CHAPTER 86

Statistical Inference and Hypothesis Testing

Don T. Phillips
Alberto Garcia-Diaz
Texas A&M University

1. INTRODUCTION

One might define the function of applied statistics as the art and science of collecting and processing data in order to make inferences about the parameters of one or more populations associated with random phenomena. These inferences are made in such a way that the conclusions reached are consistent and unbiased. When properly applied and executed, statistical procedures depend entirely on specific methodologies, definitions, and parameters required by the statistical test chosen.

The science of statistics is purely mathematical with probability theory as the cornerstone. Because all statistical methods are based on probability concepts, it is necessary for one to understand the basic concept of probability measure before undertaking statistical analysis. In order to proceed with the development of procedures for hypothesis testing and estimation, a fundamental knowledge of statistical measures, random variables, probability density functions, and statistical sampling procedures will be assumed.

2. STATISTICAL INFERENCE

Each (and every) random variable has a unique probability distribution. For the most part statisticians deal with the theory of these distributions. Engineers, on the other hand, are mostly interested in finding factual knowledge about certain random phenomena, by way of probability distributions of the variables directly involved, or other related variables.

In basic statistics we learn that probability density functions can be defined by certain constants called *distribution parameters.* These parameters in turn can be used to characterize random variables through measures of location, shape, and variability of random phenomena. The most important parameters are the mean μ and the variance σ^2. The parameter μ is a measure of the center of the distribution (an analogy is the center of gravity of a mass) while σ^2 is a measure of its spread or range (an analogy being the moment of inertia of a mass). Hence, when we speak of the mean and the variance of a random variable, we refer to two statistical parameters (constants) that greatly characterize or influence the probabilistic behavior of the random variable. The mean or expected value of a random variable x is defined as

$$\mu = \sum_x xp(x) \quad \text{or} \quad \mu = \int_x xf(x)dx$$

where $p(x)$ represents probabilities of a *discrete* random variable and $f(x)$ represents the probability density function of a *continuous* random variable. The parameters of interest are embedded in the form of the probability density functions. As an illustration, the variance of a random variable is defined as

$$\sigma^2 = \sum_x (x - \mu)^2 p(x) \quad \text{or} \quad \sigma^2 = \int_x (x - \mu)^2 f(x)dx$$

In mathematical statistics we can show that many random variables that occur in nature follow the same general form of distribution with differences only in the parameters and the statistical quantities μ and σ^2 Some of these recurring distributions have been given special names, such as:

Binomial	Beta	Uniform
Hypergeometric	Normal	Cauchy
Poisson	Chi-square	Rayleigh
Geometric	Student's-*t*	Maxwell
Negative binomial	*F* distribution	Weibull
Gamma	Exponential	Erlang

Thus, it is not difficult to see why the field of ''probability and statistics'' is a discipline within itself, nor is it difficult to see why almost every discipline in existence needs a working knowledge of statistics. Random phenomena (variables) exist in all phases of activity.

When one is interested in certain random phenomena, the first requirement seems to be that one develop some means of measuring it. Upon doing so, one often collects a number of observations of the random phenomena. Statistics deals with developing tools and techniques for choosing those observations (a sample) and manipulating them in such a way that useful information is gained about the underlying random variable(s). This information is generally derived from studying probability distributions of the random variables or functions of the random variables. The average (or mean) and/or the variance (or spread) of the probability distribution of the random variable obviously yield useful information.

While the parameters of a statistical distribution are constant, any computation based on the numerical values of the random observations in a sample may yield different quantities from sample to sample. These quantities are known as ''statistics.'' The two most widely used statistics are the mean of a sample of n observations

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

and the variance of the sample

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

Note that these descriptors are not theoretical in nature but are calculated from a set of n data points.

Mathematical developments have proven that $\bar{x}$ is usually the best single (point) estimate of μ, and S^2 is usually the best single (point) estimate of σ^2. Normally, the nature of these parameters is totally unknown, and statistics is used to draw inferences about their true values.

Since knowledge of the mean and variance is of utmost importance, statistics deals extensively with developing tools and techniques for studying their behavior. Two basic objectives will be dealt with in the remainder of this chapter:

1. Ways of testing to see whether or not some assumed value of μ or σ^2 is "reasonable" under normal operating or assumed conditions.
2. Ways of using observed values of $\bar{x}$ and S^2 so that one may state with a given measure of confidence that the population parameters of these estimates fall within a given interval. For example, we can be 95% confident that the interval from I_1 to I_2 includes the true value of μ.
   The first objective is dealt with using statistical *hypothesis testing,* while the second one gives rise to *confidence interval estimation.*

Statistical tools also exist for estimating the difference between, or testing an assumption about, the means or variances of two or more probability distributions. These tools are natural extensions of the tools developed for estimating and testing hypotheses about single populations.

All statistical methods have, as a basis, a sample of n observations on the random phenomena of interest. Such methods require that the random sample (of size n) be "representative" of the outcomes that could occur. Because of this, much is said about making sure that the sample is a random sample. Many statistical calculations become invalid when the data used are not representative.

3. STATISTICAL HYPOTHESIS TESTING

Working with a representative (random) sample of n observations, statisticians have shown that the function $Z = \sqrt{n}(\bar{x} - \mu)/\sigma$ for sufficiently large values of n is a random variable that follows (or has) a normal probability distribution with $\mu = 0$ and $\sigma^2 = 1$. This fact is the result of the well-known central limit theorem, which can be stated as follows (Bowker and Lieberman 1972; Devore 1987; Dougherty 1990; Hogg and Craig 1978):

If $\bar{x}$ is the mean of a random sample of size n taken from a population having a mean μ and a finite variance σ^2, then

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

is a random variable whose probability distribution approaches that of the standard normal distribution ($\mu = 0$, $\sigma^2 = 1$) as n approaches infinity.

This is undoubtedly the most amazing theorem in statistics for it does not require that one know anything about the shape of the probability distribution of the individual observations. It only requires that the distribution of those random observations have a finite mean, μ, and variance, σ^2. The standard normal density function is defined as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2z^2} \quad \text{for} \quad -\infty < z < \infty$$

This distribution is graphically represented in Figure 1.

If we use the notation $Z_{\alpha/2}(-Z_{\alpha/2})$ to indicate the value of Z corresponding to an area $\alpha/2$ under the distribution falling to the right (left) of $Z_{\alpha/2}$ $(-Z_{\alpha/2})$, then we can associate the cross-hatched area shown in Figure 2 with the region in which $100(1 - \alpha)\%$ of all random variables, characterized by the standard normal density function $f(Z)$ with mean $\mu_z = 0$ and variance $\sigma_z^2 = 1$, are expected to lie. Within the context of a hypothesis test, this area will be called the *acceptance region* and the

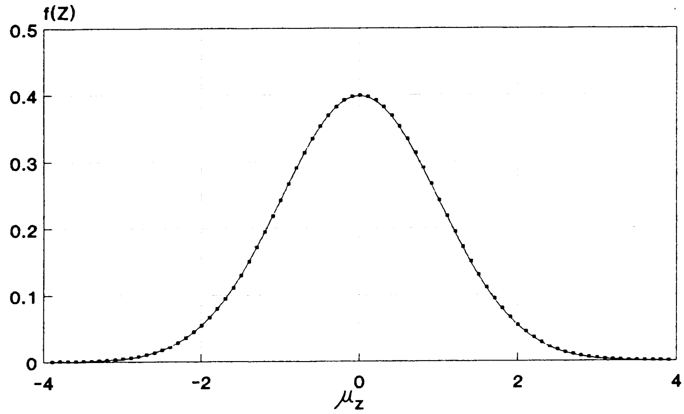

Figure 1 Standardized Normal Distribution.

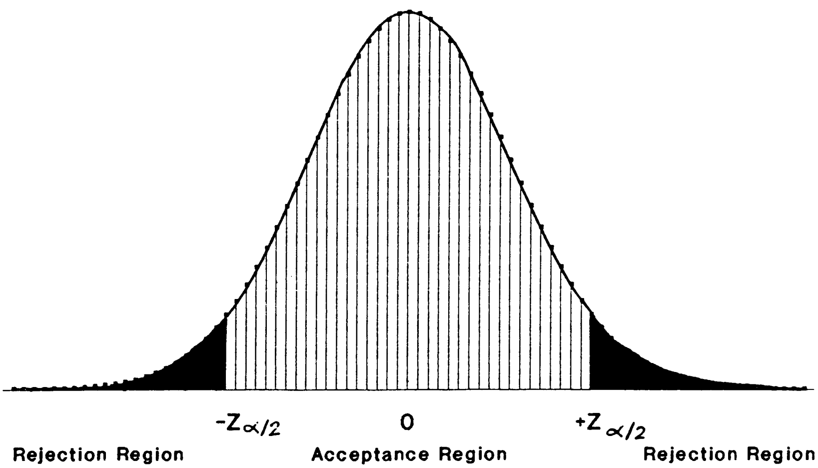

Figure 2 Acceptance and Rejection Regions.

tail areas called the *rejection region.* The points $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ will be called the *rejection points* for reasons which will shortly become evident. Although the underlying probability density function $f(Z)$ might change, these concepts will remain the same. The procedure of statistical hypothesis testing will now be described.

4. TESTING A MEAN VALUE (μ) WITH σ^2 KNOWN

Given a value of $\bar{x}$ computed using a sample of size n from an infinite population with known mean (μ) and variance (σ^2), the probability that the random variable

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

falls between the points $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ is $1 - \alpha$. Note that α is a value between zero and one and represents the probability that a random variable $\bar{x}$, which approximates the mean μ, will naturally fall outside the points $-Z_{\alpha/2}$ and $Z_{\alpha/2}$. This interpretation of the natural behavior of the random variable $\bar{x}$, along with the distribution of the (transformed) variable Z, allows one to structure a hypothesis test concerning the true mean, μ. Assume that the value of $\alpha = 0.05$, and that the following statement is to be tested:

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

The value μ_0 is the numerical value of μ which is assumed known or is hypothesized. H_0 is called the *null* or *primary* hypothesis and H_1 is called the *alternative* or *secondary* hypothesis. If a random sample of size n is extracted from the population under study, then

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

can be calculated. Since $\bar{x}$ is the best point estimator of μ, and μ is assumed to be equal to μ_0 one would expect the random variable

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

to fall between the points $-Z_{\alpha/2} = -Z_{0.025}$ and $Z_{\alpha/2} = Z_{0.025}$ 95% of the time. The values $-Z_{0.025}$ and $Z_{0.025}$ can be determined by using a standard normal table. We can see that they are equal to ± 1.96. These values are called *critical values* and obviously depend upon α. Hence, calculation of Z yields a statistic that will cause H_0 to be believed 95% of the time and H_1 to be believed only 5% of the time, when H_0 is actually *true.* Therefore, α can be interpreted as the magnitude of the error of *rejecting* the *null hypothesis* when in fact it is true. This error is often referred to as an error of type I. Additionally, if the null hypothesis is *false,* there is still a chance that the calculated value of Z will lie between ± 1.96 (when $\alpha = 0.05$). This result will cause the decision analyst to *accept* the null hypothesis when in fact it is *false.* The magnitude (likelihood) of this error is commonly denoted by β, and this error is called an error of type II. Table 1 characterizes the decision process.

In order to illustrate the basic procedures of statistical hypothesis testing, consider the following example:

An oil investment cartel is considering the purchase of an oil well from Blow Hard, Inc. in Texas. Current owners claim that the well produces on the average 100 barrels of oil per day, with a standard deviation of 10 barrels per day. In order to test this claim, the cartel chooses $\alpha = 0.05$ and observes daily production for 16 days. Total production over this period of time is 1690 barrels of oil. Can the owner's claim be disputed?

Assumptions: $\mu = 100$
$\sigma^2 = 100$
Infinite population
$\alpha = 0.05$

Hypothesis test: H_0: $\mu = 100$
H_1: $\mu \neq 100$

Test statistic: $Z - \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Critical values: $\pm Z_{0.025} = \pm 1.96$

Calculated Z statistic: $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{105.63 - 100}{10/4} = 2.252$

Since the value of $Z = 2.252$ is greater than $Z_{\alpha/2} = 1.96$, one would choose to reject H_0 in favor of

TABLE 1 Decision Based upon Sampling Evidence

True State or Nature	H_0 True	H_0 False
Hypothesis H_0 True	No Error	Type I Error
Hypothesis H_0 False	Type II Error	No Error

H_1. In other words, if H_0 is true, it is highly unlikely that a sample of size $n = 16$ would yield a value of $\bar{x}$ equal to 105.63, resulting in a value of Z as large as $Z = 2.252$. What value of $\bar{x}$ would cause the decision maker to accept H_0?

Since $\pm Z_{\alpha/2} = = \pm 1.96$ one can define the following relationship to find the limits of the acceptance region:

$$\pm 1.96 = \frac{\bar{x}_c - 100}{2.5}$$

It can be verified that the solution for $\bar{x}_c$ is given by $\bar{x}_c = 95.1$ and $\bar{x}_c = 104.9$. Therefore, any sample average between these two values will result in the acceptance of H_0.

Next, let us assume that the true population mean (daily production rate) is equal to $\mu_1 = 110$ and not 100. What is the probability that the null hypothesis will be erroneously accepted? Assuming that the variance remains constant, consider the two distributions shown in Figure 3.

Note that the probability that the null hypothesis will be accepted (erroneously) is given by the area marked β in Figure 3. This area can be calculated as follows:

Test statistic: $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Critical value: $Z_1 = \dfrac{104.9 - 110}{2.5} = -2.04$

Type II error: $\beta = P(Z \leq -2.04)$

Using a standard normal table, the probability corresponding to the β area is given by $\beta = 0.0217$. Note, however, that the β error should not include that area to the *left* of $\bar{x}_c = 95.1$. The probability that $\bar{x}$ will be less than this value given that $\mu = 110$ is determined by $P(Z \leq Z_2)$, where

$$Z_2 = \frac{95.1 - 110}{2.5} = -5.96$$

This probability is almost 0. Therefore, $\beta = 0.0217$ is the correct value to four significant digits. In other words the probability of accepting the null hypothesis when μ is actually $\mu_1 = 110$ is only 0.0217. Note that this probability is actually based on the values $\bar{x}_c = 95.1$ and $\bar{x} = 104.9$, which were uniquely determined by the chosen value of α. Clearly, one would never know what the true population mean (μ) actually is in any meaningful application. Hence, the value of β cannot be calculated except in reference to values of μ different from that specified in the null hypothesis. For this example several values of β were calculated for the alternative values of μ_1 indicated in Table 2.

For clarity, one should note that *in the limit* as μ approaches 100, the probability of a type II error approaches $\beta = 1 - \alpha$. At the exact point $\mu_0 = 100$, the null hypothesis is true and a type II error does not exist; hence, $\beta = 0$ at that single point. Figure 4 graphically depicts the behavior of the β error as a function of the true (unknown) population mean, under the original rejection criterion specified by the null hypothesis and the chosen α error. This curve is called an *operating characteristic curve,* or simply an *OC curve.*[1]

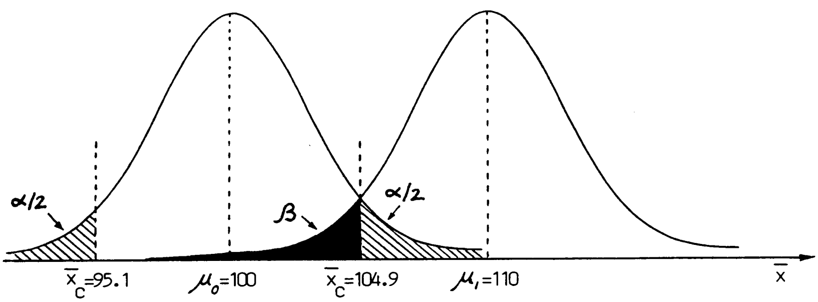

Figure 3 Probabilities of Type I and Type II Errors.

TABLE 2 Probability of Type II Error

Population Mean Values													
91	93	95	96	97	98	99	101	102	103	104	105	107	109
0.025	0.200	0.484	0.641	0.776	0.874	0.932	0.932	0.874	0.776	0.641	0.484	0.200	0.025

The calculations shown in Table 2 were performed relative to an alternative hypothesis that included two rejection points ($-Z_{\alpha/2}$, $Z_{\alpha/2}$). Such a hypothesis test is called a *two-tailed hypothesis test.*

Consider once again our numerical example. Note that the null hypothesis states that well production is *exactly* 100 barrels per day. In reality, the purchase of the well would be desirable if the daily production met or exceeded 100 barrels per day. In that case the null and alternative hypotheses would be

Case I: H_0: $\mu \geq 100$
H_1: $\mu < 100$

or

Case II: H_0: $\mu \leq 100$
H_1: $\mu > 100$

Both hypothesis statements reflect the same objective, but there are significant differences in the decision criterion utilized in each hypothesis. The null hypothesis of case I assumes that the well is producing 100 or more barrels per day unless statistical evidence proves otherwise, resulting in rejection of H_0. The null hypothesis of case II assumes that the well production is inferior unless production records indicate that daily output is more than 100 barrels per day, which will result in rejection of H_0. Both tests are valid and are called *one-tailed* hypothesis tests under a given type I error. Consider again the original data with $\alpha = 0.05$. Table 3 summarizes the calculations for the significance tests associated with cases I and II.

Both test statistics in Table 3 use the value of 100 in calculating a Z value, for it is at this single point that the type I error and the type II errors are the greatest. It should also be noted that the type II error now exists in only one direction, and hence the operating characteristic curve will be one sided. For completeness, one should note that the null hypothesis is an a priori state of nature that one chooses to believe, unless statistical evidence indicates otherwise. In statistics, one does not "prove" the null hypothesis but rather "fail to reject" the hypothesis. These concepts are consistent

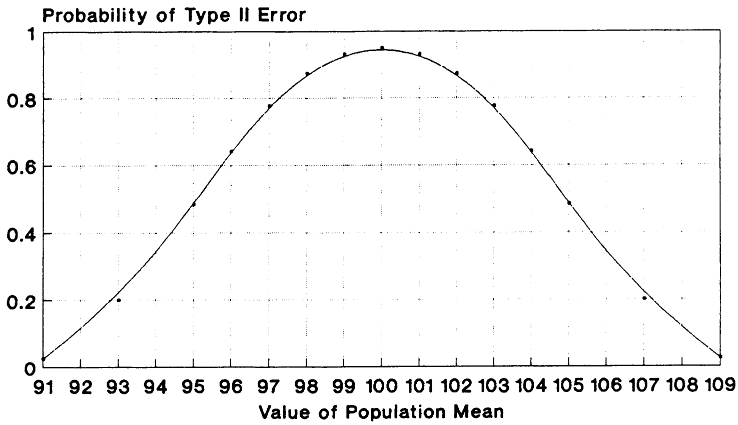

Figure 4 Operating Characteristic Curve. Level of significance = 0.05.

TABLE 3 One-Tail Hypothesis Tests for Means

Procedure	Case I	Case II
Assumptions	$\mu \geq 100$ $\sigma^2 = 100$ Normal population $\alpha = .05$	$\mu \leq 100$ $\sigma^2 = 100$ Normal population $\alpha = .05$
Hypothesis test	H_0: $\mu \geq 100$ H_1: $\mu < 100$	H_0: $\mu \leq 100$ H_1: $\mu > 100$
Test statistic	$Z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$	$Z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$
Critical value	$-Z_\alpha = -1.645$	$Z_\alpha = 1.645$
Calculated value	$Z = 2.252$	$Z = 2.252$
Conclusion	Accept H_0 Buy the well	Reject H_0 Buy the well

with the underlying uncertain (stochastic) nature under which hypothesis testing is conducted. One can never be absolutely certain of statistical inference. Along these same lines of thought, it is critical that the hypothesis test be chosen *before* statistical sampling and not after. Selection of the test (one or two tailed) and the associated a error should never be chosen a posteriori to ‘‘statistically confirm’’ any belief. Such statistical inference is obviously improper.

In summary, a decision maker can apply either a one-tailed or two-tailed hypothesis test with a chosen type I error probability equal to α. The true probability of type II error (β) is always unknown since the actual population mean is unknown. However, the β risk can be characterized by the construction of an OC curve.

5. TESTING A MEAN VALUE (μ) WITH σ^2 UNKNOWN

Consider again the oil well example. After detailed examination, it was discovered that the theoretical variance (σ^2) was really not known but had been ‘‘guessed at’’ by the seller. In order to estimate σ^2, the sample of 16 days was used to calculate S^2 in the following manner:

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

Using $n = 16$, a value of $S^2 = 225$ was calculated. Under the assumption that σ^2 is unknown, the Z statistic is no longer a valid test statistic. It can be shown that the following test statistic should be used in this case (Bowker and Lieberman 1972; Devore 1987; Hogg and Craig 1978).

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

This is called simply the student t test statistic. The rejection points $t_{\alpha/2}$ and $-t_{\alpha/2}$ for a two-tailed test can be determined from an appropriate t table, but they now depend on a parameter called the ‘‘degrees of freedom,’’ which is defined by $df = n - 1$. Consider the original two-tailed hypothesis test under the new assumption (σ^2 unknown):

Assumptions: σ^2 unknown
$\mu = 100$
Infinite population
Population is normal
$\alpha = 0.05$

Hypothesis test: H_0: $\mu = 100$
H_1: $\mu \neq 100$

Test statistic: $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

Critical values: $t_{\alpha/2,df} = t_{0.025,15} = 2.131$
$-t_{\alpha/2,df} = -t_{0.025,15} = -2.131$

Calculated t statistic: $t = \dfrac{105.63 - 100}{15/4} = 1.501$

Therefore, the null hypothesis cannot be rejected and one is led to believe that the true well production is actually 100 barrels per day.

This example illustrates the use of the t statistic for testing a hypothesis related to a population mean value (μ) when the variance (σ^2) is *unknown.* As before, both two-tailed and one-tailed tests are possible, whichever the problem situation demands. Rejection limits are set based on a chosen type I (α) error, and operating characteristic curves (OC curves) can be constructed to reflect the associated type II (β) error risk. Finally, one should note that the t test was necessitated due to the fact that σ^2 was being estimated by S^2 from a sample of size n. If n is large enough, then one would expect S^2 to closely approximate σ^2 and the Z test can be used anyway. For $n \geq 30$ this is generally an acceptable procedure.

6. HYPOTHESIS TESTING: SINGLE VARIANCE

Continuing with the example about oil well production, the potential buyer was quite perplexed at the result obtained from the two-tailed t test, since it differed from that previously obtained via the original two-tailed Z test. An engineer explained that the difference was probably a result of lack of knowledge concerning the population variance, σ^2. Since σ^2 was unknown, the induced uncertainty caused failure to reject the null hypothesis (note that the rejection points were wider).

Reflecting upon this logic, management requested a statistical examination of the sample variance to determine if the original (specified) σ^2 had indeed changed, since the estimated value of σ^2 (S^2) was greater than the original specified value. The proper statistical test is a chi-square test. The chi-square test is described as follows:

Assumptions: $\sigma^2 = 100$
Infinite population
Population is normal
$\alpha = .05$

Hypothesis test: H_0: $\sigma^2 = 100$
H_1: $\sigma^2 \neq 100$

Test statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

Critical values: $\chi^2_{df,1-\alpha/2}$; $\chi^2_{df,\alpha/2}$

Degrees of freedom: $df = n - 1$

As in previous tests, the χ^2 rejection values are both functions of the chosen type 1 error (α) and the sample size ($df = n - 1$). Critical values are easily obtained from χ^2 tables. For the numerical example the procedure is as follows:

Hypothesis test: H_0 $\sigma^2 = 100$
H_1 $\sigma^2 \neq 100$

Test statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} = \dfrac{(15)(225)}{100} = 33.75$

Critical values: $\chi^2_{15,0.975} = 6.262$, $\chi^2_{15,0.025} = 27.488$

Since $\chi^2 = 33.75$ is greater than the critical value $\chi^2_{15,0.025} = 27.488$, one is led to believe that the underlying statistical variance of well production has changed from 100 to something else. Of course, based on the value of S^2, one might conclude that it has increased to somewhere around 225. Although further investigation would be up to decision maker (buyer), it appears that a good course of action could be to obtain more sample data and reexamine the entire situation.

7. HYPOTHESIS TESTING: TWO POPULATION MEANS WITH VARIANCES KNOWN

After a rather lengthy discussion, company management reached the decision that the well should be purchased. Due to optimistic market projections, it was decided that a second well should also be purchased if one could be found that produced 100 more barrels per day on the average than the single established well. Further testing led management to believe that, for comparison purposes, the first well did indeed produce at a rate of 100 barrels per day. At this point Blow Hard, Inc. presented a new well that it claimed produced at a rate of 100 barrels per day more than the first well. Management again insisted on statistical investigation and in order to provide new, current data, a separate independent evaluation was undertaken. Samples from both wells were obtained to compare daily production rates. Blow Hard, Inc. assured the purchasing cartel that the variance in daily production for the first well was actually 240 and for the second well 276.

Additionally, a sample of production records for $n_1 = 12$ days from well 1 and $n_2 = 18$ days from well 2 provided average daily production of $\bar{x}_1 = 102$ and $\bar{x}_2 = 212$ barrels per day, respectively.

Since the variances of production, $\sigma_1^2 = 240$ and $\sigma_2^2 = 276$, are both assumed known, the proper test statistic is a Z test for the difference in the means of populations. The procedure is as follows:

Assumptions: σ_1^2 and σ_2^2 known
Infinite populations
Normal populations
$\alpha = 0.05$

Hypothesis test: H_0: $\mu_2 - \mu_1 = \delta$
H_1: $\mu_2 - \mu_1 \neq \delta$
$\delta = 100$

Test statistic: $$Z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Critical values: $-Z_{\alpha/2}, Z_{\alpha/2}$

The numerical calculations are as follows:

$$H_0: \quad \mu_2 - \mu_1 = 100$$

$$H_1: \quad \mu_2 - \mu_1 \neq 100$$

$$Z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{(212 - 102) - 100}{\sqrt{\dfrac{240}{12} + \dfrac{276}{18}}}$$

$$Z = \frac{10}{5.94} = 1.684$$

The critical value for this test is given by $\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$. Hence, there is no statistical evidence to support the rejection of the null hypothesis, and the proper decision would be to purchase both wells under the management interpretation of these results. However, the same clever engineer who questioned the first test results again questioned the validity of using σ_1^2 and σ_2^2 in the calculations. The question was then posed, "Can we perform a similar test without knowing the population variances?" The statistician responded yes as the same example used to calculate $\bar{x}_1$ and $\bar{x}_2$ could be used to estimate σ_1^2 and σ_2^2 with S_1^2 and S_2^2, respectively.

8. HYPOTHESIS TESTING WITH TWO MEANS: POPULATION VARIANCES UNKNOWN BUT ASSUMED EQUAL

The test statistic, assumptions, and hypothesis test are as follows:

Assumptions: σ_1^2 and σ_2^2 (both are unknown)
Infinite populations
Normal populations
$\alpha = .05$

Hypothesis test: H_0: $\mu_2 - \mu_1 = \delta$
H_1: $\mu_2 - \mu_1 \neq \delta$

Test statistic: $$t = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Critical values: $-t_{\alpha/2,df}, t_{\alpha/2,df}$

Degrees of freedom: $df = n_1 + n_2 - 2$

Using the same set of data, it was found that $S_1^2 = 165$ and $S_2^2 = 453$. The calculations related to this example are as follows:

Hypothesis test: H_0: $\mu_2 - \mu_1 = 100$
H_1: $\mu_2 - \mu_1 \neq 100$

The calculated value of the test statistic is given by

$$t = \frac{(212 - 102) - (100)}{\sqrt{11(165) + 17(453)}} \sqrt{\frac{(12)(18)(28)}{30}} = 1.454$$

The critical values (rejection points) for this test are $t_{0.025,28} = 2.048$ and $-t_{0.025,28} = -2.048$. Since the calculated value of $t = 1.454$ is less than the upper rejection value, there is insufficient evidence to reject the null hypothesis. Finally, one should note that under the assumption that $\sigma_1^2 = \sigma_2^2$, a pooled estimate of the variance is used from a total sample of $N = n_1 + n_2$. As in the single parameter t test, if n_1 and n_2 are both greater than 30, one can simply use the two parameter Z test directly.

Our same clever engineer now observes that these results can be reached only if it can be assumed that the two population variances are equal. At this point, the manager asks our statistician, "Can we test this assumption?" The answer is yes, using an F test for equality of variances.

9. HYPOTHESIS TESTING FOR EQUALITY OF TWO POPULATION VARIANCES

The following assumptions, hypothesis test, and test statistic should be observed when conducting an F test:

Assumptions: Normal populations
Infinite populations
Hypothesis test: H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_2^2 \neq \sigma_2^2$
Test statistic: $F = S_1^2/S_2^2$
Critical values: $F_{1-\alpha/2,v_1,v_2}$, $F_{\alpha/2,v_1,v_2}$
v_1 = degrees of freedom in numerator
$= n_1 - 1$
v_2 = degrees of freedom in denominator
$= n_2 - 1$

The critical values $F_{\alpha/2,v_1,v_2}$ for commonly used values of α are easily found in statistical tables. Once these values are obtained, the values of $F_{1-\alpha/2,v_1,v_2}$ are easily calculated for the left-hand rejection point according to the following formula

$$F_{1-\alpha/2,v_1,v_2} = (F_{\alpha/2,v_2,v_1})^{-1}$$

For example, if S^2 was calculated using a sample of size $n_1 = 13$ and S_2^2 was calculated using a sample of size $n_2 = 21$; the rejection points for the quantity $F = S_1^2/S^2$ for $\alpha = .10$ would be given by

$$F_{\alpha/2,v_1,v_2} = F_{.05,12,20} = 2.28$$

and

$$F_{1-\alpha/2,v_1,v_2} = (F_{0.05,20,12})^{-1} = (2.54)^{-1} = 0.394$$

In order to illustrate the F test, consider the previous example. Recall that $S_1^2 = 165$, $n_1 = 12$, $S_2^2 = 453$, and $n_2 = 18$.

Assumptions: $\alpha = 0.10$
Normal populations
Hypothesis test: H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$
Test statistic: $F = S_1^2/S_2^2$

Critical values: Using $\alpha = 0.10$, $n_1 = 12$, $n_2 = 18$, we obtain the critical values shown:

$F_{0.05,11,17} = 2.41$

$F_{0.95,11,17} = (F_{0.05,17,11})^{-1} = 0.372$

Calculated F statistic: $F = S_1^2/S_2^2 = 165/453 = 0.364$

Hence, the null hypothesis would be rejected and one is led to assume that $\sigma_1^2 \neq \sigma_2^2$. Our astute manager, still seeking statistical evidence, now inquires as to the availability of a statistical test when $\sigma_1^2 \neq \sigma_2^2$. Fortunately, such a test is available, and is called the t' test.

10. EQUALITY OF TWO MEANS WITH VARIANCES UNKNOWN AND NOT EQUAL

The proper statistical test for this procedure is the t' test and is based on the following:

Assumptions: $\sigma_1^2 \neq \sigma_2^2$
Normal populations
Infinite populations

Hypothesis test: H_0: $\mu_1 - \mu_2 = \delta$
H_1: $\mu_1 - \mu_2 \neq \delta$

Test statistic:
$$t' = \frac{(\bar{x}_2 - \bar{x}_1) - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

Critical values: $t_{\alpha/2,df}$, $t_{1-\alpha/2,df}$

Degrees of freedom:
$$df = \left[\frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 - 1}}\right]$$

For the oil well example, the following calculations illustrate the procedure:

Assumptions: $\sigma_1^2 \neq \sigma_2^2$
Normal populations
Infinite populations
$\alpha = 0.05$

Hypothesis test: H_0: $\mu_2 - \mu_1 = 100$
H_1: $\mu_2 - \mu_1 \neq 100$

Critical values: $t_{\alpha/2,df}$, $-t_{\alpha/2,df}$

Degrees of freedom:
$$df = \left[\frac{\left(\dfrac{165}{12} + \dfrac{453}{18}\right)^2}{\dfrac{\left(\dfrac{165}{12}\right)^2}{11} + \dfrac{\left(\dfrac{453}{18}\right)^2}{17}}\right] = [27.82] = 27$$

Critical values: $t_{0.025,27} = 2.052$

$-t_{0.025,27} = -2.052$

Calculated t statistic:
$$t' = \frac{(212 - 102) - 100}{\sqrt{\dfrac{165}{12} + \dfrac{453}{18}}} = \frac{10}{\sqrt{38.92}}$$
$$t' = 1.603$$

Since $t' = 1.603$ is less than the critical value of $t = 2.052$, statistical evidence does not support rejection of the null hypothesis.

11. CONFIDENCE INTERVAL ESTIMATION

A subject closely related to hypothesis testing is that of confidence intervals. The basic concepts are best understood by considering once again the test statistic:

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

If, as before, we use the notation $Z_{\alpha/2}$ and $-Z_{\alpha/2}$ to indicate the values of the random variable Z where $\alpha/2$ of the area lies to the right of $Z_{\alpha/2}$ and $\alpha/2$ to the left of $-Z_{\alpha/2}$, then the following probability statement must be true:

$$P\left\{-Z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right\} = 1 - \alpha$$

If we rearrange the inequality in brackets and solve for the true population parameter μ, then we obtain

$$\bar{x} - Z_{\alpha/2}\, \sigma/\sqrt{n} \le \mu \le \bar{x} + Z_{\alpha/2}\, \sigma/\sqrt{n}$$

This last expression says that if we take a random sample of n observation on the random phenomenon of interest and calculate the interval,

$$\bar{x} - Z_{\alpha/2}\, \sigma/\sqrt{n} \quad \text{to} \quad \bar{x} + Z_{\alpha/2}\, \sigma/\sqrt{n}$$

we can be $100(1 - \alpha)\%$ confident that the interval will include the true mean, μ. In other words, if we repeatedly took samples of size n and calculated the interval from each sample, in the long run $100(1 - \alpha)\%$ of the intervals will include μ. The interval is obviously calculated only once, and the resulting values of the endpoints constitute a $100(1 - \alpha)\%$ confidence interval estimate of μ. For a numerical illustration consider once again the first example given in this chapter, that is, the single oil well production problem:

Assumptions: $\alpha = 0.05$
$\sigma^2 = 100$
Critical values: $\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$
Input data: $\bar{x} = 105.63$, $n = 16$

The 95% confidence interval for the true (unknown) population mean (μ) is given by

$$105.63 - (1.96)(5/2) \le \mu \le 105.63 + (1.96)(5/2)$$

or

$$100.73 \le \mu \le 110.53$$

Note that this interval assumes that one knows σ^2, the variance of the random variable. Usually this is not the case, and we have to estimate σ^2 with S^2. If we change the expression for Z accordingly, we get a different distribution. As already mentioned in this chapter,

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

is a random variable that follows a t distribution.

Now, utilizing the knowledge that we have provided about the t distribution, we write a probability statement similar to the one for the previous case:

$$P\left\{-t_{\alpha/2} \le \frac{\bar{x} - \mu}{S/\sqrt{n}} \le t_{\alpha/2}\right\} = 1 - \alpha$$

and upon rearranging the inequality in the brackets, we obtain

$$\bar{x} - t_{\alpha/2}\, S/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2}\, S/\sqrt{n}$$

which is a $100(1 - \alpha)\%$ confidence interval estimate of μ when σ is not known.

It should be pointed out that when the sample size from which S is calculated exceeds 30, it makes little difference whether one uses the normal distribution with $\sigma = S$ or the more precise distribution. The probability distributions of t and Z become the same at $n = \infty$.

If one is interested in calculating a $100(1 - \alpha)\%$ confidence interval for the quantity $\mu_1 - \mu_2$, the difference between the means of two different distributions, one would take a random sample from each distribution, n_1 from the first and n_2 from the second. From these samples one would calculate $\bar{x}_1$ and S_1^2 as well as $\bar{x}_2$ and S_2^2. In a manner identical to that used earlier, the following $100(1 - \alpha)\%$ confidence interval could be constructed:

$$P\left\{(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right\} = 1 - \alpha$$

Clearly, this procedure could be repeated for *any* test statistic previously discussed in this section. The reader is referred to any of a number of engineering statistics texts for developments of χ^2, F, and t' confidence intervals.

12. MAXIMUM-LIKELIHOOD ESTIMATORS

A well-known procedure for finding estimators of unknown parameters is the method of maximum likelihood (Devore 1987; Dougherty 1990; Hogg and Craig 1978). Maximum-likelihood estimators are consistent and have minimum variance but are not always unbiased. A summary of the procedure follows.

Let X be a random variable with density function $f(x,\theta)$, where the parameter θ is unknown. Given a random sample of independent observations $x_1, x_2, \ldots, x_n$, the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = f(x_1, \theta) f(x_2, \theta), \ldots, f(x_n, \theta)$$

The likelihood function is actually the joint probability density function (for continuous variables) or the joint mass probability function (for discrete variables) of the n random variables. Therefore, the value of θ for which the observed sample would have the highest probability of being extracted, can be found by maximizing the likelihood function over all possible values of the parameter θ. As shown in elementary calculus, this can be achieved by setting the first derivative of the likelihood function with respect to the parameter equal to zero, and then solving for θ:

$$dL/d\theta = 0$$

The solution to this equation can be more efficiently found by considering the logarithm (base e) of the likelihood function, instead of the function itself:

$$d \ln L/d\theta = 0$$

To illustrate the fundamental steps of the procedure followed to obtain a maximum-likelihood estimator, we will consider a random variable X having the exponential density function:

$$f(x; \theta) = \theta e^{-\theta x}$$

where x is nonnegative. For a sample size n the likelihood function for this example is given by

$$L(x_1, x_2, \ldots, x_n) = \Pi_i \theta e^{-\theta x_i}$$

where the index i takes on the values $i = 1, 2, \ldots, n$. Taking the natural logarithm of the preceding likelihood function, we obtain

$$\ln L(x_1, x_2, \ldots, x_n) = n \ln \theta - \theta \sum_i x_i$$

After differentiating $\ln L$ with respect to θ, setting the derivative equal to zero, and solving for θ we obtain the final result shown:

$$\hat{\theta} = n / \sum_i x_i$$

which is equal to the inverse of the sample mean. Now let us assume that the random sample consists of the following values for $n = 5$: 0.9, 1.7, 0.4, 0.3, 2.4. In this case the maximum-likelihood estimator would be equal to $\hat{\theta} = 5/5.7 = 0.88$. It can be verified that $\bar{x}$ and $S^2 = \Sigma_i(x_i - \bar{x})^2/n$ are maximum-likelihood estimators for the mean and variance, respectively, of a normal distribution (Bowker and Lieberman 1972).

13. TESTING FOR EQUALITY OF MEANS AND VARIANCES FOR *K* POPULATIONS

13.1. Testing Means

In the case where there are k populations under consideration, and the test of hypothesis is equality of population means, a different type of procedure is necessitated. This area of statistics is called *experimental design* or *analysis of variance.* This topic is covered in Chapter 85 of this Handbook.

13.2. Testing the Homogeneity of Variances

A problem frequently encountered in applied statistics is that of testing the equality of variances of several normal populations. Let us assume that we have collected a random sample of size n_i from a normal population with mean μ_i and variance σ_i^2, repeating this basic procedure for $i = 1, 2, \ldots, k$. The hypothesis to be tested in this case can be formulated as

$$H_0: \quad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

As already explained earlier in this chapter, an F test can be used for $k = 2$. However, a different procedure is required for larger values of k.

Among several methodologies for testing this hypothesis, the following three are perhaps the best known (Snedecor and Cochran 1980): (a) Cochran's test, (b) Barlett's test, and (c) Levene's test. Each of these methods is described below.

13.3. Cochran's Test

In the case where there are k populations and the test of equality of population variances is required, the most commonly applied test is the Cochran's test for homogeneity of variances.

Assumptions: Samples are independent
Populations are normal

Test statistic: $R = \dfrac{\max\{S_1^2, S_2^2, \ldots, S_k^2\}}{\sum_{i=1}^{k} S_i^2}$

In the relationship defined for the test statistic R, S_i^2 is the unbiased point estimator of σ_i^2 for $i = 1, 2, \ldots, k$. Each S_i^2 is calculated from a sample of size n. The corresponding test of significance is conducted according to the following rule:

Test rule: Accept the null hypothesis that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$ if $R \leq RC_{\alpha,n,k}$, where $RC_{\alpha,n,k}$ is a critical value for chosen values of the type I error (α), the sample size (n), and the number of populations (k). Tables of critical values for $\alpha = 0.05$ and 0.01, for values of k up to 120 and n up to 145, are given in Bowker and Lieberman (1972).

13.4. Barlett's Test

Barlett developed in 1937 a testing procedure that can be used when all sample sizes n_i for $i = 1, 2, \ldots, k$, are not equal. The procedure is described as follows, where N indicates the total number of observations collected from the k populations:

Assumptions: Samples are independent
Populations are normal

Test statistic: $M = -\dfrac{1}{c}\,\Sigma_i(n_i - 1)\ln\dfrac{S_i^2}{S^2}$

$S^2 = \Sigma_i(n_i - 1)S_i^2/(N - k)$

$c = (\Sigma_i\, 1/f_i - 1/f)/3(k - 1) + 1$

TABLE 4 Sample Calculations for Goodness of Fit Testing

Sample	O_i	e_i	$(O_i - e_i)^2/e_i$
1	18	19.4	0.101
2	17	19.4	0.297
3	19	19.4	0.008
4	17	19.4	0.297
5	18	19.4	0.101
6	20	19.4	0.019
Total	—	—	0.823

Test rule: The statistic M has approximately a χ^2 distribution with $k - 1$ degrees of freedom. This approximation is more appropriate for sample sizes larger than 3. If the observed value of M exceeds the critical value of the chi-square statistic for a level of significance α and $k - 1$ degrees of freedom, the hypothesis is rejected.

13.5. Levene's Test

An approximate test, which is less sensitive to the lack of normality in the data than Barlett's test, was developed by H. Levene in 1960. The procedure assumes that all sample sizes are equal to n. This testing method is described as follows:

Assumptions: Samples are independent
Populations are normal (or approximately normal)

Test statistic: F

Test rule: For Levene's test we conduct an analysis of variance (ANOVA) of the absolute deviations from each sample average. Details on the ANOVA procedure are given in another chapter of this handbook. If the observed mean square ratio exceeds the appropriate critical value of the F statistic, we reject the hypothesis that all variances are equal.

14. OTHER USES OF HYPOTHESIS TESTING

Finally, it should be noted that the concepts of type 1(α) error, type II (β) error, critical values, OC curves, and one/two-tailed hypothesis tests are common to all hypothesis testing. This chapter has illustrated only tests concerning means and variances since they are most common to industrial engineering. One might also find occasions to test hypotheses in quality control applications, proportions, percentages, and in goodness-of-fit testing. These applications, and others, are well documented in a host of applied engineering textbooks.

14.1. Nonparametric Tests

One should note that for most of the hypothesis tests we have discussed, the assumption of normally distributed random variables is required. In actual practice this may not be justified. One has two choices by which this assumption can be ignored. First, one can obtain enough samples to use a Z test (normal tables) rather than a t test (t tables). If the sample is large enough, then the assumption of normality is not required. However, in the case where larger samples cannot be obtained or cost prohibits larger samples, the use of *nonparametric statistical tests* is necessitated (Hollander and Wolfe 1973; Lehmann 1975; Marascuilo and McSweeney 1977).

A nonparametric test is one that requires no assumptions regarding the form or shape of the underlying random variables. Usually, all that is required is knowledge of the scale of measurement used in the experiment and whether the random variable is discrete or continuous. The treatment of nonparametric statistics is well beyond the scope of this introductory section. However, the reader should be aware of its role in hypothesis testing. A test frequently used in industrial organizations is the goodness of fit test, which is described in the next section.

14.2. Goodness of Fit Test

The assumption of normality is not an unusual one in problems dealing with hypothesis testing about means. However, it becomes an important one when testing hypotheses about variances. For this reason, before testing the hypotheses, one should verify if there is sufficient statistical evidence indicating that the population is normal. A statistical method used for this purpose is known as "test

TABLE A Cumulative Normal Distribution $\theta(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

t	0.00	0.01	0.02	0.03	0.04
0.0	0.500 00	0.503 99	0.507 98	0.511 97	0.515 95
0.1	0.539 83	0.543 79	0.547 76	0.551 72	0.555 67
0.2	0.579 26	0.583 17	0.587 06	0.590 95	0.594 83
0.3	0.617 91	0.621 72	0.625 51	0.629 30	0.633 07
0.4	0.655 42	0.659 10	0.662 76	0.666 40	0.670 03
0.5	0.691 46	0.694 97	0.698 47	0.701 94	0.705 40
0.6	0.725 75	0.729 07	0.732 37	0.735 65	0.738 91
0.7	0.758 03	0.761 15	0.764 24	0.767 30	0.770 35
0.8	0.788 14	0.791 03	0.793 89	0.796 73	0.799 54
0.9	0.815 94	0.818 59	0.821 21	0.823 81	0.826 39
1.0	0.841 34	0.843 75	0.846 13	0.848 49	0.850 83
1.1	0.864 33	0.866 50	0.868 64	0.870 76	0.872 85
1.2	0.884 93	0.886 86	0.888 77	0.990 65	0.892 51
1.3	0.903 20	0.904 90	0.906 58	0.908 24	0.909 88
1.4	0.919 24	0.920 73	0.922 19	0.923 64	0.925 06
1.5	0.933 19	0.934 48	0.935 74	0.936 99	0.938 22
1.6	0.945 20	0.946 30	0.947 38	0.948 45	0.949 50
1.7	0.955 43	0.956 37	0.957 28	0.958 18	0.959 07
1.8	0.965 07	0.964 85	0.965 62	0.966 37	0.967 11
1.9	0.971 28	0.971 93	0.972 57	0.973 20	0.973 81
2.0	0.977 25	0.977 78	0.978 31	0.978 82	0.979 32
2.1	0.982 14	0.982 57	0.983 00	0.983 41	0.983 82
2.2	0.986 10	0.986 45	0.986 79	0.987 13	0.987 45
2.3	0.989 28	0.989 56	0.989 83	0.990 10	0.990 36
2.4	0.991 80	0.992 02	0.992 24	0.992 45	0.992 66
2.5	0.993 79	0.993 96	0.994 13	0.994 30	0.994 46
2.6	0.995 34	0.995 47	0.995 60	0.995 73	0.995 85
2.7	0.996 53	0.996 64	0.996 74	0.996 83	0.996 93
2.8	0.997 44	0.997 52	0.997 60	0.997 67	0.997 74
2.9	0.998 13	0.998 19	0.998 25	0.998 31	0.998 36
3.0	0.998 65	0.998 69	0.998 74	0.998 78	0.998 82
3.1	0.999 03	0.999 06	0.999 10	0.999 13	0.999 16
3.2	0.999 31	0.999 34	0.999 36	0.999 38	0.999 40
3.3	0.999 52	0.999 53	0.999 55	0.999 57	0.999 58
3.4	0.999 66	0.999 68	0.999 69	0.999 70	0.999 71
3.5	0.999 77	0.999 78	0.999 78	0.999 79	0.999 80
3.6	0.999 84	0.999 85	0.999 85	0.999 86	0.999 86
3.7	0.999 89	0.999 90	0.999 90	0.999 90	0.999 91
3.8	0.999 93	0.999 93	0.999 93	0.999 94	0.999 94
3.9	0.999 95	0.999 95	0.999 96	0.999 96	0.999 96
0.0	0.519 94	0.523 92	0.527 90	0.531 88	0.535 86
0.1	0.559 62	0.563 56	0.567 49	0.571 42	0.575 34
0.2	0.598 71	0.602 57	0.606 42	0.610 26	0.614 09
0.3	0.636 83	0.640 58	0.644 31	0.648 03	0.651 73
0.4	0.673 64	0.677 24	0.680 82	0.684 38	0.687 93
0.5	0.708 84	0.712 26	0.715 66	0.719 04	0.722 40
0.6	0.742 15	0.745 37	0.748 57	0.751 75	0.754 90
0.7	0.773 37	0.776 37	0.779 35	0.782 30	0.785 23
0.8	0.802 34	0.805 10	0.807 85	0.810 57	0.813 27
0.9	0.828 94	0.831 47	0.833 97	0.836 46	0.838 91
1.0	0.853 14	0.855 43	0.857 69	0.859 93	0.862 14
1.1	0.874 93	0.876 97	0.879 00	0.881 00	0.882 97
1.2	0.894 35	0.896 16	0.897 96	0.899 73	0.901 47
1.3	0.911 49	0.913 08	0.914 65	0.916 21	0.917 73
1.4	0.926 47	0.927 85	0.929 22	0.930 56	0.931 89

TABLE A (*Continued*)

t	0.00	0.01	0.02	0.03	0.04
1.5	0.939 43	0.940 62	0.941 79	0.942 95	0.944 08
1.6	0.950 53	0.951 54	0.952 54	0.953 52	0.954 48
1.7	0.959 94	0.960 80	0.961 64	0.962 46	0.963 27
1.8	0.967 84	0.968 56	0.969 26	0.969 95	0.970 62
1.9	0.974 41	0.975 00	0.975 58	0.976 15	0.976 70
2.0	0.979 82	0.980 30	0.980 77	0.981 24	0.981 69
2.1	0.984 22	0.984 61	0.985 00	0.985 37	0.985 74
2.2	0.987 78	0.988 09	0.988 40	0.988 70	0.988 99
2.3	0.990 61	0.990 86	0.991 11	0.991 34	0.991 58
2.4	0.992 86	0.993 05	0.993 24	0.993 43	0.993 61
2.5	0.994 61	0.994 77	0.994 92	0.995 06	0.995 20
2.6	0.995 98	0.996 09	0.996 21	0.996 32	0.996 43
2.7	0.997 02	0.997 11	0.997 20	0.997 28	0.997 36
2.8	0.997 81	0.997 88	0.997 95	0.998 01	0.998 07
2.9	0.998 41	0.998 46	0.998 51	0.998 56	0.998 61
3.0	0.998 86	0.998 89	0.998 93	0.998 97	0.999 00
3.1	0.999 18	0.999 21	0.999 24	0.999 26	0.999 29
3.2	0.999 42	0.999 44	0.999 46	0.999 48	0.999 50
3.3	0.999 60	0.999 61	0.999 62	0.999 64	0.999 65
3.4	0.999 72	0.999 73	0.999 74	0.999 75	0.999 76
3.5	0.999 81	0.999 81	0.999 82	0.999 83	0.999 83
3.6	0.999 87	0.999 87	0.999 88	0.999 83	0.999 89
3.7	0.999 91	0.999 92	0.999 92	0.999 92	0.999 92
3.8	0.999 94	0.999 94	0.999 95	0.999 95	0.999 95
3.9	0.999 96	0.999 96	0.999 96	0.999 97	0.999 97

Source: From W. M. Hines and D. C. Montgomery, *Probability and Statistics in Engineering Science,* 2nd Ed., John Wiley & Sons, New York, 1980, pp. 474–475. Reprinted by permission.

of goodness of fit.'' Actually, this test can be used to verify if a random sample comes from a specified theoretical distribution (such as the binomical, Poisson, uniform, and normal distributions). The procedure can be described as follows:

1. A theoretical distribution is specified in the null hypothesis H_0.
2. A level of significance α is assumed.
3. An empirical distribution with m frequency classes is formed with the observations in a random sample of size n. In this distribution 0_i is the observed frequency (number of observations) in class i, for $i = 1, 2, \ldots, m$.
4. The test statistic used is defined as

$$G = \sum_{i=1}^{m} (O_i - e_i)^2 / e_i$$

where e_i is the expected frequency for class i according to the theoretical distribution specified in the null hypothesis H_0.

It is shown that this test statistic approximately follows a chi-square distribution, with $m - 1 - r$ degrees of freedom, where r is the number of population parameters estimated by sample statistics.

5. The decision rule is simply stated as follows:

H_0 is rejected if the calculated value of G exceeds the critical value $\chi^2_{m-1-r,\alpha}$

The following numerical example illustrates these steps of the goodness of fit procedure. In a test of industrial welding robots 6 samples of 20 robots each were taken at random. Each robot in the sample operated continuously until it either failed or reached 1000 hr of operation. It is desired to verify if the true proportion of the entire population of robots that can operate over 1000 hr is 0.97.

TABLE B Percentage Points of the *t* Distribution

	α										
v^a	0.45	0.40	0.35	0.30	0.25	0.125	0.05	0.025	0.0125	0.005	0.0025
1	0.158	0.325	0.510	0.727	1.000	2.414	6.314	12.71	25.45	63.66	127.3
2	0.142	0.289	0.445	0.617	0.817	1.604	2.920	4.303	6.205	9.925	14.09
3	0.137	0.277	0.424	0.584	0.765	1.423	2.353	3.183	4.177	5.841	7.453
4	0.134	0.271	0.414	0.569	0.741	1.344	2.132	2.776	3.495	4.604	5.598
5	0.132	0.267	0.408	0.559	0.727	1.301	2.015	2.571	3.163	4.032	4.773
6	0.131	0.265	0.404	0.553	0.718	1.273	1.943	2.447	2.969	3.707	4.317
7	0.130	0.263	0.402	0.549	0.711	1.254	1.895	2.365	2.841	3.500	4.029
8	0.130	0.262	0.399	0.546	0.706	1.240	1.860	2.306	2.752	3.355	3.833
9	0.129	0.261	0.398	0.543	0.703	1.230	1.833	2.262	2.685	3.250	3.690
10	0.129	0.260	0.397	0.542	0.700	1.221	1.813	2.228	2.634	3.169	3.581
11	0.129	0.260	0.396	0.540	0.697	1.215	1.796	2.201	2.593	3.106	3.500
12	0.128	0.259	0.395	0.539	0.695	1.209	1.782	2.179	2.560	3.055	3.428
13	0.128	0.259	0.394	0.538	0.694	1.204	1.771	2.160	2.533	3.012	3.373
14	0.128	0.258	0.393	0.537	0.692	1.200	1.761	2.145	2.510	2.977	3.326
15	0.128	0.258	0.393	0.536	0.691	1.197	1.753	2.132	2.490	2.947	3.286
20	0.127	0.257	0.391	0.533	0.687	1.185	1.725	2.086	2.423	2.845	3.153
25	0.127	0.256	0.390	0.531	0.684	1.178	1.708	2.060	2.385	2.787	3.078
30	0.127	0.256	0.389	0.530	0.683	1.173	1.697	2.042	2.360	2.750	3.030
40	0.126	0.255	0.388	0.529	0.681	1.167	1.684	2.021	2.329	2.705	2.971
60	0.126	0.254	0.387	0.527	0.679	1.162	1.671	2.000	2.299	2.660	2.915
120	0.126	0.254	0.386	0.526	0.677	1.156	1.658	1.980	2.270	2.617	2.860
∞	0.126	0.253	0.385	0.524	0.674	1.150	1.645	1.960	2.241	2.576	2.807

[a] v = degrees of freedom.
Source: From W. M. Hines and D. C. Montgomery, *Probability and Statistics in Engineering Science,* 2nd Ed., John Wiley & Sons, New York, 1980, p. 477. Reprinted by permission.

In other words, we want to test the hypothesis that the number of robots satisfying the specified requirements follows a binomial distribution with parameter equal to 0.97.

The computations for the appropriate chi-square test with $\alpha = 0.05$, $m = 6$, and $r = 0$ are summarized in Table 4, where O_i is the observed number of robots in each sample of 20 operating over 1000 hr, and $e_i = (20)\ (0.97) = 19.4$. Since $G = 0.823$ is significantly smaller than $\chi^2_{5,.05} = 11.1$, we accept the null hypothesis that the true proportion of robots operating continuously over 1000 hr is indeed 0.97.

Other non-parametric techniques frequently used are:

1. The sign test to compare two treatments. We assume that there are several independent pairs of observations on the two treatments. The hypothesis to be tested states that each difference has a probability distribution having mean equal to zero. For each difference the algebraic sign is noted and then the number of times the less frequent sign is considered as the test statistic. There are specialized tables for the critical value of this quantity once a level of significance is chosen.
2. The Wilcox signed-rank test is used to test the hypothesis that observations come from symmetrical populations having a specified common median. For each observation the hypothesized median is substracted and then all differences are ranked from lowest to highest in order of magnitude, omitting the algebraic sign. The test statistic is the sum of ranks for all differences originally having positive signs. The significance test is performed by means of a statistic known as the signed-rank statistic.
3. Run tests to test the hypothesis that observations have been randomly collected from a single population. In this procedure, positive signs are assigned to observations above the median, and negative signs to those below. If the number of runs associated with plus and minus signs is larger or smaller than expected by chance, the hypothesis is rejected. The critical values for the number of runs comes from specialized tables.

There are entire books devoted to nonparametric statistical testing. We have only tried to alert the reader to several common examples.

TABLE C Percentage Points of the χ^2 Distribution

v^a α	0.995	0.990	0.975	0.950	0.500	0.050	0.025	0.010	0.005
1	0.00+	0.00+	0.00+	0.00+	0.45	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	18.34	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	24.23	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	90.53	95.07	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17

[a] v = degrees of freedom.
Source: From W. M. Hines and D. C. Montgomery, *Probability and Statistics in Engineering Science,* 2nd Ed., John Wiley & Sons, New York, 1980, p. 476. Reprinted by permission.

15. HYPOTHESIS TESTING IN THE ANALYSIS OF DESIGNED EXPERIMENTS

15.1. One-Factor Experiments

The simplest experiment corresponds to one factor and no restrictions on randomization. The statistical model of the experiment response is formulated as

$$Y_{if} = \mu + \tau_j + \varepsilon_{ij}, \qquad j = 1, 2, \ldots, k; \qquad i = 1, 2, \ldots, n_j$$

where μ = common effect, μ_j = population mean for jth treatment, and τ_j = effect of jth treatment, defined as $\mu_j - \mu$. There are two types of completely randomized one-factor models:

Fixed-effect model, H_0: $\tau_j = 0$, for all j
Random-effect model, H_0: $\sigma_\tau^2 = 0$, τ_j distributed as $N(0, \sigma_\tau^2)$

The hypothesis testing is carried out in each case by means of an F-test.

Other one-factor experiments can include one or more restrictions on randomization. Examples of these designs are the randomized block design, the Latin square design, the Graeco-Latin square design, and the Youden square design. In all these design the effect of the restrictions on randomization can be tested using F-tests.

TABLE D Percentage Points of the *F* Distribution ($\alpha = 0.10$)

v_2	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.13	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Source: From W. M. Hines and D. C. Montgomery, *Probability and Statistics in Engineering Science,* 2nd Ed., John Wiley & Sons, New York, 1980, pp. 482–483. Reprinted by permission.

15.2. After-ANOVA Range Tests

The purpose of the range test is to investigate which *pairs* of treatments are significantly different to cause H_0 to be rejected. Consider the sample averages $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_k$, computed from k random samples drawn from the k populations corresponding to the k levels of the factor. Furthermore, let us assume that $n_j = n$ for all j. These sample averages can be rearranged in increasing order of magnitude, from smallest to largest, as $\overline{Y}_{(1)}, \overline{Y}_{(2)}, \ldots, \overline{Y}_{(k)}$. The hypothesis to be tested is formulated as H_0: $\mu_{(j)} = \mu_{(i)}$, for any two values of i and j such that $j > i$.

The statistic to be used is the *studentized range statistic* as $Q_{rf} = \dfrac{\overline{Y}_{(j)} - \overline{Y}_{(i)}}{(MS_{\text{error}}/n)^{1/2}}$ where $r = j - i + 1$ is number of steps (on an ordered scale) associated with the range defined by the ith and the jth treatments, or more specifically, treatments (i) and (j). There are several versions of the range test, the following ones being the most popular. These tests use specialized tables of critical values of the studentized range statistic:

1. Newman-Keusl range test, $r = 2, \ldots, k$
2. Tukey's Range Procedure, $r = k$
3. Duncan's Multiple Range Test, $r = 2, \ldots, k$

TABLE D (*Continued*)

10	12	15	20	24	30	40	60	120	∞	v_2
60.20	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.83	1
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	2
5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13	3
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	4
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10	5
2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	6
2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	7
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	8
2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	9
2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06	10
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97	11
2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90	12
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85	13
2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80	14
2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76	15
2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72	16
2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69	17
1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66	18
1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63	19
1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61	20
1.92	1.88	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59	21
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57	22
1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55	23
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53	24
1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	25
1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	26
1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	27
1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	28
1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	29
1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	30
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	40
1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	60
1.65	1.60	1.54	1.48	1.45	1.41	1.37	1.32	1.26	1.19	120
1.60	1.55	1.49	1.42	1.33	1.34	1.30	1.24	1.17	1.00	∞

15.3. Factorial Experiments

Instead of assuming that the experimental response can be affected by one factor, a factorial experiment assumes that it is affected by two or more factors, each having an arbitrary number of levels. In a factorial experiment a level combination (one level per factor) is known as an experimental condition. All experimental conditions are sampled, and all observations are collected at random. It is possible to test the significance of each main effect (due to a factor) or and interaction effect (due to a group of factors) by means of an F-test.

15.4. Hypothesis Testing in Regression Analysis

A regression model is a *fitting* relationship that allows the estimation of a dependent variable or experimental response for given settings of a specified group of independent variables or factors. The parameters of the model are known as regression coefficients. Typical tests include the following:

1. Testing the hypothesis that a regression coefficient is equal to zero. The test statistic is the t-statistic.
2. Testing for linearity of regression using an analysis-of-variance technique. The test statistic is the F-statistic.

REFERENCES

Bowker, A. H., and Lieberman, G. J. (1972), *Engineering Statistics,* 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ.

Devore, J. L. (1987), *Probability and Statistics for Engineering and the Sciences,* 2nd Ed., Brooks/Cole, Monterey, CA.

Dougherty, E. R. (1990), *Probability and Statistics for the Engineering, Computing, and Physical Sciences,* Prentice-Hall, Englewood Cliffs, NJ.

Hogg, R. V., and Craig, A. T. (1978), *Introduction to Mathematical Statistics,* 4th Ed., Macmillan, New York.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods,* 7th Ed., Iowa State University Press, Ames, IA.

Hollander, M., and Wolfe, D. (1973), *Nonparametric Statistical Methods,* John Wiley & Sons, New York.

Lehmann, E. (1975), *Nonparametrics: Statistical Methods Based on Ranks,* Holden-Day, San Francisco.

Marascuilo, L., and McSweeney, M. (1977), *Nonparametric and Distribution-Free Methods for the Social Sciences,* Brooks/Cole, Monterey, CA.