PROBABILITY AND STATISTICS FOR ENGINEERS

IRWIN MILLER
ARTHUR D. LITTLE, INC.

AND

JOHN E. FREUND
PROFESSOR OF MATHEMATICS
ARIZONA STATE UNIVERSITY

Prentice-Hall Mathematics Series
DR. ALBERT A. BENNETT, EDITOR

PRENTICE-HALL, INC., ENGLEWOOD CLIFFS, NEW JERSEY

PRENTICE-HALL INTERNATIONAL, INC. *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD. *Sydney*
PRENTICE-HALL OF CANADA, LTD. *Toronto*
PRENTICE-HALL OF INDIA (PRIVATE) LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC. *Tokyo*

Library of Congress Catalog Card Number 65–10075

Current printing (last digit):
13 12 11

PRINTED IN THE UNITED STATES OF AMERICA, *C71166*

8.4 Hypotheses Concerning One Mean

In this section we shall consider more generally the problem of testing the hypothesis that the mean of a population equals a specified value against a suitable alternative; that is, we shall test

$$H_0\colon \mu = \mu_0$$

against one of the alternatives

$$H_1\colon \mu < \mu_0, \quad H_1\colon \mu > \mu_0, \quad \text{or} \quad H_1\colon \mu \neq \mu_0$$

and the critical region we shall use will be of the form $\bar{x} < C$, $\bar{x} > C$, or $\bar{x} < C_1$ *or* $\bar{x} > C_2$, respectively. Since none of these alternative hypotheses actually specifies a unique value of μ, it is impossible to compute β (the probability of a Type II error) for any of these tests, and it would seem reasonable to describe them as tests of whether $\bar{x}$ is *significantly less* than μ_0, *significantly greater* than μ_0, or *significantly different* from μ_0.

A test like this, in which the probability of a false acceptance of H_0 cannot be uniquely determined, is commonly called a *significance test.* The probability α of a Type I error, also called the *level of significance*, can be calculated because μ is uniquely specified by H_0, and the rejection of H_0 is "safe" in this sense. On the other hand, there is a danger inherent in the acceptance of H_0 because the probability of its false acceptance cannot be obtained. Thus, whenever possible, the hypothesis H_0 is chosen so that we shall be willing to "reserve judgment" about its validity, unless there is clear evidence that leads to its rejection. Also for this reason, H_0 will be called a *null hypothesis;* it is set up as a "straw man" with the objective of determining whether or not it can be rejected.

The idea of setting up a null hypothesis is not an uncommon one, even in nonstatistical thinking. In fact, this is exactly what is done in an American court of law, where an accused is assumed to be innocent unless he is proven guilty "beyond a reasonable doubt." The null hypothesis states that the accused is *not guilty*, and the probability expressed subjectively by the phrase "beyond a reasonable doubt" leads to the level of significance α. Thus, the "burden of proof" is always on the prosecution in the sense that the accused is found not guilty unless the null hypothesis of innocence is clearly disproved. Note that this does not imply that the defendant has been proved innocent if found not guilty; it implies only that he has not been proved guilty. Of course, since we cannot legally "reserve judgment" if proof of guilt is not established, the accused is freed and we act as if the null hypothesis of innocence were accepted. Note that this is what we do sometimes in tests of statistical hypotheses, when we cannot afford the luxury of reserving judgment.

To establish a parallel between this argument and the kind of practical problem in which significance tests are ordinarily applied, let us consider the following examples. Suppose a decision has to be made whether to buy an automatic welding machine to do a job formerly done by hand, and it is felt that the machine would be economical only if μ, the average number of defective welds per hundred, were less than 5. Thus, we have a situation which calls for a test of the null hypothesis H_0: $\mu = 5$, although we are not told directly which alternative hypothesis to use. If the burden of proof is to be placed on the machine, it would be appropriate to test H_0 against the *one-sided alternative* H_1: $\mu < 5$, installing the welder only if it can be proved statistically that it produces less than 5 defectives per 100 welds. Note that if $\alpha = 0.05$, there is only a 5 per cent chance of erroneously rejecting H_0 and installing the automatic welder. On the other hand, if the burden of proof were placed on the existing method, the one-sided alternative H_1: $\mu > 5$ would be appropriate. In this case H_0 would be accepted and the automatic welder installed *unless* it were proved (say, with $\alpha = 0.05$) that the automatic welder produces too many defectives. The choice of which one-sided alternative to use in this and in similar situations is a practical rather than a statistical problem; it simply depends on where we wish to place the burden of proof.

To give an example leading to a *two-sided alternative*, suppose a canner wants to test whether the correct amount of fruit juice is being packed in his 20-ounce cans. Since the label reads "20 ounces," the canner cannot afford to pack much less than 20 ounces for fear of losing customer acceptance or running afoul of the law, nor can he afford to pack much more than 20 ounces for fear of losing a substantial part of his profit. Thus, the canner is concerned with the two-sided alternative H_1: $\mu \neq 20$, and the can-filling process will be left undisturbed unless the mean weight of the contents of a sample of cans is significantly different from 20 ounces.

Returning now to the general problem of testing the null hypothesis $\mu = \mu_0$, we find that this problem has already been solved in Section 8.3, provided the sample size is large and the population standard deviation is known. In that case the critical region is given by $\bar{x} < C$ if the alternative hypothesis is $\mu < \mu_0$, $\bar{x} > C$ if the alternative hypothesis is $\mu > \mu_0$, and $\bar{x} < C_1$ *or* $\bar{x} > C_2$ if the alternative hypothesis is $\mu \neq \mu_0$, and the formulas for computing C, C_1, and C_2 are given on pages 154 and 158. An equivalent, but simpler, method of specifying the critical region is to base it on the statistic

◆ $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$ ◆

instead of $\bar{x}$. If the level of significance is α, and z_α is, as before, such that the area under the standard normal curve to its right is equal to α, the critical regions for testing H_0: $\mu = \mu_0$ can be expressed as in the following table:

CRITICAL REGIONS FOR TESTING H_0: $\mu = \mu_0$
(Large Sample, σ Known)

Alternative hypothesis	*Reject H_0 if*
$\mu < \mu_0$	$z < -z_\alpha$
$\mu > \mu_0$	$z > z_\alpha$
$\mu \neq \mu_0$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

To illustrate, let us return to the problem concerning the automatic welder. The null hypothesis is $\mu = 5$ and we shall use the alternative $\mu < 5$, putting the burden of proof on the automatic welder. Suppose that the decision is to be based on a sample of 64 sets, each of which contains 100 welds, and that the mean and the standard deviation of the number of defective welds per set are, respectively, 4.8 and 1.2. Although σ is actually unknown, the sample is large enough to approximate it with $s = 1.2$, and we thus obtain

162-177

$$z = \frac{4.8 - 5}{1.2/\sqrt{64}} = -1.33$$

If the level of significance is to be $\alpha = 0.05$, we find from Table III that the *critical value* is $-z_{.05} = -1.645$; since the calculated value of z is not less than -1.645, the null hypothesis cannot be rejected and we decide, in fact, that the machine is not to be installed. (The reader will be asked to graph the OC curve of this test in Exercise 9 on page 172.)

If the sample size is small and σ is unknown, the tests just described cannot be used. However, if the sample comes from a normal population (to within a reasonable degree of approximation), we can make use of the theory discussed in Section 7.3 and base the test of the hypothesis H_0: $\mu = \mu_0$ on the statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

The resulting critical regions are as shown in the table on page 165. In this table t_α is as defined on page 137 (the area to its right under the t distribution with $n - 1$ degrees of freedom is equal to α).

To illustrate, let us reconsider the problem of deciding whether changes have to be made in the fruit-juice canning process, namely, the problem in which the null hypothesis $\mu = 20$ is to be tested against the alternative hypothesis $\mu \neq 20$. Suppose that the level of significance is to be $\alpha = 0.01$, and that the net weights of the contents of a sample of 25 cans have a

Testing H_0: $\mu = \mu_0$
, σ Known)

Reject H_0 if
$z < -z_\alpha$
$z > z_\alpha$
$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

problem concerning the automatic
5 and we shall use the alternative
the automatic welder. Suppose that
le of 64 sets, each of which contains
standard deviation of the number of
y, 4.8 and 1.2. Although σ is actually
to approximate it with $s = 1.2$, and

$$= -1.33$$

$= 0.05$, we find from Table III that
since the calculated value of z is not
s cannot be rejected and we decide,
installed. (The reader will be asked
Exercise 9 on page 172.)
is unknown, the tests just described
ple comes from a normal population
oximation), we can make use of the
e the test of the hypothesis H_0: $\mu = \mu_0$

$$- \mu_0 \over /\sqrt{n}$$

shown in the table on page 165. In
.37 (the area to its right under the
reedom is equal to α).
problem of deciding whether changes
ning process, namely, the problem in
to be tested against the alternative
level of significance is to be $\alpha = 0.01$,
tents of a sample of 25 cans have a

Critical Regions for Testing H_0: $\mu = \mu_0$
(Normal Population, σ Unknown)

Alternative hypothesis	*Reject H_0 if for $n - 1$ degrees of freedom*
$\mu < \mu_0$	$t < -t_\alpha$
$\mu > \mu_0$	$t > t_\alpha$
$\mu \neq \mu_0$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

mean of $\bar{x} = 20.03$ and a standard deviation of $s = 0.04$ ounces. To decide whether to adjust the process, we calculate

$$t = \frac{20.03 - 20}{0.04/\sqrt{25}} = 3.75$$

and since this exceeds 2.797, the value of $t_{.005}$ with 24 degrees of freedom (see Table IV), the null hypothesis will have to be rejected. (It is difficult to graph the *OC* curve for this test, because the sampling distribution of the test statistic is not the t distribution unless $\mu = 20$. However, in the *Biometrika Table* mentioned in the Bibliography there is a special table from which the necessary probabilities can be obtained.)

In spite of the result obtained in this test, the manufacturer may not wish to adjust his machinery, since the loss due to overfilling the cans by a very small amount may actually be less than the cost of experimenting with adjustments. This illustrates the important fact that a result which is *statistically significant* may not be *commercially significant*. Under the circumstances, it might be more appropriate to test the null hypothesis $\mu = 20$ against an alternative such as $\mu < 19.95$ or $\mu > 20.05$, if it is felt that either case will definitely call for an adjustment.

8.5 Hypotheses Concerning Two Means

When dealing with population means, we are frequently faced with the problem of making decisions about the relative values of two or more means. Leaving the general problem until Chapter 13, we shall devote this section to tests concerning the difference between *two* means. For example, if two kinds of steel are being considered for use in certain structural steel beams, we may take samples and decide which is better by comparing their mean strengths; also, if an achievement test is given to a group of industrial engineers and to a group of civil engineers, we may want to decide whether any observed difference between the means of their scores is significant or whether it may be attributed to chance.

Formulating the problem more generally, we shall consider two populations having the means μ_1 and μ_2 and the variances σ_1^2 and σ_2^2, and we shall want to test the null hypothesis $\mu_1 - \mu_2 = \delta$, where δ is a specified constant, on the basis of independent random samples of size n_1 and n_2. Analogous to the tests concerning one mean, we shall consider tests of this null hypothesis against each of the alternatives $\mu_1 - \mu_2 < \delta$, $\mu_1 - \mu_2 > \delta$, and $\mu_1 - \mu_2 \neq \delta$. The test, itself, will depend on the difference between the sample means, $\bar{x}_1 - \bar{x}_2$, and if both samples are large and the population variances are known, it can be based on the statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

whose sampling distribution is (approximately) the standard normal distribution. Here $\sigma_{\bar{x}_1 - \bar{x}_2}$ is the standard deviation of the sampling distribution of the difference between the sample means, and its value for random samples from infinite populations may be obtained with the use of the following theorem, which we shall state without proof:

THEOREM 8.1. *If the distributions of two independent random variables have the means μ_1 and μ_2 and the variances σ_1^2 and σ_2^2, then the distribution of their* sum (*or* difference) *has the mean* $\mu_1 + \mu_2$ (or $\mu_1 - \mu_2$) *and the variance* $\sigma_1^2 + \sigma_2^2$.

To find the variance of the difference between the means of two independent random samples of size n_1 and n_2 from infinite populations, note first that the variances of the two means, themselves, are

$$\sigma_{\bar{x}_1}^2 = \frac{\sigma_1^2}{n_1} \qquad \text{and} \qquad \sigma_{\bar{x}_2}^2 = \frac{\sigma_2^2}{n_2}$$

where σ_1^2 and σ_2^2 are the variances of the respective populations. Thus, by Theorem 8.1

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

and the test statistic can be written as

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Analogous to the table on page 164, the critical regions for testing the null hypothesis H_0: $\mu_1 - \mu_2 = \delta$ are as follows:

162-174

CRITICAL REGIONS FOR TESTING H_0: $\mu_1 - \mu_2 = \delta$
(Large Samples, σ_1 and σ_2 Known)

Alternative hypothesis	*Reject H_0 if*
$\mu_1 - \mu_2 < \delta$	$z < -z_\alpha$
$\mu_1 - \mu_2 > \delta$	$z > z_\alpha$
$\mu_1 - \mu_2 \neq \delta$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

To illustrate this kind of test, suppose that an achievement test is given to 50 industrial engineers (Group 1) and to 60 civil engineers (Group 2), and that the results are as follows:

$$\bar{x}_1 = 89, \qquad s_1 = 7$$
$$\bar{x}_2 = 87, \qquad s_2 = 5$$

If we wish to test at the 0.05 level of significance whether the observed difference of 2 points between the two means is significant or whether it can be attributed to chance, the appropriate null hypothesis and alternative hypothesis are H_0: $\mu_1 - \mu_2 = 0$ and H_1: $\mu_1 - \mu_2 \neq 0$. Accordingly, we put $\delta = 0$ in the formula for z and the test statistic becomes

$$z = \frac{89 - 87}{\sqrt{\frac{49}{50} + \frac{25}{60}}} = 1.69$$

(Note that we have approximated the population variances with s_1^2 and s_2^2, which is justifiable since both samples are fairly large.) Since the value which we obtained for the test statistic lies between the critical values of -1.96 and 1.96, the null hypothesis cannot be rejected; thus, we conclude that the observed difference between the means is *not significant* at the 0.05 level or, in other words, that it can well be attributed to chance.

If either (or both) samples are small and the population variances are unknown, we can base tests of the null hypothesis H_0: $\mu_1 - \mu_2 = \delta$ on a suitable t statistic, provided it is reasonable to assume that *both populations are normal with* $\sigma_1 = \sigma_2$. Under these conditions it can be shown that the sampling distribution of the statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s_{\bar{x}_1 - \bar{x}_2}}$$

is the t distribution with $n_1 + n_2 - 2$ degrees of freedom. In this formula the denominator involves a "pooled estimate" of the population variance.

To clarify what we mean here by a "pooled estimate" of the population variance, let us first consider the problem of estimating the variance of the distribution of the difference between two sample means. Under the assumption that $\sigma_1^2 = \sigma_2^2\ (= \sigma^2)$, this variance is given by

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

and we now estimate σ^2 by "pooling" the two sums of squared deviations from the respective sample means. In other words, we estimate σ^2 by means of

$$\frac{\Sigma\,(x_1 - \bar{x}_1)^2 + \Sigma\,(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where $\Sigma\,(x_1 - \bar{x}_1)^2$ is the sum of the squared deviations from the mean for the first sample, while $\Sigma\,(x_2 - \bar{x}_2)^2$ is the sum of the squared deviations from the mean for the second sample. We divide by $n_1 + n_2 - 2$, since there are $n_1 - 1$ independent deviations from the mean in the first sample, $n_2 - 1$ in the second, and we thus have $n_1 + n_2 - 2$ independent deviations from the mean to estimate the population variance. Substituting this estimate of σ^2 into the above expression for $\sigma^2_{\bar{x}_1 - \bar{x}_2}$ and then substituting the square root of the result into the denominator of the formula for t on page 167, we finally obtain

$$\blacklozenge \qquad t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}\sqrt{\frac{n_1 n_2(n_1 + n_2 - 2)}{n_1 + n_2}} \qquad \blacklozenge$$

for the statistic on which we shall base the test. The corresponding critical regions for testing the null hypothesis H_0: $\mu_1 - \mu_2 = \delta$ are as shown in the following table:

CRITICAL REGIONS FOR TESTING H_0: $\mu_1 - \mu_2 = \delta$
(Normal Populations, $\sigma_1 = \sigma_2 = \sigma$, σ Unknown)

Alternative hypothesis	*Reject H_0 if for $n_1 + n_2 - 2$ degrees of freedom*
$\mu_1 - \mu_2 < \delta$	$t < -t_\alpha$
$\mu_1 - \mu_2 > \delta$	$t > t_\alpha$
$\mu_1 - \mu_2 \neq \delta$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

To illustrate this kind of test, let us assume that a sample of 10 steel beams from Mill A has a mean tensile strength of 54,000 psi with a standard deviation of 2100 psi, and that a sample of 12 beams from Mill B has a mean tensile strength of 49,000 psi with a standard deviation of 1900 psi. The beams from Mill B cost less than those from Mill A, and we are inclined to buy from Mill B unless the beams from Mill A are at least 2000 psi stronger on the average than those from Mill B. Consequently, we shall test the null hypothesis $H_0: \mu_A - \mu_B = 2000$, against the one-sided alternative $H_1: \mu_A - \mu_B > 2000$, and we shall choose a level of significance of $\alpha = 0.01$. The value of the test statistic is

$$t = \frac{(54{,}000 - 49{,}000) - 2000}{\sqrt{9(2100)^2 + 11(1900)^2}} \sqrt{\frac{10(12)(20)}{22}} = 3.52$$

and since this exceeds 2.528, the value of $t_{.01}$ for 20 degrees of freedom, the null hypothesis will have to be rejected and the beams purchased from Mill A. (Note that by choosing the alternative hypothesis $\mu_A - \mu_B > 2000$, we place the burden of proof on Mill A.)

In this last example we arbitrarily went ahead and performed a *two-sample t test,* tacitly assuming that the population variances were equal. Fortunately, the test is not overly sensitive to small differences between the population variances, and the procedure used in this instance is quite justifiable. To be on safer grounds, however, we should first have tested whether the difference between the sample variances may be attributed to chance; a procedure for performing such a test will be given in Chapter 9.

If the difference between the sample variances is large or if it is otherwise unreasonable to treat the population variances as being equal, we cannot use the two-sample t test just described. However, there are several alternative methods that can be used instead, which do not require the assumption of equal population variances. One of these, the *paired-sample t test,* applies to two random samples of the same size, which need not be independent. Briefly, the procedure is to work with the differences of paired observations, where the first member of each pair comes from the first sample and the second member comes from the second sample, and to use the one-sample t test described in Section 8.4 to determine whether the mean of the differences is significantly different from δ. Sometimes, as in the case where two examinations are given to each of n persons, the pairing is "natural"; in all other cases the pairing should be random.

To illustrate the paired-sample t test, suppose that a dye is to be tested for resistance to fading by exposing 8 dyed specimens of various kinds to sunlight for a specified period of time. The reflectivity of light of the same color as the dye is measured for each specimen (in arbitrary units) before and after exposure to sunlight, and it will be concluded that the dye is *not* resistant to fading if the difference in reflectivity indices is significantly greater than 1. The following are the results obtained in this experiment:

	Before exposure x_1	*After exposure* x_2
Specimen 1	19	14
Specimen 2	5	4
Specimen 3	24	20
Specimen 4	8	8
Specimen 5	10	9
Specimen 6	11	9
Specimen 7	7	5
Specimen 8	16	15

The differences between these paired observations are 5, 1, 4, 0, 1, 2, 2, 1, their mean is 2.00, and their standard deviation is 1.69. Assuming that the differences may be treated as a sample from a normal population with $\mu_0 = 1$, the test statistic for the one-sample t test has the value

162-174

$$t = \frac{2.00 - 1}{1.69/\sqrt{8}} = 1.67$$

If the level of significance is to be 0.05, we find that $t_{.05}$ for 7 degrees of freedom equals 1.895 and, hence, that the null hypothesis cannot be rejected. We could conclude that the dye is resistant to fading or we could reserve judgment until more data are obtained.

Although this paired-sample t test can be used when sampling from normal populations *regardless of whether the samples are independent or the population variances are equal*, it has two disadvantages. First, the sample sizes must be equal, and second, there is a serious loss of information in the sense that the test is performed as if there were only n observations instead of $2n$ observations. An alternate test which avoids these disadvantages when the samples are independent is given in Exercise 20 below.

EXERCISES

1. The management of a food processing plant is considering the installation of new equipment for sorting eggs. If μ_1 is the average number of eggs sorted per hour by their old machine and μ_2 is the corresponding average for the new machine, the null hypothesis they shall want to test is $\mu_1 - \mu_2 = 0$.
   (a) What alternative hypothesis should they use if the burden of proof is to be put on the new equipment and the old equipment will be kept unless the null hypothesis is rejected?
   (b) What alternative hypothesis should they use if the burden of proof is to be put on the old equipment?

After exposure x_2
14
4
20
8
9
9
5
15

rvations are 5, 1, 4, 0, 1, 2, 2, 1,
iation is 1.69. Assuming that the
a normal population with $\mu_0 =$
st has the value

1.67

we find that $t_{.05}$ for 7 degrees of
the null hypothesis cannot be
is resistant to fading or we could
ained.

n be used when sampling from
he samples are independent or the
disadvantages. First, the sample
serious loss of information in the
were only n observations instead
hich avoids these disadvantages
in Exercise 20 below.

s

plant is considering the installation
is the average number of eggs sorted
s the corresponding average for the
shall want to test is $\mu_1 - \mu_2 = 0$.
they use if the burden of proof is to
he old equipment will be kept unless

d they use if the burden of proof is

(c) What alternative hypothesis should they use so that the rejection of the null hypothesis could lead either to buying the new machine or keeping the old one?

2. A producer of extruded plastic products finds that his mean daily inventory is 1148 pieces. A new marketing policy has been put into effect and it is desired to test the null hypothesis that the mean daily inventory remains unchanged. What alternative hypothesis should be used if
   (a) it is desired to *prove* that the new policy reduces inventory?
   (b) it is desired to know whether or not the new policy changes the mean daily inventory?
   (c) the new policy will remain in effect unless it can be *proved* that it causes an increase in inventory?

3. A random sample of boots worn by 50 soldiers in a desert region showed an average life of 1.24 years with a standard deviation of 0.55 years. Under standard conditions, such boots are known to have an average life of 1.40 years. Is there reason to assert at a level of significance of 0.05 that use in the desert causes the average life of such boots to decrease?

4. A sample of 9 measurements of the percentage of manganese in ferromanganese has a mean of 84.0 and a standard deviation of 1.2. Assuming that the sample has been selected at random from a normal population, test the null hypothesis that the true percentage is 80.0 against the alternative that it exceeds 80.0 at the 0.05 level of significance.

5. Test runs with 5 models of an experimental engine showed that they operated, respectively, for 20, 19, 22, 17, and 18 minutes with 1 gallon of a certain kind of fuel. Is this evidence at the 0.01 level of significance that the models are not operating at a desired standard (average) of 22 minutes per gallon? What assumptions are required to perform this test?

6. A quick and inexpensive analytical procedure for the determination of titanium has been developed by a chemist. To show its accuracy, the developer presented 50 independent determinations, having a mean of 0.0095 ppm and a variance of $81.0 \cdot 10^{-8}$. The material tested by the new procedure was carefully checked by a virtually exact but very tedious method, and it was believed that the titanium in this material was in fact 0.0093 ppm. Using a level of significance of 0.05, decide whether there is any reason to doubt the accuracy of the new procedure.

7. A testing laboratory wants to check whether the average lifetime of a certain kind of cutting tool is 2000 pieces against the alternative that it is less than 2000 pieces. What conclusion will they reach at a level of significance of 0.01, if 6 tests showed tool lives of 2010, 1980, 1920, 2005, 1975, and 1950 pieces?

8. A random sample of 100 tires produced by a certain firm lasted on the average 21,000 miles with a standard deviation of 1500 miles. Can it be claimed

162-174

that the true mean life of tires produced by this firm exceeds 20,000 miles? Use $\alpha = 0.05$.

9. Calculate some of the necessary probabilities and graph the *OC* curve for the test used as an illustration on page 164.

10. Graph the *OC* curve for the test described in Exercise 6.

11. The diameters of rotor shafts in a lot have a mean of 0.249 in. and a standard deviation of 0.003 in. The inner diameters of bearings in another lot have a mean of 0.255 in. and a standard deviation of 0.002 in.
    (a) What are the mean and the standard deviation of the clearances between shafts and bearings selected from these lots?
    (b) If a shaft and a bearing are selected at random, what is the probability that the shaft will not fit inside the bearing? (Assume that both dimensions are normally distributed.)

12. An investigation of the relative merits of two types of flashlight batteries showed that a sample of 100 batteries made by Company A had a mean lifetime of 24 hours with a standard deviation of 4 hours. If a sample of 80 batteries from Company B had a mean lifetime of 40 hours with a standard deviation of 6 hours, can it be concluded at the 0.05 level of significance that the batteries made by Company B have a mean lifetime at least 10 hours longer than those made by Company A?

13. A company wants to compare the lifetimes of two stones used in an abrasive process and it finds that the average lifetime of 10 stones of the first kind is 58 pieces with a standard deviation of 6 pieces, and that the average lifetime of 12 stones of the second kind is 66 pieces with a standard deviation of 4 pieces. Test the null hypothesis that there is no difference between the true average lifetimes of the two stones against the alternative that the second is superior. Use $\alpha = 0.01$. What assumption must be met to perform the test?

14. Members of an army evaluation team are attempting to evaluate the relative merits of two designs of antitank projectiles. A sample of 10 projectiles of type A are fired at maximum range, with a mean target error of 24 feet and a variance of 16 feet. A sample of 8 projectiles of type B are fired, with a mean target error of 30 feet and a variance of 25 feet. Is there a significant difference between the mean target errors of the two kinds of projectiles at the 0.01 level? (Assume that the target errors are normally distributed.)

15. Two randomly selected groups of 50 undergraduate engineering students are taught an assembly operation by two different methods and then tested for performance. The first group averaged 120 points with a standard deviation of 12 points while the second group averaged 112 points with a standard deviation of 9 points. If μ_1 is the true mean performance of students taught by the first method and μ_2 is the true mean performance of students taught by the second method, test the null hypothesis $\mu_1 = \mu_2$ at the 0.05 level against the two-sided alternative $\mu_1 \neq \mu_2$.

d by this firm exceeds 20,000 miles?

ilities and graph the *OC* curve for
164.

ed in Exercise 6.

e a mean of 0.249 in. and a standard
ers of bearings in another lot have a
ion of 0.002 in.

deviation of the clearances between
ese lots?

i at random, what is the probability
he bearing? (Assume that both di-

of two types of flashlight batteries
made by Company A had a mean
iation of 4 hours. If a sample of 80
lifetime of 40 hours with a standard
at the 0.05 level of significance that
e a mean lifetime at least 10 hours
?

es of two stones used in an abrasive
time of 10 stones of the first kind is
pieces, and that the average lifetime
eces with a standard deviation of 4
ere is no difference between the true
st the alternative that the second is
on must be met to perform the test?

e attempting to evaluate the relative
ctiles. A sample of 10 projectiles of
h a mean target error of 24 feet and
ojectiles of type B are fired, with a
nce of 25 feet. Is there a significant
rs of the two kinds of projectiles at
et errors are normally distributed.)

lergraduate engineering students are
ifferent methods and then tested for
120 points with a standard deviation
veraged 112 points with a standard
ean performance of students taught
ean performance of students taught
ypothesis $\mu_1 = \mu_2$ at the 0.05 level
2.

16. It is claimed that the resistance of electric wire can be reduced at least 0.050 ohm by alloying. Twenty-five tests each on alloyed wire and standard wire produced the following results:

	Mean	*Standard deviation*
Alloyed wire	0.089 ohm	0.003 ohm
Standard wire	0.141 ohm	0.002 ohm

Using a level of significance of 0.05, determine whether the claim has been substantiated.

17. Tests are run on the performance of samples of 4 plastic and 4 wooden bowling pins, with special attention paid to the number of lines for which they can be used before showing dents or other imperfections. The results obtained for the 4 plastic pins are 2650, 2770, 2480, and 2660 lines, while those for the 4 wooden pins are 1420, 1600, 1545, and 1395 lines. If μ_1 and μ_2 are the respective true means for the two kinds of pins, test at $\alpha = 0.01$ whether plastic pins last on the average 1000 lines longer. What assumptions are required to perform this test?

18. To determine the effectiveness of an industrial safety program, the following data were collected on lost-time accidents (the figures given are mean man-hours lost per month over a period of 1 year):

Plant no.	1	2	3	4	5	6	7	8
Before program	38.5	69.2	15.3	9.7	120.9	47.6	78.8	52.1
After program	28.7	62.2	28.9	0.0	93.5	49.6	86.5	40.2

Test at the 0.10 level of significance whether the safety program was effective in reducing lost-time accidents.

19. The following data were obtained in an experiment designed to check whether there is a systematic difference in the blood pressure readings yielded by two different instruments:

	Reading obtained with Instrument A	*Reading obtained with Instrument B*
Patient 1	136	141
Patient 2	115	117
Patient 3	142	141
Patient 4	140	145
Patient 5	123	127
Patient 6	147	146
Patient 7	133	135
Patient 8	150	152
Patient 9	138	135
Patient 10	147	152

Use a level of significance of 0.05 to test whether there is a difference in the true average readings obtained with the two instruments.

20. When dealing with two independent random samples from normal populations whose variances are not necessarily equal, the following *Smith-Satterthwaite* test can be used to test the null hypothesis $\mu_1 - \mu_2 = \delta$. The test statistic is given by

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

and its sampling distribution can be approximated by the t distribution with

$$\frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

degrees of freedom. Use this test for the data of Exercise 14 and compare the answer with the one previously obtained.

21. Use the formula for t on page 168 to construct a $1 - \alpha$ confidence interval for δ, the difference between the two population means.

22. Use the formula obtained in Exercise 21 to construct a 0.95 confidence interval for the difference between the mean lifetimes of the two abrasive stones of Exercise 13.

162-174