Procedure	Example
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 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 $(\bar{x}_A - \bar{x}_B) \pm u$ is a $1-\alpha$ confidence interval estimate of $(m_A - m_B)$.	

Problem 1.6.3.1 Case I - Operating Characteristic Curves and Determination of Sample Size.

Operating Characteristics of the Test - Figures 1.3.1 and 1.3.2 give the operating characteristic (OC) curves of the above test for $\alpha = .05$ and $\alpha = .01$ respectively, for various values of $n = n_A + n_B - 1$. Although we have assumed we do not know the standard deviation of the performances of A and B, in order to use the OC curve, we would have to have a value for $\sigma_A = \sigma_B = \sigma$, the common standard deviation of the performance of A and B. This may be possible since we often know the range in which σ lies. (See Section 1.2.3). If we use too large a value for σ, the effect is to make our estimates more conservative.

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

$$\Delta' = \frac{m_A - m_B}{\sigma} \frac{1}{\sqrt{n_A + n_B - 1}} \sqrt{\frac{n_A n_B}{n_A + n_B}}$$

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

Problem 1.6.3.1 Case I -

Selection of Sample Size n.

If we specify α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of size $\pm (m_A - m_B)$, and if we wish to put $n_A = n_B$, we can get a good approximation to $n_A = n_B$ from the following formula:

$$n_A = n_B = \frac{b + \sqrt{b^2 - 8a}}{2a}$$

where

$$a = \left(\frac{\Delta}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2,$$

$$b = 2 + a\left(1 + \frac{z^2_{1-\alpha/2}}{4} \right),$$

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma}$$

Problem 1.6.3.1 - Do A and B differ in average performance?

1.6.3.1.2 Case II - Variability of A and B unknown, cannot be assumed equal.

Test Procedure	Example
i) Choose α, the significance level of the test. (Actually the procedure outlined will give a significance level of only approximately α).	
ii) 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.	
iii) Compute $$f = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{s_B}{n_B}\right)^2}{n_B - 1}} - 2$$	

Procedure	Example
iv) Look up $t^*_{1-\alpha/2}$ for f^1 degrees of freedom in Table II, where f^1 is the integer nearest to f.	
v) Compute $u = t^*_{1-\alpha/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$	
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.	
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	

Problem 1.6.3.1 - Do products A and B differ in average performance?

1.6.3.1.3 Case III - The variability in performance of each of A and B is known from previous experience, and the standard deviations are σ_A and σ_B respectively.

Test Procedure	Example
i) Choose α, the level of significance of the test.	
ii) Look up $z_{1-\alpha/2}$ in Table I.	
iii) Compute $\bar{x}_A$ and $\bar{x}_B$, the sample means of the n_A and n_B from 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 $\lvert\bar{x}_A - \bar{x}_B\rvert > u$, decide that A and B differ with regard to their average performance	

<u>Procedure</u>

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 $(\bar{x}_A - \bar{x}_B) \pm u$ is a $1-\alpha$ confidence interval estimate of $(m_A - m_B)$.

<u>Example</u>

Problem 1.6.3.1.3 Case III

Operating Characteristics of the Test - Figures 1.3.5 and 1.3.6 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 β, the probability of failing to detect a difference of $(m_A - m_B)$.

If $n_B = n_A / c$, we can put $d^* = \frac{(m_A - m_B)}{\sqrt{\sigma_A^2 + c\sigma_B^2}}$ and again

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

Selection of Sample Size n - If we specify α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of $\pm (m_A - m_B)$, then we can use the above mentioned OC curves in reverse to obtain the proper sample size.

For the sample sizes to be equal, a good approximation to $n_A = n_B$ can be obtained from the following formula:

$$n_A = n_B = \left(\frac{z_{1-\alpha/2} - z_\beta}{d} \right)^2$$

For the sample sizes to be proportional, ($n_A = cn_B$), a good approximation can be obtained from the following formula:

$$n_A = cn_B = \left(\frac{z_{1-\alpha/2} - z_\beta}{d^*} \right)^2$$

The values for $z_{1-\alpha/2}$ and z_β can be obtained from Table I.

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

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

Test Procedure

Procedure	Example
(i) Choose α, 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: $s_p = \sqrt{\frac{(n_A-1)s^2_A+(n_B-1)s^2_B}{n_A + n_B - 2}}$	
(v) Compute: $u = t_{1-\alpha}\, s_p \sqrt{\frac{n_A + n_B}{n_A n_B}}$	

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 A and B differ in average performance.	
(vii) Let m_A and m_B be the true averages of A and B. Note that the interval from $\{(x_A - x_B) - u\}$ to ∞ is a $1-\alpha$ one-sided confidence interval estimate of the true difference (m_A-m_B).	

1.6.3.2.1 Case I - Operating Characteristics and Determination of Sample Size

Operating Characteristics of the Test - Figs. 1.3.3 and 1.3.4 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 σ lies (See Section 1.2.3). If we use too large a value for σ, the effect is to make our estimate more conservative.

Determination of sample size n

If we wish to have $n_A = n_B$, the following formula gives a good approximation:

$$n_A = n_B = \frac{b + \sqrt{b^2 - 8a}}{2a}$$

where

$$a = \left(\frac{\Delta}{z_{1-\alpha} + z_{1-B}} \right)^2$$

$$b = 2 + a \left(1 + \frac{z^2_{1-\alpha}}{4} \right)$$

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

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

1.6.3.2.2 Case II - Variability of A and B unknown, cannot be assumed equal.

Test Procedure

Procedure	Example
(i) Choose α, the significance level of the test.	
(ii) 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.	
(iii) Compute: $f = \dfrac{\left(\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}\right)^2}{\dfrac{\left(\dfrac{s_A}{n_A}\right)^2}{n_A - 1} + \dfrac{\left(\dfrac{s_B}{n_B}\right)^2}{n_B - 1}} - 2$	
(iv) Look up $t^*_{1-\alpha}$ for f' degrees of freedom in Table II, where f' is the integer nearest to f.	

Procedure	Example
(v) Compute: $$u = t^{*}_{1-\alpha}\left(\frac{s^2_A}{n_A} + \frac{s^2_B}{n_B}\right)$$	
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 A and B differ.	
vii) Let m_A and m_B be the true averages of A and B. Note that the interval from $\{(\bar{x}_A - \bar{x}_B) - u\}$ to ∞ is a $1-\alpha$ one-sided confidence interval estimate of the true difference $(m_A - m_B)$.	

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

1.6.3.2.3 Case III - The variability in performance of each of A and B is known from previous experience and the standard deviations are σ_A and σ_B respectively.

Procedure	Example
(i) Choose α, 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: $u = z_{1-\alpha}\sqrt{\frac{\sigma^2_A}{n_A} + \frac{\sigma^2_B}{n_B}}$	
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 they differ.	

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

Problem 1.6.3.2.3 - Operating Characteristics and Determination of Sample Size

Operating Characteristics of the Test - Figs. 1.3.7 and 1.3.8 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

$$d = \frac{(m_A - m_B)}{\sqrt{\sigma^2_A + \sigma^2_B}}$$

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

If $n_A = Cn_B$, we can put $d^* = \dfrac{(m_A - m_B)}{\sqrt{\sigma^2_A + C\sigma^2_B}}$

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

Selection of Sample Size n.

If we specify α, our significance level, and β, the probability or risk we are willing to take of not detecting a difference of $(m_A - m_B)$, then we can use the above mentioned OC curves in reverse to obtain the proper sample size.

A more accurate value of n_A can be obtained from the following formulas:

If we wish to have $n_A = n_B$,

$$n_A = n_B = \left(\frac{z_{1-\alpha} - z_\beta}{d} \right)^2$$

If we wish to have $n_A = Cn_B$,

$$n_A = Cn_B = \left(\frac{z_{1-\alpha} - z_\beta}{d*} \right)^2$$

1.6.4 Comparison of the Averages of Several Products - Single Measured Characteristic.

We wish to test whether there is a difference in the average performance of several products (k in number). The technique outlined below was developed for the case where we have an equal number of observations on each product (i.e., $n_1 = n_2 = \dots n_k = n$). When the n's are not all equal, but are not very different, the procedure may still be employed, using their harmonic mean

$$\bar{n}_H = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_k}}$$

in the formula of step vi below.

Procedure	Example
i) Choose α, 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, ..., s_k^2, the sample variance for each of the k products.	

Procedure	Example
iii) Compute: $s_e^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+\ldots+(n_k-1)s_k^2}{(n_1+n_2+\ldots+n_k)-k}$ and $s_e = \sqrt{s_e^2}$	
iv) Compute w, the largest difference between any two sample means.	
v) Look up $q^*_{1-\alpha}$ in Table IV entering the table with k and ν where $\nu = [n_1 + n_2 + \ldots + n_k] - k$.	
vi) Compute: $q = \frac{w\sqrt{n}}{s_e}$ if $n_1 = n_2 = \ldots = n_k = n$ or $q = \frac{w\sqrt{\bar{n}_H}}{s_e}$ if $n_1 \simeq n_2 \simeq \ldots \simeq n_k$ where $\bar{n}_H = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}+\ldots+\frac{1}{n_k}}$	

Procedure	Example
vii) If $q > q^*_{1-\alpha}(k,\nu)$, decide that the averages differ. If $q \leq q^*_{1-\alpha}(k,\nu)$, decide that we have no reason to believe that the averages differ.	

It is worth-while noting that we can simultaneously make confidence interval estimates for each of the $\frac{k(k-1)}{2}$ pairs of differences between product averages, with a confidence of $1-\alpha$ that all of the estimates are correct (if $n_1 = n_2 = \dots = n_k$). The confidence intervals are

$$(\bar{x}_i - \bar{x}_j) \pm q_{1-\alpha}(k,\nu) \quad s_e\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

, where $\bar{x}_i$, $\bar{x}_j$, are sample means of the i^{th} and j^{th} products.

1.6.5 Comparison of a Given Product with a Standard with Respect to Average Values of Two Possibly Correlated Characteristics.

<u>Problem</u> - The average values of 2 characteristics of a standard product are known to be equal to m_1 and m_2 (e.g. for a certain standard alloy, m_1 = average tensile strength, m_2 = average hardness).

We wish to determine whether a new product differs from the standard in average performance with respect to either or both of these characteristics. (NOTE: The method can be generalized to any number of characteristics, but the amount of computation required increases rapidly. For the general method see Ref. [ ]- Hicks or Ref. [ ] - Hotelling).

1.6.5.1 Case I.

The variabilities of items with respect to each characteristic, i.e., σ_1 and σ_2, are unknown; the relationship (correlation) between the two measurement characteristics is unknown.

Procedure

i) Record the observations in a table as follows:

Item No.	Characteristic No. 1	Characteristic No. 2
1	X_{11}	X_{21}
2	X_{12}	X_{22}
⋮	⋮	⋮
n	X_{1n}	X_{2n}

---

ii) Then compute:

Sum	$\Sigma X_1 =$	$\Sigma X_2 =$
Sum of Squares	$\Sigma X^2_1 =$	$\Sigma X^2_2 =$
Sum of cross Products	$\Sigma X_1 X_2 =$	

Procedure (Continued)

iii) Compute the correlation coefficient

$$r = \frac{(\Sigma x_1 x_2)}{\sqrt{(\Sigma x^2_1)(\Sigma x^2_2)}}$$

where

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - \frac{(\Sigma X_1)(\Sigma X_2)}{n}$$

$$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n}$$

$$\Sigma x_2^2 = \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n}$$

---

iv) Compute $\bar{X}_1$, $\bar{X}_2$, s_1^2, and s_2^2, the means and variances of the n observations for each characteristic.

---

v) Compute:

$$d_1^2 = (\bar{X}_1 - m_1)^2$$

$$d_2^2 = (\bar{X}_2 - m_2)^2$$

m_1 and m_2 are the known means of the standard.

---

vi) Compute:

$$T^2 = \frac{n}{1-r^2}\left[\frac{d_1^2}{s_1^2} - \frac{2rd_1d_2}{s_1s_2} + \frac{d_2^2}{s_2^2}\right]$$

---

vii) Compute $F = \left[\frac{n-2}{2(n-1)}\right] T^2$

---

viii) Using chosen α (the significance level of the test) look up $F^*_{1-\alpha}$ for $(2,n-2)$ degrees of freedom in Table III.

---

ix) If $F > F^*$, decide that the new product differs from the standard in average performance. (Otherwise, there is no reason to believe they differ.)

What can be said about the location of the true averages of the new type? We can assert, with $100(1-\alpha)$ percent confidence that the true averages μ_1 and μ_2 lie inside the ellipse with center at $(\bar{X}_1, \bar{X}_2)$, and axes and angle of rotation determined by s_1^2, s_2^2, r and α, that is obtained by setting the above expression for T^2 equal to $\left[\frac{2(n-1)}{n-2}\right] F^*_{1-\alpha}$ for $(2,n-2)$ degrees of freedom. (In a long series of samples each containing n observations from the same population, we could calculate a corresponding series of ellipses which would have varying centers, axes, and angles of rotation, by virtue of the varying values of $\bar{X}_1$, $\bar{X}_2$, s_1^2, s_2^2, and r. We would expect $100(1-\alpha)$ percent of these ellipses to include the true point (μ_1, μ_2)).

Note also that we might use the T^2 procedure to test whether $m_1 = m_2$. In this case, the T^2 test becomes

$$T^2 = \frac{n(\overline{X}_1 - \overline{X}_2)^2}{(s_1^2 + s_2^2 - 2rs_1s_2)}$$

which is a special case of the ordinary two-sided test of difference between averages (unknown σ) - i.e., the t test. (See section 1.6.3.1.1).

1.6.5.2 Case II

The variabilities of items with respect to the 2 characteristics are known to be equal to σ_1 and σ_2; there is a known relationship between measurements of the 2 characteristics (i.e. their correlation coefficient = ρ).

Procedure

i) Record the observations in a table as follows:

Item No.	Characteristic No. 1	2
1	X_{11}	X_{21}
2	X_{12}	X_{22}
⋮	⋮	⋮
n	X_{1n}	X_{2n}

---

ii) Compute $\bar{X}_1$ and $\bar{X}_2$, the means of the n observations for each characteristic.

---

iii) Compute $d_1 = \bar{X}_1 - m_1$ m_1 and m_2 are the known means of the standard.

$d_2 = \bar{X}_2 - m_2$

---

iv) Compute:

$$T^2 = \frac{n}{1-\rho^2}\left[\frac{d_1^{\,2}}{\sigma^2} - \frac{2\rho\, d_1 d_2}{\sigma_1\,\sigma_2} + \frac{d_2^{\,2}}{\sigma_2^{\,2}}\right]$$

---

v) Using the chosen α (the significance level of the test,) look up $\chi^2_{1-\alpha}$ for 2 degrees of freedom in Table V.

---

vi) If $T^2 > \chi^2_{1-\alpha}$, decide that the new product differs from the standard in average performance. Otherwise, there is no reason to believe that they differ.

What can be said about the location of the true averages of the new type? We can assert with $100(1-\alpha)$ percent confidence that the true averages μ_1 and μ_2 lie inside the ellipse with center at $(\bar{X}_1, \bar{X}_2)$, and axes and angles of rotation determined by $\sigma_1^{\,2}$, $\sigma_2^{\,2}$, ρ and α, that is obtained by setting the above expression for T^2 equal to $\chi^2_{1-\alpha}$ for 2 degrees of freedom. (In a long series of samples each containing n observations

from the same population, we could calculate a corresponding series of ellipses, which would have varying centers. We would expect $100(1-\alpha)$ percent of these ellipses to include the true point (μ_1, μ_2).

Note also that we might use the present T^2 procedure to test whether $m_1 = m_2$. In this case the T^2 test becomes

$$T^2 = \frac{n(\overline{X}_1 - \overline{X}_2)^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} ,$$

which is a special case of the ordinary two-sided test of difference between averages (known σ). See section 1.6.3.1.3.

1.6.6 Comparison of Two Products with Respect to Average Values of Two Possibly Correlated Characteristics.

Problem - We take a sample of n_1 items of product A, and n_2 items of product B, and measure each item with regard to 2 different properties (e.g. tensile strength and hardness). We wish to determine whether product A differs from product B in average performance with respect to either or both of these characteristics. (NOTE: the method can be generalized to any number of characteristics, but the amount of computation required increases rapidly. For the general method, see Ref. [ ] - Hicks or Ref. [ ] - Hotelling).

1.6.6.1 Case I

The variabilities of items of product A and product B with respect to characteristic No. 1 are unknown, but are assumed to be the same for product A and product B. i.e., for characteristic No. 1 $\sigma_1(A) = \sigma_1(B) = \sigma_1$ (unknown); and, similarly, for characteristic No. 2, $\sigma_2(A) = \sigma_2(B) = \sigma_2$ (unknown). The correlations between the two measurement characteristics among items of product A and among items of product B are unknown, but such correlations as may exist are assumed to be the same for product A and for product B, i.e., $\rho_{12}(A) = \rho_{12}(B)$.

Procedure

i) Record the observations in a table as follows:

Item No.	Product A Characteristic 1	Product A Characteristic 2	Product B Characteristic 1	Product B Characteristic 2
1	X_{11}	Y_{11}	X_{21}	Y_{21}
2	X_{12}	Y_{12}	X_{22}	Y_{22}
⋮	⋮	⋮	⋮	⋮
n_1	X_{1n}	Y_{1n}	⋮	⋮
⋮			⋮	⋮
n_2			X_{2n}	Y_{2n}

ii) Then compute:

Sum	$\sum_j X_{1j} =$ $\Sigma Y_{1j} =$	$\Sigma X_{2j} =$ $\Sigma Y_{2j} =$
Sum of Squares	$\Sigma X_{1j}^2 =$ $\Sigma Y_{1j}^2 =$	$\Sigma X_{2j}^2 =$ $\Sigma Y_{2j}^2 =$
Sum of Cross Products	$\sum_j X_{1j} Y_{1j}$	$\sum_j X_{2j} Y_{2j}$

---

iii) Compute $\bar{X}_1, \bar{Y}_1; \bar{X}_2, \bar{Y}_2; \; s^2_{X_1}, s^2_{Y_1}; s^2_{X_2}, s^2_{Y_2}$,

the means and variances of the columns of observations.

Compute $d_1 = \bar{X}_1 - \bar{X}_2$; $d_2 = \bar{Y}_1 - \bar{Y}_2$.

---

iv) Compute:

$$s_1^2 = \frac{(n_1-1)s^2_{X1} + (n_2-1)s^2_{X2}}{n_1 + n_2 - 2} , \quad s_1 =$$

$$s_2^2 = \frac{(n_1-1)s^2_{Y1} + (n_2-1)s^2_{Y2}}{n_1 + n_2 - 2} , \quad s_2 =$$

---

v) Compute:

$$s_{X1Y1} = \frac{\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{n_1}}{n_1 - 1}$$

$$s_{X2Y2} = \frac{\Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{n_2}}{n_2 - 1}$$

$$s_{12} = \frac{(n_1-1)s_{X1Y1} + (n_2-1)s_{X2Y2}}{n_1 + n_2 - 2}$$

$$r = \frac{s_{12}}{s_1 \; s_2}$$

---

vi) Compute:

$$T^2 = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)}{1-r^2} \left[\frac{d_1^2}{s_1^2} - \frac{2 r d_1 d_2}{s_1 \; s_2} + \frac{d_2^2}{s_2^2} \right]$$

---

vii) Compute:

$$F = \left[\frac{n_1 + n_2 - 3}{2(n_1 + n_2 - 2)} \right] T^2$$

---

viii) Using the chosen α (significance level of the test), look up $F^*_{1-\alpha}$ for $(2, n_1 + n_2 - 3)$ degrees of freedom in Table III.

---

ix) If $F > F^*$, decide that the overall average performance of products X and Y is different. Otherwise, there is no reason to believe that they differ.

NOTE: The test is a generalization of the t test given in section 1.6.3.1.1. If we applied a t-test to each of the characteristics separately, we obtain:

$$t_1 = \frac{\bar{X}_1 - \bar{X}_2}{s_1\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{d_1}{s_{d_1}}$$

$$t_2 = \frac{\bar{Y}_1 - \bar{Y}_2}{s_2\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{d_2}{s_{d_2}}$$

Therefore, in terms of t we may write

$$T^2 = \frac{1}{1-r^2}\left(t_1^{\,2} - 2rt_1t_2 + t_2^{\,2}\right)$$

1.6.6.2 Case II

The variabilities of items with respect to the 2 characteristics are known to be equal to σ_1 and σ_2 for both products A and product B. There is a known relationship between measurements of the 2 characteristics - i.e., their correlation coefficient is $= \rho 12$ (again true for both product A and product B).

Procedure

i) Record the observations in a table as follows:

Item No.	PRODUCT A Characteristic 1	PRODUCT A Characteristic 2	PRODUCT B Characteristic 1	PRODUCT B Characteristic 2
1	X_{11}	Y_{11}	X_{21}	Y_{21}
2	X_{12}	Y_{12}	X_{22}	Y_{22}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n_1	X_{1n}	Y_{1n}	.	.
.			.	.
.			.	.
n_2			X_{2n}	Y_{2n}

ii) Compute $\bar{X}_1$, $\bar{Y}_1$; $\bar{X}_2$, $\bar{Y}_2$

---

iii) Compute: $d_1 = \bar{X}_1 - \bar{X}_2$; $d_2 = \bar{Y}_1 - \bar{Y}_2$

---

iv) Compute:

$$T^2 = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)}{1-\rho^2}\left[\frac{d_1^{\,2}}{\sigma_1^{\,2}} - \frac{2\rho d_1 d_2}{\sigma_1 \sigma_2} + \frac{d_2^{\,2}}{\sigma_2^{\,2}}\right]$$

---

v) Using the chosen α (significance level of the test), look up $\chi^2_{1-\alpha}$ for 2 degrees of freedom in Table V.

---

vi) If $T^2 > \chi^2_{1-\alpha}$, decide that the overall average performance of products X and Y is different. Otherwise, there is no reason to believe that they differ.

1.6.7 Comparison of the Variability in Performance of a Given Material, Product or Process with that of a Standard - Single Measured Characteristic.

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

1.6.7.1 - Does the variability of the product <u>differ</u> from that of the standard?

1.6.7.2 - Does the variability of the product <u>exceed</u> that of the standard?

1.6.7.3 - Is the variability of the product <u>less</u> 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.

If the departure from normality is moderately large the operating characteristics of the test may be seriously

altered. In cases where serious non-normality is suspected, the methods of Section ___ are recommended.

Problem 1.6.7.1 - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be σ. 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 α, the level of significance of the test.	
ii) Look up $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$, both for $n-1$ degrees of freedom, in Table V.	
iii) Compute s^2, the sample variance of the n observations.	

Procedure	Example
iv) Compute: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	
v) If the computed χ^2 exceeds the tabular $\chi^2_{1-\alpha/2}$, or is less than $\chi^2_{\alpha/2}$, 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 $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\, n-1}}$ to $\frac{(n-1)s^2}{\chi^2_{\alpha/2,\, n-1}}$ is a 100 (1-α) percent confidence interval estimate of σ^2, the variance of the new product. (see section 1.2.2).

Problem 1.6.7.2 - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be σ. 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 α, the level of significance of the test.	
ii) Look up $\chi^2_{1-\alpha}$ for n-1 degrees of freedom in Table V.	
iii) Compute s^2, the sample variance of the n observations.	

Procedure	Example
iv) Compute $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$.	
v) If the computed χ^2 exceeds $\chi^2_{1-\alpha}$, 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 $\frac{(n-1)s^2}{\chi^2_{1-\alpha,\ n-1}}$ is a 100 (1-α) percent confidence interval estimate of σ^2, the variance of the new product. (See section 1.2.2).

Problem 1.6.7.3 - The variability in the performance of a standard material, product or process as measured by its standard deviation is known to be σ. 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 less than 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 α, the level of significance of the test.	
ii) Look up χ^2_α for n-1 degrees of freedom in Table V.	
iii) Compute s^2, the sample variance of the n observations.	
iv) Compute $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	

Procedure	Example
v) If the computed χ^2 is less than χ^2_α 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 $\frac{(n-1)s^2}{\chi^2_{\alpha, n-1}}$ is a $1-\alpha$ confidence interval estimate of σ^2, the variance of the new product. (See section 1.2.2)

1.6.8 Comparison of the Variability of Two Materials, Products or Processes.

We shall consider 2 problems:

1.6.8.1 - Does the variability of product A <u>differ</u> from that of product B? (We are not concerned which is larger).

1.6.8.2 - Does the variability of product A <u>exceed</u> 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. If the departure from normality is moderately large, the operating characteristics of the test may be seriously altered. In cases where serious non-normality is suspected, the methods of Section ___ are recommended.

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 1.6.8.1 - 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 α, 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.	

Procedure	Example
iv) Compute $F = s_A^2 / s_B^2$.	
v) If $F > F^*_{1-\alpha/2;n_A-1,n_B-1}$ or $F < \frac{1}{F^*_{1-\alpha/2;n_B-1,n_A-1}}$ 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

$$\frac{1}{F_{1-\alpha/2;\, n_B-1,\, n_A-1}}$$

and $F_{1-\alpha/2;\, n_A-1,\, n_B-1}$ is a 100 $(1-\alpha)$ percent confidence interval estimate of the ratio σ_A^2 / σ_B^2 (See section 1.2.2).

Problem 1.6.8.2 - 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 α, 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.	

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 front 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. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS