START

0009966

9 2 1 2 5 6 3 1 7 5 3

PROBABILITY AND STATISTICS FOR MODERN ENGINEERING

LAWRENCE L. LAPIN

San José State University

JUL 1990
RECEIVED
EDMC

Brooks/Cole Engineering Division
Monterey, California

Brooks/Cole Publishing Company
A Division of Wadsworth, Inc.

© 1983 by Wadsworth, Inc., Belmont, California 94002. All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher, Brooks/Cole Engineering Division, Monterey, California 93940, a division of Wadsworth, Inc.

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Library of Congress Cataloging in Publication Data:

Lapin, Lawrence L.
Probability and statistics for modern engineering.
Bibliography: p.
Includes index.
1. Engineering—Statistical methods.
2. Probabilities. I. Title.
TA340.L36 1983 519.2'02462 82-25286

ISBN 0-534-01460-7

Sponsoring Editor: Ray Kingman
Manuscript Editor: Lillian R. Rodberg
Production: Del Mar Associates
Interior Design: John Odam
Illustrations: Alice Harmon, Kim Fraley
Typesetting: Allservice Phototypesetting Company
Production Services Manager: Stacey C. Sawyer

CHAPTER NINE

HYPOTHESIS TESTING

LIKE ESTIMATION, **HYPOTHESIS TESTING** is one of the basic forms of statistical inference. This broad area of statistics usually culminates in an immediate decision, making it the more dynamic form of statistical inference. Testing is based on two complementary assumptions, or **hypotheses,** regarding unknown populations. Taking various forms, the most common hypotheses involve assumed levels for a population parameter. Later portions of this book describe other types of hypotheses that help to compare two or more populations or are concerned with properties such as the randomness of data, the manner in which two variables are related (if at all), or the shape of the population distribution.

This chapter begins with hypothesis tests for the unknown value of the population mean. These procedures are useful for evaluating decisions involving two complementary actions when the unknown level of μ is the pivotal factor. For example, the decision whether or not to repair equipment might depend on how much the mean reading deviates from some nominal figure. Statistical inferences are needed in such evaluations because μ can be evaluated only indirectly from sample data. For this purpose the sample mean serves as the **test statistic.** Thus, repairs might be requested only if the computed level of $\bar{X}$ exceeds a specified level.

Of course samples may give false readings of reality. The potential for such sampling error can be controlled but never eliminated. Sampling error is managed differently in hypothesis testing than in estimating, where the major concern is with precision and reliability. This control is achieved by establishing a **decision rule** that achieves some optimal balance among the probabilities for taking incorrect actions. For example, the incidence of unnecessary repairs must be kept small, while failure to make needed repairs must be avoided. Hypothesis testing procedures focus on such errors of commission and omission, which may be controlled by judicious determination of the decision rule or by selecting a sample large enough to keep the error probabilities acceptably small.

9-1 BASIC CONCEPTS OF HYPOTHESIS TESTING

To introduce hypothesis testing, we will consider a decision involving the population mean. The sample mean $\bar{X}$ will serve as the test statistic. The following example illustrates the underlying concepts of hypothesis testing. A general procedure that may be applied in most testing situations is introduced later.

ILLUSTRATION: Bacterial Leaching in Ore Processing

A mining engineer is studying ways to increase the production of metal from a large copper deposit. A substantial amount of copper-bearing ore is presently bypassed because its quality is too low for economical processing at current world prices. The engineer is interested in a new process based on bacterial leaching. This technique isolates key minerals in poor ores through a chemical reaction caused by strains of the bacterium *Thiobacillus*. These microbes convert the iron in the ore from a ferrous form into ferric iron that then acts in oxidizing the insoluble copper sulfides present in the ore's pyrites. The process culminates in collection of a liquid solution from which copper may be recovered.

The effectiveness of bacterial leaching depends on the particular strain of bacteria used and the composition of the ore. Since there is little history of successful application, a statistical sampling study will be performed to determine whether bacterial leaching should be adopted on a major scale. A sample of 100 loads of ore will be removed from random coordinates throughout the unworkable deposit site and processed by a promising strain of bacteria at a test leaching dump. The final yield, in pounds of copper per ton of ore, will then be determined for each sample load. These data will represent the yields obtainable from bacterial leaching of the more than 100 million tons of target ore.

The success of large-scale bacterial leaching will depend on the level of the mean copper yield μ that will ultimately be achieved from the entire deposit. On the basis of current prices and present processing costs, the break-even level has been established at 36 pounds of recoverable copper per ton of presently unworkable ore. This quantity is the pivotal level for the statistical investigation.

The Structure of a Hypothesis Test

Statistical testing involves two complementary hypotheses. Here these are defined in terms of levels for the unknown population mean μ. The engineer has established the following hypotheses:

Null Hypothesis (process is uneconomical)

$$H_0: \mu \leq 36 \text{ lb/ton}$$

Alternative Hypothesis (process is economical)

$$H_1: \mu > 36 \text{ lb/ton}$$

Although the adjective "null" has historically been applied to the hypothesis representing no change, this designation is arbitrary. Most often, *the null hypothesis will be the one for which erroneous rejection is the more serious consequence*. For notational convenience, the null hypothesis is represented by H_0, and the opposite or alternative hypothesis by H_1.

Hypothesis testing involves two complementary actions or choices. Our mining engineer must decide whether to adopt full-scale bacterial leaching or to abandon the process for now (and perhaps do further testing). These two actions may be expressed in terms of the foregoing hypotheses as

Accept H_0 (Abandon bacterial leaching for now)
Reject H_0 (Adopt bacterial leaching)

(Rejecting the null hypothesis is the same as accepting the alternative, and ac-

cepting H_0 is tantamount to rejecting H_1.) The engineer's choice will be based on the results of the sample investigation.

The decision must be made without knowing which of the two hypotheses is true. A determination must be made whether the sample evidence tends to favor the null hypothesis or refute it. This will be based on the following

TEST STATISTIC

Sample mean copper yield $\overline{X}$ (lb/ton)

A large level for $\overline{X}$ (say 50 lb/ton) will give support to the efficacy of bacterial leaching, whereas a small value (such as 10) will deny it. Some critical level must serve as the point of demarcation between adopting the process or abandoning it.

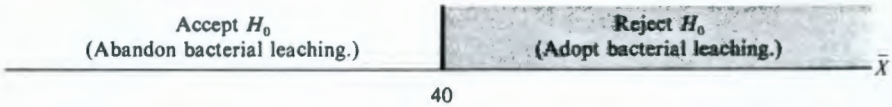
After the sample data are obtained, the engineer will compute the above and apply the following

DECISION RULE

Accept H_0 (abandon bacterial leaching for now) if $\overline{X} \leq 40$ lb/ton.

Reject H_0 (adopt bacterial leaching) if $\overline{X} > 40$.

The pivotal value of 40 lb/ton serves as the **critical value** of the test statistic. It is sometimes convenient to portray the decision rule in terms of **acceptance** and **rejection regions,** as in the following:



The critical value of $\overline{X}$ lies above the 36 lb/ton break-even level for μ that the engineer used to establish the hypotheses. The test statistic's critical value will ordinarily lie outside the range of the null hypothesis. This provides some leeway and reduces the chance that an atypical sample result will cause the wrong decision.

At this point it may be helpful to summarize the structure of the mining engineer's investigation. Table 9-1 arranges the problem in terms of a decision table. There is a column for each of the decision-maker's two choices or acts and a row for the two uncertain population events, representing the truth or falsity of the null hypothesis. Notice that each act-event combination culminates in a particular outcome. Two of these are correct decisions:

To accept H_0 when H_0 is true
(abandon an uneconomical process).

To reject H_0 when H_0 is false
(adopt an economical process).

Table 9-1 Decision Table for Bacterial Leaching Illustration

Events	Acts: Accept H_0 (Abandon Process) $\bar{X} \leq 40$	Acts: Reject H_0 (Adopt Process) $\bar{X} > 40$
H_0 True (Process Uneconomical) $\mu \leq 36 \quad (= \mu_0)$	**Correct Decision** (Abandon Uneconomical Process)	**Type I Error** (Adopt Uneconomical Process) Probability α
H_0 False (Process Economical) $\mu > 36 \quad (= \mu_0)$	**Type II Error** (Abandon Economical Process) Probability β	**Correct Decision** (Adopt Economical Process)

Two other outcomes result in **errors.** These are:

Type I Error: To reject H_0 when it is true (adopt an uneconomical process).

Type II Error: To accept H_0 when it is false (abandon an economical process).

Finding the Error Probabilities

A major concern in hypothesis testing is controlling the incidence of the two kinds of errors. The degree of such control may be summarized by the error probabilities:

$$\alpha = \Pr[\text{type I error}] = \Pr[\text{reject } H_0 \mid H_0 \text{ true}]$$

$$\beta = \Pr[\text{type II error}] = \Pr[\text{accept } H_0 \mid H_0 \text{ false}]$$

Although both errors are undesirable, rejecting a true null hypothesis is typically the more significant. For this reason the type I error probability α is also referred to as the **significance level.**

The sampling distribution of $\bar{X}$ may be assumed to be normally distributed with mean μ and standard error $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. Appropriate parameters must be established for the particular normal curve used in finding the error probabilities.

Using the decision rule established earlier for the bacterial leaching study, we may evaluate the error probabilities. Under the null hypothesis $\mu \leq 36$, so a range of values applies. Any value 36 or less is permitted under H_0, such as $\mu = 12.3$, $\mu = 27.5$, or $\mu = 34.99$. Probability evaluations are usually made for the worst case, when μ falls exactly at the H_0 limit, which is denoted by μ_0. In the present illustration $\mu_0 = 36$ lb/ton. In computing α it is assumed that $\mu = \mu_0$.

The mining engineer does not know the value for σ any more than he knows μ, but he was able to arrive at a "ballpark" guess. A yield almost certain to be exceeded, 1 pound per ton (which is what would be expected by just washing the ore with water and processing the rinse waste) was subtracted from the present yield of higher-quality, conventionally processed ores—150 lb/ton. A

range of 6 standard deviations was thereby established. Thus, for planning purposes he found

$$\sigma = \frac{99.9\text{th} - .1\text{st percentiles}}{6} = \frac{150 - 1}{6} = 24.83$$

which for convenience was rounded to 25 lb/ton. (Keep in mind that the above is simply the best guess available for σ, and any probability evaluations based on it may vary considerably from the truth. But *some* value must be used, and σ can never be precisely known anyway until the entire deposit has been processed by bacteria.) Using this quantity, we have $\sigma_{\bar{X}} = 25/\sqrt{100} = 2.5$.

We use the engineer's decision rule to compute the type I error probability

$$\begin{aligned}\alpha &= \Pr[\text{reject } H_0 \mid H_0 \text{ true}] \\ &= \Pr[\bar{X} > 40 \mid \mu = 36\ (= \mu_0)] \\ &= 1 - \Phi\left(\frac{40 - 36}{2.5}\right) = 1 - .9452 = .0548\end{aligned}$$

Thus, there is a better than 5% chance that the engineer would get a sample result so untypically great that he would adopt an uneconomical bacterial leaching program. (This is actually a worst-case figure, since when $\mu = \mu_0 = 36$, the adopted program would exactly break even; for any lower level, such as $\mu = 20$, the chance of incorrect rejection of H_0 would be more remote.)

This appears to be very good. But what about that other error, accepting H_0 when it is false?

Similar computations provide type II error probabilities. There are many ways for the null hypothesis to be false—this will happen in the present illustration whenever the true μ exceeds μ_0. Suppose we assume that $\mu = 42$ lb/ton. We then have

$$\begin{aligned}\beta &= \Pr[\text{accept } H_0 \mid H_0 \text{ false}] \\ &= \Pr[\bar{X} \le 40 \mid \mu = 42] \\ &= \Phi\left(\frac{40 - 42}{2.5}\right) = .2119\end{aligned}$$

There is a substantial probability that the engineer will get an untypically small level for $\bar{X}$ when indeed the bacterial leaching process would be quite economical.

Figure 9-1 shows the two separate normal curves that apply for the type I and II error probabilities we have just computed. Notice that α is represented by an upper-tail area in the top curve, whereas β is represented by a lower-tail area in the bottom curve.

Similar type II error probabilities could be obtained for other economical levels for population mean copper yield. Figure 9-2 shows several cases, including $\mu = 50$ (which gives $\beta < .00001$) and $\mu = 39$ (giving $\beta = .6554$).

The dilemma of hypothesis testing—achieving an acceptable balance between α and the spectrum of possible β's—may be resolved by selecting an appropriate decision rule. And, as in statistical estimation, where increasing the sample size will provide greater precision and reliability, a larger n will reduce both the type I and type II error probabilities.

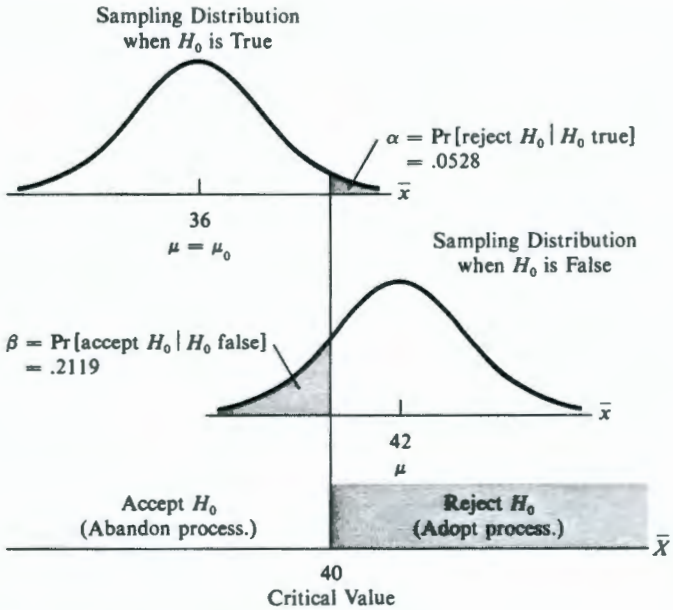


Figure 9-1 Illustration of hypothesis testing for bacterial leaching decision.

Determining an Appropriate Decision Rule

A new decision rule may be found simply by shifting the critical value for $\bar{X}$ to a new position. Two new decision rules for the bacterial leaching decision are illustrated in Figure 9-3.

In (a) the critical value has been reduced to 38.5 lb/ton. Notice under this rule that it will be more likely to reject a true H_0 (adopt an uneconomical process) with $\alpha = .1587$. But there is less chance of accepting a false H_0 (abandoning an economical process) with $\beta = .0808$. In (b) the critical value for $\bar{X}$ is at 41 lb/ton, with reverse changes providing a lower α (.0228) and a higher β (.3446).

Either of these rules might be an improvement over the original one. That will depend on the attitudes of the engineer and mining company management. The null hypothesis was formulated in such a way that the type I error is the more serious one of adopting an uneconomical bacterial leaching process. This is typical of statistical testing, where a target level (such as 5%) for the probability of this outcome is often prescribed in advance; that in turn establishes a unique critical value and decision rule. Unfortunately, a decision-maker must live with whatever β's the chosen decision rule brings. As we have seen, when the engineer's critical value is 40 there will be a considerable chance that the bacterial leaching process will be abandoned when it might indeed be economical—perhaps considerably so.

Choice of critical value is really a matter of deciding where to position the "fulcrum" so that a proper balance is achieved between α and the β's. If none of the equilibria prove satisfactory, the decision-maker must improve the "leverage," which can be achieved only by increasing the sample size.

The Effect of Sample Size

Like precision and reliability in estimation, the type I and type II error probabilities of hypothesis testing are competing ends, so that *one* cannot be im-

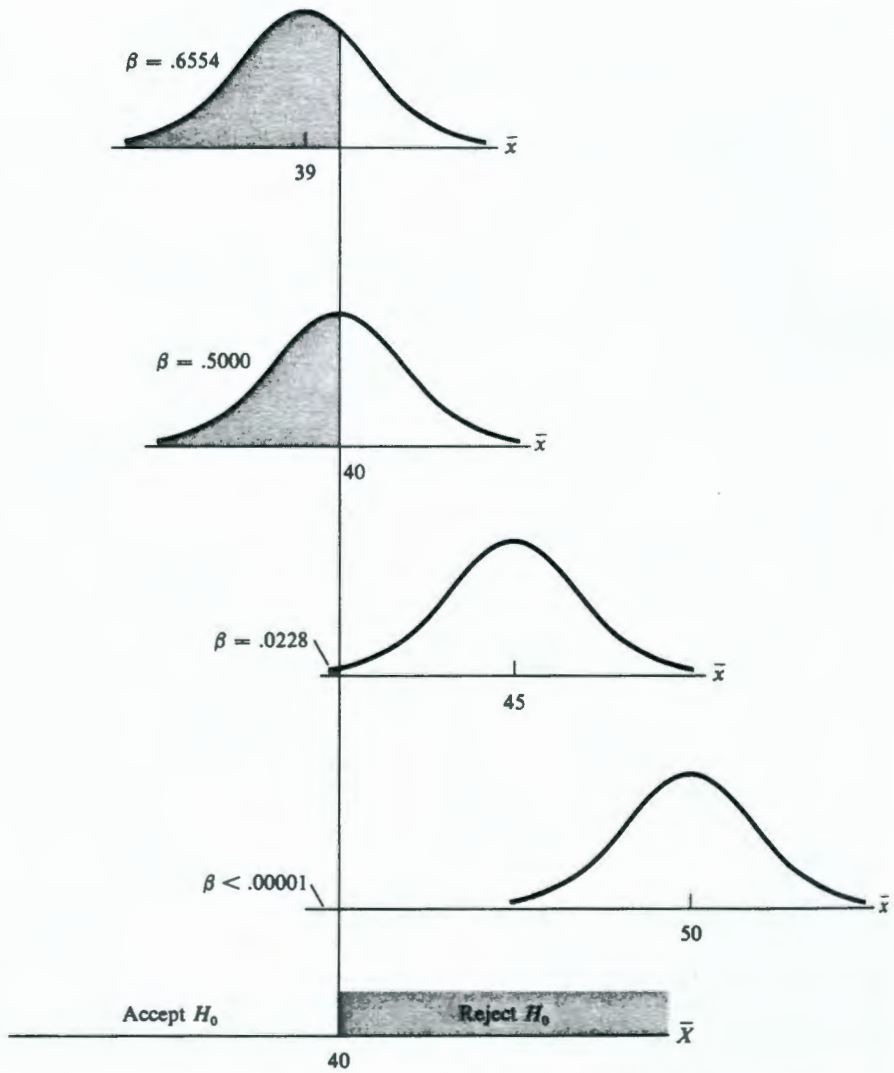


Figure 9-2 Type II error probabilities for accepting H_0 when it is false, for several different possible means.

proved except at the expense of the other. The only way to improve *both* α and β is to reduce potential sampling error by increasing the sample size. Figure 9-4 shows how the mining engineer's situation improves when n is raised from 100 to 200 while the original decision rule is retained. Notice that the normal curves cluster more tightly about the mean, with new levels of $\alpha = .0119$ (when $\mu = \mu_0 = 36$) and $\beta = .1292$ (when $\mu = 42$).

Making the Decision

The preceding discussion applies entirely to the planning stage of a hypothesis testing study. At that time—before data are collected—the hypotheses are formulated, the sample size is established (usually by budgetary considerations), and the decision rule is established (often being dictated by the chosen significance level α). The sample data are then collected. The actual decision is more or less automatic, depending on the computed level of the test statistic.

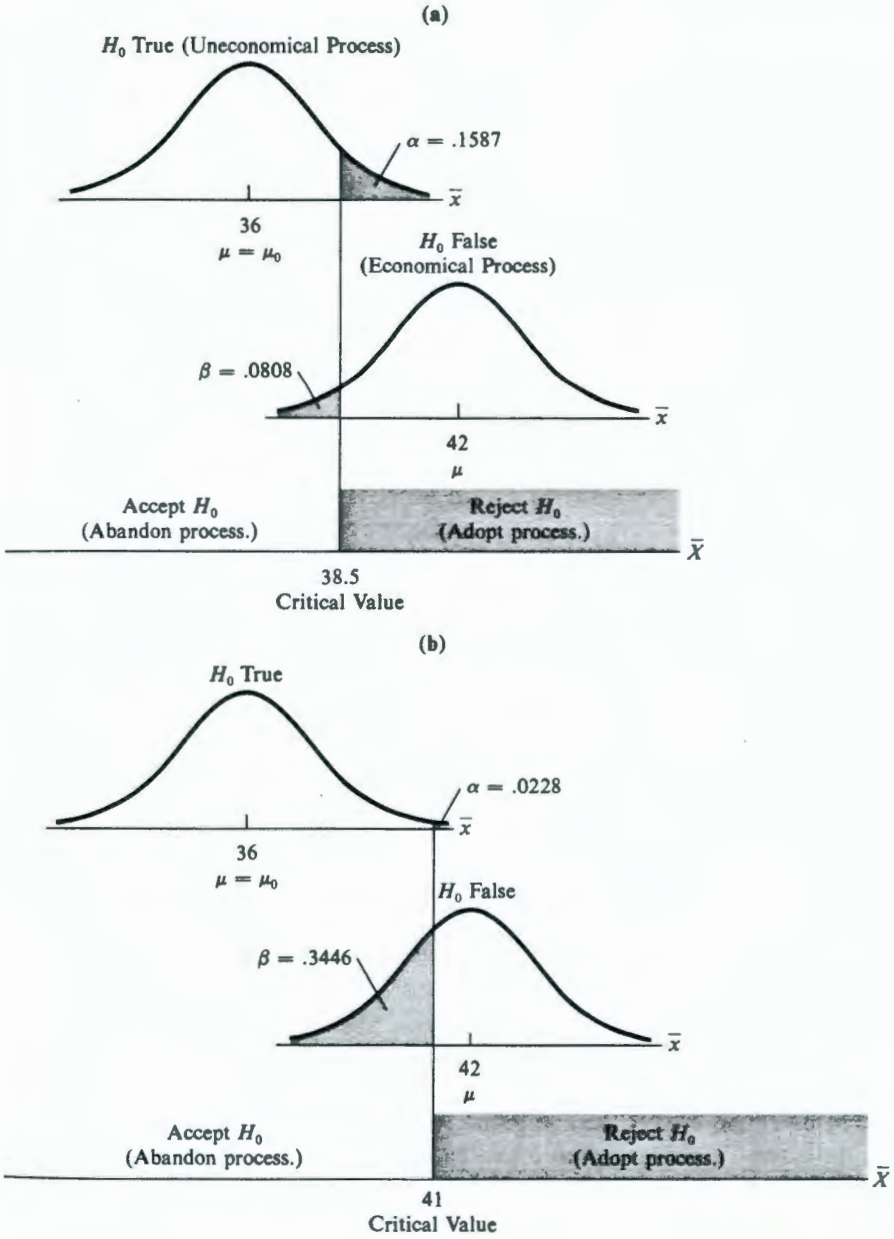


Figure 9-3 Error probabilities as they change with a shift in critical value for bacterial leaching decision.

We will conclude the bacterial leaching illustration by considering two hypothetical scenarios that might result after the data are in.

SCENARIO ONE. Suppose that the mean yield per ton from the 100 test loads brought to the leaching dump turns out to be $\bar{X} = 47.6$ lb/ton. Since this quantity falls in the rejection region, the null hypothesis that $\mu \leq 36$ must be rejected; the sample results are statistically significant. The engineer concludes that large-scale bacterial leaching will be economical and recommends to man-

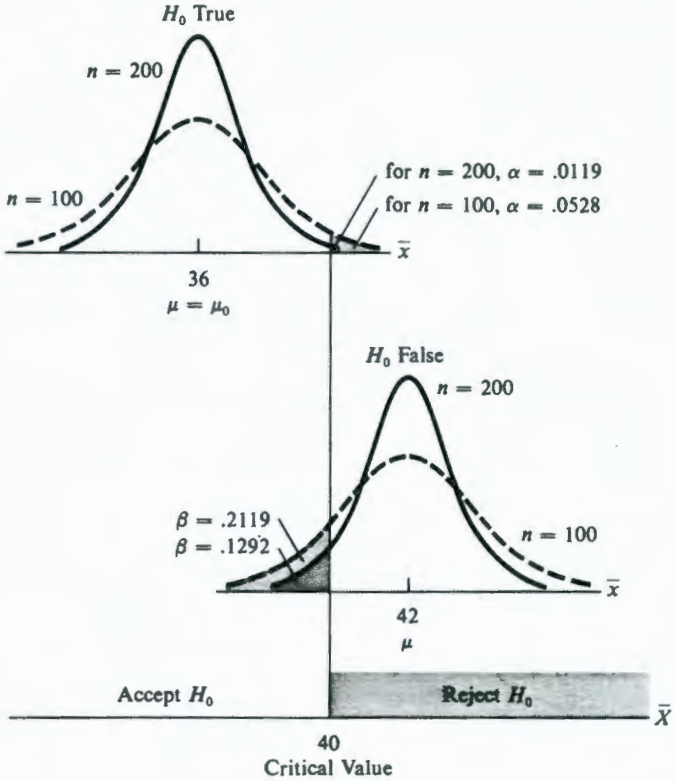


Figure 9-4 Both error probabilities are reduced when the sample size is increased from 100 to 200.

agement that it be adopted. The engineer takes comfort in the magnitude by which the computed mean exceeds the break-even level of 36 lb/ton. This quantity exceeds that hypothetical level for μ by more than 4 times $\sigma_{\bar{X}}$.

SCENARIO TWO. Suppose instead that the resulting mean copper yield from the test batches is only $\bar{X} = 32.1$ lb/ton. This value falls in the acceptance region, and the engineer concludes that bacterial leaching is not economical at this time. (Indeed, under the established decision rule the engineer would take the same action even if the computed $\bar{X}$ were much closer to the critical value of 40 lb/ton—say 38, or even 39.9.) In accepting H_0 the test results are found not to be statistically significant. This does not mean the findings are unimportant, only that H_0 cannot be rejected. The engineer realizes that there is a substantial probability that this decision is incorrect because μ might indeed be greater than the postulated break-even point. Unfortunately, there is no way to know. The case for bacterial leaching can always be reopened should conditions change, perhaps because of rising copper prices or development of an improved strain of bacteria.

Formulating the Hypotheses

As noted, H_0 is usually designated in such a way that the type I error is the most important one to avoid. Consider the following example.

EXAMPLE: Does a New Type of Battery Last Longer?

A design engineer for consumer products is evaluating a new battery substance based on different chemistry than is presently in use. It gives batteries a mean lifetime of approximately 100 hours. The new battery may last at least as long, on the average, as the present one—or it may die more rapidly. A sampling study will be used, and a decision whether or not to change to the new battery will be based on the observed mean time between failures (MTBF) of test cells composed under the new chemistry. Either of the following assumptions regarding the mean lifetime μ of the new battery might serve as the null hypothesis:

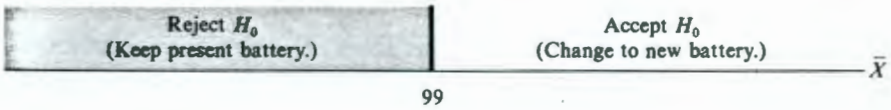
$$\mu \leq 100 \text{ hours} \quad \text{or} \quad \mu \geq 100 \text{ hours}$$

Since the company wants to keep its reputation for innovation, the more serious error was judged to be keeping the present battery when the new one lasts longer. (Certainly it would be undesirable to introduce a new battery that doesn't have a longer life, but this was judged far less damaging.) Thus, the following designation was made:

H_0: $\mu \geq 100$ hours (New battery is at least as good.)

H_1: $\mu < 100$ hours (New battery has a shorter lifetime.)

A decision rule based on $\overline{X}$ was established so that H_0 would be rejected for shorter computed MTBFs and accepted for longer ones:



The test described in the battery example is a **lower-tailed test,** since the rejection region coincides with the lower tail of the normal curve for $\overline{X}$ that applies under the null hypothesis. The reversed rejection region in the bacterial leaching illustration indicates that an **upper-tailed test** applies in that application. Some tests involve hypotheses in which the rejection region is split in two, with an intervening acceptance region. The following is an example of such a **two-sided test.**

EXAMPLE: Meeting Labeling Requirements

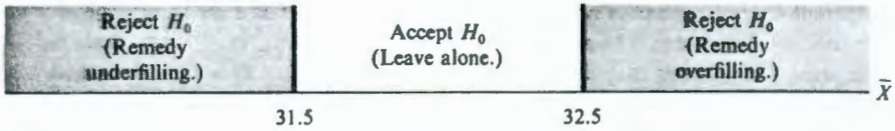
Packaging items for public consumption can be an exacting task because of regulations or trade practices regarding labeling of the volume or weight of the contents. Too much ingredient is as much to be avoided as too little. Sampling is often used in quality-control investigations concerning weights and measures. Typically the null hypothesis takes the form that the labeled mean is being met exactly, with the opposite alternative. For example,

H_0: $\mu = 32$ grams (Labeling specifications are met)

H_1: $\mu \neq 32$ grams (Labeling specifications are violated)

There are two ways for the null hypothesis to be untrue—if there is overfilling

($\mu > 32$ grams) or if there is underfilling ($\mu < 32$ grams). The following form is used for the decision rule:



Problems

9-1 Suppose that copper prices have dropped so that the break-even level for the mean yield from bacterial leaching rises to 37 lb/ton. Answer the following:

(a) Reformulate the engineer's hypotheses.

(b) Using the critical value of 40, determine (1) the type I error probability α and (2) the type II error probability β when $\mu = 42$.

9-2 Suppose a mining engineer must decide whether to use a new steel alloy instead of the one presently used to reinforce the mineshaft linings. The null hypothesis is that the present bars are at least as strong as the new ones. He will either accept this assumption, retaining the present reinforcing bars, or reject it and replace them.

State in words (a) the type I error, and (b) the type II error.

9-3 Consider each of the hypothesis-testing applications below. Indicate for each outcome whether it is a correct decision, a type I error, or a type II error:

(a) H_0: New power cell's lifetime does not exceed old one's.
- (1) Change to new when old lasts as long or longer.
- (2) Keep old when new lasts longer.
- (3) Keep old when old lasts as long or longer.
- (4) Change to new when new lasts longer.

(b) H_0: Memory chips are satisfactory.
- (1) Reject satisfactory shipment.
- (2) Accept satisfactory shipment.
- (3) Reject unsatisfactory shipment.
- (4) Accept unsatisfactory shipment.

(c) H_0: New design is safe.
- (1) Approve an unsafe design.
- (2) Disapprove an unsafe design.
- (3) Disapprove a safe design.
- (4) Approve a safe design.

9-4 The following hypotheses are to be tested, with $\mu_0 = 100$.

$$H_0\colon \mu \le \mu_0 \qquad H_1\colon \mu > \mu_0$$

Assume that the population standard deviation is $\sigma = 28$ and the sample size is $n = 100$. The following decision rule applies:

$$\textbf{Accept } H_0 \text{ if } \bar{X} \le 104$$

$$\textbf{Reject } H_0 \text{ if } \bar{X} > 104$$

Determine the type I error probability α when $\mu = 100$ and the type II error probability β when $\mu = 110$.

9-5 Repeat Problem 9-4 using 106 instead as the critical value. Is the new α larger or smaller? Is the new β larger or smaller?

9-6 Repeat Problem 9-4 using a larger sample size of $n = 150$. Is the new α larger or smaller? Is the new β larger or smaller?

9-7 The following hypotheses are to be tested, with $\mu_0 = 900$.

$$H_0\colon \mu \ge \mu_0 \qquad H_1\colon \mu < \mu_0$$

Assume that the population standard deviation is $\sigma = 50$ and the sample size is $n = 100$. The following decision rule applies:

$$\textbf{Accept } H_0 \text{ if } \bar{X} \geq 885$$

$$\textbf{Reject } H_0 \text{ if } \bar{X} < 885$$

Determine the type I error probability α when $\mu = 900$ and the type II error probability β when $\mu = 875$.

9-2 TESTING THE MEAN USING LARGE SAMPLES

We are now ready to apply the basic concepts of the preceding section to testing the mean. A general procedure has been developed for this application that may be adapted to a wide variety of hypothesis testing situations. Some of these pertain to other parameters than the mean and will be described later in this chapter.

The Hypothesis Testing Steps

All statistical tests involve the same basic steps outlined below.

STEP 1. *Formulate the hypotheses.* In all cases involving the mean, the decision parameter is the unknown μ. A limiting value μ_0 must be identified for defining H_0. In the bacterial leaching illustration, μ_0 was the break-even point for the procedure. Although the choice of μ_0 is often an economic consideration, other reasons may prevail. The level of μ_0 may be prescribed, as in a labeling specification for package contents. Or, μ_0 might be a benchmark level that corresponds to a present procedure, process, material, or part; we saw such an example in the battery example of Section 9-1.

In one-sided tests, error considerations usually determine which side of μ_0 to include under H_0. The null hypothesis will be selected from one of the two forms,

$$\mu \leq \mu_0 \quad \text{or} \quad \mu \geq \mu_0$$

The two errors are:

1. Taking the wrong action when in fact $\mu \leq \mu_0$.
2. Taking the wrong action when in fact $\mu \geq \mu_0$.

The null hypothesis will be H_0: $\mu \leq \mu_0$ if (1) is more serious and H_0: $\mu \geq \mu_0$ if (2) is more serious.

STEP 2. *Select the test statistic.* In testing the mean, $\bar{X}$ is the natural test statistic. By the central limit theorem we may ordinarily use the normal curve to represent the sampling distribution of $\bar{X}$. When the population standard deviation is known, $\sigma_{\bar{X}}$ may be computed, and for any given level of μ we may establish probabilities for possible values of $\bar{X}$.

But when σ is unknown, the appropriate normal curve may not be specified. In those cases, testing procedures employ instead the

NORMAL DEVIATE TEST STATISTIC
(σ UNKNOWN)

$$z = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

The above quantity is approximately normally distributed and should be used only when n is large. The analogous small sample procedure uses instead the Student t statistic and is described in Section 9-4.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* As we have seen, the common practice in statistical testing is to base the procedure on avoidance of the more serious type I error. The probability α of this, also referred to as the *significance level,* may then be prescribed in advance. With one-sided applications, the resulting experiment will be a lower- or upper-tailed test, depending on whether small or large values of the test statistic refute H_0. Figure 9-5 summarizes those two situations.

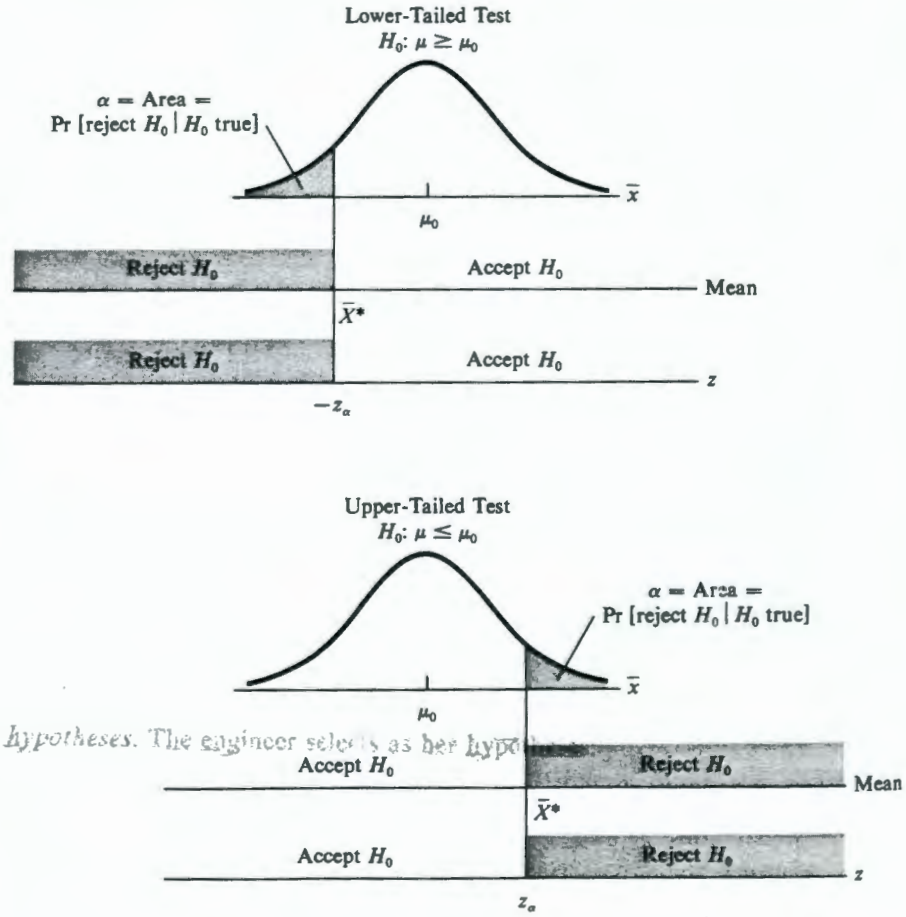


Figure 9-5 One-sided hypothesis tests will be lower- or upper-tailed, depending on the orientation of H_0.

STEP 4. *Formulate the hypotheses.* The engineer selects as her hypo

The critical value of the test statistic defines the acceptance and rejection regions. In testing the mean, this is determined by the **critical normal deviate** z_α for which the normal-curve upper-tail area is α.

When $\overline{X}$ serves directly as the test statistic, the foregoing normal deviates are used to compute the

CRITICAL VALUE FOR THE SAMPLE MEAN (σ KNOWN)

Upper-Tailed Test:

$$\overline{X}^* = \mu_0 + z_\alpha \sigma_{\overline{X}}$$

Lower-Tailed Test:

$$\overline{X}^* = \mu_0 - z_\alpha \sigma_{\overline{X}}$$

STEP 4. *Collect the sample data and compute the value of the test statistic.* This step incorporates the time-consuming and expensive second stage of the statistical sampling study. No revisions to the earlier steps should be based on the particular results obtained here.

STEP 5. *Make the decision.* The final choice is automatic, in accordance with the decision rule established in Step 3. H_0 must be accepted if the computed value of the test statistic falls in the acceptance region and rejected otherwise. By convention, should the computed value round precisely to the critical value, H_0 must be accepted.

An Upper-Tailed Test

ILLUSTRATION: Axial Stress Testing of a New Alloy

A mechanical engineer is considering a new nickel–chrome-iron alloy. She has ordered 100 sample castings, which are to be tested at a materials laboratory for endurance under axial stress. The engineer is seeking a metal strong enough to meet customer specifications for parts in a new stamping machine. These require that the mean number of cycles to failure μ obtained in vibration testing exceed 500,000. The target population represents the endurance measurements that might be obtained if every possible casting—not just the sample 100—were subjected to the test procedure. So that quick results may be obtained, a lower endurance figure was specified than is ordinarily found in stress testing. Accordingly, unusually extreme levels were set for the test displacement and force parameters.

The following hypothesis testing steps were taken.

STEP 1. *Formulate the hypotheses.* The engineer selects as her hypotheses:

H_0: $\mu \leq 500{,}000$ (Endurance specifications are not exceeded)

H_1: $\mu > 500{,}000$ (Endurance specifications are exceeded)

The designations reflect her strong desire to avoid choosing the alloy when it does not exceed customer specifications. These specifications establish $\mu_0 = 500{,}000$ as the limiting value for the mean number of cycles to failure.

STEP 2. *Select the test statistic.* The sample mean $\bar{X}$ will serve as the test statistic. Its sampling distribution may be represented by the normal curve centered at μ with standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. From previous testing with other materials, it has been established that the standard deviation in endurance should be around 50,000 cycles to failure. For planning purposes that value will be assumed for σ, regardless of the level for μ. Thus, $\sigma_{\bar{X}} = 50{,}000/\sqrt{100} = 5{,}000$.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* Because she regards so seriously the type I error of rejecting H_0 when it is true, the engineer chooses a 1% significance level, so that $\alpha = .01$. From Appendix Table E this corresponds to a critical normal deviate of $z_{.01} = 2.33$. Large positive levels for z will refute H_0, and the test is upper-tailed. The critical value for the sample mean is

$$\bar{X}^* = 500{,}000 + 2.33(5{,}000) = 511{,}650$$

The decision rule is shown in Figure 9-6.

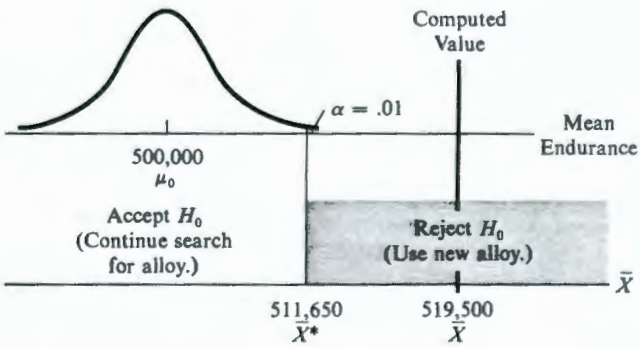


Figure 9-6

STEP 4. *Collect the sample data and compute the value of the test statistic.* The test results provide a computed sample mean of $\bar{X} = 519{,}500$ cycles to failure.

STEP 5. *Make the decision.* The computed mean exceeds the critical value and falls in the rejection region. The engineer must therefore reject H_0 and adopt the new alloy in the stamping machine.

The computed $\bar{X}$ lies about 4 standard deviations above μ_0, so the chance is remote for getting a result as extreme as that obtained should the true μ actually be equal to μ_0.

Figure 9-7

ILLUSTRATION: Selecting the Faster Fermentation Temperature

A Lower-Tailed Test

A chemical engineer is pilot-testing a fermentation process for the manufacture of a pharmaceutical product. All design parameters have been selected except for the temperature at which fermentation itself will take place. Current temperature-control hardware is designed to operate at a low nominal level, although theoretically an environment 10°C higher should provide a faster fermentation rate. Since the higher-temperature operation will require specially

designed control devices, that environment will be incorporated into the final design only if sample batches ferment significantly faster than the nominal mean time of 30 hours. A series of 50 pilot batches are to be run at the hotter temperature.

The following hypothesis testing steps apply:

STEP 1. *Formulate the hypotheses.* Expressing by μ the mean fermentation time of all future batches that might be processed under high temperatures, the engineer establishes:

H_0: $\mu \geq 30$ hours (Higher-temperature fermentation is not faster)

H_1: $\mu < 30$ hours (Higher-temperature fermentation is faster)

The limiting case is the nominal processing time $\mu_0 = 30$ hours. The $\geq$ orientation was selected for H_0 because the more serious (type I) error is to choose the higher-temperature process when it is not actually the faster one.

STEP 2. *Select the test statistic.* Because the population standard deviation is not known, the test statistic is the normal deviate z.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* The engineer establishes a significance level of $\alpha = .05$. This level reflects his conservatism regarding the possible type II error (accepting H_0 when H_0 is false). The probabilities (β's) of this second error (not designing the higher-temperature process when in fact it is faster) would be uncomfortably large with a smaller α. The critical normal deviate is $z_{.05} = 1.64$. The test is lower-tailed, since a fast mean processing time $\overline{X}$ under the higher-temperature environment will refute H_0. Thus, negative values for z lying below -1.64 will lead to rejection. Figure 9-7 applies.

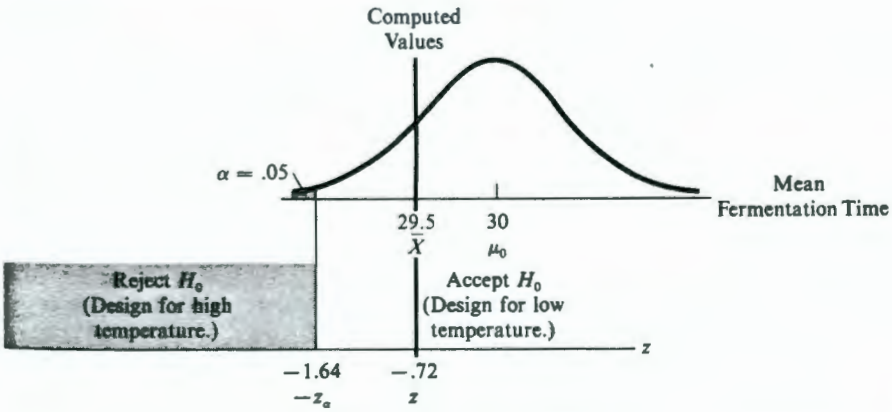


Figure 9-7

STEP 4. *Collect the sample data and compute the value of the test statistic.* The following test results were obtained:

$$\overline{X} = 29.5 \text{ hours}$$

$$s = 4.91$$

The above provide the computed normal deviate

$$z = \frac{29.5 - 30}{4.91 / \sqrt{50}} = -.72$$

STEP 5. *Make the decision.* The above quantity falls inside the acceptance region, so the engineer must accept H_0 and design the process for low fermentation temperatures.

Type II Error Considerations

In accepting the null hypothesis that the low-temperature environment results in faster fermentation, there is the danger that the type II error has been committed. It is good statistical procedure to consider the potential for such error. The engineer might be incorrectly designing for low fermentation temperatures when the actual value of μ is faster than nominal. Of course that could not be established without further testing. As we have seen, the incidence of such type II errors may be controlled by using a larger α or by increasing the sample size. We will discuss this point further in Section 9-7.

Problems

9-8 Answer the following for each of the hypothesis testing situations given below:
(a) Find the critical value(s) for the sample mean.
(b) For the given computed sample mean indicate in each case whether the null hypothesis should be accepted or rejected.

	(1)	(2)	(3)	(4)
n	100	50	200	150
σ	5	4	20	15
H_0	$\mu \leq 50$	$\mu \geq 60$	$\mu \geq 100$	$\mu \leq 100$
α	.05	.01	.01	.05
$\overline{X}$	55.3	63.5	95.8	101.2

9-9 The mechanical engineer from the text illustration performs another stress test on a new material, where $\sigma = 600{,}000$ cycles to failure and H_0: $\mu \leq 10{,}000{,}000$. Suppose a sample size of $n = 50$ castings is tested at a significance level of $\alpha = .05$.
(a) Assuming that the same choices must be made, identify the acceptance and rejection regions.
(b) The computed mean is 10,115,000. What action should the engineer take?

9-10 Suppose that the chemical engineer in the fermentation-temperature illustration had reported identical sample results, but for a larger sample size of $n = 100$ instead. Perform hypothesis testing steps 3–5.

9-11 Consider a sample of size $n = 100$ taken at random from a small population of size $N = 500$. The null hypothesis is that $\mu \leq 5{,}000$. Suppose the population standard deviation is known to be $\sigma = 150$ and that $\overline{X}$ will serve as the test statistic.
(a) Assuming that a significance level of $\alpha = .01$ is desired, find the critical value for the sample mean and determine the decision rule.
(b) Should H_0 be accepted or rejected if the computed sample mean turns out to be $\overline{X} = 5{,}060$?

9-12 A sample of size $n = 100$ is taken from a large population where σ is unknown. The null hypothesis is that $\mu \leq 5{,}000$.

(a) Assuming a 1% significance level and that the normal deviate will serve as the test statistic, determine the decision rule.

(b) Suppose that the sample provides $\overline{X} = 5{,}011$ and $s = 75.3$. For this outcome should H_0 be accepted or rejected?

9-13 A biological reduction process is being evaluated as an alternative to a conventional technique for which the mean yield of active ingredients is $\mu = 10$ grams/liter with a standard deviation of $\sigma = .5$ gram/liter. The biological method will be an effective replacement for the present procedure if it can be assumed to provide a yield of more than 12 grams/liter. A total of 100 batches will be pilot-tested under bacterial reduction.

(a) If test results indicate that significant improvement in yield might be made, the biological procedure will be adopted. There should be only a 5% chance that this will be done when the new technique's yield is not high enough to be effective as a replacement. Assuming the same standard deviation in yield applies whichever procedure is used, perform hypothesis testing steps 1–3.

(b) The test results provide a yield of 13.3 grams/liter. What action should be taken?

9-14 A new test stand might be acquired by the inspection department of GizMo, an instrumentation fabricator, if the mean time to check out a standard "black box" is significantly less than that for the present unit. The new test stand will save enough operating costs to pay for itself only if it can "burn in" all new instrument boxes in 15 hours or less, on the average. The equipment vendor allows GizMo to time-test a sample of 50 boxes. No educated guesses have been made regarding the target population's characteristics.

(a) Assuming that GizMo wants no more than a 1% chance of buying the new test stand when it will not save enough to justify its cost, perform hypothesis testing steps 1–3.

(b) The test results provide a mean burn-in time of 13.5 hours with a standard deviation of 3.4 hours. What action should be taken?

9-3 TWO-SIDED HYPOTHESIS TESTS OF THE MEAN

Statistical investigations sometimes involve hypotheses of the following form

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

Notice that the alternative hypothesis is true both when $\mu < \mu_0$ and when $\mu > \mu_0$. Since H_0 has two sides, investigations involving the preceding hypotheses are called **two-sided tests.** These are frequently encountered in statistical quality control, where requirements are that the mean achieve some specified level exactly.

The decision rule in testing hypotheses of the above form must involve two critical values, establishing a two-sided rejection region. The type I error of rejecting a true H_0 may be committed in two ways, depending on whether the test statistic is untypically large or small. Each case is therefore assigned half of the total type I error probability.

When testing the mean with large samples or with a known standard deviation, a normal curve serves as the sampling distribution for $\overline{X}$. Figure 9-8 illustrates the procedure. As with one-sided tests, the computed normal deviate z serves as the test statistic when σ is unknown. The critical normal deviate is

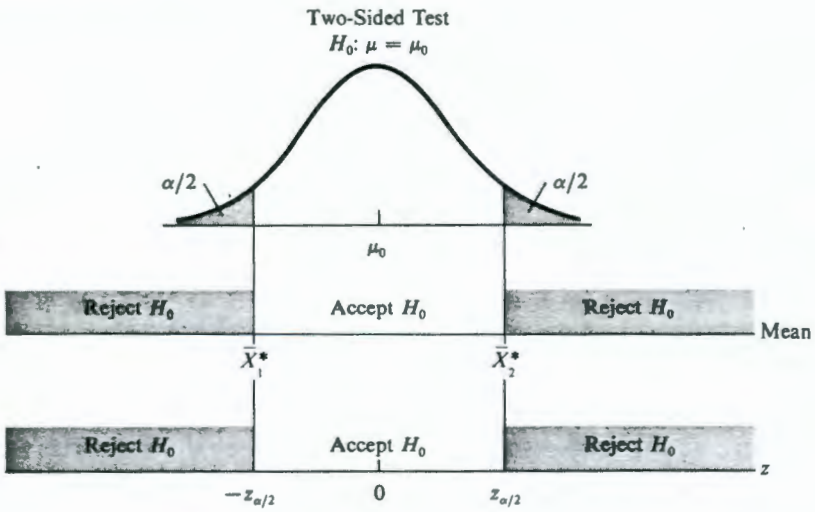


Figure 9-8 Decision rule for two-sided test.

denoted by $z_{\alpha/2}$. Above this point the upper tail of the standard normal curve includes an area equal to $\alpha/2$; the same area lies below $-z_{\alpha/2}$. When $\bar{X}$ serves directly as the test statistic, the following expressions provide the

CRITICAL VALUES FOR THE SAMPLE MEAN (σ KNOWN)

Two-Sided Test:

$$\bar{X}_1^* = \mu_0 - z_{\alpha/2}\sigma_{\bar{X}}$$

$$\bar{X}_2^* = \mu_0 + z_{\alpha/2}\sigma_{\bar{X}}$$

Conducting the Test

ILLUSTRATION: Replacing a pH Regulator

The superintendent of a new chemical plant is establishing operational policies, such as when to perform critical maintenance. When practical, those actions will be based on sample data. This is the case for a process in which the pH of an intermediate stage must be maintained at a neutral level. A control valve keeps the mean batch pH at 7.0. Owing to the caustic nature of the fluids flowing through it, the valve must occasionally be replaced. This action will be based on the mean pH of 100 sample vials taken at random times from the separation tank containing the output from that stage.

Hypothesis steps 1–3 show how the policy for valve replacement was established. Steps 4 and 5 apply to a particular sample outcome.

STEP 1. *Formulate the hypotheses.* The following hypotheses apply:

$$H_0: \mu = 7.0 \text{ (Process is neutral)}$$

$$H_1: \mu \neq 7.0 \text{ (Process is not neutral)}$$

STEP 2. *Select the test statistic.* The computed mean pH for the sample vials serves as the test statistic. The standard deviation for individual test vials has long been established at .50 on the pH scale. This establishes the standard error for $\bar{X}$: $\sigma_{\bar{X}} = .50/\sqrt{100} = .05$.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* It is very expensive to shut down the process to replace the control valve, so plant policy permits just a 1% chance of doing this unnecessarily. The significance level for each sample test is therefore $\alpha = .01$. The critical normal deviate is $z_{.005} = 2.57$, and the critical values for the sample mean are

$$\bar{X}_1^* = 7.0 - 2.57(.05) = 6.87 \text{ pH}$$

$$\bar{X}_2^* = 7.0 + 2.57(.05) = 7.13 \text{ pH}$$

The acceptance and rejection regions are shown in Figure 9-9.

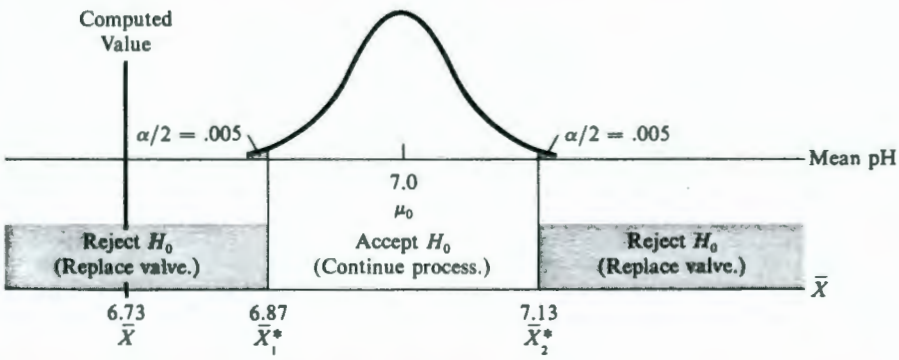


Figure 9-9

STEP 4. *Collect the sample data and compute the value of the test statistic.* The computed mean pH for a particular test turned out to be $\bar{X} = 6.73$.

STEP 5. *Make the decision.* Since this value falls in one of the two rejection regions, H_0 must be rejected. At the present time it must be concluded that the process mean pH is not neutral (and is too acidic). The process will be shut down to replace the control valve.

Hypothesis Testing and Confidence Intervals

A two-sided hypothesis test may be explained in terms of statistical estimation. An equivalent procedure for conducting such a test would be to construct a $100(1 - \alpha)\%$ confidence interval from the sample results. Then if μ_0 falls inside the interval, H_0 must be accepted; otherwise it must be rejected. α is the proportion of such intervals that will fail to bracket μ_0 when it is the true mean. In other words, when this procedure is used, the type I error of rejecting H_0 when it is true will occur at frequency α.

For the pH regulator decision, $1 - \alpha = 1 - .01 = .99$. Using the same

sample results from before, we construct the 99% confidence interval estimate for μ:

$$\mu = \overline{X} \pm z_{.005}\frac{\sigma}{\sqrt{n}}$$

$$= 6.73 \pm 2.57\frac{.50}{\sqrt{100}} = 6.73 \pm .13$$

or

$$6.60 \leq \mu \leq 6.86$$

Because this interval does not contain $\mu_0 = 7.0$, H_0 must be rejected. This is the same conclusion reached earlier. If the plant superintendent were to compute a similar 99% confidence interval for each sample outcome, then μ_0 would fall inside of 99% of those computed when the true mean was in fact equal to $\mu_0 = 7.0$; in each case H_0 would be accepted. But 1% of the intervals would lie either totally above or totally below 7.0 when that is the true mean, resulting in incorrect rejection (unnecessary valve replacement).

When σ is unknown, the computed sample standard deviation s is used instead of σ in computing the confidence interval.

Problems

9-15 Answer the following for each of the hypothesis testing situations given below:

(a) Find the critical values for the sample mean.

(b) For the given computed sample mean, indicate in each case whether the null hypothesis should be accepted or rejected.

	(1)	(2)	(3)	(4)
n	100	150	200	50
σ	15	14	20	5
H_0	$\mu = 45$	$\mu = 60$	$\mu = 150$	$\mu = 90$
α	.01	.05	.01	.05
$\overline{X}$	46.5	57.3	154.1	90.4

9-16 The plant superintendent in the pH regulator illustration found that too much supplemental processing was required because faulty control valves were not replaced often enough. He raised the significance level to $\alpha = .05$.

(a) Perform hypothesis testing step 3.

(b) What action should be taken for each of the following sample results?

(1) $\overline{X} = 6.85$ (2) $\overline{X} = 7.24$ (3) $\overline{X} = 6.59$ (4) $\overline{X} = 7.07$

9-17 A plant produces rods with a specified 1 cm mean diameter. The standard deviation is .01 cm. A decision rule must be established for determining when to correct for over- or undersized output. That choice will be based on a sample of 100 rods, under the stipulation that there be only a 1% chance of taking corrective action when the mean diameter is exactly on target.

(a) Perform hypothesis testing steps 1–3.

(b) What act should be taken if $\overline{X} = .993$ cm?

(c) Suppose that $\overline{X} = 1.0023$ cm. Construct a 99% confidence interval estimate for the mean diameter. Does this contain μ_0? What action should be taken?

9-18 The above plant will soon be producing a .5 cm rod. The standard deviation is unknown. Again, corrective action will be based on a sample of 100 rods, with the same error goal.

(a) Perform hypothesis testing steps 1–3.
(b) What act should be taken if $\bar{X} = .497$ cm and $s = .0075$ cm?
(c) Construct a 99% confidence interval for the mean diameter when $\bar{X} = .508$ cm and $s = .013$ cm. Does this contain μ_0? What action should be taken?

9-19 A quality-control inspector for a microwave transmitter manufacturer is assessing a shipment of 400 crystal controls. The actual broadcast frequency will depend on the resonant frequency, which will vary slightly from crystal to crystal, but the mean level should achieve the rated target of .55 mHz. The exact standard deviation for the individual crystal frequencies is unknown. A random sample of 50 crystals from the shipment will be tested and the resonant frequency determined for each. The entire shipment will be rejected if the observed mean is significantly above or below the rated level and accepted otherwise. The inspector wants just a 1% chance of rejecting a shipment in which the mean frequency exactly matches the rated level.
(a) Formulate the inspector's hypotheses.
(b) Sample results provide $\bar{X} = .5503$ mHz and $s = 485$ Hz. Construct a 99% confidence interval estimate of the mean resonant frequency for the entire shipment. What action should the inspector take?

9-20 The final stage of a chemical process is sampled and the level of impurities determined. The final product is recycled if there are too many impurities and the controls are readjusted if there are too few (which is an indication that too much catalyst is being added). If it is concluded that the mean impurity level is the nominal level of $\mu = .01$ gram/liter, the process is continued without interruption. The standard deviation of impurities has been established to be $\sigma = .005$ gram/liter. A sample of $n = 100$ specimens is measured every 24 hours and the level of impurities is determined for each.
(a) Perform hypothesis testing steps 1–3, assuming a significance level of 5%.
(b) What action should be taken for each of the following sample results for the mean level of impurities?
(1) $\bar{X} = .0095$. (2) $\bar{X} = .0088$. (3) $\bar{X} = .0112$. (4) $\bar{X} = .0104$.
(c) Find the type II error probability when the true mean impurity level is:
(1) $\mu = .009$. (2) $\mu = .012$. (3) $\mu = .008$. (4) $\mu = .014$.

TESTING THE MEAN USING SMALL SAMPLES 9-4

Although it is generally desirable to base statistical evaluations on large sample sizes, this is not always possible. Almost universally, budgetary constraints place limitations on n. And we have seen that investigations of rare phenomena or those involving human subjects must often be made with an arbitrarily small number of observations.

Using the Student t Test Statistic

Similar procedures apply when small samples are used in testing the mean. Invoking the central limit theorem, we may still ordinarily assume a normal sampling distribution for $\bar{X}$. But the population standard deviation must be known to specify the normal curve completely. (And even then—except when the population itself is normal—the resulting sampling distribution is only approximate.) When σ is unknown, as we have seen, inferences regarding μ may be based instead on the **Student t distribution.**

To test the mean when σ is unknown and n is small, the same basic 5 steps are employed. But rather than basing the decision directly on $\overline{X}$ we use the

STUDENT t TEST STATISTIC (σ UNKNOWN)

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

The above may be employed in exactly the same manner as when z is the test statistic. For a prescribed significance level α, the one-sided test critical value is denoted by t_α. Using $n - 1$ degrees of freedom, this quantity may be read from Appendix Table G. The following decision rules apply.

Lower-Tailed Test (H_0: $\mu \geq \mu_0$):

Accept H_0 if $t \geq -t_\alpha$

Reject H_0 if $t < -t_\alpha$

Upper-Tailed Test (H_0: $\mu \leq \mu_0$):

Accept H_0 if $t \leq t_\alpha$

Reject H_0 if $t > t_\alpha$

In the case of two-sided tests the type I error probability is shared equally by the two tails of the t curve, and the two-sided rejection region involves the critical value $t_{\alpha/2}$.

Two-Sided Test (H_0: $\mu = \mu_0$):

Accept H_0 if $-t_{\alpha/2} \leq t \leq t_{\alpha/2}$

Reject H_0 if $t < -t_{\alpha/2}$ or if $t > t_{\alpha/2}$

ILLUSTRATION: Modifying Air Traffic Controller CRT Displays

A common complaint of air traffic controllers is fatigue arising from watching CRT (cathode-ray tube) displays over extended periods of time. As part of a summer job, one student working in a human-engineering section at the FAA considered modification of the CRT itself. He noted that the scopes show the "noise" from periodic scans of the search sector; this backscattering appears as a flash that gradually disappears until the next scan. This phenomenon gives controller display screens an annoying pulsation that is not only distracting but might also contribute to eye fatigue. The student proposed using a filter to eliminate most of this pulsating illumination.

Test modifications to existing hardware were expensive, and a crude apparatus was devised for three terminals in a nearby control center. With these, it was possible to monitor a random sample of 12 controllers. Although fatigue is hard to measure, the student devised an ingenious yardstick. At the midmorning and midafternoon breaks, each test subject was given a stack of quarters and directed to play the then wildly popular video arcade game "ZapMan," placed there especially for the test. Recorded game scores would be used to represent cumulative on-job fatigue. Getting these data would be simple, and any impairment of game-playing skill would likely reflect on-job alertness, be-

cause this particular game simulates concentration and reactivity levels needed in controlling aircraft.

A subject's afternoon score was expected to be lower than the morning's. During a week's trial, the test subjects would work at their regular terminals. Their individual aggregate scores (per minute played) would be recorded and the drop computed. During the following week, the same persons would use the experimental CRTs, and an analogous "drop" figure would be subtracted from the earlier one to find a mean difference in diurnal score-drop. The presumption is that if the experimental CRT is less fatiguing, this should be reflected in smaller afternoon drops in ZapMan scores, yielding a positive mean difference μ. The following steps apply.

STEP 1. *Formulate the hypotheses.* The student chose the following:

H_0: $\mu \leq 0$ (Diurnal score drops off at least as much using experimental CRTs, so they don't reduce fatigue.)

H_1: $\mu > 0$ (Diurnal score drops off less using experimental CRTs, so they do reduce fatigue.)

STEP 2. *Select the test statistic.* The target population standard deviation is an unknowable quantity. Because of the small sample size, the Student t serves as the test statistic.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* The student plans to recommend further detailed study of CRT modification if the evidence is strong that it would reduce fatigue. To protect himself from the embarrassment of making that recommendation only to have the CRT modification actually found to be ineffectual, he has chosen $\alpha = .05$ as the significance level. The resulting test is upper-tailed, as portrayed in Figure 9-10.

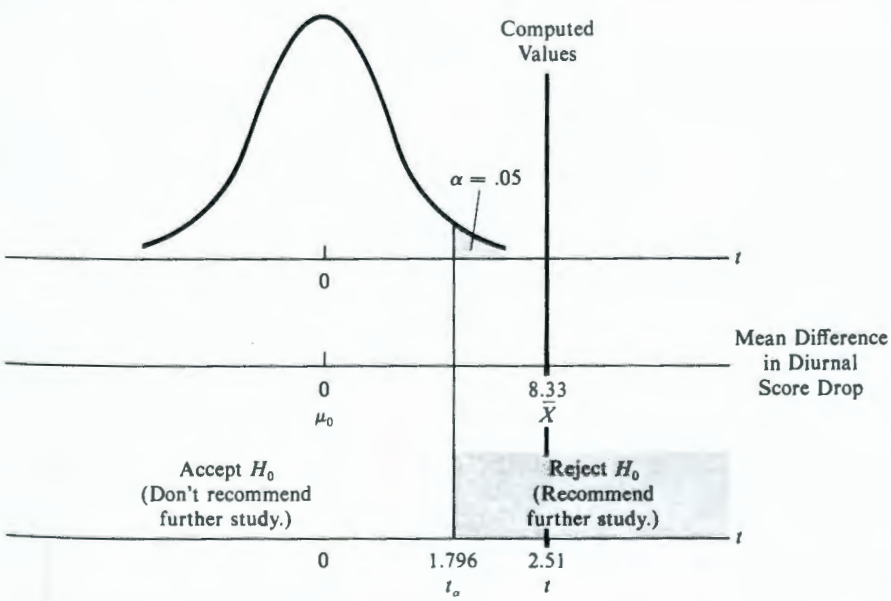


Figure 9-10

STEP 4. *Collect the sample data and compute the value of the test statistic.* The following results were obtained:

ZapMan Score (per minute of play)

	Regular CRT			Experimental CRT			
Controller	A.M.	P.M.	Drop	A.M.	P.M.	Drop	Difference
A	150	125	25	155	140	15	10
B	185	189	−4	222	220	2	−6
C	299	302	−3	344	350	−6	3
D	25	13	12	47	35	12	0
E	123	115	8	134	129	5	3
F	87	89	−2	94	93	1	−3
G	535	488	47	624	615	9	38
H	406	395	11	521	514	7	4
I	212	190	22	249	242	7	15
J	99	81	18	98	92	6	12
K	111	97	14	96	92	4	10
L	253	206	47	382	349	33	14
							100

The above data provide

$$\bar{X} = 8.33$$
$$s = 11.50$$

The computed value for the test statistic is

$$t = \frac{8.33 - 0}{11.50/\sqrt{12}} = 2.51$$

STEP 5. *Make the decision.* The computed t exceeds the critical value and falls into the rejection region. The student should therefore reject the null hypothesis. He will conclude that controllers experience greater drop-off in videogame scores after spending most of their working days at the regular CRT and that this might be attributable to greater fatigue than results from the experimental CRT. He will therefore recommend further study be made of possible CRT improvements.

Problems

9-21 Answer the following for each of the hypothesis testing situations given below:

(a) Find the critical value for the Student t statistic and indicate the acceptance and rejection regions.

(b) For the given sample results in each case indicate whether the null hypothesis should be accepted or rejected.

(1)	(2)	(3)	(4)
H_0:	H_0:	H_0:	H_0:
$\mu \leq 100$ m	$\mu \leq 14''$	$\mu = .45$ g	$\mu \geq 2{,}250$ psi
$n = 20$	$n = 25$	$n = 13$	$n = 18$
$\alpha = .05$	$\alpha = .01$	$\alpha = .10$	$\alpha = .05$
$\bar{X} = 125.3$ m	$\bar{X} = 13.55''$	$\bar{X} = .52$ g	$\bar{X} = 2{,}156$ psi
$s = 8.1$	$s = .97$	$s = .14$	$s = 171.4$

9-22 Refer to the data in the physical therapy apparatus example on page 263. Should the mechanical engineer accept or reject the null hypothesis that $\mu \leq 20$ days? Use $\alpha = .05$.

9-23 Refer to the data in the logic circuit wafer example on page 253. Should the industrial engineer accept or reject the null hypothesis that $\mu \geq .20$ oz? Use $\alpha = .01$.

9-24 An engineering department manager wants to decide whether or not to include a computer-aided design (CAD) software package in his budget. He is skeptical about the vendor's time-savings claims. A test has been conducted using pairs of nearly equally skillful engineers who independently design the same item—one using CAD and the other unassisted. For $n = 8$ parts the following percentage time-savings by CAD over the unassisted design were obtained:

80	10	37	26
45	29	44	5

At the 5% significance level, can the manager conclude that CAD will indeed yield savings in time over present design methods?

9-25 The following sample-failure data (thousands of miles) were obtained for a type of catalytic converter:

62.3	44.4	49.2	63.3	47.6	60.1
37.4	55.8	57.5	58.3	56.2	54.3

(a) Construct a 95% confidence interval estimate for the mean distance to failure.
(b) At the 5% significance level, must you accept or reject the null hypothesis that the catalytic converter will last, on the average, 50 thousand miles?

9-26 The makers of ZapMan have expressed an interest in the student's data from the CRT experiment. They wish to determine the effect a week's steady play will have on score levels. After subtracting the first week's A.M. score (regular) from the second week's A.M. score (experimental), the sample mean increase may be computed. The null hypothesis is that one week's practice will improve a person's ZapMan score by at most 10 points/minute. At the 1% significance level, what conclusion should be reached?

9-5 TESTING THE PROPORTION

Tests of the proportion are next in importance to those for the mean. The proportion is especially important in statistical quality control, where a major concern is whether the level of defectives in a shipment is high or low; similarly, the overall quality of a production process might be measured by the proportion of defective output. Proportions are also fundamental in work measurement, where work sampling is often used instead of stopwatch procedures. The proportion is also a key parameter in specifying a Bernoulli process, and a considerable amount of probability analysis might be based upon an assumed level.

The testing procedures established for the mean are easily extended to the proportion. The sample proportion P is a natural test statistic for experiments regarding π, but sometimes the number of successes R is employed equivalently. The appropriate sampling distribution for P or R is either the binomial or the hypergeometric—the latter applying when sampling without replacement from small populations. A normal curve closely approximates these underlying

distributions over a wide range of π's and n's, and tests of the proportion are most often based on that approximation.

The Sample Proportion Test Statistic

The sample proportion is a convenient test statistic for testing the value of the population proportion. Tests fall into one of the same three categories as in testing the mean. The following respective decision rules apply:

Upper-Tailed Test (H_0: $\pi \leq \pi_0$):

Accept H_0 if $P \leq P^*$

Reject H_0 if $P > P^*$

Lower-Tailed Test (H_0: $\pi \geq \pi_0$):

Accept H_0 if $P \geq P^*$

Reject H_0 if $P < P^*$

Two-Sided Test (H_0: $\pi = \pi_0$):

Accept H_0 if $P_1^* \leq P \leq P_2^*$

Reject H_0 if $P < P_1^*$ or if $P > P_2^*$

The sampling distribution of P is approximated by a normal curve having mean $E(P) = \pi$ and standard error $\sigma_P = \sqrt{\pi(1-\pi)/n}$ (which must be multiplied by the finite population correction factor when N is small). Critical normal deviates are used in the same manner as in testing the mean. The following expressions are used to find the

CRITICAL VALUES OF THE PROPORTION

Upper-Tailed Tests:

$$P^* = \pi_0 - \frac{.5}{n} + z_\alpha \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Lower-Tailed Tests:

$$P^* = \pi_0 + \frac{.5}{n} - z_\alpha \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Two-Sided Tests:

$$P_1^* = \pi_0 + \frac{.5}{n} - z_{\alpha/2} \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

and

$$P_2^* = \pi_0 - \frac{.5}{n} + z_{\alpha/2} \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

[illegible] of $\alpha = .05$ was [illegible], so the critical deviate denoted as z_α [illegible] the hypothesis provide a lower-tailed test and the critical value [illegible]

(Notice that a continuity correction of size $.5/n$ is subtracted in computing critical values in the upper tail and added for those in the lower tail.)

ILLUSTRATION: Is Double Data Entry Necessary?

A computer systems analyst for a large electric utility wondered if it might be practical to cut the tremendous expense of the data processing staff. Hundreds of data clerks presently encode meter-reader tags for entry into the customer-billing data base. To minimize billing errors, two clerks independently enter the data from each tag, and if the entries do not coincide, the tag is reprocessed. Although the possibility that errors will occur is extremely remote when tags are processed in this manner, the number of clerks required is twice that of a single data entry system. Furthermore, extra data processing is required to make the verification comparisons.

The analyst wonders if significant cost savings might be achieved by abandoning double-entry processing in favor of entering the data just once—obviously a much more error-prone procedure. Under the proposed scheme, large errors would be filtered out by comparing usage with the average of previous months; only those customer tags on which usage is above or below average by a factor of 3 would be recycled once for verification. Any subsequent over- or underbilling would ultimately be rectified by the cumulative readings obtained for a customer in future months.

Because of the expenses that would arise in handling customer complaints for overbilling and owing to the cost of uncollected funds from underbilled accounts, the efficacy of this procedure depends on a fairly low single-entry error rate. Management consensus was that any new-procedure error rate not exceeding .5% would result in savings for the company. The pivotal value for the proportion π of all erroneous entries was therefore set at $\pi_0 = .005$.

A sampling experiment was initiated and performed under conditions that duplicated very closely those anticipated under the new procedure. For the test period the clerks were asked to be especially careful, as their work was not going to be totally verified. To preclude lackadaisical performance, each clerk's work would still be checked on a random basis for accuracy—and the test subjects were so informed. The following steps were performed using a sample of $n = 2{,}000$ meter tags:

STEP 1. *Formulate the hypotheses.* The following hypotheses were selected by the analyst:

H_0: $\pi \geq .005$ (π_0) (New procedure is impracticable)

H_1: $\pi < .005$ (New procedure may be an improvement)

STEP 2. *Select the test statistic.* The sample proportion of errors P serves as the test statistic.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* The analyst plans to recommend adoption of the new procedure if the observed error rate is sufficiently small. A significance level of $\alpha = .05$ was selected, so the critical normal deviate is $z_\alpha = 1.64$. The hypotheses provide a lower-tailed test, and the critical value of the error proportion is

$$P^* = .005 + \frac{.5}{2{,}000} - 1.64\sqrt{\frac{.005(.995)}{2{,}000}} = .003$$

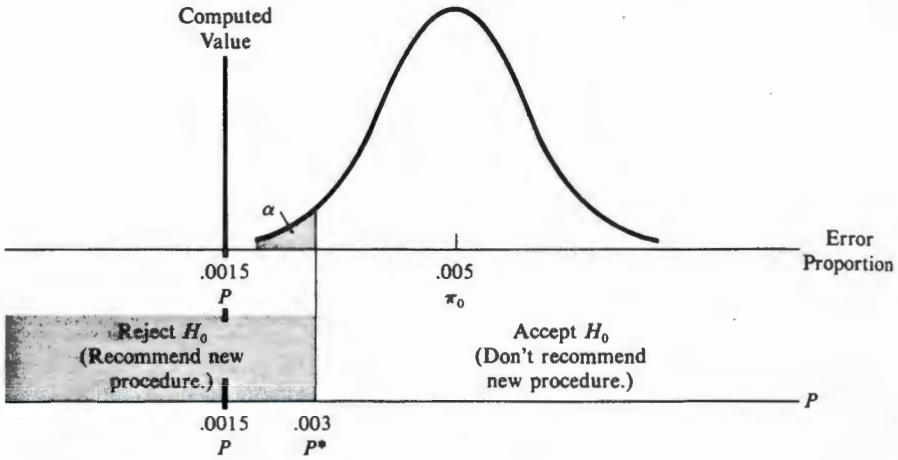


Figure 9-11

The decision rule is summarized in Figure 9-11.

STEP 4. *Collect the sample data and compute the value of the test statistic.* Only 3 of the sample tags were in error, so that the computed error proportion was only $P = 3/2{,}000 = .0015$.

STEP 5. *Make the decision.* The computed value falls below the critical value, and the analyst must reject the null hypothesis and conclude that the new procedure might provide savings in operational costs. The analyst therefore recommends the new procedure.

Further Remarks

The foregoing procedure need be modified only slightly when the number of successes R serves as the test statistic, multiplying the right-hand side of each critical value expression by n. It is also possible to use the binomial distribution directly in finding the decision rule, although our capability for doing this will be constrained by the limitations of tables (or computer time). Both of these complications, and the case of the small population, are left as exercises.

The type II error plays an analogous role in testing the proportion. Probabilities for this outcome may be computed for any level of π where H_0 is false. But the value of σ_P will change for each case.

Problems

9-27 Suppose that the systems analyst described in the text used instead $\pi_0 = .0035$ and a 1% significance level. Assuming the same sample results, what action should the analyst take?

9-28 In testing the null hypothesis that $\pi = .05$ at a significance level of $\alpha = .01$, indicate for each of the following possible results whether it should be accepted or rejected when $n = 500$.

(a) $P = .03$ (b) $P = .06$ (c) $P = .025$ (d) $P = .055$

9-29 For the data-entry illustration in the text, suppose that $\pi_0 = .007$ had been used with a 10% significance level. The test statistic is the number of errors R in a sample of size 4,000.

(a) Determine the critical value and resulting decision rule.
(b) What decision will be reached if the number of sample defectives were
(1) 25? (2) 6? (3) 10? (4) 31?

9-30 Consider again the data-entry decision as described in Section 9-5 and assume that the given decision rule will be used. Using the normal approximation, determine the type II error probability for incorrectly accepting H_0 (not recommending the new procedure when it might be an improvement) when single entry of data involves a true error rate of only
(a) .0040. (b) .0030. (c) .0025. (d) .0015.

9-31 A quality-control inspector must decide about a very large shipment of 10 ohm amplifier resistors. A sample of $n = 100$ resistors will be tested and the number of defectives determined. Depending on that result, the entire shipment will be purchased or returned. The inspector wants at most a 5% chance of purchasing a poor-quality batch (having 10% or more defectives).
(a) Using the number of defectives in the sample as the test statistic, perform hypothesis testing steps 1 and 3. Use the binomial distribution directly instead of the normal approximation.
(b) What action should be taken when the number of sample defectives is
(1) 4? (2) 6? (3) 5? (4) 7?

9-32 A statistician is testing the null hypothesis that exactly half of all engineers will have remained in the profession 10 years after receiving their bachelor's. He took a random sample of 200 graduates from the class of 1974 and determined their occupations in 1984. He found that 111 persons were still employed primarily as engineers.
(a) Construct a 95% confidence interval estimate for the proportion of engineers remaining in the profession. (Ignore the continuity correction.)
(b) At the 5% significance level, should the statistician accept or reject his hypothesis?

9-33 Consider fuses of a particular type that are inspected by both the producer and the consumer.
(a) The manufacturer defines as marketable any batch of fuses in which the proportion of defectives does not exceed .07. The inspection rules are based on the requirement that there should be no more than a 1% chance of scrapping any marketable batch. This outcome is referred to as the **producer's risk.** A sample of 500 fuses will be tested from a production run of 4,000. Perform hypothesis testing steps 1–3. What action should be taken if the sample is found to contain 8% defectives?
(b) One of the users of the fuses is a laboratory, where the policy is that any batch of fuses of a particular type is satisfactory if it contains less than 5% defectives. The inspection rules stipulate that there should be only a 5% chance of accepting an unsatisfactory batch. This outcome is referred to as the **consumer's risk.** A sample of 100 fuses is to be tested from a shipment of 500. Perform hypothesis testing steps 1–3. What action should be taken if the sample is found to contain 2 defectives?

9-34 The number of switches placed on line in a communications system depends on the level of traffic. At 20 random times over a brief interval it is determined what proportion of the time all switches are busy. An additional switch is then placed on line if this quantity exceeds a critical level. A similar rule is applied to determine when to remove one of the switches, which is done whenever the sample proportion falls below a smaller critical value.
(a) Assume that there should be only a 5% chance of unnecessarily adding a switch when all the switches are busy no more than 50% of the time.
(1) Perform hypothesis testing steps 1–3.

(2) What action should be taken if all switches were busy in 65% of the observed sample times?

(b) Assume that there should be only a 10% chance of incorrectly removing a switch when all the switches are busy at least 30% of the time.

(1) Perform hypothesis testing steps 1–3.

(2) What action should be taken if all switches were busy in 25% of the observed sample times?

9-6 TESTING THE VARIANCE

Although far less frequently encountered than tests of the mean or proportion, similar procedures may be applied to the variance. Recall that this parameter is important to situations where consistency, or the lack of it, is crucial. For example, although the mean queueing time is important to waiting persons, they feel more comfortable when they can closely predict how much time must be sacrificed. People are more tolerant of long lines that move at a fairly predictable pace—lines in which the variability in waiting time is small and the length of wait may be gauged by the number of persons in line.

Testing with the Chi-Square Statistic

Testing procedures for the variance are analogous to those for the parameters already discussed. As in those tests, the analogous sample statistic—in this case the sample variance s^2—will be the basis for accepting or rejecting a null hypothesis regarding σ^2. As we have seen, the chi-square distribution may be used to find probabilities for level of s. But it is more convenient instead to make the following transformation in getting the

CHI-SQUARE TEST STATISTIC FOR THE VARIANCE

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

The above random variable has the standard chi-square sampling distribution.

Hypotheses regarding σ^2 take the same form as those for μ. A pivotal level σ_0^2 is specified for the population variance. The following testing situations apply.

Lower-Tailed Test (H_0: $\sigma^2 \geq \sigma_0^2$):

Accept H_0 if $\chi^2 \geq \chi^2_{1-\alpha}$

Reject H_0 if $\chi^2 < \chi^2_{1-\alpha}$

Upper-Tailed Test (H_0: $\sigma^2 \leq \sigma_0^2$):

Accept H_0 if $\chi^2 \leq \chi^2_{\alpha}$

Reject H_0 if $\chi^2 > \chi^2_{\alpha}$

Two-Sided Test (H_0: $\sigma^2 = \sigma_0^2$):

Accept H_0 if $\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}$

Reject H_0 if $\chi^2 < \chi^2_{1-\alpha/2}$ or if $\chi^2 > \chi^2_{\alpha/2}$

Recall that the chi-square density is asymmetrical, and that the table entries all involve the $>$ orientation. For this reason the one-sided critical values are designated by the subscript $1 - \alpha$ in the lower tail and α in the upper tail, with analogous values for two-sided tests. The quantities may be read from Appendix Table H using $n - 1$ for the number of degrees of freedom.

ILLUSTRATION: Consistency in Chemical Reaction Times

A chemical engineer wants to determine if a low-pressure purification procedure for a synthetic fiber will result in an improvement over the present high-pressure process. Although the latter is believed considerably faster and has a low mean processing time, the time needed to complete a batch varies considerably—creating difficulties in scheduling plant facilities. The high-pressure process has taken as long as 112 hours to complete but has terminated as quickly as 33 hours, and the standard deviation has been established at $\sigma = 15$ hours. The chi-square procedure is based on the variance, which for the present process is $\sigma^2 = (15)^2 = 225$. It will be worthwhile to change to low-pressure processing if the new procedure provides a σ^2 half as large.

Using $\sigma_0^2 = 112.5$, the following hypothesis testing steps were taken.

STEP 1. *Formulate the hypotheses.* Wanting to protect most against the error of selecting the low-pressure process when in fact its processing times are not sufficiently consistent, the following apply:

H_0: $\sigma^2 \geq 112.5$ (The low-pressure process is too varied to justify a switch.)

H_1: $\sigma^2 < 112.5$ (The low-pressure process times are consistent enough to warrant switching.)

The above indicate that the test will be a lower-tailed one.

STEP 2. *Select the test statistic.* The chi-square statistic is used.

STEP 3. *Establish the significance level and the acceptance and rejection regions for the decision rule.* The engineer chooses a significance level of .01. A small value for s^2, and hence χ^2, will tend to refute the null hypothesis. The decision rule in Figure 9-12 was determined on the basis of a sample of 10 pilot batches processed at the lower pressure.

STEP 4. *Collect the sample data and compute the value of the test statistic.* The following processing times in hours were obtained:

35.4	44.7	58.9	28.6	37.5
49.1	52.6	55.3	36.2	50.0

From the above, the sample mean is $\overline{X} = 44.8$ hours and the sample variance is $s^2 = 99.28$. The computed value for the chi-square test statistic is

$$\chi^2 = \frac{(10 - 1)(99.28)}{112.5} = 7.94$$

STEP 5. *Make the decision.* The computed test statistic falls in the acceptance region. Although the computed variance is smaller than σ_0^2, it is not significantly

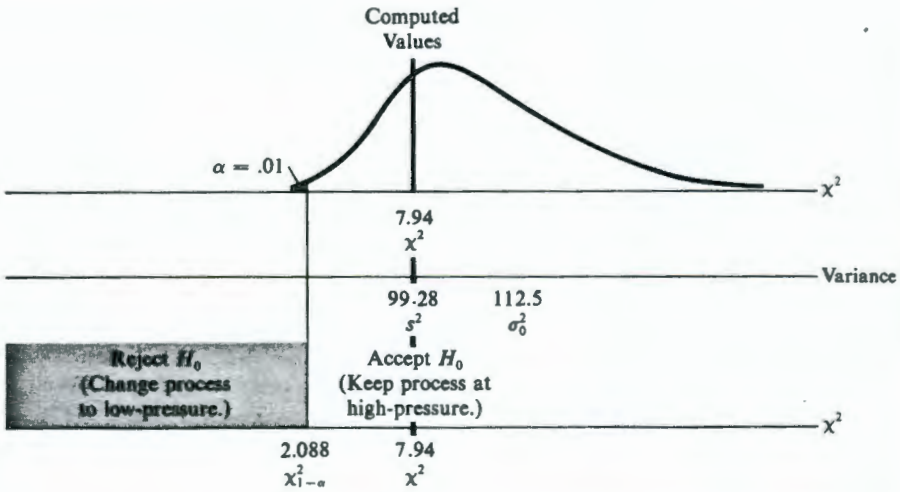


Figure 9-12

low enough to reject the null hypothesis. The engineer must therefore accept H_0 that the low-pressure process is not sufficiently consistent to warrant changes. The high-pressure procedure will be retained.

Testing with the Normal Approximation

Recall that the chi-square density becomes symmetrical for large sample sizes, approaching the shape of the normal curve. A rule of thumb is that the normal approximation may be used whenever $n > 30$. (Table H stops at 30 degrees of freedom.) In those cases the chi-square test statistic is transformed into the

NORMAL DEVIATE TEST STATISTIC FOR VARIANCE (LARGE n)

$$z = \frac{\chi^2 - (n-1)}{\sqrt{2n-1}} = \frac{(n-1)(s^2 - \sigma_0^2)}{\sigma_0^2\sqrt{2n-1}}$$

The critical normal deviate z_α applies in one-sided tests, and $z_{\alpha/2}$ is used in two-sided tests.

As an illustration, suppose that an industrial engineer wishes to test the null hypothesis that the standard deviation or variance in waiting times by partially completed assemblies at a particular station is the same as at another one where the standard deviation is known to be 10 minutes (σ_0). His null hypothesis is that $\sigma^2 = (10)^2 = 100$. For this two-sided test he chooses a significance level of $\alpha = .05$. A random sample of 100 assemblies are tagged and timed as they flow through the work station. The following decision rule applies.

The computed value of the sample standard deviation is $s = 12.5$, so that $s^2 = (12.5)^2 = 156.25$, and

$$z = \frac{99(156.25 - 100)}{100\sqrt{199}} = 3.95$$

which falls in the upper rejection region shown in Figure 9-13. The engineer concludes that waiting times at the test station involve greater variability than at the reference location.

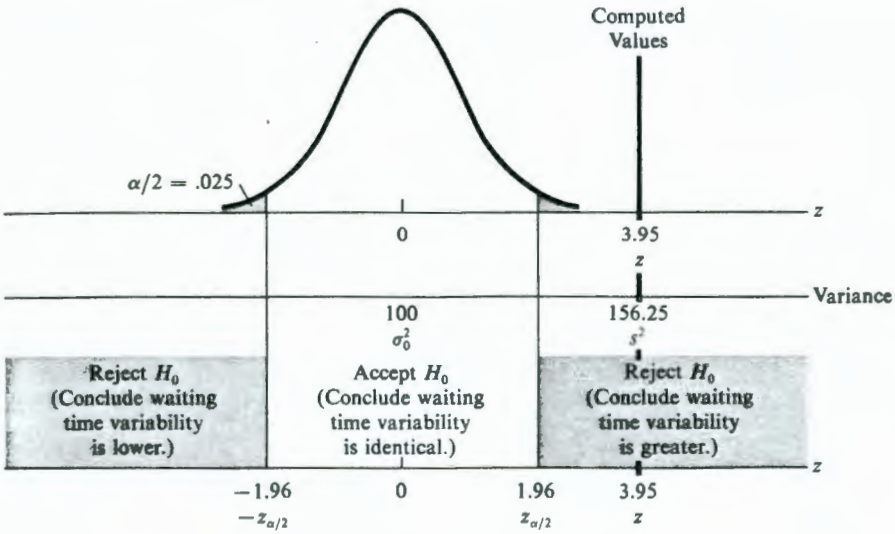


Figure 9-13

Problems

9-35 Determine for each of the following whether H_0 should be accepted or rejected. In each the sample size is $n = 20$.

(a)	(b)	(c)	(d)
H_0: $\sigma^2 \geq 100$	H_0: $\sigma^2 \leq 200$	H_0: $\sigma^2 = 150$	H_0: $\sigma^2 \leq 15$
$s^2 = 50$	$s^2 = 150$	$s^2 = 120$	$s^2 = 25$
$\alpha = .05$	$\alpha = .01$	$\alpha = .04$	$\alpha = .05$

9-36 An educational researcher believes that even though foreign engineering students on the average score lower than all students on a particular graduate achievement examination, their drastically different backgrounds should account for greater variability. As her null hypothesis she assumes that this is not so, and that the variance in foreign students' scores is $\leq 2{,}000$ (σ^2), the variance of all test scores. Using a random sample of 25 foreign-student scores, the computed variance is $s^2 = 2{,}749$. At a 5% significance level, can she conclude that foreign students' test scores are more varied than those of all students?

9-37 Refer to the physical-therapy-apparatus example on page 263. Should the mechanical engineer accept or reject the null hypothesis that $\sigma^2 = 100$? Use $\alpha = .10$.

9-38 Refer to the logic circuit wafer example on page 253. Should the industrial engineer accept or reject the null hypothesis that $\sigma^2 \leq .0015$? Use $\alpha = .01$.

9-39 An industrial engineer wishes to test the null hypothesis that the variance in waiting time by machinists at a tool cage is less than or equal to 25. A random sample of 100 machinist times were logged, and the sample variance was computed to be 41.4. At the 5% significance level, should the null hypothesis be accepted or rejected?

9-7 SELECTING THE TEST PROCEDURE

Our introduction to hypothesis testing would not be complete without discussing some of the considerations involved in selecting the procedure itself. One major concern is how efficiently the selected test makes use of available sample data. This is partly determined by the form of the hypotheses and the particular test statistic employed. In comparing procedures, the better test is the one providing lower probabilities for the two types of errors. Ordinarily, this aspect may be improved by increasing the sample size, but investigators usually are required to keep sampling expenses within budgetary limitations.

Some Important Questions

Several fundamental questions must be resolved in planning a statistical testing experiment.

What particular test should be used? The choice of test depends partly on the kind of decision to be made. Often this involves comparing a new procedure with an existing method, and the objective is to select the better way. Often quality may be expressed in terms of a mean μ, the better method involving the larger or smaller level. But there may be other ways to measure the same things. For example, instead of using the mean copper yield as the basis for evaluating the bacterial leaching decision at the outset of the chapter, the mining engineer could conceivably use the metabolizing rate or some other biological yardstick. A queueing evaluation might be based on the mean waiting time, but the determining factor might instead be the proportion of time the service facility is busy. In many applications—for instance, work measurement—either the proportion or the mean might serve as the measure of effectiveness; the test statistic $\bar{X}$ or P coincides. The better test—the more efficient one—will be the one that provides lower error probabilities for the same cost. Often, two or more populations must be compared when the parameters of each are unknown. More complex controlled experiments are then necessary; we will encounter procedures for evaluating some of these in later chapters. In some cases the populations are better represented by **nonparametric tests.**

How do we evaluate a test in terms of its protection against erroneous decisions? Both the type I error (rejecting H_0 when it is true) and the type II error (accepting H_0 when it is false) are undesirable. But we must tolerate some probability for each as the unavoidable price of basing our choice on a sample. Ordinarily the probability (α) of the more serious type I error is specified in advance, and the decision rule is selected to match. But if the resulting rule provides unacceptably large chances (β's) for the type II error, an investigator may choose to revise it. One useful tool employed by statisticians for this purpose is the operating characteristic curve.

Are there any underlying assumptions that might limit the scope of a testing procedure? A test statistic may have limited validity arising from the probability theory and assumptions used to derive its sampling distribution. For instance, we have been using the normal curve to find error probabilities for $\bar{X}$ and to set decision rules. But doing so is nearly always an approximation. The quality of that approximation may be very poor for small sample sizes (and it may be totally inappropriate for certain unusual populations). When n is small

and σ is unknown, more accurate procedures might be based on the Student t distribution. But even then, the population itself is theoretically assumed to be normal (which is usually not the case). Similar problems arise with the chi-square distribution. When the theoretical underpinnings of a traditional procedure are seriously violated by circumstances, statisticians may substitute a nonparametric test that presupposes very little about the sampled population. Fortunately, such violations often are not serious; a procedure that is workable anyway is **robust** with respect to a violation of that assumption.

The Operating Characteristic Curve

In all the testing situations encountered so far, there is a range of possible values for the decision parameter. When sufficient information is known about the underlying population, it is possible to compute the conditional probability for accepting H_0 (or rejecting H_0) for any given value in that range. It may be useful to plot several of these probabilities on a graph, as in Figure 9-14, which provides the **operating characteristic curve** (or OC curve) for the bacterial leaching decision from the beginning of the chapter.

The ordinate provides $\Pr[\text{accept } H_0 \mid \mu]$ for levels of the population mean copper yield μ on the abscissa. For any population mean > 36 $(= \mu_0)$ pounds per ton, the null hypothesis is false and the process is economical. The corresponding points on the curve are the type II error probabilities β. When $\mu = \mu_0$, the probability of accepting H_0 is exactly $1 - \alpha$. Below 36, the probability of accepting H_0 (then a correct action) approaches 1. Notice that the β's are quite large, even when μ considerably exceeds the break-even level of 36. These indicate that the mining engineer faces a high probability of incorrectly accepting a

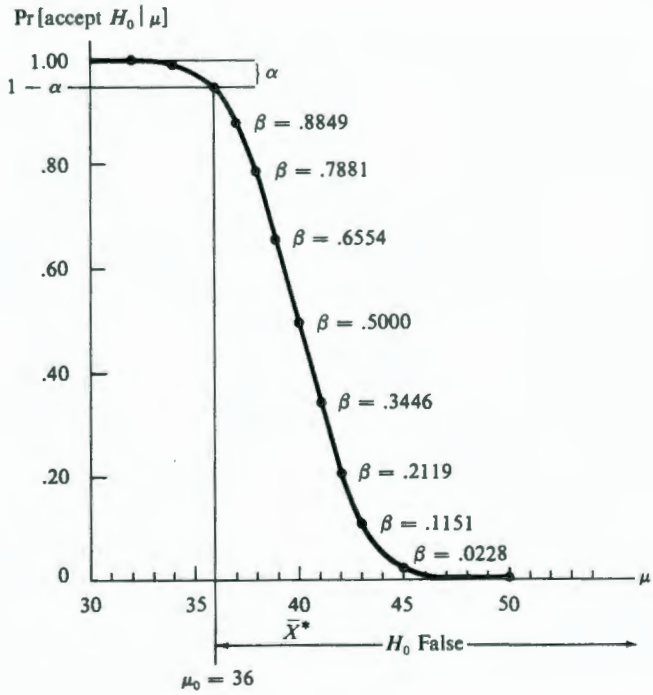


Figure 9-14 Operating characteristic curve for bacterial leaching decision.

false null hypothesis—not recommending an economical bacterial leaching procedure. This stems from the choice of decision rule, with the critical value of $\overline{X}^* = 40$.

If the mining engineer finds these error probabilities too high, he has two choices. (1) He may drop the critical value to a lower level. The effect of this would be to shift the OC curve to the left, but that results in a lower "accept" probability when $\mu = 36$, reducing $1 - \alpha$ and raising α. This increase in the type I error probability may be even more unappealing than the original β's. (2) A second remedy would be to increase the sample size. Although this increases the cost of the experiment, the reduction in error probabilities may be worthwhile.

Figure 9-15 shows the original OC curve for $n = 100$ and the OC curve when the sample size is raised to $n = 200$. Notice that except where the curves cross, $n = 200$ provides higher probabilities for accepting H_0 for low levels of μ and lower probabilities for high μ's. Both curves achieve the same .50 accept probability when $\mu = \overline{X}^*$. The limiting case is provided by the rectangular portion, where n is infinite and there is no uncertainty. (When the decision rule is shifted and $\overline{X}^* = \mu_0$, no error can occur in that limiting case.)

Figure 9-16 shows the shape of OC curves for the three types of test orientations encountered in this chapter. Some statisticians prefer to place the probability of rejecting on the ordinate, which provides the **power curve.** That display provides exactly the same information as the OC curve.

The OC curve may prove to be a valuable tool for evaluating a proposed test and decision rule. But as a practical matter, it is not always possible to construct such a curve. There may be many unspecified ways for the type II error to occur, and it may be difficult or impossible to compute β for each. For instance, unless σ is known, the β probabilities in testing μ are indeterminate.

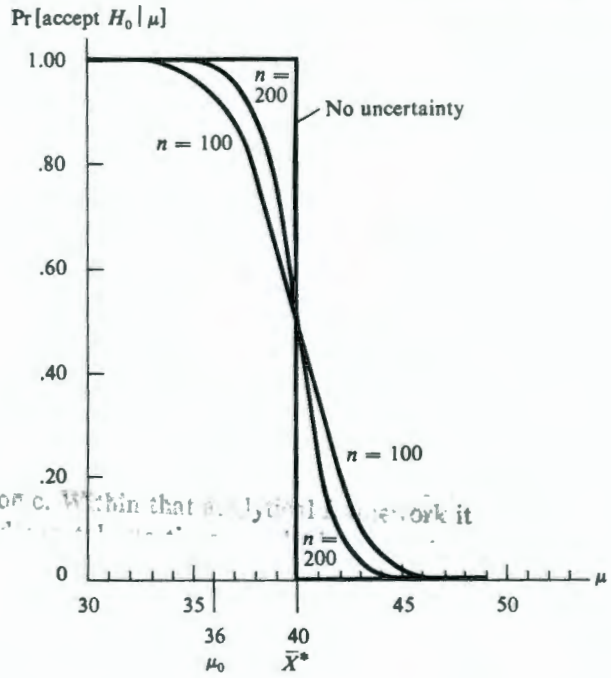


Figure 9-15 Bacterial leaching OC curves when $n = 100$ and $n = 200$. Also shown is the limiting case, when there is no uncertainty.

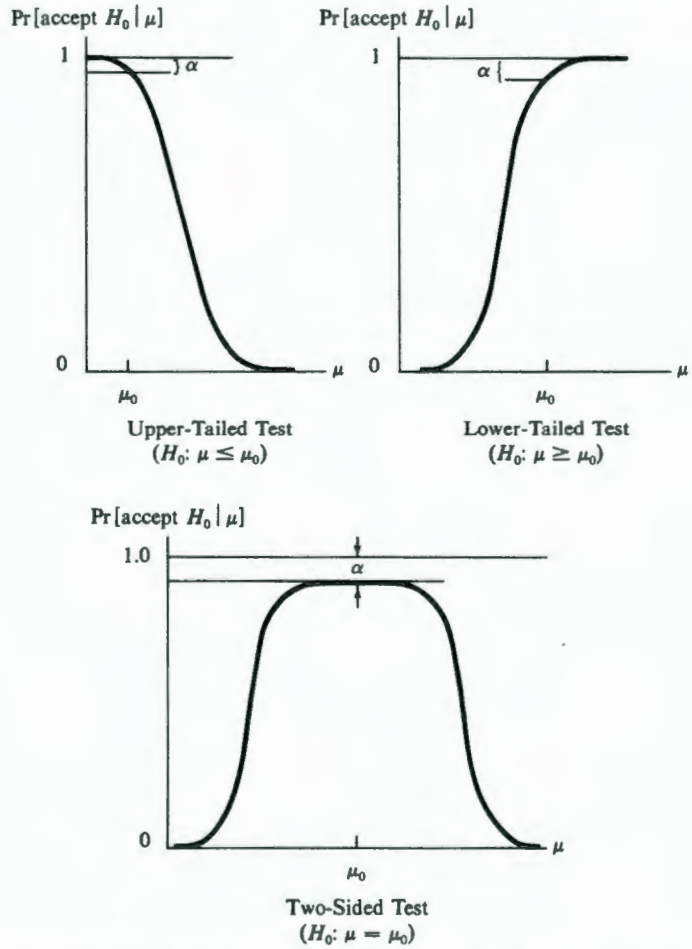


Figure 9-16 Basic OC curve shapes.

Some procedures, such as the nonparametric ones, do not even have abscissa units for constructing a graph.

Limitations of Hypothesis Testing

Hypothesis testing as described in this chapter was originally designed for applications in biology and medicine. Within budgetary limitations, avoidance of error is an overriding consideration in those fields. Thus, choosing a procedure that limits the incidence of the more serious error is perhaps the best decision-making criterion. But that criterion may not be best for decision-making in other areas.

Later in this book we will encounter Bayesian decision-making procedures where a wider focus is made. Error is just one aspect of Bayesian evaluations, which are based on a broader **statistical decision theory.** These more general procedures allow us to establish probabilities for the level of μ (or π). It is possible to expand the dimensions of the statistical evaluation to consider a payoff measure that quantifies each outcome. Within that analytical framework it is even possible to consider such fundamental questions as whether or not to sample at all. (It may be better to decide without the bother and benefit of sampling!)

Problems

9-40 Suppose that the mining engineer in the bacterial-leaching illustration changes his significance level to $\alpha = .01$ and uses a sample size of 150 batches.

(a) Determine his new decision rule when $\overline{X}$ is the test statistic.

(b) Construct an operating-characteristic curve using the decision rule from (a).

9-41 Construct the OC curve for the biological reduction process decision in Problem 9-13.

9-42 Construct the OC curve for the rod decision in Problem 9-17.

9-43 Construct the OC curve for the data processing decision illustrated on page 305.

Review Problems

9-44 For each of the following situations, (1) specify the test statistic, (2) give its critical value, and (3) indicate whether the test is lower-tailed, upper-tailed, or two-sided. In each case use $\alpha = .05$ and assume a large population unless otherwise indicated.

(a) H_0: $\mu \leq 50, \sigma = 10, n = 100$

(b) H_0: $\mu \geq 100, \sigma = ?, n = 100$

(c) H_0: $\mu = 75, \sigma = ?, n = 20$

(d) H_0: $\pi \leq .20, n = 100$

(e) H_0: $\pi \geq .40, n = 50, N = 200$

(f) H_0: $\mu \leq .50, \sigma = .15, n = 25, N = 150$

(g) H_0: $\mu \geq 1.46, \sigma = ?, n = 25$

(h) H_0: $\mu = 25, \sigma = 2, n = 200, N = 1{,}000$

9-45 The following data (minutes) apply for a random sample of processing times for a chemical reaction:

2.3	6.7	3.8	5.0	4.9	6.1
4.4	5.2	3.9	4.8	4.6	5.7
5.3	4.7	4.2	4.7	5.7	4.8

Can you conclude that the mean of all processing times exceeds 4 minutes? Use $\alpha = .05$.

9-46 The following data were obtained for the amount of time (seconds) taken by a proposed computer system to compile a sample of short FORTRAN programs:

.6	1.0	1.8	4.8	4.2
4.4	2.4	3.5	3.9	3.9
4.7	3.0	2.3	2.9	1.0

(a) At the 5% significance level, test the null hypothesis that the mean compilation time for all short FORTRAN programs run on the system is at least 4 seconds. Should it be accepted or rejected?

(b) At the 5% significance level, test the null hypothesis that the standard deviation in compilation time is less than or equal to .9 second ($\sigma^2 \leq .81$). Should it be accepted or rejected?

9-47 A pharmaceutical loader has been set to insert exactly 5 milligrams into each capsule of a particular drug. Periodically a sample of 100 capsules is taken and the contents of each are measured. The standard deviation has been established at .2 milligram per capsule. Corrective action will be taken whenever the computed mean is significantly great or small.

(a) Determine a decision rule when a 1% chance is permitted for taking unnecessary action.

(b) Applying the above, what is the probability of failing to take needed correction when the true mean is (1) 5.05 milligrams? (2) 4.97 milligrams? (3) 5.07 milligrams? (4) 4.98 milligrams?

9-48 A program director for a large computer maker is evaluating a circuit-designing software package. Not only is the program purported to ease the engineers' tasks, but its author also claims that the finished products will be more economical because fewer logic elements will be needed. To test this theory, the software was borrowed and used to redesign 20 prototype circuits already completed the traditional way. The resulting computer-generated designs all worked and involved a mean percentage reduction in components of 3.4% ($s = 4.2\%$).

(a) At the 1% significance level, can the project manager reject the null hypothesis that the software designs involve at least as many elements as do traditional ones?

(b) What is the smallest significance level at which the null hypothesis can be rejected?

9-49 Refer to the data in Problem 9-45. What is the lowest significance level at which you can conclude that the variance in processing time exceeds .25 minute^2?

9-50 Refer to the data in Problem 9-46.

(a) Construct a 99% confidence interval estimate for the mean. Should the null hypothesis that $\mu = 3.50$ minutes be accepted or rejected at the 1% significance level?

(b) Construct a 98% confidence interval estimate for the variance. Should the null hypothesis that $\sigma^2 = 1.0$ be accepted or rejected at the 2% significance level?

9-51 Only 18 persons in a sample of 100 students indicated a desire for a full-scale summer program. Can it be concluded that the majority of students don't want such a program? Use $\alpha = .01$ for incorrectly making that conclusion. Assume a large population and apply the normal approximation.

9-52 Specifications for a component require that the MTBF exceed 100 hours. The actual MTBF of any shipment is unknown, although it is assumed for each that the lifetime standard deviation is 10 hours. There should be only a 5% chance of concluding that a shipment meets specifications when it does not. A statistician takes a random sample of 50 items from a particular shipment of 300 and computes the MTBF to be 102 hours. What conclusion does he reach?

9-53 The failure rate of jet engine fan blades subjected to a particular test condition has been specified to be less than .05. An entire production batch found to violate that requirement will be scrapped; otherwise they will be assembled into engines. Assuming just a 10% chance of erroneously assembling with a poor-quality batch of fan blades, determine the maximum number of failed blades in a sample of 100 that would still allow the batch to be used in assembly. (Use the binomial distribution directly, assuming the batch size is large.)

9-54 It is sometimes possible to find a required sample size so that the targeted levels for both α and β may be specified. Consider a one-sided hypothesis test of the mean where σ is known and N is large.

(a) Letting μ_1 represent the assumed level for μ for which β is specified and z_β the corresponding normal deviate, derive an expression for finding the required sample size.

(b) Find the level for n that should be used in each of the following:

(1) $\mu_0 = 90,\ \mu_1 = 100,\ \sigma = 25,\ \alpha = .05,\ \beta = .10$
(2) $\mu_0 = 50,\ \mu_1 = 30,\ \sigma = 15,\ \alpha = .01,\ \beta = .05$
(3) $\mu_0 = 1.52,\ \mu_1 = 1.55,\ \sigma = .5,\ \alpha = .05,\ \beta = .10$
(4) $\mu_0 = 1{,}100,\ \mu_1 = 1{,}000,\ \sigma = 100,\ \alpha = .10,\ \beta = .20$

9-55 In a similar fashion to Problem 9-54 the required sample size may be found for testing the proportion when N is large and the normal approximation applies.

(a) Letting π_1 denote the assumed level for the population proportion for which β is specified and z_β the corresponding normal deviate, derive an expression for finding the required sample size. (Ignore the continuity correction.)

(b) Find the level for n that should be used in each of the following:

(1) $\pi_0 = .15$, $\pi_1 = .20$, $\alpha = .05$, $\beta = .10$

(2) $\pi_0 = .10$, $\pi_1 = .05$, $\alpha = .01$, $\beta = .05$

(3) $\pi_0 = .50$, $\pi_1 = .60$, $\alpha = .10$, $\beta = .15$

(4) $\pi_0 = .51$, $\pi_1 = .49$, $\alpha = .05$, $\beta = .10$