

3

Inferences about Process Quality

CHAPTER OUTLINE

3-1 STATISTICS AND SAMPLING DISTRIBUTIONS

- 3-1.1 Sampling from a Normal Distribution
- 3-1.2 Sampling from a Bernoulli Distribution
- 3-1.3 Sampling from a Poisson Distribution

3-2 POINT ESTIMATION OF PROCESS PARAMETERS

3-3 STATISTICAL INFERENCE FOR A SINGLE SAMPLE

- 3-3.1 Inference on the Mean of a Population, Variance Known
- 3-3.2 The Use of P -Values for Hypothesis Testing
- 3-3.3 Inference on the Mean of a Normal Distribution, Variance Unknown
- 3-3.4 Inference on the Variance of a Normal Distribution
- 3-3.5 Inference on a Population Proportion
- 3-3.6 The Probability of Type II Error and Sample Size Decisions

3-4 STATISTICAL INFERENCE FOR TWO SAMPLES

- 3-4.1 Inference for a Difference in Means, Variances Known

- 3-4.2 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

- 3-4.3 Inference on the Variances of Two Normal Distributions

- 3-4.4 Inference on Two Population Proportions

3-5 WHAT IF THERE ARE MORE THAN TWO POPULATIONS? THE ANALYSIS OF VARIANCE

- 3-5.1 An Example
- 3-5.2 The Analysis of Variance
- 3-5.3 Checking Assumptions: Residual Analysis

Supplemental Material for Chapter 3

- S3-1 Random Samples
- S3-2 Expected Value and Variance Operators
- S3-3 Proof that $E(\bar{x}) = \mu$ and $E(s^2) = \sigma^2$
- S3-4 More about Parameter Estimation
- S3-5 Proof that $E(s) \neq \sigma$
- S3-6 More about Checking Assumptions in the t -Test
- S3-7 Expected Mean Squares in the Single-Factor Analysis of Variance

LEARNING OBJECTIVES

1. Explain the concept of random sampling
2. Explain the concept of a sampling distribution
3. Explain the general concept of estimating the parameters of a population or probability distribution
4. Know how to explain the precision with which a parameter is estimated
5. Construct and interpret confidence intervals on a single mean and on the difference in two means
6. Construct and interpret confidence intervals on a single variance or the ratio of two variances
7. Construct and interpret confidence intervals on a single proportion and on the difference in two proportions
8. Test hypotheses on a single mean and on the difference in two means
9. Test hypotheses on a single variance and on the ratio of two variances
10. Test hypotheses on a single proportion and on the difference in two proportions
11. Use the P -value approach for hypothesis testing
12. Understand how the analysis of variance (ANOVA) is used to test hypotheses about the equality of more than two means

3-1 STATISTICS AND SAMPLING DISTRIBUTIONS

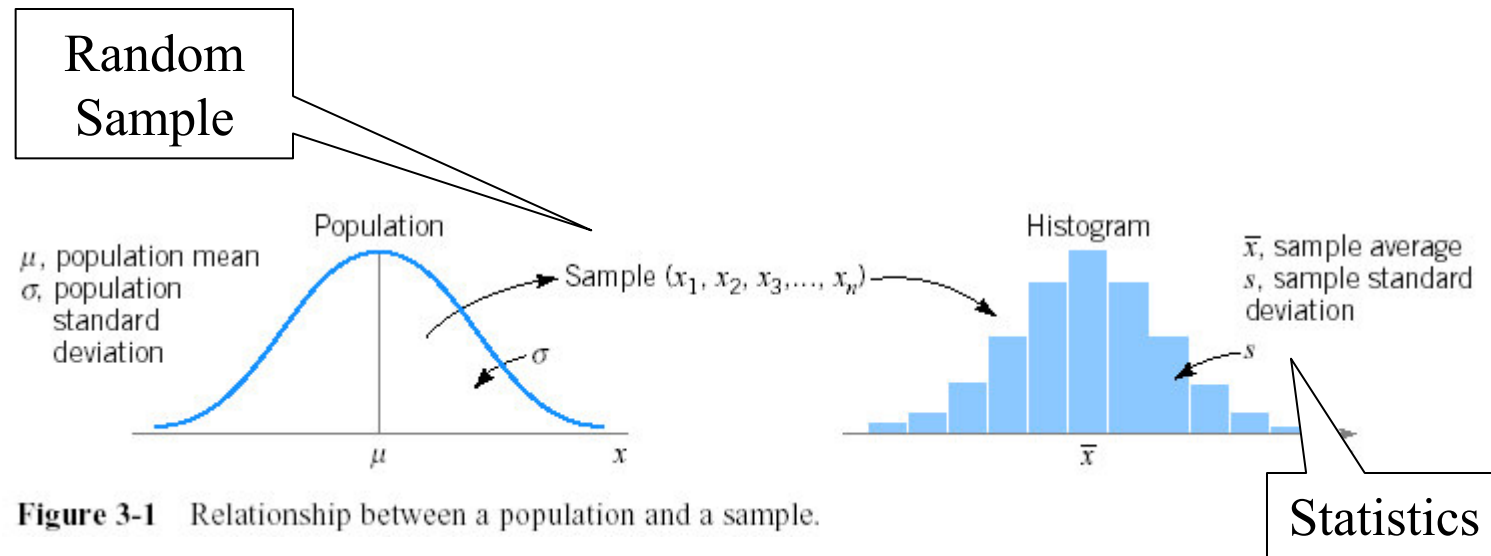


Figure 3-1 Relationship between a population and a sample.

If we know the probability distribution of the population from which the sample was taken, we can often determine the probability distribution of various statistics computed from the sample data. The probability distribution of a statistic is called a **sampling distribution**. We now present the sampling distributions associated with three common sampling situations.

3-1.1 Sampling from a Normal Distribution

Chi-square (χ^2) Distribution

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$y = x_1^2 + x_2^2 + \cdots + x_n^2$$

$$f(y) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} y^{(n/2)-1} e^{-y/2} \quad y > 0 \quad (3-4)$$

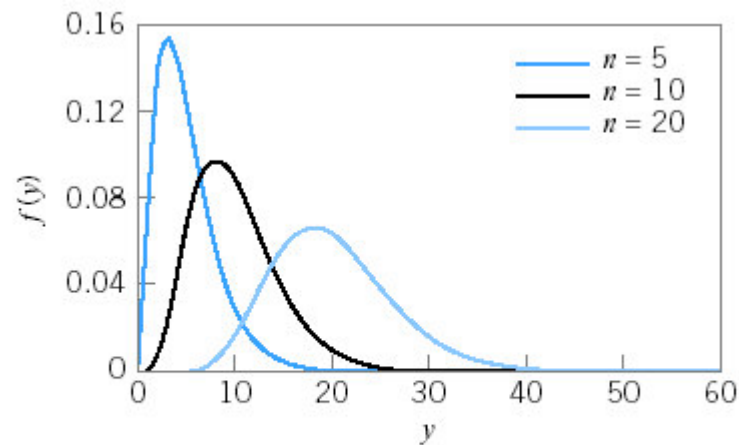


Figure 3-2 Chi-square distribution for selected values of n (number of degrees of freedom).

To illustrate the use of the chi-square distribution, suppose that $x_1, x_2, \dots, x_n$ is a random sample from an $N(\mu, \sigma^2)$ distribution. Then the random variable

$$y = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \quad (3-5)$$

has a chi-square distribution with $n - 1$ degrees of freedom. However, using equation 3-2, which defines the sample variance, we may rewrite equation 3-5 as

$$y = \frac{(n-1)s^2}{\sigma^2}$$

That is, the sampling distribution of $(n-1)s^2/\sigma^2$ is χ_{n-1}^2 when sampling from a normal distribution.

t Distribution

$$t = \frac{x}{\sqrt{y/k}} \quad (3-6)$$

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\Gamma(k/2)} \left(\frac{t^2}{k} + 1 \right)^{-(k+1)/2} \quad -\infty < t < \infty \quad (3-7)$$

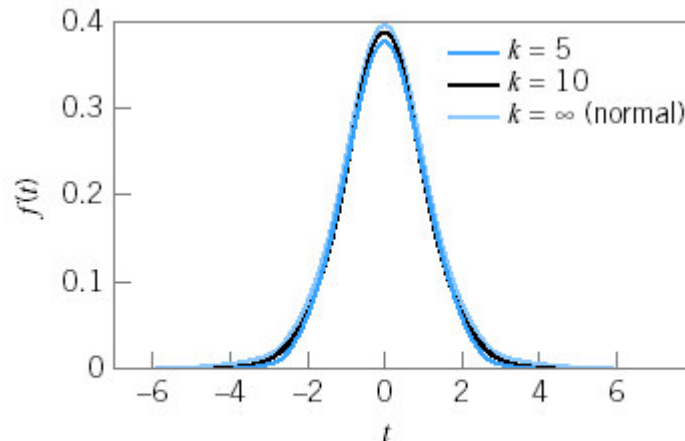


Figure 3-3 The t distribution for selected values of k (number of degrees of freedom).

As an example of a random variable that is distributed as t , suppose that $x_1, x_2, \dots, x_n$ is a random sample from the $N(\mu, \sigma^2)$ distribution. If $\bar{x}$ and s^2 are computed from this sample, then

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{s/\sigma} \sim \frac{N(0,1)}{\sqrt{\chi_{n-1}^2/(n-1)}}$$

using the fact that $(n-1)s^2/\sigma^2 \sim \chi_{n-1}^2$. Now, $\bar{x}$ and s^2 are independent, so the random variable

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{3-8}$$

has a t distribution with $n - 1$ degrees of freedom.

F Distribution

$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)} \frac{x^{(u/2)-1}}{\left[\left(\frac{u}{v}\right)x + 1\right]^{(u+v)/2}} \quad 0 < x < \infty \quad (3-10)$$

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

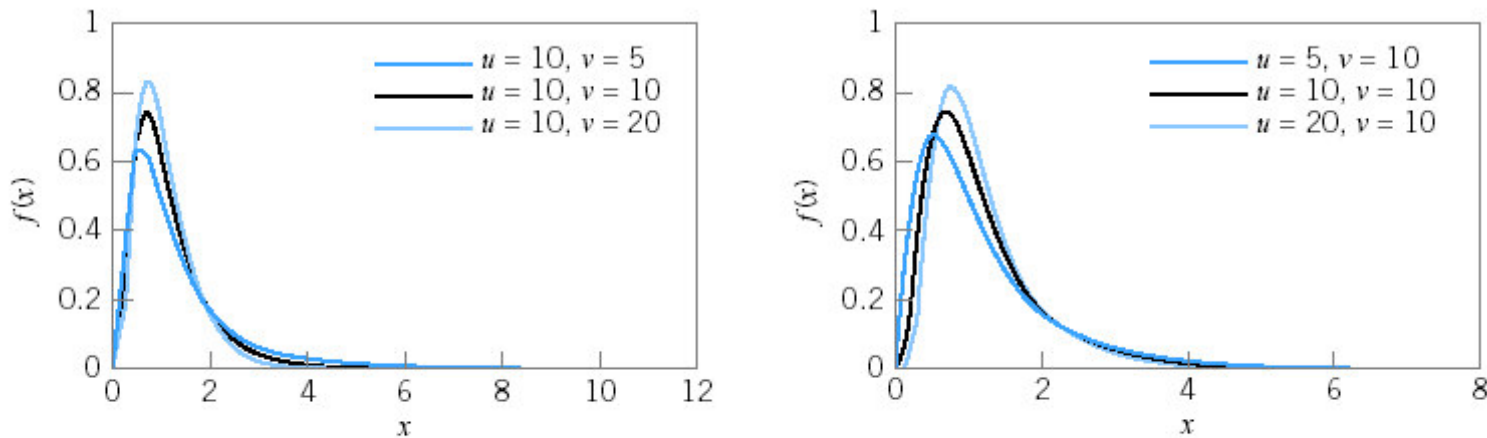


Figure 3-4 The F distribution for selected values of u (numerator degrees of freedom) and v (denominator degrees of freedom).

3-1.2 Sampling from a Bernoulli Distribution

Suppose that a random sample of n observations—say, $x_1, x_2, \dots, x_n$ —is taken from a Bernoulli process with constant probability of success p . Then the sum of the sample observations

$$x = x_1 + x_2 + \cdots + x_n \quad (3-11)$$

has a binomial distribution with parameters n and p . Furthermore, since each x_i is either 0 or 1, the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3-12)$$

is a discrete random variable with range space $\{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$. The distribution of $\bar{x}$ can be obtained from the binomial since

$$P\{\bar{x} \leq a\} = P\{x \leq an\} = \sum_{k=0}^{[an]} \binom{n}{k} p^k (1-p)^{n-k}$$

where $[an]$ is the largest integer less than or equal to an . The mean and variance of $\bar{x}$ are

$$\mu_{\bar{x}} = p$$

and

$$\sigma_{\bar{x}}^2 = \frac{p(1-p)}{n}$$

3-1.3 Sampling from a Poisson Distribution

Consider a random sample of size n from a Poisson distribution with parameter λ —say, $x_1, x_2, \dots, x_n$. The distribution of the sample sum

$$x = x_1 + x_2 + \dots + x_n \quad (3-13)$$

is also Poisson with parameter $n\lambda$. More generally, the sum of n independent Poisson random variables is distributed Poisson with parameter equal to the sum of the individual Poisson parameters.

Now consider the distribution of the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3-14)$$

This is a discrete random variable that takes on the values $\{0, 1/n, 2/n, \dots\}$, and with probability distribution found from

$$P\{\bar{x} \leq a\} = P\{x \leq an\} = \sum_{k=0}^{[an]} \frac{e^{-n\lambda} (n\lambda)^k}{k!} \quad (3-15)$$

where $[an]$ is the largest integer less than or equal to an . The mean and variance of $\bar{x}$ are

$$\mu_{\bar{x}} = \lambda$$

and

$$\sigma_{\bar{x}}^2 = \frac{\lambda}{n}$$

Sometimes more general linear combinations of Poisson random variables are used in quality-engineering work. For example, consider the linear combination

$$L = a_1x_1 + a_2x_2 + \cdots + a_mx_m = \sum_{i=1}^m a_ix_i \quad (3-16)$$

where the $\{x_i\}$ are independent Poisson random variables each having parameter $\{\lambda_i\}$, respectively, and the $\{a_i\}$ are constants. This type of function occurs in situations where a unit of product can have m different types of defects or nonconformities (each modeled with a Poisson distribution with parameter λ_i) and the function used for quality monitoring purposes is a linear combination of the number of observed nonconformities of each type. The constants $\{a_i\}$ in equation 3-16 might be chosen to weight some types of nonconformities more heavily than others. For example, functional defects on a unit would receive heavier weight than appearance flaws. These schemes are sometimes called **demerit** procedures (see Section 6-3.3). In general, the distribution of L is not Poisson unless all $a_i = 1$ in equation 3-16; that is, sums of independent Poisson random variables are Poisson distributed, but more general linear combinations are not.

3-2 POINT ESTIMATION OF PROCESS PARAMETERS

We may define an estimator of an unknown parameter as a statistic that corresponds to the parameter. A particular numerical value of an estimator, computed from sample data, is called an **estimate**. A **point estimator** is a statistic that produces a single numerical value as the estimate of the unknown parameter.

To illustrate, consider the random variable x with probability distribution $f(x)$ shown in Fig. 3-1 on p. 88. Suppose that the mean μ and variance σ^2 of this distribution are both unknown. If a random sample of n observations is taken, then the **sample mean** $\bar{x}$ and **sample variance** s^2 are point estimators of the **population mean** μ and **population variance** σ^2 , respectively. Suppose that this distribution represents a process producing bearings and the random variable x is the inside diameter. We want to obtain point estimates of the mean and variance of the inside diameter of bearings produced by this process. We could measure the inside diameters of a random sample of $n = 20$ bearings (say). Then the sample mean and sample variance could be computed. If this yields $\bar{x} = 1.495$ and $s^2 = 0.001$, then the point estimate of μ is $\hat{\mu} = \bar{x} = 1.495$ and the point estimate of σ^2 is $\hat{\sigma}^2 = s^2 = 0.001$. Recall that the “^” symbol is used to denote an estimate of a parameter.

A number of important properties are required of good point estimators. Two of the most important of these properties are the following:

1. The point estimator should be **unbiased**. That is, the expected value of the point estimator should be the parameter being estimated.
2. The point estimator should have **minimum variance**. Any point estimator is a random variable. Thus, a minimum variance point estimator should have a variance that is smaller than the variance of any other point estimator of that parameter.

The *sample* mean and variance $\bar{x}$ and s^2 are unbiased estimators of the *population* mean and variance μ and σ^2 , respectively. That is,

$$E(\bar{x}) = \mu \quad \text{and} \quad E(s^2) = \sigma^2$$

where the operator E is simply the expected value operator, a shorthand way of writing the process of finding the mean of a random variable. (See the supplemental material for this chapter for more information about mathematical expectation.)

The sample standard deviation s is *not* an unbiased estimator of the population standard deviation σ . It can be shown that

$$\begin{aligned} E(s) &= \left(\frac{2}{n-1} \right)^{1/2} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} \sigma \\ &= c_4 \sigma \end{aligned} \tag{3-17}$$

Appendix Table VI gives values of c_4 for sample sizes $2 \leq n \leq 25$. We can obtain an unbiased estimate of the standard deviation from

$$\hat{\sigma} = \frac{s}{c_4} \tag{3-18}$$

- As n gets large the bias goes to zero

In many applications of statistics to quality-engineering problems, it is convenient to estimate the standard deviation by the **range method**. Let $x_1, x_2, \dots, x_n$ be a random sample of n observations from a normal distribution with mean μ and variance σ^2 . The **range** of the sample is

$$\begin{aligned} R &= \max(x_i) - \min(x_i) \\ &= x_{\max} - x_{\min} \end{aligned} \tag{3-19}$$

That is, the range R is simply the difference between the largest and smallest sample observations.

The random variable $W = R/\sigma$ is called the **relative range**. The distribution of W has been well studied. The mean of W is a constant d_2 that depends on the size of the sample. That is, $E(W) = d_2$. Therefore, an unbiased estimator of the standard deviation σ of a normal distribution is

$$\hat{\sigma} = \frac{R}{d_2} \quad (3-20)$$

Values of d_2 for sample sizes $2 \leq n \leq 25$ are given in Appendix Table VI.

Using the range to estimate σ dates from the earliest days of statistical quality control, and it was popular because it is very simple to calculate. With modern calculators and computers, this isn't a major consideration today. Generally, the "quadratic estimator" based on s is preferable. However, if the sample size n is relatively small, the range method actually works very well. The relative efficiency of the range method compared to s is shown here for various sample sizes:

Sample Size n	Relative Efficiency
2	1.000
3	0.992
4	0.975
5	0.955
6	0.930
10	0.850

For moderate values of n —say, $n \geq 10$ —the range loses efficiency rapidly, as it ignores all of the information in the sample between the extremes. However, for small sample sizes—say, $n \leq 6$ —it works very well and is entirely satisfactory. We will use the range method to estimate the standard deviation for certain types of control charts in Chapter 5. The **supplemental text material** contains more information about using the range to estimate variability. Also see Woodall and Montgomery (2000–01).

3-3 STATISTICAL INFERENCE FOR A SINGLE SAMPLE

The techniques of statistical inference can be classified into two broad categories: **parameter estimation** and **hypothesis testing**.

A **statistical hypothesis** is a statement about the values of the parameters of a probability distribution.

Alternative Hypothesis

Null Hypothesis

$$\begin{aligned} H_0: & \mu = 1.500 \\ H_1: & \mu \neq 1.500 \end{aligned} \quad (3-21)$$

- In this example, H_1 is a **two-sided alternative hypothesis**

To test a hypothesis, we take a random sample from the population under study, compute an appropriate **test statistic**, and then either reject or fail to reject the null hypothesis H_0 . The set of values of the test statistic leading to rejection of H_0 is called the **critical region** or **rejection region** for the test.

Two kinds of errors may be committed when testing hypotheses. If the null hypothesis is rejected when it is true, then a type I error has occurred. If the null hypothesis is not rejected when it is false, then a type II error has been made. The probabilities of these two types of errors are denoted as

$$\alpha = P\{\text{type I error}\} = P\{\text{reject } H_0 | H_0 \text{ is true}\}$$

$$\beta = P\{\text{type II error}\} = P\{\text{fail to reject } H_0 | H_0 \text{ is false}\}$$

Sometimes it is more convenient to work with the **power** of the test, where

$$\text{Power} = 1 - \beta = P\{\text{reject } H_0 | H_0 \text{ is false}\}$$

Thus, the power is the probability of *correctly* rejecting H_0 .

Thus, the power is the probability of *correctly* rejecting H_0 . In quality control work, α is sometimes called the **producer's risk**, because it denotes the probability that a good lot will be rejected, or the probability that a process producing acceptable values of a particular quality characteristic will be rejected as performing unsatisfactorily.

In addition, β is sometimes called the **consumer's risk**, because it denotes the probability of accepting a lot of poor quality, or allowing a process that is operating in an unsatisfactory manner relative to some quality characteristic to continue in operation.

The general procedure in hypothesis testing is to specify a value of the probability of type I error α , and then to design a test procedure so that a small value of the probability of type II error β is obtained. Thus, we speak of directly controlling or choosing the α risk. The β risk is generally a function of sample size and is controlled indirectly. The larger is the sample size(s) used in the test, the smaller is the β risk.

In this section we will review hypothesis testing procedures when a **single sample** of n observations has been taken from the process. We will also show how the information about the values of the process parameters that is in this sample can be expressed in terms of an interval estimate called a **confidence interval**. In Section 3-4 we will consider statistical inference for two samples from two possibly different processes.

3-3.1 Inference on the Mean of a Population, Variance Known

$$\begin{aligned}H_0: \mu &= \mu_0 \\H_1: \mu &\neq \mu_0\end{aligned}\tag{3-22}$$

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\tag{3-23}$$

- H_1 in equation 3-22 is a **two-sided alternative hypothesis**
- The procedure for testing this hypothesis is to:
 - take a random sample of n observations on the random variable x ,
 - compute the test statistic, and
 - reject H_0 if $|Z_0| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage of the standard normal distribution.

One-Sided Alternative Hypotheses

- In some situations we may wish to reject H_0 only if the true mean is larger than μ_0
 - Thus, the one-sided alternative hypothesis is $H_1: \mu > \mu_0$, and we would reject $H_0: \mu = \mu_0$ only if $Z_0 > Z_\alpha$
- If rejection is desired only when $\mu < \mu_0$
 - Then the alternative hypothesis is $H_1: \mu < \mu_0$, and we reject H_0 only if $Z_0 < -Z_\alpha$

..... **EXAMPLE 3-1**

The response time of a distributed computer system is an important quality characteristic. The system manager wants to know whether the mean response time to a specific type of command exceeds 75 millisecc. From previous experience, he knows that the standard deviation of response time is 8 millisecc. The appropriate hypotheses are

$$H_0: \mu = 75$$

$$H_1: \mu > 75$$

The command is executed 25 times and the response time for each trial is recorded. We assume that these observations can be considered as a random sample of the response times. The sample average response time is $\bar{x} = 79.25$ millisecc. The value of the test statistic is

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{79.25 - 75}{8/\sqrt{25}} = 2.66$$

If we specify a type I error of $\alpha = 0.05$, then from Appendix Table II we find $Z_\alpha = Z_{0.05} = 1.645$. Therefore, we reject $H_0: \mu = 75$ and conclude that the mean response time exceeds 75 millisecc.

.....

Confidence Intervals

An interval estimate of a parameter is the interval between two statistics that includes the true value of the parameter with some probability. For example, to construct an interval estimator of the mean μ , we must find two statistics L and U such that

$$P\{L \leq \mu \leq U\} = 1 - \alpha \quad (3-24)$$

The resulting interval

$$L \leq \mu \leq U$$

is called a **100(1 - α)% confidence interval** for the unknown mean μ . L and U are called the lower and upper confidence limits, respectively, and $1 - \alpha$ is called the confidence coefficient. Sometimes the half-interval width $U - \mu$ or $\mu - L$ is called the **accuracy** of the confidence interval. The interpretation of a confidence interval is that if a large number of such intervals are constructed, each resulting from a random sample, then 100(1 - α)% of these intervals will contain the true value of μ . Thus, confidence intervals have a frequency interpretation.

The confidence interval (3-24) might be more properly called a **two-sided** confidence interval, as it specifies both a lower and an upper bound on μ . Sometimes in quality-control applications, a **one-sided** confidence bound might be more appropriate. A one-sided lower $100(1 - \alpha)\%$ confidence bound on μ would be given by

$$L \leq \mu \quad (3-25)$$

where L , the lower confidence bound, is chosen so that

$$P\{L \leq \mu\} = 1 - \alpha \quad (3-26)$$

A one-sided upper $100(1 - \alpha)\%$ confidence bound on μ would be

$$\mu \leq U \quad (3-27)$$

where U , the upper confidence bound, is chosen so that

$$P\{\mu \leq U\} = 1 - \alpha \quad (3-28)$$

Confidence Interval on Mean, Variance Known

Consider the random variable x , with unknown mean μ and known variance σ^2 . Suppose a random sample of n observations is taken—say, $x_1, x_2, \dots, x_n$ —and $\bar{x}$ is computed. Then the $100(1 - \alpha)\%$ two-sided confidence bound on μ is

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (3-29)$$

where $Z_{\alpha/2}$ is the percentage point of the $N(0, 1)$ distribution such that $P\{z \geq Z_{\alpha/2}\} = \alpha/2$.

Furthermore, a $100(1 - \alpha)\%$ upper confidence bound on μ is

$$\mu \leq \bar{x} + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad (3-30)$$

whereas a $100(1 - \alpha)\%$ lower confidence bound on μ is

$$\bar{x} - Z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \mu \quad (3-31)$$

..... **EXAMPLE 3-2**

Reconsider the computer response time scenario from Example 3-1. Since $\bar{x} = 79.25$ millisecc, we know that a reasonable point estimate of the mean response time is $\hat{\mu} = \bar{x} = 79.25$ millisecc. We can also find a $100(1 - \alpha)\%$ confidence interval for μ . Suppose a 95% two-sided confidence interval is specified. Then from equation 3-29 we can compute

$$\begin{aligned}\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ 79.25 - 1.96 \frac{8}{\sqrt{25}} &\leq \mu \leq 79.25 + 1.96 \frac{8}{\sqrt{25}} \\ 76.114 &\leq \mu \leq 82.386\end{aligned}$$

Another way to express this result is that our estimate of mean response time is 79.25 millisecc $\pm$ 3.136 millisecc with 95% confidence.

Notice that the confidence interval in Example 3-2 does *not* include the value $\mu = 75$. Furthermore, in Example 3-1 the hypothesis $H_0: \mu = 75$ was rejected at $\alpha = 0.05$. This is not a coincidence. In general, the test of significance for a parameter at level of significance will lead to rejection of H_0 if, and only if, the parameter value specific in H_0 is not included in the $100(1 - \alpha)\%$ confidence interval.

3-3.2 The Use of P -Values for Hypothesis Testing

Definition

The **P -value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 .

For the normal distribution tests discussed above, it is relatively easy to compute the P -value. If z_0 is the computed value of the test statistic, then the P -value is

$$P = \begin{cases} 2[1 - \Phi(|Z_0|)] & \text{for a two-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & \text{for an upper-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \\ \Phi(Z_0) & \text{for a lower-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0 \end{cases}$$

Here, $\Phi(Z)$ is the standard normal cumulative distribution function defined in Chapter 2.

To illustrate this, consider the computer response time problem in Example 3-1. The computed value of the test statistic is $Z_0 = 2.66$ and since the alternative hypothesis is one-tailed, the P -value is

$$P = 1 - \Phi(2.66) = 0.0039$$

Thus, $H_0: \mu = 75$ would be rejected at any level of significance $\alpha \geq P = 0.0039$. For example, H_0 would be rejected if $\alpha = 0.01$, but it would not be rejected if $\alpha = 0.001$.

3-3.3 Inference on the Mean of a Normal Distribution, Variance Unknown

$$\begin{aligned}H_0: \mu &= \mu_0 \\H_1: \mu &\neq \mu_0\end{aligned}\tag{3-32}$$

As σ^2 is unknown, it may be estimated by s^2 . If we replace σ in equation 3-23 by s , we have the **test statistic**

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\tag{3-33}$$

- For the two-sided alternative hypothesis, reject H_0 if $|t_0| > t_{\alpha/2, n-1}$, where $t_{\alpha/2, n-1}$, is the upper $\alpha/2$ percentage of the t distribution with $n - 1$ degrees of freedom
- For the one-sided alternative hypotheses,
 - If $H_1: \mu_1 > \mu_0$, reject H_0 if $t_0 > t_{\alpha, n-1}$, and
 - If $H_1: \mu_1 < \mu_0$, reject H_0 if $t_0 < -t_{\alpha, n-1}$
- One could also compute the P -value for a t -test

..... **EXAMPLE 3-3**

Rubber can be added to asphalt to reduce road noise when the material is used as pavement. Table 3-1 shows the stabilized viscosity (cP) of 15 specimens of asphalt paving material.

Table 3-1 Stabilized Viscosity of Rubberized Asphalt

Specimen	Stabilized Viscosity
1	3193
2	3124
3	3153
4	3145
5	3093
6	3466
7	3355
8	2979
9	3182
10	3227
11	3256
12	3332
13	3204
14	3282
15	3170

To be suitable for the intended pavement application, the mean stabilized viscosity should be equal to 3200. Suppose that we wish to test this hypothesis using $\alpha = 0.05$. Based on experience we are willing to initially assume that stabilized viscosity is normally distributed. The appropriate hypotheses are

$$H_0: \mu = 3200$$

$$H_1: \mu \neq 3200$$

The sample mean and sample standard deviation are

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{48,161}{15} = 3210.73$$

$$s = \sqrt{\frac{\sum_{i=1}^{15} x_i^2 - \frac{\left(\sum_{i=1}^{15} x_i\right)^2}{15}}{15 - 1}} = \sqrt{\frac{154,825,783 - \frac{(48,161)^2}{15}}{14}} = 117.61$$

and the test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3210.73 - 3200}{117.61/\sqrt{15}} = 0.35$$

Since the calculated value of the test statistic does not exceed $t_{0.025, 14} = 2.145$ or $-t_{0.025, 14} = -2.145$, we cannot reject the null hypothesis. Therefore, there is no strong evidence to conclude that the mean stabilized viscosity is different from 3200 cP.

The assumption of normality for the t -test can be checked by constructing a normal probability plot of the stabilized viscosity data. Figure 3-5 shows the normal probability plot. Because the observations lie along the straight line, there is no problem with the normality assumption.

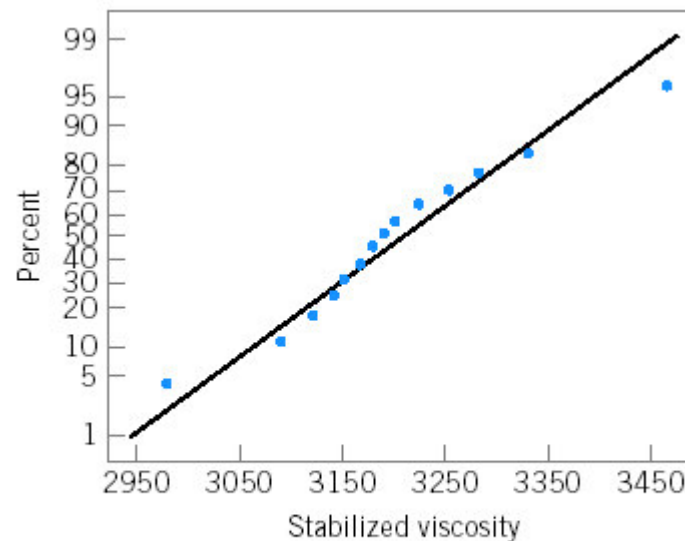


Figure 3-5 Normal probability plot of the stabilized viscosity data.

Minitab can conduct the one-sample t -test. The output from this software package is shown in the following display:

One-Sample T: Example 3-3				
Test of $\mu = 3200$ vs $\mu \text{ not } = 3200$				
Variable	N	Mean	StDev	SE Mean
Example 3-3	15	3210.7	117.6	30.4
Variable	95.0% CI		T	P
Example 3-3	(3145.6, 3275.9)		0.35	0.729

Confidence Interval on the Mean of a Normal Distribution with Variance Unknown

Suppose that x is a normal random variable with unknown mean μ and unknown variance σ^2 . From a random sample of n observations the sample mean $\bar{x}$ and sample variance s^2 are computed. Then a $100(1 - \alpha)\%$ two-sided confidence interval on the true mean is

$$\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \quad (3-34)$$

where $t_{\alpha/2, n-1}$ denotes the percentage point of the t distribution with $n - 1$ degrees of freedom such that $P\{t_{n-1} \geq t_{\alpha/2, n-1}\} = \alpha/2$.

- Section 3-3.4 describes hypothesis testing and confidence intervals on the variance of a normal distribution

3-3.5 Inference on a Population Proportion

Hypothesis Testing

$$\begin{aligned}H_0: & p = p_0 \\H_1: & p \neq p_0\end{aligned}\tag{3-42}$$

$$Z_0 = \begin{cases} \frac{(x + 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} & \text{if } x < np_0 \\ \frac{(x - 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} & \text{if } x > np_0 \end{cases}\tag{3-43}$$

The null hypothesis $H_0: p = p_0$ is rejected if $|Z_0| > Z_{\alpha/2}$. The one-sided alternative hypotheses are treated similarly.

..... **EXAMPLE 3-5**

A foundry produces steel castings used in automobile manufacturing. We wish to test the hypothesis that the fraction conforming or fallout from this process is 10%. In a random sample of 250 castings, 41 were found to be nonconforming. To test

$$H_0: p = 0.1$$

$$H_1: p \neq 0.1$$

we calculate the test statistic

$$Z_0 = \frac{(x - 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{(41 - 0.5) - (250)(0.1)}{\sqrt{250(0.1)(1 - 0.1)}} = 3.27$$

Using $\alpha = 0.05$ we find $Z_{0.025} = 1.96$, and therefore $H_0: p = 0.1$ is rejected (the P -value here is $P = 0.00108$). That is, the process fraction nonconforming or fallout is not equal to 10%.

.....

3-3.5 Inference on a Population Proportion

Confidence Intervals on a Population Proportion

There are several approaches to constructing the confidence interval on p . If n is large and $p \geq 0.1$ (say), then the normal approximation to the binomial can be used, resulting in the $100(1 - \alpha)\%$ confidence interval:

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (3-44)$$

If n is small, then the binomial distribution should be used to establish the confidence interval on p . If n is large but p is small, then the Poisson approximation to the binomial is useful in constructing confidence intervals

..... **EXAMPLE 3-6**

In a random sample of 80 automotive crankshaft bearings, 15 of the bearings have a surface finish that is rougher than the specifications will allow. The point estimate of the fraction nonconforming in the process is

$$\hat{p} = \frac{15}{80} = 0.1875$$

Assuming that the normal approximation to the binomial is appropriate, a 95% confidence interval on the process fraction nonconforming is found from equation 3-44 as

$$0.1875 - 1.96\sqrt{\frac{0.1875(0.8125)}{80}} \leq p \leq 0.1875 + 1.96\sqrt{\frac{0.1875(0.8125)}{80}}$$

which reduces to

$$0.1020 \leq p \leq 0.2730$$

.....

3-3.6 The Probability of Type II Error and Sample Size Decisions

$$\beta = \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \quad (3-46)$$

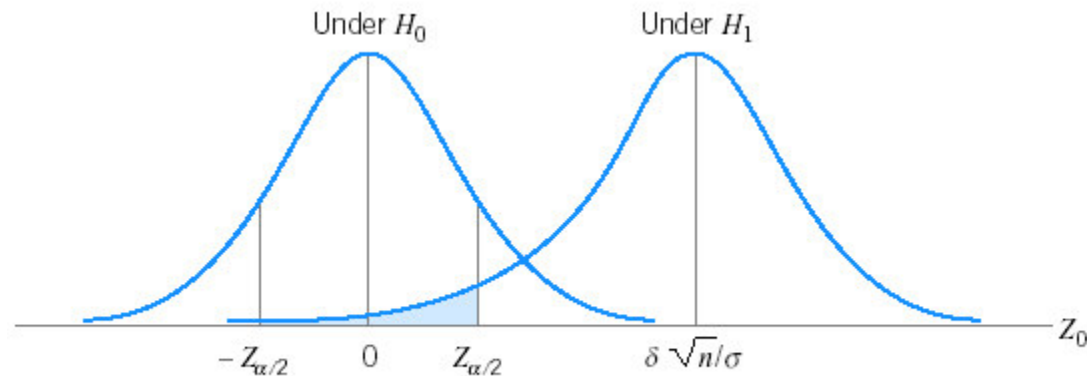


Figure 3-6 The distribution of Z_0 under H_0 and H_1 .

..... **EXAMPLE 3-7**

The mean contents of coffee cans filled on a particular production line are being studied. Standards specify that the mean contents must be 16.0 oz, and from past experience it is known that the standard deviation of the can contents is 0.1 oz. The hypotheses are

$$H_0: \mu = 16.0$$

$$H_1: \mu \neq 16.0$$

A random sample of nine cans is to be used, and the type I error probability is specified as $\alpha = 0.05$. Therefore, the test statistic is

$$Z_0 = \frac{\bar{x} - 16.0}{0.1\sqrt{9}}$$

and H_0 is rejected if $|Z_0| > Z_{0.025} = 1.96$.

Suppose that we wish to find the probability of type II error if the true mean contents are $\mu_1 = 16.1$ oz. Since this implies that $\delta = \mu_1 - \mu_0 = 16.1 - 16.0 = 0.1$, we have

$$\begin{aligned}\beta &= \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(1.96 - \frac{(0.1)(3)}{0.1}\right) - \Phi\left(-1.96 - \frac{(0.1)(3)}{0.1}\right) \\ &= \Phi(-1.04) - \Phi(-4.96) \\ &= 0.1492\end{aligned}$$

That is, the probability that we will incorrectly fail to reject H_0 if the true mean contents are 16.1 oz is 0.1492. Equivalently, we can say that the power of the test is $1 - \beta = 1 - 0.1492 = 0.8508$.

We note from examining equation 3-46 and Fig. 3-6 that β is a function of n , δ , and α . It is customary to plot curves illustrating the relationship between these parameters. Such a set of curves is shown in Fig. 3-7 for $\alpha = 0.05$. Graphs such as these are usually called **operating-characteristic (OC) curves**. The parameter on the vertical axis of these curves is β , and the parameter on the horizontal axis is $d = |\delta|/\sigma$. From examining the operating-characteristic curves, we see that

1. The further the true mean μ_1 is from the hypothesized value μ_0 (i.e., the larger the value of δ), the smaller is the probability of type II error for a given n and α . That is, for a specified sample size and α , the test will detect large differences more easily than small ones.
2. As the sample size n increases, the probability of type II error gets smaller for a specified δ and α . That is, to detect a specified difference we may make the test more powerful by increasing the sample size.

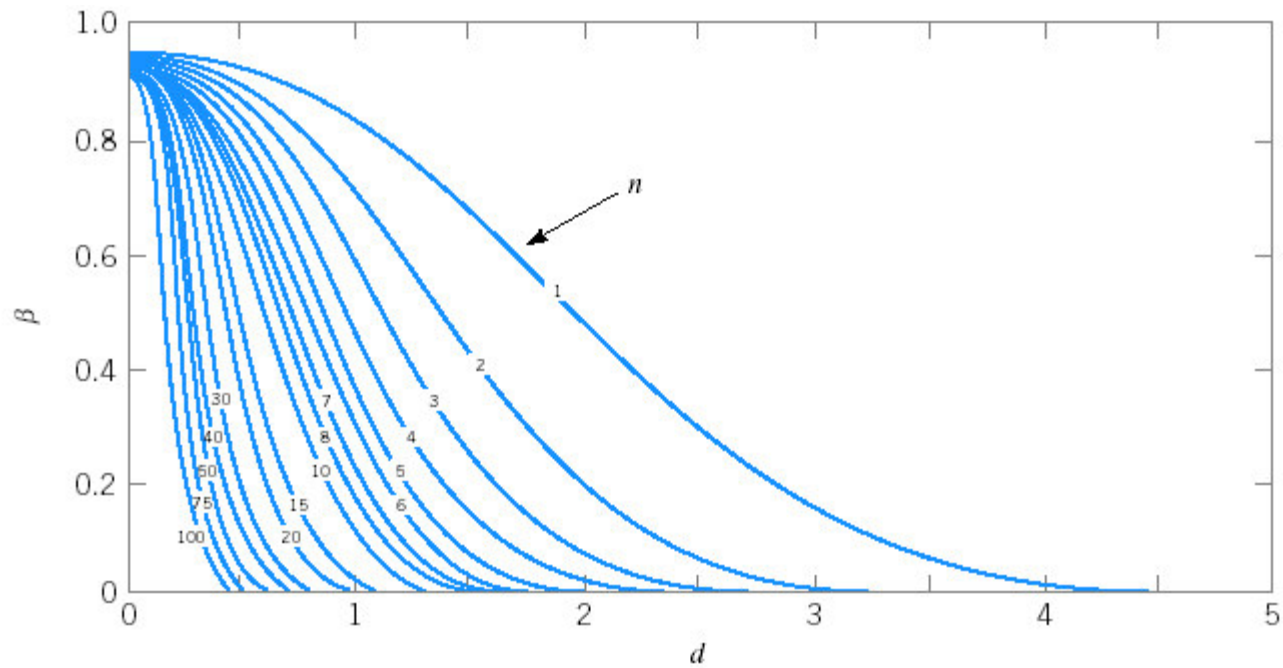


Figure 3-7 Operating-characteristic curves for the two-sided normal test with $\alpha = 0.05$.
(Reproduced with permission from C. L. Ferris, F. E. Grubbs, and C. L. Weaver, "Operating Characteristic Curves for the Common Statistical Tests of Significance," *Annals of Mathematical Statistics*, June 1946.)

Minitab can also perform power and sample size calculations for several hypothesis testing problems. The following Minitab display reproduces the power calculations from the coffee can-filling problem in Example 3-7.

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)

Calculating power for mean = null + difference

Alpha = 0.05 Sigma = 0.1

Difference	Sample Size	Power
0.1	9	0.8508

The following display shows several sample size and power calculations based on the rubberized asphalt problem in Example 3-3.

Power and Sample Size			
1-Sample t Test			
Testing mean = null (versus not = null)			
Calculating power for mean = null + difference			
Alpha = 0.05 Sigma = 117.61			
Difference	Sample Size	Power	
50	15	0.3354	
1-Sample t Test			
Testing mean = null (versus not = null)			
Calculating power for mean = null + difference			
Alpha = 0.05 Sigma = 117.61			
Difference	Sample Size	Target Power	Actual Power
50	46	0.8000	0.8055
1-Sample t Test			
Testing mean = null (versus not = null)			
Calculating power for mean = null + difference			
Alpha = 0.05 Sigma = 117.61			
Difference	Sample Size	Power	
100	15	0.8644	

In the first portion of the display, Minitab calculates the power of the test in Example 3-3, assuming that the engineer would wish to reject the null hypothesis if the true mean stabilized viscosity differed from 3200 by as much as 50, using $s = 117.61$ as an estimate of the true standard deviation. The power is 0.3354, which is low. The next calculation determines the sample size that would be required to produce a power of 0.8, a much better value. Minitab reports that a considerably larger sample size, $n = 46$, would be required. The final calculation determines the power with $n = 15$ if a larger difference between the true mean stabilized viscosity and the hypothesized value is of interest. For a difference of 100, Minitab reports the power to be 0.8644.

3-4 STATISTICAL INFERENCE FOR TWO SAMPLES

3-4.1 Inference for a Difference in Means, Variances Known

Assumptions

1. $x_{11}, x_{12}, \dots, x_{1n_1}$ is a random sample from population 1.
2. $x_{21}, x_{22}, \dots, x_{2n_2}$ is a random sample from population 2.
3. The two populations represented by x_1 and x_2 are independent.
4. Both populations are normal, or if they are not normal, the conditions of the central limit theorem apply.

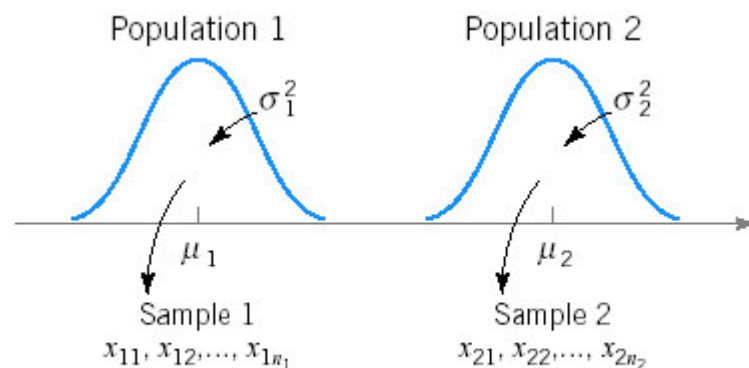


Figure 3-8 Two independent populations.

Based on the assumptions and the preceding results, we may state the following.

The quantity

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (3-47)$$

has an $N(0, 1)$ distribution.

Hypothesis Tests for a Difference in Means, Variances Known

Testing Hypotheses on $\mu_1 - \mu_2$, Variances Known

Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic:
$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (3-48)$$

Alternative Hypotheses

$$H_1: \mu_1 - \mu_2 \neq \Delta_0$$

$$H_1: \mu_1 - \mu_2 > \Delta_0$$

$$H_1: \mu_1 - \mu_2 < \Delta_0$$

Rejection Criterion

$$Z_0 > Z_{\alpha/2} \text{ or } Z_0 < -Z_{\alpha/2}$$

$$Z_0 > Z_{\alpha}$$

$$Z_0 < -Z_{\alpha}$$

..... **EXAMPLE 3-8**

A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\bar{x}_1 = 121$ min and $\bar{x}_2 = 112$ min, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

The hypotheses of interest here are

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

or equivalently,

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Now since $\bar{x}_1 = 121$ min and $\bar{x}_2 = 112$ min, the test statistic is

$$Z_0 = \frac{121 - 112}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$

Because the test statistic $Z_0 = 2.52 > Z_{0.05} = 1.645$, we reject $H_0: \mu_1 = \mu_2$ at the $\alpha = 0.05$ level and conclude that adding the new ingredient to the paint significantly reduces the drying time. Alternatively, we can find the P -value for this test as

$$P\text{-value} = 1 - \Phi(2.52) = 0.0059$$

Therefore, $H_0: \mu_1 = \mu_2$ would be rejected at any significance level $\alpha \geq 0.0059$.



Confidence Interval on a Difference in Means, Variances Known

$$\bar{x}_1 - \bar{x}_2 - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (3-49)$$

3-4.2 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Hypotheses Tests for the Difference in Means

Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\begin{aligned}H_0: \mu_1 - \mu_2 &= \Delta_0 \\H_1: \mu_1 - \mu_2 &\neq \Delta_0\end{aligned}\tag{3-50}$$

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

The **pooled estimator** of σ^2 , denoted by s_p^2 , is defined by

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\tag{3-51}$$

The Two-Sample Pooled t -Test¹

Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic:
$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3-53)$$

Alternative Hypotheses

$$H_1: \mu_1 - \mu_2 \neq \Delta_0$$

$$H_1: \mu_1 - \mu_2 > \Delta_0$$

$$H_1: \mu_1 - \mu_2 < \Delta_0$$

Rejection Criterion

$$t_0 > t_{\alpha/2, n_1 + n_2 - 2} \text{ or } t_0 < -t_{\alpha/2, n_1 + n_2 - 2}$$

$$t_0 > t_{\alpha, n_1 + n_2 - 2}$$

$$t_0 < -t_{\alpha, n_1 + n_2 - 2}$$

..... **EXAMPLE 3-9**

Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently in use, but catalyst 2 is acceptable. Since catalyst 2 is cheaper, it should be adopted, providing it does not change the process yield. An experiment is run in the pilot plant and results in the data shown in Table 3-2. Is there any difference between the mean yields? Use $\alpha = 0.05$ and assume equal variances.

Table 3-2 Catalyst Yield Data, Example 3-9

Observation Number	Catalyst 1	Catalyst 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75
$\bar{x}_1 = 92.255$		$\bar{x}_2 = 92.733$
$s_1 = 2.39$		$s_2 = 2.98$

The hypotheses are

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

From Table 3-2 we have $\bar{x}_1 = 92.255$, $s_1 = 2.39$, $n_1 = 8$, $\bar{x}_2 = 92.733$, $s_2 = 2.98$, and $n_2 = 8$. Therefore,

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(7)(2.39)^2 + (7)(2.98)^2}{8 + 8 - 2} = 7.30$$

$$s_p = \sqrt{7.30} = 2.70$$

and

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{2.70 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{92.255 - 92.733}{2.70 \sqrt{\frac{1}{8} + \frac{1}{8}}} = -0.35$$

Because $t_{0.025,14} = -2.145 < -0.35 < 2.145$, the null hypothesis cannot be rejected. That is, at the 0.05 level of significance, we do not have strong evidence to conclude that catalyst 2 results in a mean yield that differs from the mean yield when catalyst 1 is used.

Figure 3-9 shows comparative box plots for the yield data for the two types of catalysts. These comparative box plots indicate that there is no obvious difference in the median of the two samples, although the second sample has a slightly larger sample dispersion or variance. There are no exact rules for comparing two samples with box plots; their primary value is in the visual impression they provide as a tool for explaining the results of a hypothesis test, as well as in verification of assumptions.

Figure 3-10 presents a Minitab normal probability plot of the two samples of yield data. Note that both samples plot approximately along straight lines, and the straight lines for each sample have similar slopes. (Recall that the slope of the line is proportional to the standard deviation.) Therefore, we conclude that the normality and equal variances assumptions are reasonable.

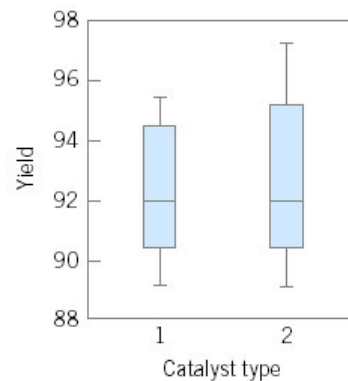


Figure 3-9 Comparative box plots for the catalyst yield data.

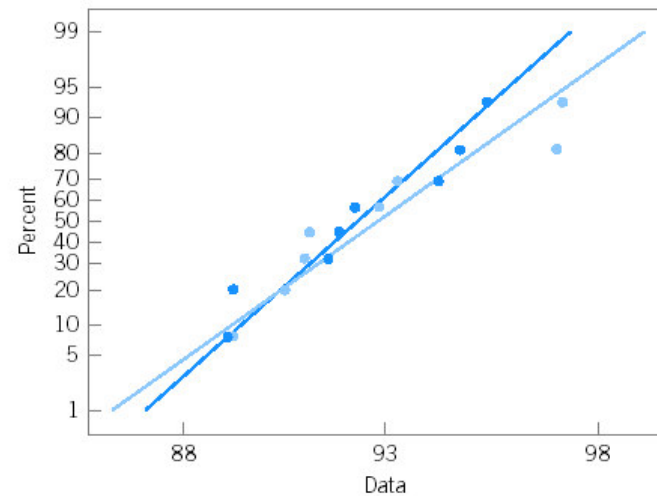


Figure 3-10 Minitab normal probability plot of the catalyst yield-data.

3-4.2 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Hypotheses Tests for the Difference in Means

Case 2: $\sigma_1^2 \neq \sigma_2^2$

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (3-54)$$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{(s_1^2/n_1)^2}{n_1 + 1} + \frac{(s_2^2/n_2)^2}{n_2 + 1}} - 2 \quad (3-55)$$

Confidence Interval on the Difference on Means, Variances Unknown

Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

If $\bar{x}_1$, $\bar{x}_2$, s_1^2 , and s_2^2 are the means and variances of two random samples of sizes n_1 and n_2 , respectively, from two independent normal populations with unknown but equal variances, then a $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\begin{aligned} \bar{x}_1 - \bar{x}_2 - t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{aligned} \quad (3-56)$$

where $s_p = \sqrt{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)}$ is the pooled estimate of the common population standard deviation, and $t_{\alpha/2, n_1+n_2-2}$ is the upper $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Confidence Interval on the Difference on Means, Variances Unknown

Case 2: $\sigma_1^2 \neq \sigma_2^2$

If $\bar{x}_1$, $\bar{x}_2$, s_1^2 , and s_2^2 are the means and variances of two random samples of sizes n_1 and n_2 , respectively, from two independent normal populations with unknown and unequal variances, then an approximate $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3-57)$$

where v is given by equation 3-55 and $t_{\alpha/2, v}$ is the upper $\alpha/2$ percentage point of the t distribution with v degrees of freedom.

Computer Solution

Two-sample statistical tests can be performed using most statistics software packages. The following display presents the output from the Minitab two-sample t -test routine for the catalyst yield data in Example 3-9.

Two-Sample t -test and CI: Catalyst 1, Catalyst 2				
Two-sample T for Catalyst 1 vs Catalyst 2				
	N	Mean	StDev	SE Mean
Catalyst 1	8	92.26	2.39	0.84
Catalyst 2	8	92.73	2.98	1.1
Difference = μ Catalyst 1 - μ Catalyst 2				
Estimate for difference: -0.48				
95% CI for difference: (-3.39, 2.44)				
t -test of difference = 0 (vs not =): T-value = -0.35				
P-Value = 0.729 DF = 14				

Minitab will also perform power and sample size calculations for the two-sample pooled t -test. The following display from Minitab illustrates some calculations for the catalyst yield problem in Example 3-9.

In the first part of the display, Minitab calculates the power of the test in Example 3-9 assuming that we want to reject the null hypothesis if the true mean difference in yields for the two catalysts were as large as 2, using the pooled estimate of the standard deviation $s_p = 2.70$. For the sample size of $n_1 = n_2 = 8$ for each catalyst, the power is reported as 0.2816, which is quite low. The next calculation determines the sample size that would be required to produce a power of 0.75, a much better value. Minitab reports that a considerably larger sample size for each catalyst type, $n_1 = n_2 = 27$, would be required.

Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)

Calculating power for mean 1 = mean 2 + difference

Alpha = 0.05 Sigma = 2.7

Difference	Sample Size	Power
2	8	0.2816

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)

Calculating power for mean 1 = mean 2 + difference

Alpha = 0.05 Sigma = 2.7

Difference	Sample Size	Target Power	Actual Power
2	27	0.7500	0.7615

Paired Data

It should be emphasized that we have assumed that the two samples used in the above tests are independent. In some applications, *paired* data are encountered. Observations in an experiment are often paired to prevent extraneous factors from inflating the estimate of the variance; hence, this method can be used to improve the precision of comparisons between means. For a further discussion of paired data, see Montgomery and Runger (2003). The analysis of such a situation is illustrated in the following example.

..... **EXAMPLE 3-11**

Two different types of machines are used to measure the tensile strength of synthetic fiber. We wish to determine whether or not the two machines yield the same average tensile strength values. Eight specimens of fiber are randomly selected, and one strength measurement is made using each machine on each specimen. The coded data are shown in Table 3-3.

Table 3-3 Paired Tensile Strength Data for Example 3-11

Specimen	Machine 1	Machine 2	Difference
1	74	78	-4
2	76	79	-3
3	74	75	-1
4	69	66	3
5	58	63	-5
6	71	70	1
7	66	66	0
8	65	67	-2

The data in this experiment have been paired to prevent the difference between fiber specimens (which could be substantial) from affecting the test on the difference between machines. The test procedure consists of obtaining the differences of the pair of observations on each of the n specimens—say, $d_j = x_{1j} - x_{2j}$, $j = 1, 2, \dots, n$ —and then testing the hypothesis that the mean of the difference μ_d is zero. Note that testing $H_0: \mu_d = 0$ is equivalent to testing $H_0: \mu_1 = \mu_2$; furthermore, the test on μ_d is simply the one-sample t -test discussed in Section 3-3.3. The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

where

$$\bar{d} = \frac{1}{n} \sum_{j=1}^n d_j$$

and

$$s_d^2 = \frac{\sum_{j=1}^n (d_j - \bar{d})^2}{n-1} = \frac{\sum_{j=1}^n d_j^2 - \frac{\left(\sum_{j=1}^n d_j\right)^2}{n}}{n-1}$$

and $H_0: \mu_d = 0$ is rejected if $|t_0| > t_{\alpha/2, n-1}$.

In our example we find that

$$\bar{d} = \frac{1}{n} \sum_{j=1}^n d_j = \frac{1}{8}(-11) = -1.38$$

$$s_d^2 = \frac{\sum_{j=1}^n d_j^2 - \frac{\left(\sum_{j=1}^n d_j\right)^2}{n}}{n-1} = \frac{65 - \frac{(-11)^2}{8}}{7} = 7.13$$

Therefore, the test statistic is

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-1.38}{2.67/\sqrt{8}} = -1.46$$

Choosing $\alpha = 0.05$ results in $t_{0.025,7} = 2.365$, and we conclude that there is no strong evidence to indicate that the two machines differ in their mean tensile strength measurements (the P -value is $P = 0.18$).

3-4.3 Inference on the Variances of Two Normal Distributions

- Section 3-4.3 describes hypothesis testing and confidence intervals on the variance of two normal distributions

3-4.4 Inference on Two Population Proportions

Large-Sample Test for $H_0: p_1 = p_2$

Suppose that the two independent random samples of sizes n_1 and n_2 are taken from two populations, and let x_1 and x_2 represent the number of observations that belong to the class of interest in samples 1 and 2, respectively. Furthermore, suppose that the normal approximation to the binomial is applied to each population, so that the estimators of the population proportions $\hat{p}_1 = x_1/n_1$ and $\hat{p}_2 = x_2/n_2$ and have approximate normal distributions. We are interested in testing the hypotheses

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2$$

The statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \quad (3-62)$$

is distributed approximately as standard normal and is the basis of a test for $H_0: p_1 = p_2$.

Null hypothesis: $H_0: p_1 = p_2$

Test statistic:
$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (3-63)$$

<u>Alternative Hypotheses</u>	<u>Rejection Criterion</u>
$H_1: p_1 \neq p_2$	$Z_0 > Z_{\alpha/2}$ or $Z_0 < -Z_{\alpha/2}$
$H_1: p_1 > p_2$	$Z_0 > Z_{\alpha}$
$H_1: p_1 < p_2$	$Z_0 < -Z_{\alpha}$

3-4.4 Inference on Two Population Proportions

Confidence Interval on the Difference in Two Population Proportions

If there are two population proportions of interest—say, p_1 and p_2 —it is possible to construct a $100(1 - \alpha)\%$ confidence interval on their difference. The confidence interval is as follows.

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} &\leq p_1 - p_2 \\ &\leq \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \end{aligned} \quad (3-64)$$

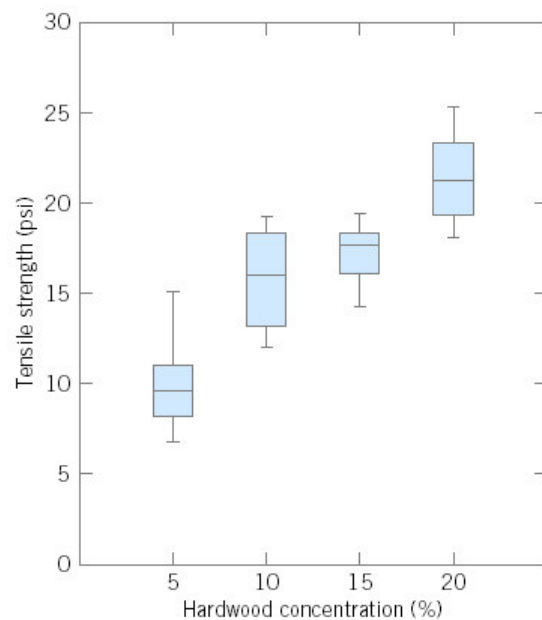
This result is based on the normal approximation to the binomial distribution.

3-5 WHAT IF THERE ARE MORE THAN TWO POPULATIONS? THE ANALYSIS OF VARIANCE

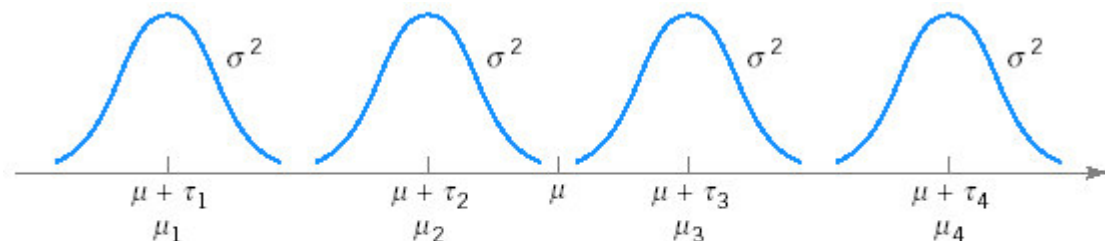
3-5.1 An Example

Table 3-4 Tensile Strength of Paper (psi)

Hardwood Concentration (%)	Observations						Totals	Averages
	1	2	3	4	5	6		
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96



(a)



(b)

Figure 3-11 (a) Box plots of hardwood concentration data. (b) Display of the model in equation 3-65 for the completely randomized single-factor experiment.

3-5.2 The Analysis of Variance

Table 3-5 Typical Data for a Single-Factor Experiment

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\dots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\dots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
.	.	.	$\dots$	.	.	.
.	.	.	$\dots$	.	.	.
.	.	.	$\dots$	.	.	.
a	y_{a1}	y_{a2}	$\dots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{..}$	$\bar{y}_{..}$

We may describe the observations in Table 3-5 by the **linear statistical model**

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (3-65)$$

where y_{ij} is a random variable denoting the (ij) th observation, μ is a parameter common to all treatments called the **overall mean**, τ_i is a parameter associated with the i th treatment called the i th **treatment effect**, and ε_{ij} is a random error component.

Let $y_{i.}$ represent the total of the observations under the i th treatment and $\bar{y}_{i.}$ represent the average of the observations under the i th treatment. Similarly, let $y_{..}$ represent the grand total of all observations and $\bar{y}_{..}$ represent the grand mean of all observations. Expressed mathematically,

$$\begin{aligned} y_{i.} &= \sum_{j=1}^n y_{ij} & \bar{y}_{i.} &= y_{i.}/n \quad i = 1, 2, \dots, a \\ y_{..} &= \sum_{i=1}^a \sum_{j=1}^n y_{ij} & \bar{y}_{..} &= y_{..}/N \end{aligned} \quad (3-67)$$

where $N = an$ is the total number of observations. Thus, the “dot” subscript notation implies summation over the subscript that it replaces.

We are interested in testing the equality of the a treatment means $\mu_1, \mu_2, \dots, \mu_a$. Using equation 3-66, we find that this is equivalent to testing the hypotheses

$$\begin{aligned} H_0: \quad & \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1: \quad & \tau_i \neq 0 \text{ for at least one } i \end{aligned} \quad (3-68)$$

Thus, if the null hypothesis is true, each observation consists of the overall mean μ plus a realization of the random error component ε_{ij} . This is equivalent to saying that all N observations are taken from a normal distribution with mean μ and variance σ^2 . Therefore, if the null hypothesis is true, changing the levels of the factor has no effect on the mean response.

The total variability in the data is described by the **total sum of squares**

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

The basic ANOVA partition of the total sum of squares is given in the following definition.

The **sum of squares identity** is

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad (3-69)$$

Therefore, we write equation 3-69 symbolically as

$$SS_T = SS_{\text{Treatments}} + SS_E \quad (3-71)$$

The expected value of the treatment sum of squares is

$$E(SS_{\text{Treatments}}) = (a - 1)\sigma^2 + n \sum_{i=1}^a \tau_i^2$$

Now if the null hypothesis in equation 3-68 is true, each τ_i is equal to zero and

$$E\left(\frac{SS_{\text{Treatments}}}{a - 1}\right) = \sigma^2$$

If the alternative hypothesis is true, then

$$E\left(\frac{SS_{\text{Treatments}}}{a - 1}\right) = \sigma^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a - 1}$$

The ratio $MS_{\text{Treatments}} = SS_{\text{Treatments}}/(a - 1)$ is called the **mean square for treatments**. Thus, if H_0 is true, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2 , whereas if H_1 is true, $MS_{\text{Treatments}}$ estimates σ^2 plus a positive term that incorporates variation due to the systematic difference in treatment means. (Refer to the supplemental material for this chapter for the proofs of these two statements.)

We can also show that the expected value of the error sum of squares is $E(SS_E) = a(n - 1)\sigma^2$. Therefore, the **error mean square** $MS_E = SS_E/[a(n - 1)]$ is an unbiased estimator of σ^2 regardless of whether or not H_0 is true.

The **error mean square**

$$MS_E = \frac{SS_E}{a(n - 1)}$$

is an unbiased estimator of σ^2 .

There is also a partition of the number of degrees of freedom that corresponds to the sum of squares identity in equation 3-69. That is, there are $an = N$ observations; thus, SS_T has $an - 1$ degrees of freedom. There are a levels of the factor, so $SS_{\text{Treatments}}$ has $a - 1$ degrees of freedom. Finally, within any treatment there are n replicates providing $n - 1$ degrees of freedom with which to estimate the experimental error. Since there are a treatments, we have $a(n - 1)$ degrees of freedom for error. Therefore, the degrees of freedom partition is

$$an - 1 = a - 1 + a(n - 1)$$

Now assume that each of the a populations can be modeled as a normal distribution. Using this assumption we can show that if the null hypothesis H_0 is true, the ratio

$$F_0 = \frac{SS_{\text{Treatments}} / (a - 1)}{SS_E / [a(n - 1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \quad (3-72)$$

has an F distribution with $a - 1$ and $a(n - 1)$ degrees of freedom.

Efficient computational formulas for the sums of squares may be obtained by expanding and simplifying the definitions of $SS_{\text{Treatments}}$ and SS_T . This yields the following results.

Definition

The sums of squares computing formulas for the analysis of variance with equal sample sizes in each treatment are

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N} \quad (3-73)$$

and

$$SS_{\text{Treatments}} = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \quad (3-74)$$

The error sum of squares is obtained by subtraction as

$$SS_E = SS_T - SS_{\text{Treatments}} \quad (3-75)$$

The computations for this test procedure are usually summarized in tabular form as shown in Table 3-6. This is called an **analysis of variance (or ANOVA) table**.

Table 3-6 The Analysis of Variance for a Single-Factor Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a - 1$	$MS_{\text{Treatments}}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Error	SS_E	$a(n - 1)$	MS_E	
Total	SS_T	$an - 1$		

..... **EXAMPLE 3-12**

Consider the paper tensile strength experiment described in Section 3-5.1. We can use the analysis of variance to test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper.

The hypotheses are

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_1: \tau_i \neq 0 \text{ for at least one } i$$

Table 3-7 Minitab Analysis of Variance Output for the Paper Tensile Strength Experiment

One-Way Analysis of Variance					
Analysis of Variance					
Source	DF	SS	MS	F	P
Factor	3	382.79	127.60	19.61	0.000
Error	20	130.17	6.51		
Total	23	512.96			
				Individual 95% CIs For Mean Based on Pooled StDev	
Level	N	Mean	StDev	---+-----+-----+-----+-	
5	6	10.000	2.828	(--*--)	
10	6	15.667	2.805	(--*--)	
15	6	17.000	1.789	(--*--)	
20	6	21.167	2.639	(--*--)	
Pooled StDev = 2.551				---+-----+-----+-----+-	
				10.0 15.0 20.0 25.0	

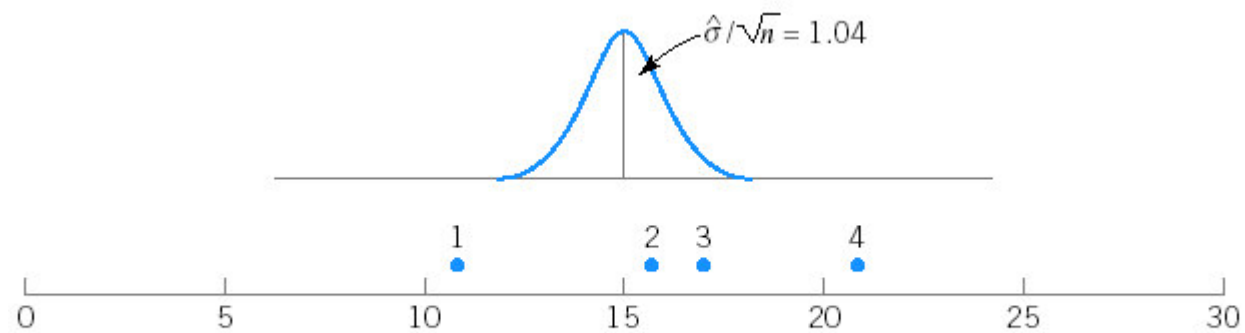


Figure 3-12 Tensile strength averages from the hardwood concentration experiment in relation to a normal distribution with standard deviation $\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$

3-5.3 Checking Assumptions: Residual Analysis

Therefore, the residual is $e_{ij} = y_{ij} - \bar{y}_{i.}$; that is, the difference between an observation and the corresponding factor-level mean. The residuals for the hardwood percentage experiment are shown in Table 3-8.

Table 3-8 Residuals for the Hardwood Experiment

Hardwood Concentration	Residuals					
5%	-3.00	-2.00	5.00	1.00	-1.00	0.00
10%	-3.67	1.33	-2.67	2.33	+3.33	-0.67
15%	-3.00	1.00	2.00	0.00	-1.00	1.00
20%	-2.17	3.83	0.83	1.83	-3.17	-1.17

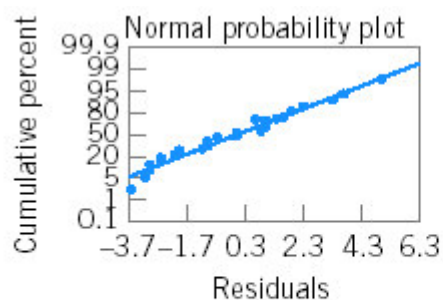


Figure 3-13 Normal probability plot of residuals from the hardwood concentration experiment.

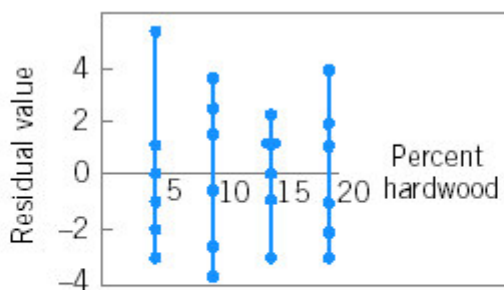


Figure 3-14 Plot of residuals versus factor levels.

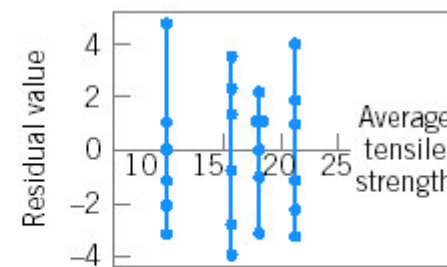


Figure 3-15 Plot of residuals versus $\bar{y}_{i.}$

IMPORTANT TERMS AND CONCEPTS

Alternative hypothesis	Parameters of a distribution
Analysis of variance (ANOVA)	Point estimator
Binomial distribution	Poisson distribution
Checking assumptions for statistical inference procedures	Power of a statistical test
Chi-square distribution	Random sample
Confidence interval	Sampling distribution
Confidence intervals on means, known variance(s)	Statistic
Confidence intervals on means, unknown variance(s)	t -distribution
Confidence intervals on proportions	Test statistic
Confidence intervals on the variance of a - normal distribution	Tests of hypotheses on means, known variance(s)
Confidence intervals on the variances of two normal distributions	Tests of hypotheses on means, unknown variance(s)
Critical region for a test statistic	Tests of hypotheses on proportions
F -distribution	Tests of hypotheses on the variance of a normal distribution
Hypothesis testing	Tests of hypotheses on the variances of two normal distributions
Minimum variance estimator	Type I error
Null hypothesis	Type II error
	Unbiased estimator

LEARNING OBJECTIVES

1. Explain the concept of random sampling
2. Explain the concept of a sampling distribution
3. Explain the general concept of estimating the parameters of a population or probability distribution
4. Know how to explain the precision with which a parameter is estimated
5. Construct and interpret confidence intervals on a single mean and on the difference in two means
6. Construct and interpret confidence intervals on a single variance or the ratio of two variances
7. Construct and interpret confidence intervals on a single proportion and on the difference in two proportions
8. Test hypotheses on a single mean and on the difference in two means
9. Test hypotheses on a single variance and on the ratio of two variances
10. Test hypotheses on a single proportion and on the difference in two proportions
11. Use the P -value approach for hypothesis testing
12. Understand how the analysis of variance (ANOVA) is used to test hypotheses about the equality of more than two means