1

The Role of Statistics in Engineering

4-6 Normal Distribution

5-5 Linear Combinations of Random Variables

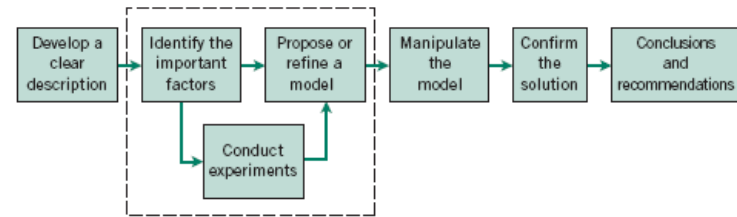## 1-1 The Engineering Method and Statistical Thinking

Figure 1.1 The engineering method

1-1 The Engineering Method and Statistical Thinking

The field of statistics deals with the collection, presentation, analysis, and use of data to

- Make decisions
- Solve problems
- Design products and processes

1-1 The Engineering Method and Statistical Thinking

- Statistical techniques are useful for describing and understanding variability.

- By variability, we mean successive observations of a system or phenomenon do *not* produce exactly the same result.

- Statistics gives us a framework for describing this variability and for learning about potential sources of variability.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

An engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3/32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine. Eight prototype units are produced and their pull-off forces measured (in pounds): 12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

•The **dot diagram** is a very useful plot for displaying a small body of data - say up to about 20 observations.
• This plot allows us to see easily two features of the data; the **location,** or the middle, and the **scatter** or **variability.**
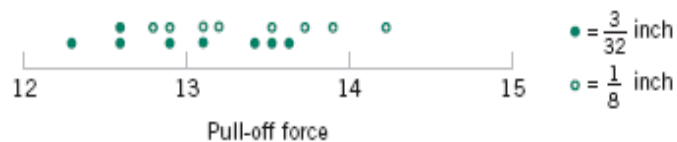

Figure 1-2 Dot diagram of the pull-off force data when wall thickness is 3/32 inch.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

• The engineer considers an alternate design and eight prototypes are built and pull-off force measured.
• The dot diagram can be used to compare two sets of data

Figure 1-3 Dot diagram of pull-off force for two wall thicknesses.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

• Since pull-off force varies or exhibits variability, it is a random variable.

• A random variable, X, can be model by

$$X = \mu + \varepsilon$$

where μ is a constant and ε a random disturbance.
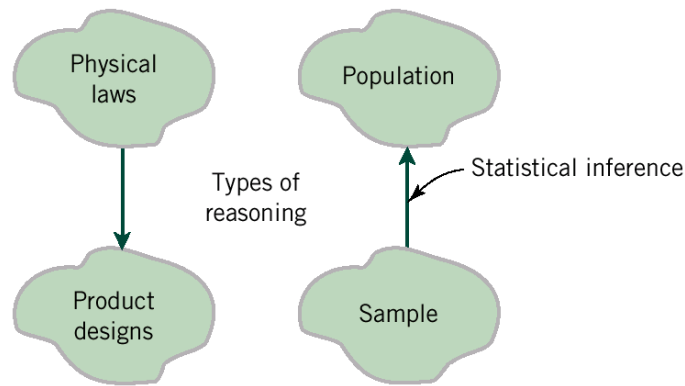
1-1 The Engineering Method and Statistical Thinking

Figure 1-4   Statistical inference is one type of reasoning.

1-2 Collecting Engineering Data

Three basic methods for collecting data:

- A **retrospective** study using historical data
  - May not be useful
- An **observational** study
  - Cannot tell the cause-effect
- A **designed experiment**
  - Make deliberate changes to observe response
  - Can tell the cause-effect

1-3 Mechanistic and Empirical Models

A **mechanistic model** is built from our underlying knowledge of the basic physical mechanism that relates several variables.

Ohm's Law: Current = voltage/resistance

$$I = E/R \quad or \quad I = E/R + \varepsilon$$

An **empirical model** is built from our engineering and scientific knowledge of the phenomenon, but is not directly developed from our theoretical or first-principles understanding of the underlying mechanism.

Table 1-2   Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400
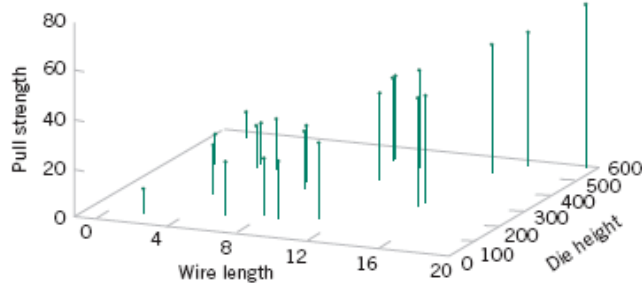
Figure 1-15 Three-dimensional plot of the wire and pull strength data.
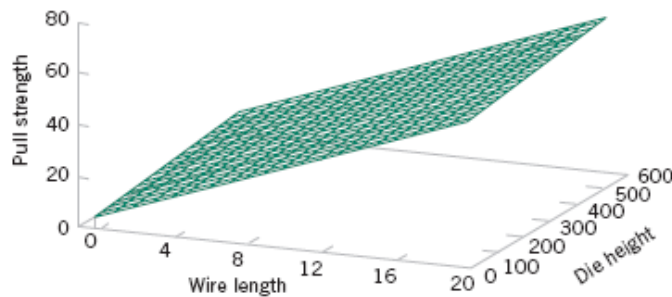
Figure 1-16 Plot of the predicted values of pull strength from the empirical model.

1-3 Mechanistic and Empirical Models

$$\text{Pull strength} = \beta_0 + \beta_1(\text{wire length}) + \beta_2(\text{die height}) + \epsilon$$

In general, this type of empirical model is called a **regression model**.

The **estimated** regression line is given by

$$\widehat{\text{Pull strength}} = 2.26 + 2.74(\text{wire length}) + 0.0125(\text{die height})$$

4-6 Normal Distribution

Definition

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \qquad (4\text{-}8)$$
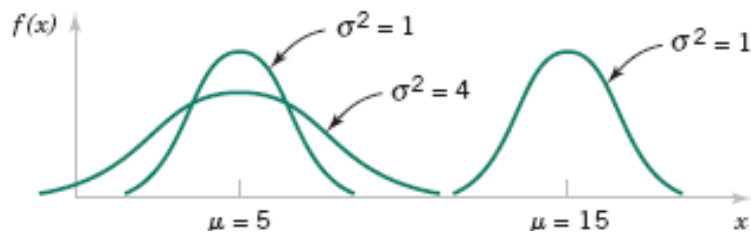
is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \qquad (4\text{-}9)$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of X are shown to equal μ and σ^2, respectively, at the end of this Section 5-6.
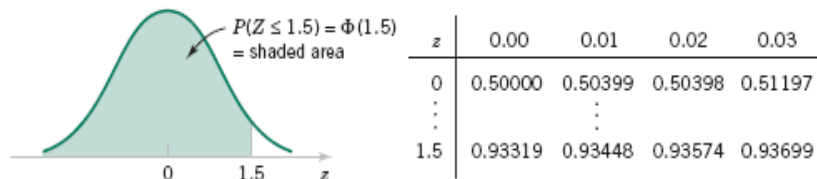
4-6 Normal Distribution

Figure 4-10 Normal probability density functions for selected values of the parameters μ and σ^2.

Definition : Standard Normal

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z.
    The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = P(Z \le z)$$

Example 4-11

Assume Z is a standard normal random variable. Appendix Table II provides probabilities of the form $P(Z \le z)$. The use of Table II to find $P(Z \le 1.5)$ is illustrated in Fig. 4-13. Read down the z column to the row that equals 1.5. The probability is read from the adjacent column, labeled 0.00, to be 0.93319.

    The column headings refer to the hundredth's digit of the value of z in $P(Z \le z)$. For example, $P(Z \le 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.

z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
$\vdots$		$\vdots$		
1.5	0.93319	0.93448	0.93574	0.93699

$P(Z \le 1.5) = \Phi(1.5)$ = shaded area

Figure 4-13 Standard normal probability density function.

Standardizing

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4-10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Example 4-13

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement will exceed 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Fig. 4-15. We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table II,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$$

Rather than using Fig. 4-15, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{(X - 10)}{2} > \frac{(13 - 10)}{2}\right) = P(Z > 1.5) = 0.06681$$
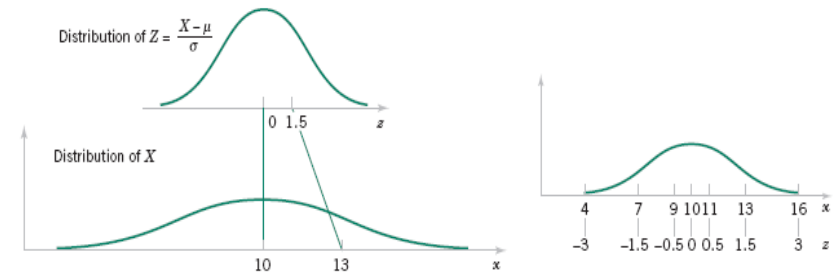
Figure 4-15 Standardizing a normal random variable.

To Calculate Probability

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z) \qquad (4\text{-}11)$$

where Z is a **standard normal random variable**, and $z = \dfrac{(x - \mu)}{\sigma}$ is the z-value obtained by **standardizing** X.
The probability is obtained by entering Appendix Table II with $z = (x - \mu)/\sigma$.

Example 4-14 (continued)

Determine the value for which the probability that a current measurement is below this value is 0.98. The requested value is shown graphically in Fig. 4-16. We need the value of x such that $P(X < x) = 0.98$. By standardizing, this probability expression can be written as

$$\begin{aligned} P(X < x) &= P((X - 10)/2 < (x - 10)/2) \\ &= P(Z < (x - 10)/2) \\ &= 0.98 \end{aligned}$$

Appendix Table II is used to find the z-value such that $P(Z < z) = 0.98$. The nearest probability from Table II results in
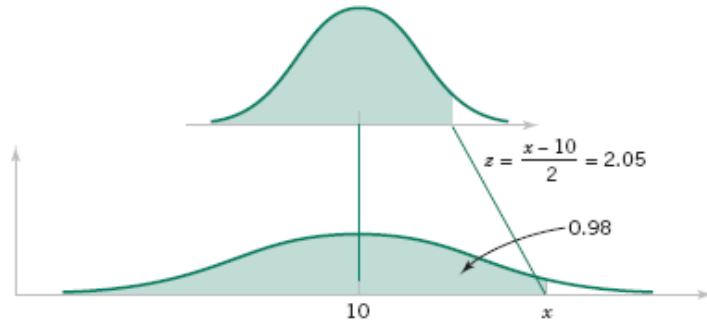
$$P(Z < 2.05) = 0.97982$$

Therefore, $(x - 10)/2 = 2.05$, and the standardizing transformation is used in reverse to solve for x. The result is

$$x = 2(2.05) + 10 = 14.1 \text{ milliamperes}$$

4-6 Normal Distribution

$$z = \frac{x - 10}{2} = 2.05$$

0.98

Figure 4-16 Determining the value of x to meet a specified probability.

5-5 Linear Combinations of Random Variables

Definition

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_1, c_2, \ldots, c_p$,

$$Y = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p \qquad (5\text{-}34)$$

is a **linear combination** of $X_1, X_2, \ldots, X_p$.

Mean of a Linear Combination

If $Y = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p$,

$$E(Y) = c_1 E(X_1) + c_2 E(X_2) + \cdots + c_p E(X_p) \qquad (5\text{-}35)$$

5-5 Linear Combinations of Random Variables

Variance of a Linear Combination

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p$, then in general

$$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \cdots + c_p^2 V(X_p) + 2 \sum_{i<j} \sum c_i c_j \, \text{cov}(X_i, X_j) \qquad (5\text{-}36)$$

If $X_1, X_2, \ldots, X_p$ are **independent**,

$$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \cdots + c_p^2 V(X_p) \qquad (5\text{-}37)$$

5-5 Linear Combinations of Random Variables

Example 5-33

An important use of equation 5-37 is in **error propagation** that is presented in the following example.

A semiconductor product consists of three layers. If the variances in thickness of the first, second, and third layers are 25, 40, and 30 nanometers squared, what is the variance of the thickness of the final product.

Let X_1, X_2, X_3, and X be random variables that denote the thickness of the respective layers, and the final product. Then

$$X = X_1 + X_2 + X_3$$

The variance of X is obtained from equaion 5-39

$$V(X) = V(X_1) + V(X_2) + V(X_3) = 25 + 40 + 30 = 95 \text{ nm}^2$$

Consequently, the standard deviation of thickness of the final product is $95^{1/2} = 9.75$ nm and this shows how the variation in each layer is propagated to the final product.

5-5 Linear Combinations of Random Variables

Mean and Variance of an Average

If $\bar{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\bar{X}) = \mu \qquad (5\text{-}38a)$$

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\bar{X}) = \frac{\sigma^2}{p} \qquad (5\text{-}38b)$$

5-5 Linear Combinations of Random Variables

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$Y = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p$$

is a normal random variable with

$$E(Y) = c_1 \mu_1 + c_2 \mu_2 + \cdots + c_p \mu_p$$

and

$$V(Y) = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + \cdots + c_p^2 \sigma_p^2 \qquad (5\text{-}39)$$

5-5 Linear Combinations of Random Variables

Example 5-34

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E(X_1) = 2$ centimeters and standard deviation 0.1 centimeter and that X_2 is normal with $E(X_2) = 5$ centimeters and standard deviation 0.2 centimeter. Also, assume that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 centimeters.

Then, $Y = 2X_1 + 2X_2$ is a normal random variable that represents the perimeter of the part. We obtain, $E(Y) = 14$ centimeters and the variance of Y is

$$V(Y) = 4 \times 0.1^2 + 4 \times 0.2^2 = 0.2$$

Now,

$$P(Y > 14.5) = P[(Y - \mu_Y)/\sigma_Y > (14.5 - 14)/\sqrt{0.2}]$$
$$= P(Z > 1.12) = 0.13$$

Some useful results to remember
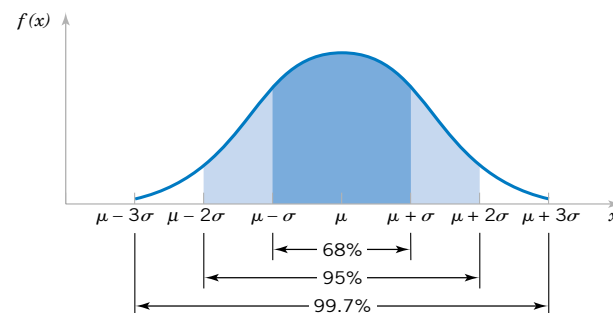
For any normal random variable

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

6
Random Sampling and Data Description

CHAPTER OUTLINE

6-1 Numerical Summaries

Definition: Sample Range

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample range** is

$$r = \max(x_i) - \min(x_i) \qquad (6\text{-}6)$$

Definition: Sample Mean

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (6\text{-}1)$$
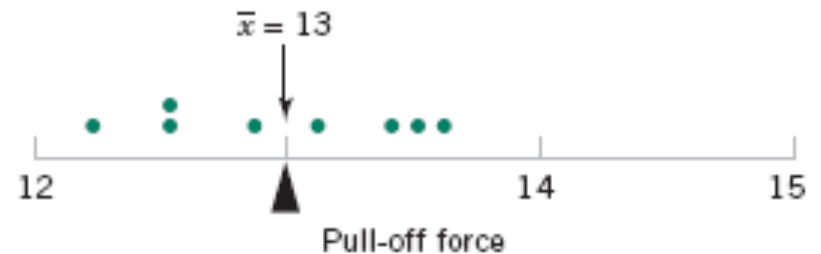
6-1 Numerical Summaries

Example 6-1

Let's consider the eight observations collected from the prototype engine connectors from Chapter 1. The eight observations are $x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. The sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{12.6 + 12.9 + \cdots + 13.1}{8}$$

$$= \frac{104}{8} = 13.0 \text{ pounds}$$

A physical interpretation of the sample mean as a measure of location is shown in the dot diagram of the pull-off force data. See Figure 6-1. Notice that the sample mean $\bar{x} = 13.0$ can be thought of as a "balance point." That is, if each observation represents 1 pound of mass placed at the point on the x-axis, a fulcrum located at $\bar{x}$ would exactly balance this system of weights.

6-1 Numerical Summaries

Figure 6-1 The sample mean as a balance point for a system of weights.

6-1 Numerical Summaries

Population Mean

For a finite population with N (equally likely) measurements, the mean is

$$\mu = \sum_{i=1}^{N} x_i f(x_i) = \frac{\sum_{i=1}^{N} x_i}{N} \qquad (6\text{-}2)$$

The sample mean is a reasonable estimate of the population mean.

6-1 Numerical Summaries

Definition: Sample Variance

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the **sample variance** is

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} \qquad (6\text{-}3)$$

The **sample standard deviation**, s, is the positive square root of the sample variance.

- n-1 is referred to as the **degrees of freedom**.

6-1 Numerical Summaries

How Does the Sample Variance Measure Variability?
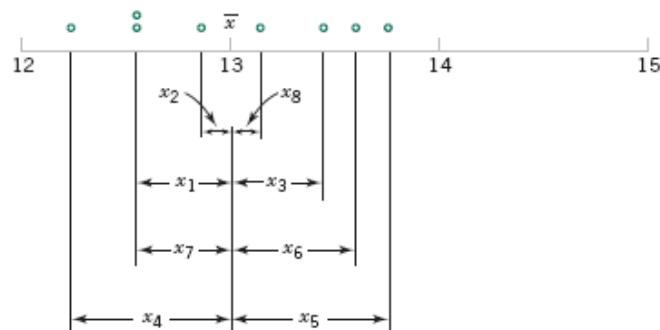
Figure 6-2 How the sample variance measures variability through the deviations $x_i - \bar{x}$.

6-1 Numerical Summaries

Example 6-2

Table 6-1 displays the quantities needed for calculating the sample variance and sample standard deviation for the pull-off force data. These data are plotted in Fig. 6-2. The numerator of s^2 is

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1.60$$

so the sample variance is

$$s^2 = \frac{1.60}{8 - 1} = \frac{1.60}{7} = 0.2286 \ (\text{pounds})^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.48 \ \text{pounds}$$

6-1 Numerical Summaries

Computation of s²

$$s^2 = \frac{\sum\limits_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}}{n - 1} \qquad (6\text{-}4)$$

6-1 Numerical Summaries

Population Variance

When the population is finite and consists of N (equally likely) values, we may define the population variance as

$$\sigma^2 = \frac{\sum\limits_{i=1}^{N} (x_i - \mu)^2}{N} \qquad (6\text{-}5)$$

The sample variance is a reasonable estimate of the population variance.

6-2 Stem-and-Leaf Diagrams

A **stem-and-leaf diagram** is a good way to obtain an informative visual display of a data set $x_1, x_2, \ldots, x_n$, where each number x_i consists of at least two digits. To construct a stem-and-leaf diagram, use the following steps.

Steps for Constructing a Stem-and-Leaf Diagram

(1) Divide each number x_i into two parts: a **stem,** consisting of one or more of the leading digits and a **leaf,** consisting of the remaining digit.

(2) List the stem values in a vertical column.

(3) Record the leaf for each observation beside its stem.

(4) Write the units for stems and leaves on the display.

6-2 Stem-and-Leaf Diagrams

Example 6-4

To illustrate the construction of a stem-and-leaf diagram, consider the alloy compressive strength data in Table 6-2. We will select as stem values the numbers $7, 8, 9, \ldots, 24$. The resulting stem-and-leaf diagram is presented in Fig. 6-4. The last column in the diagram is a frequency count of the number of leaves associated with each stem. Inspection of this display immediately reveals that most of the compressive strengths lie between 110 and 200 psi and that a central value is somewhere between 150 and 160 psi. Furthermore, the strengths are distributed approximately symmetrically about the central value. The stem-and-leaf diagram enables us to determine quickly some important features of the data that were not immediately obvious in the original display in Table 6-2.

6-2 Stem-and-Leaf Diagrams

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

6-2 Stem-and-Leaf Diagrams

Figure 6-4 Stem-and-leaf diagram for the compressive strength data in Table 6-2.

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem : Tens and hundreds digits (psi); Leaf: Ones digits (psi)

6-2 Stem-and-Leaf Diagrams

Figure 6-6 Stem-and-leaf diagram from Minitab.

```
Character Stem-and-Leaf Display

Stem-and-leaf of Strength
N = 80    Leaf Unit = 1.0
  1       7     6
  2       8     7
  3       9     7
  5      10     1 5
  8      11     0 5 8
 11      12     0 1 3
 17      13     1 3 3 4 5 5
 25      14     1 2 3 5 6 8 9 9
 37      15     0 0 1 3 4 4 6 7 8 8 8 8
(10)     16     0 0 0 3 3 5 7 7 8 9
 33      17     0 1 1 2 4 4 5 6 6 8
 23      18     0 0 1 1 3 4 6
 16      19     0 3 4 6 9 9
 10      20     0 1 7 8
  6      21     8
  5      22     1 8 9
  2      23     7
  1      24     5
```

6-2 Stem-and-Leaf Diagrams

Data Features

• The **median** is a measure of central tendency that divides the data into two equal parts, half below the median and half above. If the number of observations is even, the median is halfway between the two central values.

From Fig. 6-6, the 40th and 41st values of strength as 160 and 163, so the median is $(160 + 163)/2 = 161.5$. If the number of observations is odd, the median is the *central* value.

The **range** is a measure of variability that can be easily computed from the ordered stem-and-leaf display. It is the maximum minus the minimum measurement. From Fig.6-6 the range is 245 - 76 = 169.

6-2 Stem-and-Leaf Diagrams

Data Features

•When an ordered set of data is divided into four equal parts, the division points are called **quartiles.**

•The **first** or **lower quartile**, q_1 , is a value that has approximately one-fourth (25%) of the observations below it and approximately 75% of the observations above.

•The **second quartile**, q_2, has approximately one-half (50%) of the observations below its value. The second quartile is *exactly* equal to the **median**.

•The **third** or upper quartile, q_3, has approximately three-fourths (75%) of the observations below its value. As in the case of the median, the quartiles may not be unique.

6-2 Stem-and-Leaf Diagrams

Data Features

• The compressive strength data in Figure 6-6 contains $n =$ 80 observations. Minitab software calculates the first and third quartiles as the $(n + 1)/4$ and $3(n + 1)/4$ ordered observations and interpolates as needed.

For example, $(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$.

Therefore, Minitab interpolates between the 20th and 21st ordered observation to obtain $q_1 = 143.50$ and between the 60th and 61st observation to obtain $q_3 = 181.00$.

6-2 Stem-and-Leaf Diagrams

Data Features

• The **interquartile range** is the difference between the upper and lower quartiles, and it is sometimes used as a measure of variability.

• In general, the $100k$th **percentile** is a data value such that approximately $100k$% of the observations are at or below this value and approximately $100(1 - k)$% of them are above it.

6-4 Box Plots

• The **box plot** is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and identification of observations that lie unusually far from the bulk of the data.

• **Whisker**
• **Outlier**
• **Extreme outlier**
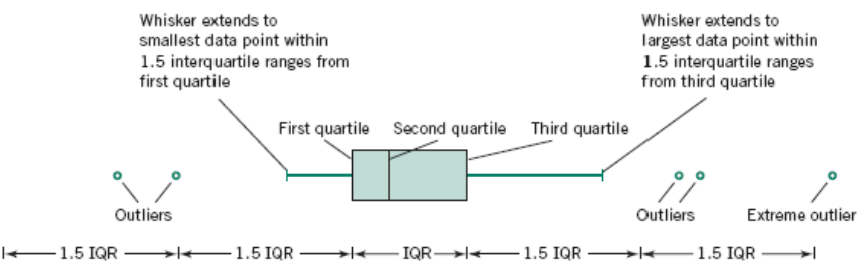
6-4 Box Plots

Figure 6-13 Description of a box plot.

6-4 Box Plots

Example: The ordered data in Example 6-1 are
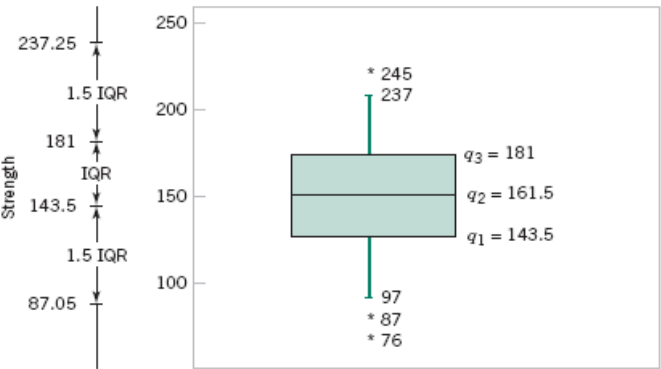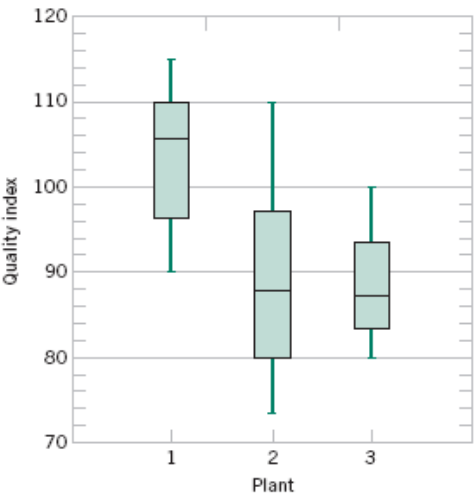12.3, 12.6, 12.6, 12.9, 13.1, 13.4, 13.5, 13.6

6-4 Box Plots

Figure 6-14 Box plot for compressive strength data in Table 6-2.

6-4 Box Plots

**Figure 6-15** Comparative box plots of a quality index at three plants.

6-3 Frequency Distributions and Histograms

- A **frequency distribution** is a more compact summary of data than a stem-and-leaf diagram.

- To construct a frequency distribution, we must divide the range of the data into intervals, which are usually called **class intervals**, **cells**, or **bins**.

Constructing a Histogram (Equal Bin Widths):

(1) Label the bin (class interval) boundaries on a horizontal scale.
(2) Mark and label the vertical scale with the frequencies or the relative frequencies.
(3) Above each bin, draw a rectangle where height is equal to the frequency (or relative frequency) corresponding to that bin.
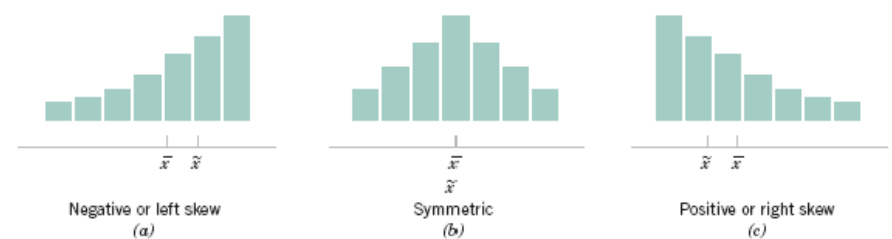
6-3 Frequency Distributions and Histograms
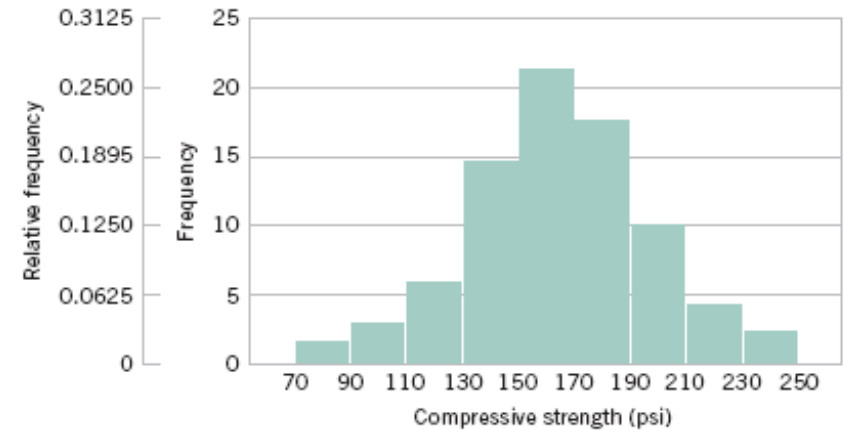
Figure 6-7 Histogram of compressive strength for 80 aluminum-lithium alloy specimens.

6-3 Frequency Distributions and Histograms
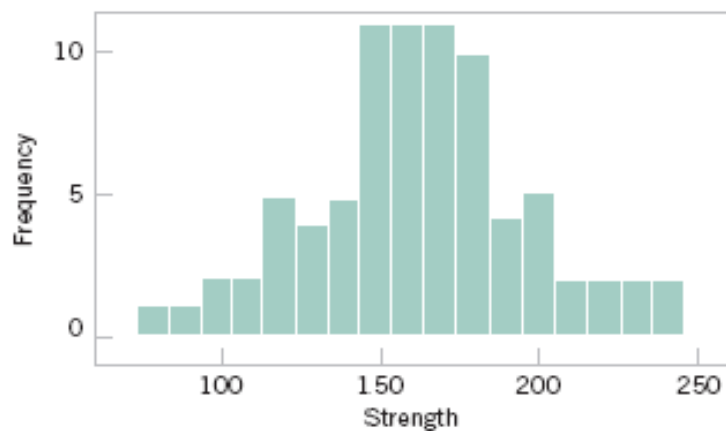
Figure 6-8 A histogram of the compressive strength data from Minitab with 17 bins.

6-3 Frequency Distributions and Histograms

Figure 6-11 Histograms for symmetric and skewed distributions.

6-6 Probability Plots

• **Probability plotting** is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data.

• Probability plotting typically uses special graph paper, known as **probability paper,** that has been designed for the hypothesized distribution. Probability paper is widely available for the normal, lognormal, Weibull, and various chi-square and gamma distributions.
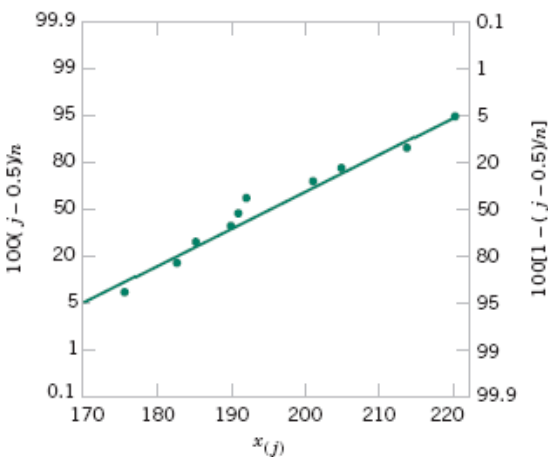
6-6 Probability Plots

Example 6-7

Ten observations on the effective service life in minutes of batteries used in a portable personal computer are as follows: 176, 191, 214, 220, 205, 192, 201, 190, 183, 185. We hypothesize that battery life is adequately modeled by a normal distribution. To use probability plotting to investigate this hypothesis, first arrange the observations in ascending order and calculate their cumulative frequencies $(j - 0.5)/10$ as shown in Table 6-6.

Table 6-6 Calculation for Constructing a Normal Probability Plot

j	$x_{(j)}$	$(j - 0.5)/10$	z_j
1	176	0.05	−1.64
2	183	0.15	−1.04
3	185	0.25	−0.67
4	190	0.35	−0.39
5	191	0.45	−0.13
6	192	0.55	0.13
7	201	0.65	0.39
8	205	0.75	0.67
9	214	0.85	1.04
10	220	0.95	1.64

6-6 Probability Plots

Example 6-7 (continued)

The pairs of values $x_{(j)}$ and $(j - 0.5)/10$ are now plotted on normal probability paper. This plot is shown in Fig. 6-19. Most normal probability paper plots $100(j - 0.5)/n$ on the left vertical scale and $100[1 - (j - 0.5)/n]$ on the right vertical scale, with the variable value plotted on the horizontal scale. A straight line, chosen subjectively, has been drawn through the plotted points. In drawing the straight line, you should be influenced more by the points near the middle of the plot than by the extreme points. A good rule of thumb is to draw the line approximately between the 25th and 75th percentile points. This is how the line in Fig. 6-19 was determined. In assessing the "closeness" of the points to the straight line, imagine a "fat pencil" lying along the line. If all the points are covered by this imaginary pencil, a normal distribution adequately describes the data. Since the points in Fig. 6-19 would pass the "fat pencil" test, we conclude that the normal distribution is an appropriate model.
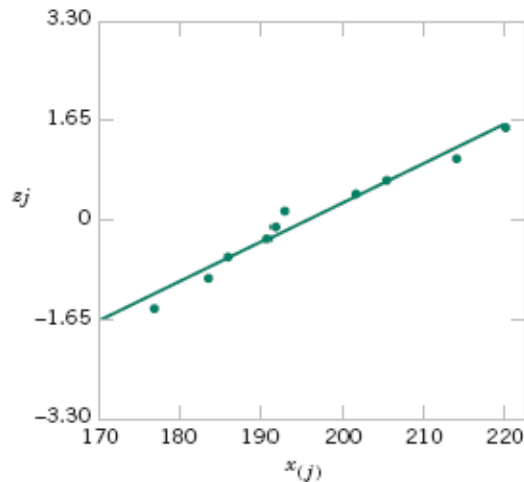
6-6 Probability Plots

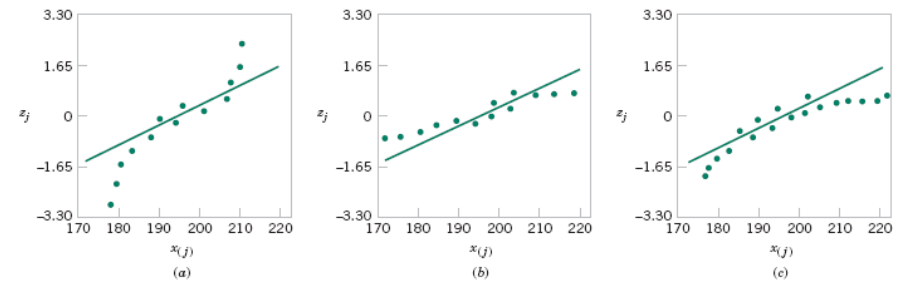**Figure 6-19** Normal probability plot for battery life.

6-6 Probability Plots

Figure 6-20 Normal probability plot obtained from standardized normal scores.

6-6 Probability Plots

Figure 6-21 Normal probability plots indicating a nonnormal distribution. (a) Light-tailed distribution. (b) Heavy-tailed distribution. (c) A distribution with positive (or right) skew.

6-5 Time Sequence Plots

• A **time series** or **time sequence** is a data set in which the observations are recorded in the order in which they occur.
• A **time series plot** is a graph in which the vertical axis denotes the observed value of the variable (say x) and the horizontal axis denotes the time (which could be minutes, days, years, etc.).
• When measurements are plotted as a time series, we often see
    •**trends,**
    •**cycles, or**
    •**other broad features of the data**

6-5 Time Sequence Plots
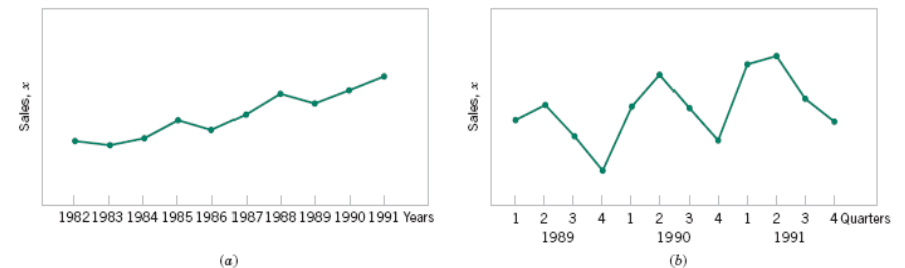
Figure 6-16 Company sales by year (a) and by quarter (b).

6-5 Time Sequence Plots

Leaf	Stem	Time series plot
8	9e	
6	9s	
45	9f	
2333	9t	
0010000	9z	
99998	8e	
66677	8s	
45	8f	
23	8t	
1	8z	
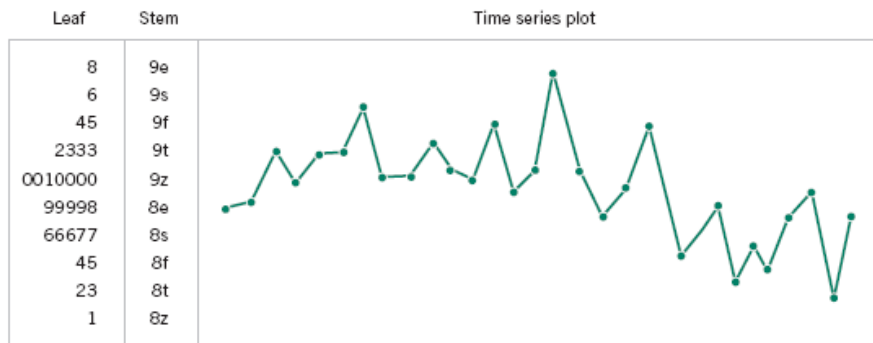
Figure 6-18 A digidot plot of chemical process concentration readings, observed hourly.

Some Useful Comments

- Locations: mean and median
- Spreads: standard deviation (s.d.) and IQR
  - Mean and s.d. are **sensitive** to extreme values (**outliers**)
  - Median and IQR are **resistant** to extreme values and are better for skewed distributions
  - Use mean and s.d. for symmetrical distributions without outliers
- Software has defaults, which may not be the best choice
  - How many stems or bins?
  - The reference line in a normal probability plot.

7

Sampling Distributions and Point Estimation of Parameters

Cramer-Rao Inequality
Fisher Information

CHAPTER OUTLINE

7-1 Introduction

• The field of statistical inference consists of those methods used to make decisions or to draw conclusions about a **population**.

• These methods utilize the information contained in a **sample** from the population in drawing conclusions.

• Statistical inference may be divided into two major areas:

- Parameter estimation

- Hypothesis testing

7-1 Introduction

Suppose that we want to obtain a point estimate of a population parameter. We know that before the data is collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$. Therefore, any function of the observation, or any **statistic,** is also a random variable. For example, the sample mean $\overline{X}$ and the sample variance S^2 are statistics and they are also random variables.

Since a statistic is a random variable, it has a probability distribution. We call the probability distribution of a statistic a **sampling distribution.** The notion of a sampling distribution is very important and will be discussed and illustrated later in the chapter.

Definition

> A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

7-1 Introduction

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

7-1 Introduction

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \overline{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \overline{x}_1 - \overline{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

7.2 Sampling Distributions and the Central Limit Theorem

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.
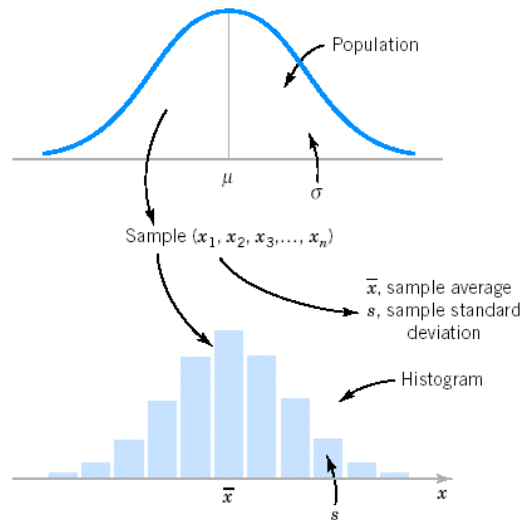
Definitions:

> The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

> A **statistic** is any function of the observations in a random sample.

> The probability distribution of a statistic is called a **sampling distribution.**

7.2 Sampling Distributions

Figure 6-3
Relationship between a population and a sample.

7.2 Sampling Distributions

Suppose $X_1, \ldots, X_n$ are a random sample from a population with mean μ and variance σ^2.

(a) What are the mean and variance of the sample mean?

(b) What is the sampling distribution of the sample mean if the population is normal.

7.2 Sampling Distributions and the Central Limit Theorem

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2/n, if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem.** The statement is as follows:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad (7\text{-}1)$$

as $n \to \infty$, is the standard normal distribution.

If the population is normal, the sampling distribution of Z is exactly standard normal.
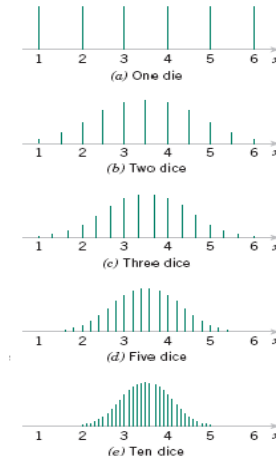
7.2 Sampling Distributions and the Central Limit Theorem

Figure 7-1 Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

CLT Simulation

7.2 Sampling Distributions and the Central Limit Theorem

Example 7-1

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-1. Standardizing the point $\overline{X} = 95$ in Fig. 7-2, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5)$$
$$= 0.0062$$

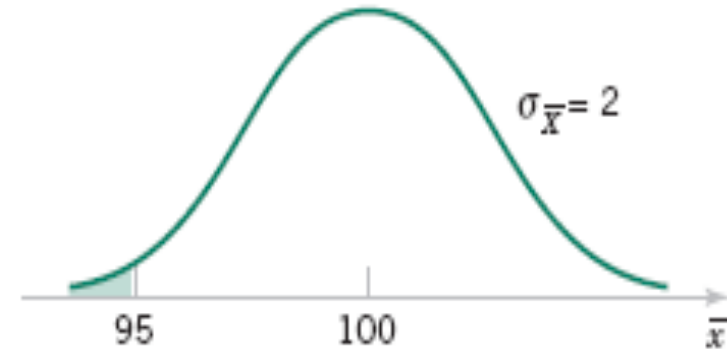# 7.2 Sampling Distributions and the Central Limit Theorem

Figure 7-2 Probability for Example 7-1

7.2 Sampling Distributions and the Central Limit Theorem

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

7-3 General Concepts of Point Estimation

7-3.1 Unbiased Estimators

Definition

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \qquad (7\text{-}5)$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \qquad (7\text{-}6)$$

is called the **bias** of the estimator $\hat{\Theta}$.

Example 7-4

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\overline{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

First consider the sample mean. In Equation 5.40a in Chapter 5, we showed that $E(\overline{X}) = \mu$. Therefore, the sample mean $\overline{X}$ is an unbiased estimator of the population mean μ.

Now consider the sample variance. We have

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1} E \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1} E \sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

Example 7-4 (continued)

The last equality follows from Equation 5-37 in Chapter 5. However, since $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

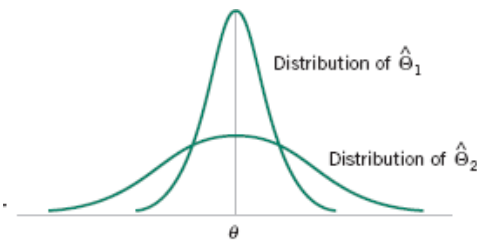$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2)$$

$$= \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

7-3.2 Variance of a Point Estimator

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

Figure 7-5 The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

7-3.3 Standard Error: Reporting a Point Estimate

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Suppose we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the standard error of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the above equation, the estimated standard error of $\overline{X}$ would be

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

7-3.3 Standard Error: Reporting a Point Estimate

Example 7-5

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

$$41.60, 41.48, 42.34, 41.95, 41.86,$$
$$42.18, 41.72, 42.26, 41.81, 42.04$$

A point estimate of the mean thermal conductivity at 100°F and 550 watts is the sample mean or

$$\bar{x} = 41.924 \text{ Btu/hr-ft-°F}$$

Example 7-5 (continued)

The standard error of the sample mean is $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, and since σ is unknown, we may replace it by the sample standard deviation $s = 0.284$ to obtain the estimated standard error of $\bar{X}$ as

$$\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

Notice that the standard error is about 0.2 percent of the sample mean, implying that we have obtained a relatively precise point estimate of thermal conductivity. If we can assume that thermal conductivity is normally distributed, 2 times the standard error is $2\hat{\sigma}_{\bar{X}} = 2(0.0898) = 0.1796$, and we are highly confident that the true mean thermal conductivity is with the interval 41.924 ± 0.1756, or between 41.744 and 42.104.

7-3.4 Mean Square Error of an Estimator

The **mean squared error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 \qquad (7\text{-}7)$$

The mean squared error is an important criterion for comparing two estimators. Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimators of the parameter θ, and let MSE ($\hat{\Theta}_1$) and MSE ($\hat{\Theta}_2$) be the mean squared errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. Then the **relative efficiency** of $\hat{\Theta}_2$ to $\hat{\Theta}_1$ is defined as

$$\frac{\text{MSE}(\hat{\Theta}_1)}{\text{MSE}(\hat{\Theta}_2)} \qquad (7\text{-}8)$$

If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$, in the sense that it has a smaller mean square error.

Figure 7-6 A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

7-4 Methods of Point Estimation

- Problem: To find p=P(heads) for a biased coin.
- Procedure: Flip the coin n times.
- Data (a random sample) : X_1, X_2, ...,X_n
  - where X_i=1 or 0 if the ith outcome is heads or tails.
- Question: How to estimate p using the data?

7-4 Methods of Point Estimation

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

7-4 Methods of Point Estimation

Example 7-7: Consider normal distribution $N(\mu,\sigma^2)$.

Find the moment estimators of μ and σ^2.

7-4 Methods of Point Estimation

7-4.2 Method of Maximum Likelihood

Definition

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \qquad (7\text{-}9)$$

Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator (MLE)** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

7-4 Methods of Point Estimation

Example 7-9

Let X be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$L(p) = p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}\cdots p^{x_n}(1-p)^{1-x_n}$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

7-4 Methods of Point Estimation

Example 7-9 (continued)

We observe that if $\hat{p}$ maximizes $L(p)$, $\hat{p}$ also maximizes $\ln L(p)$. Therefore,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

Now

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

Equating this to zero and solving for p yields $\hat{p} = (1/n)\sum_{i=1}^{n} x_i$. Therefore, the maximum likelihood estimator of p is

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

7-4 Methods of Point Estimation

Examples 7-6 and 7-11

The time to failure of an electronic module used in an automobile engine controller is tested at an elevated temperature to accelerate the failure mechanism. The time to failure is **exponentially distributed**. Eight units are randomly selected and tested, resulting in the following failure time (in hours): 11.96, 5.03, 67.40, 16.07, 31.50, 7.73, 11.10, 22.38.

Here X is exponentially distributed with parameter λ.

(a) What is the moment estimate of λ?

(b) What is the MLE estimate of λ?

7-4 Methods of Point Estimation
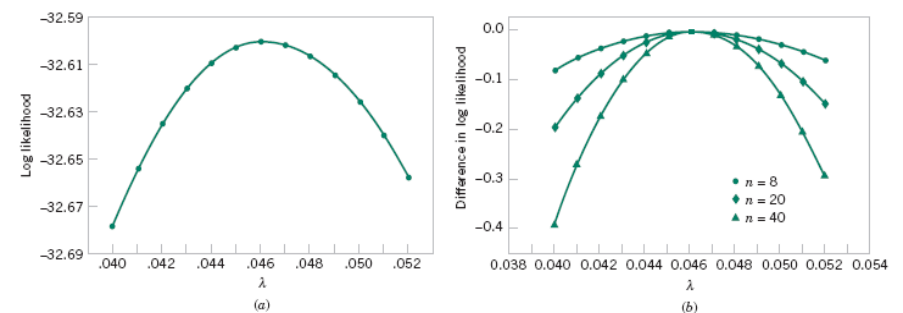
Figure 7-7 Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Difference in Log likelihood if n = 8, 20, and 40.

Example 7-12

Let X be normally distributed with mean μ and variance σ^2, where both μ and σ^2 are unknown. The likelihood function for a random sample of size n is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i - \mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i - \mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Example 7-12 (continued)

Now

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Once again, the maximum likelihood estimators are equal to the moment estimators.

Cramer-Rao Inequality (extra!)

Let $X_1, X_2, \cdots, X_n$ be a random sample with pdf $f(x, \theta)$.

If $\hat{\Theta}$ is an unbiased estimator of θ, then

$$\text{var}(\hat{\Theta}) \geq \frac{1}{nI(\theta)}$$

where

$$I(\theta) = E\left[\frac{\partial}{\partial \theta} \ln f(X; \theta)\right]^2 = -E\left[\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta)\right]$$

is the Fisher information.

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ $[E(\hat{\Theta}) \simeq \theta]$,

(2) the variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\Theta}$ has an approximate normal distribution.

The Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

In the normal distribution case, the maximum likelihood estimators of μ and σ^2 were $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function h, which yields

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}$$

Thus, the maximum likelihood estimator of the standard deviation σ is *not* the sample standard deviation S.

Complications in Using Maximum Likelihood Estimation

• It is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve.

• It may not always be possible to use calculus methods directly to determine the maximum of $L(\theta)$.

• See Example 7-14.

8

Statistical Intervals for a Single Sample

CHAPTER OUTLINE

8-1 Introduction

• In the previous chapter we illustrated how a parameter can be estimated from sample data. However, it is important to understand how good is the estimate obtained.

• Bounds that represent an interval of plausible values for a parameter are an example of an interval estimate.

• Three types of intervals will be presented:

  • Confidence intervals

  • Prediction intervals

  • Tolerance intervals

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. From the results of Chapter 5 we know that the sample mean $\overline{X}$ is normally distributed with mean μ and variance σ^2/n. We may **standardize** $\overline{X}$ by subtracting the mean and dividing by the standard deviation, which results in the variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{8-3}$$

Now Z has a standard normal distribution.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

A **confidence interval** estimate for μ is an interval of the form $l \leq \mu \leq u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

$$P\{L \leq \mu \leq U\} = 1 - \alpha \tag{8-4}$$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \leq \mu \leq u \tag{8-5}$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

- The endpoints or bounds l and u are called lower- and upper-confidence limits, respectively.

- Since Z follows a standard normal distribution, we can write:

$$P\left\{-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - \alpha$$

Now manipulate the quantities inside the brackets by (1) multiplying through by $\sigma/\sqrt{n}$, (2) subtracting $\overline{X}$ from each term, and (3) multiplying through by -1. This results in

$$P\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \tag{8-6}$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

Definition

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1 - \alpha)\%$ CI on μ is given by

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \tag{8-7}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-1

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60°C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1J$. We want to find a 95% CI for μ, the mean impact energy. The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 10$, $\sigma = 1$, and $\bar{x} = 64.46$. The resulting 95% CI is found from Equation 8-7 as follows:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \leq \mu \leq 65.08$$

That is, based on the sample data, a range of highly plausible vaules for mean impact energy for A238 steel at 60°C is $63.84J \leq \mu \leq 65.08J$.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Interpreting a Confidence Interval

- The confidence interval is a random interval

- The appropriate interpretation of a confidence interval (for example on μ) is: The observed interval $[l, u]$ brackets the true value of μ, with confidence $100(1-\alpha)$.

- Examine Figure 8-1 on the next slide.

- Simulation on CI

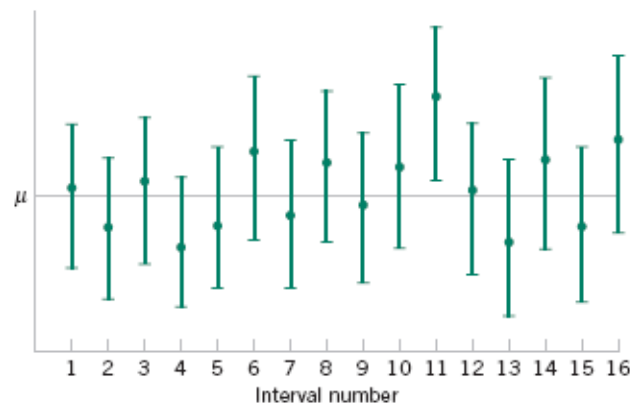## 8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Figure 8-1 Repeated construction of a confidence interval for μ.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Confidence Level and Precision of Error

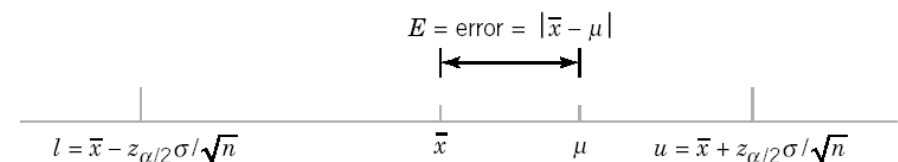The length of a confidence interval is a measure of the precision of estimation.

Figure 8-2 Error in estimating μ with $\bar{x}$.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.2 Choice of Sample Size

If $\bar{x}$ is used as an estimate of μ, we can be $100(1 - \alpha)\%$ confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \qquad (8\text{-}8)$$

Example 8-2

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most $1.0J$. Since the bound on error in estimation E is one-half of the length of the CI, to determine n we use Equation 8-8 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 0.025$. The required sample size is 16

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

and because n must be an integer, the required sample size is $n = 16$.

8-2.3 One-Sided Confidence Bounds

Definition

A $100(1 - \alpha)\%$ **upper-confidence bound** for μ is

$$\mu \leq u = \bar{x} + z_\alpha \sigma/\sqrt{n} \qquad (8\text{-}9)$$

and a $100(1 - \alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} = l \leq \mu \qquad (8\text{-}10)$$

8-2.4 General Method to Derive a Confidence Interval

It is easy to give a general method for finding a confidence interval for an unknown parameter θ. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations. Suppose we can find a statistic $g(X_1, X_2, \ldots, X_n; \theta)$ with the following properties:

1. $g(X_1, X_2, \ldots, X_n; \theta)$ depends on both the sample and θ.
2. The probability distribution of $g(X_1, X_2, \ldots, X_n; \theta)$ does not depend on θ or any other unknown parameter.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.4 General Method to Derive a Confidence Interval

In the case considered in this section, the parameter $\theta = \mu$. The random variable $g(X_1, X_2, \ldots, X_n; \mu) = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ and satisfies both conditions above; it depends on the sample and on μ, and it has a standard normal distribution since σ is known. Now one must find constants C_L and C_U so that

$$P[C_L \le g(X_1, X_2, \ldots, X_n; \theta) \le C_U] = 1 - \alpha \qquad (8\text{-}11)$$

Because of property 2, C_L and C_U do not depend on θ. In our example, $C_L = -z_{\alpha/2}$ and $C_U = z_{\alpha/2}$. Finally, you must manipulate the inequalities in the probability statement so that

$$P[L(X_1, X_2, \ldots, X_n) \le \theta \le U(X_1, X_2, \ldots, X_n)] = 1 - \alpha \qquad (8\text{-}12)$$

This gives $L(X_1, X_2, \ldots, X_n)$ and $U(X_1, X_2, \ldots, X_n)$ as the lower and upper confidence limits defining the $100(1 - \alpha)\%$ confidence interval for θ. The quantity $g(X_1, X_2, \ldots, X_n; \theta)$ is often called a "pivotal quantity" because we pivot on this quantity in Equation 8-11 to produce Equation 8-12. In our example, we manipulated the pivotal quantity $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ to obtain $L(X_1, X_2, \ldots, X_n) = \overline{X} - z_{\alpha/2}\sigma/\sqrt{n}$ and $U(X_1, X_2, \ldots, X_n) = \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}$.

8-2.5 A Large-Sample Confidence Interval for μ

Definition

When n is large, the quantity

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \qquad (8\text{-}13)$$

is a **large sample confidence interval** for μ, with confidence level of approximately $100(1 - \alpha)\%$.

Example 8-4

An article in the 1993 volume of the *Transactions of the American Fisheries Society* reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

1.230	0.490	0.490	1.080	0.590	0.280	0.180	0.100	0.940
1.330	0.190	1.160	0.980	0.340	0.340	0.190	0.210	0.400
0.040	0.830	0.050	0.630	0.340	0.750	0.040	0.860	0.430
0.044	0.810	0.150	0.560	0.840	0.870	0.490	0.520	0.250
1.200	0.710	0.190	0.410	0.500	0.560	1.100	0.650	0.270
0.270	0.500	0.770	0.730	0.340	0.170	0.160	0.270	

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)
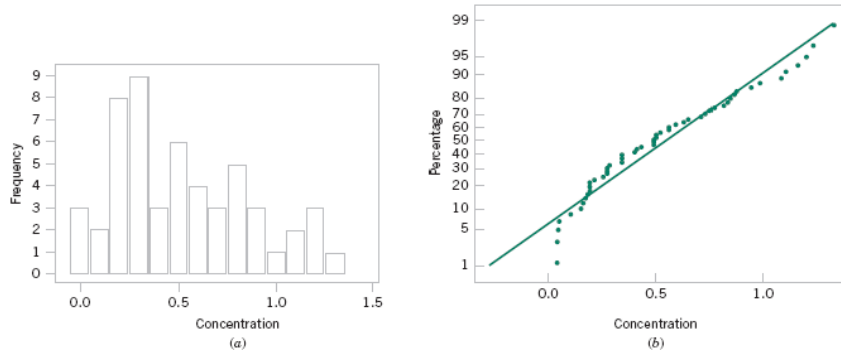
Figure 8-3 Mercury concentration in largemouth bass
(a) Histogram. (b) Normal probability plot

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)

Figure 8-3(a) and (b) presents the histogram and normal probability plot of the mercury concentration data. Both plots indicate that the distribution of mercury concentration is not normal and is positively skewed. We want to find an approximate 95% CI on μ. Because $n > 40$, the assumption of normality is not necessary to use Equation 8-13. The required quantities are $n = 53$, $\bar{x} = 0.5250$, $s = 0.3486$, and $z_{0.025} = 1.96$. The approximate 95% CI on μ is

$$\bar{x} - z_{0.025}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \leq \mu \leq 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \leq \mu \leq 0.6189$$

This interval is fairly wide because there is a lot of variability in the mercury concentration measurements.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

A General Large Sample Confidence Interval

$$\hat{\theta} - z_{\alpha/2}\,\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\,\sigma_{\hat{\theta}} \qquad (8\text{-}14)$$

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown
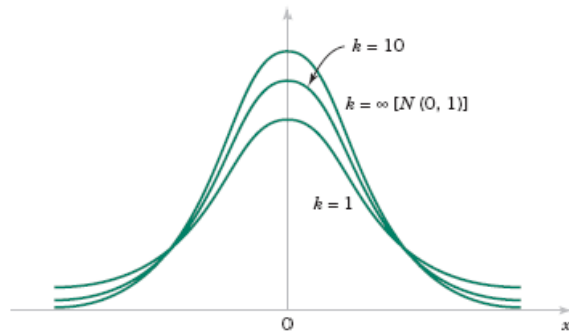
8-3.1 The t distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \qquad (8\text{-}15)$$

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

8-3.1 The t distribution

Figure 8-4 Probability density functions of several t distributions.
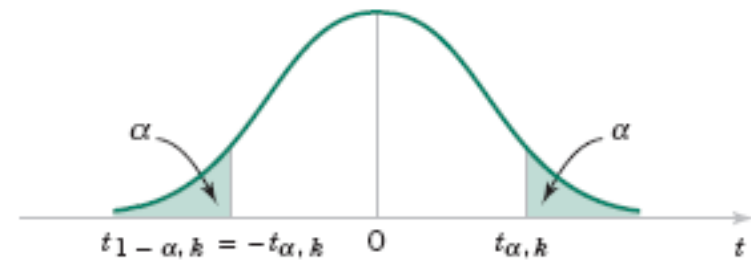
Figure 8-5 Percentage points of the t distribution.

8-3.2 The t Confidence Interval on μ

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)$ percent confidence interval on μ is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \qquad (8\text{-}18)$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

One-sided confidence bounds on the mean are found by replacing $t_{\alpha/2,n-1}$ in Equation 8-18 with $t_{\alpha,n-1}$.

Example 8-5

An article in the journal *Materials Engineering* (1989, Vol. II, No. 4, pp. 275–281) describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows (in megapascals):

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	
19.5	8.8	13.6	11.9	11.4	

The sample mean is $\bar{x} = 13.71$, and the sample standard deviation is $s = 3.55$. Figures 8-6 and 8-7 show a box plot and a normal probability plot of the tensile adhesion test data, respectively. These displays provide good support for the assumption that the population is normally distributed. We want to find a 95% CI on μ. Since $n = 22$, we have $n - 1 = 21$ degrees of freedom for t, so $t_{0.025,21} = 2.080$. The resulting CI is

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n}$$

$$13.71 - 2.080(3.55)/\sqrt{22} \le \mu \le 13.71 + 2.080(3.55)/\sqrt{22}$$

$$13.71 - 1.57 \le \mu \le 13.71 + 1.57$$

$$12.14 \le \mu \le 15.28$$

The CI is fairly wide because there is a lot of variability in the tensile adhesion test measurements.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown
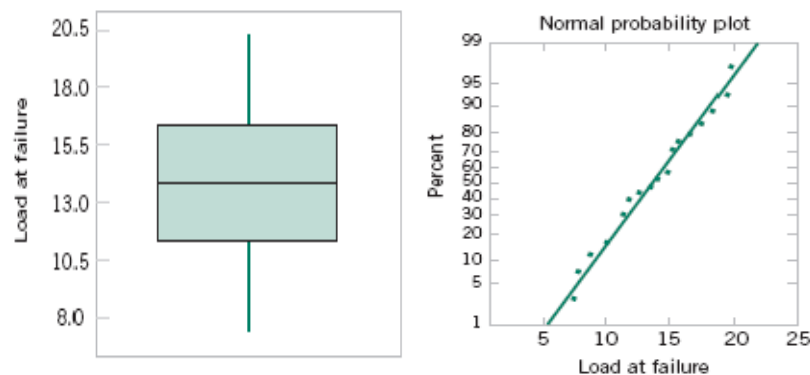
Figure 8-6/8-7 Box and Whisker plot and Normal probability plot for the load at failure data in Example 8-5.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \tag{8-19}$$

has a chi-square (χ^2) distribution with $n-1$ degrees of freedom.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Figure 8-8 Probability density functions of several χ^2 distributions.
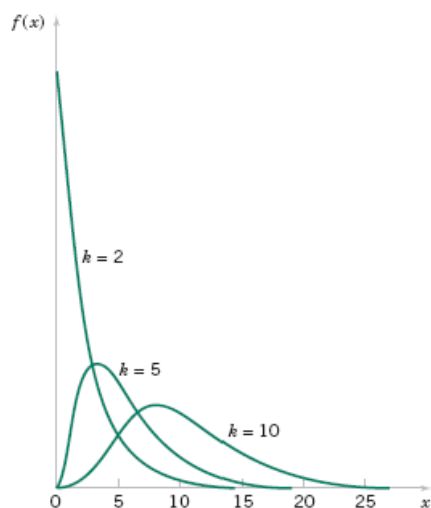
8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{8-21}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A **confidence interval for σ** has lower and upper limits that are the square roots of the corresponding limits in Equation 8-21.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

One-Sided Confidence Bounds

The $100(1 - \alpha)\%$ lower and upper confidence bounds on σ^2 are

$$\frac{(n - 1)s^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n - 1)s^2}{\chi^2_{1-\alpha,n-1}} \qquad (8\text{-}22)$$

respectively.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Example 8-6

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. A 95% upper-confidence interval is found from Equation 8-22 as follows:

$$\sigma^2 \leq \frac{(n - 1)s^2}{\chi^2_{0.95,19}}$$

or

$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

This last expression may be converted into a confidence interval on the standard deviation σ by taking the square root of both sides, resulting in

$$\sigma \leq 0.17$$

Therefore, at the 95% level of confidence, the data indicate that the process standard deviation could be as large as 0.17 fluid ounce.

8-5 A Large-Sample Confidence Interval For a Population Proportion

Normal Approximation for Binomial Proportion

If n is large, the distribution of

$$Z = \frac{X - np}{\sqrt{np(1 - p)}} = \frac{\hat{P} - p}{\sqrt{\dfrac{p(1 - p)}{n}}}$$

is approximately standard normal.

The quantity $\sqrt{p(1 - p)/n}$ is called the standard error of the point estimator $\hat{P}$.

8-5 A Large-Sample Confidence Interval For a Population Proportion

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1 - \alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \qquad (8\text{-}25)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

8-5 A Large-Sample Confidence Interval For a Population Proportion

Example 8-7

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\hat{p} = x/n = 10/85 = 0.12$. A 95% two-sided confidence interval for p is computed from Equation 8-25 as

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

or

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \leq p \leq 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

which simplifies to

$$0.05 \leq p \leq 0.19$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Choice of Sample Size

The sample size for a specified value E is given by

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \tag{8-26}$$

An upper bound on n is given by

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \tag{8-27}$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Example 8-8

Consider the situation in Example 8-7. How large a sample is required if we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05? Using $\hat{p} = 0.12$ as an initial estimate of p, we find from Equation 8-26 that the required sample size is

$$n = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.12(0.88) \cong 163$$

If we wanted to be *at least* 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p, we would use Equation 8-27 to find the sample size

$$n = \left(\frac{z_{0.025}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.05}\right)^2 (0.25) \cong 385$$

Notice that if we have information concerning the value of p, either from a preliminary sample or from past experience, we could use a smaller sample while maintaining both the desired precision of estimation and the level of confidence.

8-5 A Large-Sample Confidence Interval For a Population Proportion

One-Sided Confidence Bounds

The approximate $100(1-\alpha)$% lower and upper confidence bounds are

$$\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \quad \text{and} \quad p \leq \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{8-28}$$

respectively.

8-6 Guidelines for Constructing Confidence Intervals

The most difficult step in constructing a confidence interval is often the match of the appropriate calculation to the objective of the study. Common cases are listed in Table 8-1 along with the reference to the section that covers the appropriate calculation for a confidence interval test. Table 8-1 provides a simple road map to help select the appropriate analysis. Two primary comments can help identify the analysis:

1. Determine the parameter (and the distribution of the data) that will be bounded by the confidence interval or tested by the hypothesis.

2. Check if other parameters are known or need to be estimated.

8-7 Tolerance and Prediction Intervals

8-7.1 Prediction Interval for Future Observation

A $100(1 - \alpha)\%$ prediction interval on a single future observation from a normal distribution is given by

$$\bar{x} - t_{\alpha/2,n-1}s\sqrt{1 + \frac{1}{n}} \leq X_{n+1} \leq \bar{x} + t_{\alpha/2,n-1}s\sqrt{1 + \frac{1}{n}} \qquad (8\text{-}29)$$

The prediction interval for X_{n+1} will always be longer than the confidence interval for μ.

8-7 Tolerance and Prediction Intervals

Example 8-9

Reconsider the tensile adhesion tests on specimens of U-700 alloy described in Example 8-5. The load at failure for $n = 22$ specimens was observed, and we found that $\bar{x} = 13.71$ and $s = 3.55$. The 95% confidence interval on μ was $12.14 \leq \mu \leq 15.28$. We plan to test a twenty-third specimen. A 95% prediction interval on the load at failure for this specimen is

$$\bar{x} - t_{\alpha/2,n-1}s\sqrt{1 + \frac{1}{n}} \leq X_{n+1} \leq \bar{x} + t_{\alpha/2,n-1}s\sqrt{1 + \frac{1}{n}}$$

$$13.71 - (2.080)3.55\sqrt{1 + \frac{1}{22}} \leq X_{23} \leq 13.71 + (2.080)3.55\sqrt{1 + \frac{1}{22}}$$

$$6.16 \leq X_{23} \leq 21.26$$

Notice that the prediction interval is considerably longer than the CI.

8-7 Tolerance and Prediction Intervals

8-7.2 Tolerance Interval for a Normal Distribution

Consider a population of semiconductor processors. Suppose that the speed of these processors has a normal distribution with mean $\mu = 600$ megahertz and standard deviation $\sigma = 30$ megahertz. Then the interval from $600 - 1.96(30) = 541.2$ to $600 + 1.96(30) = 658.8$ megahertz captures the speed of 95% of the processors in this population because the interval from -1.96 to 1.96 captures 95% of the area under the standard normal curve. The interval from $\mu - z_{\alpha/2}\sigma$ to $\mu + z_{\alpha/2}\sigma$ is called a **tolerance interval**.

If μ and σ are unknown, we can use the data from a random sample of size n to compute $\bar{x}$ and s, and then form the interval $(\bar{x} - 1.96s, \bar{x} + 1.96s)$. However, because of sampling variability in $\bar{x}$ and s, it is likely that this interval will contain less than 95% of the values in the population. The solution to this problem is to replace 1.96 by some value that will make the proportion of the distribution contained in the interval 95% with some level of confidence. Fortunately, it is easy to do this.

8-7 Tolerance and Prediction Intervals

8-7.2 Tolerance Interval for a Normal Distribution

Definition

A **tolerance interval** for capturing at least $\gamma\%$ of the values in a normal distribution with confidence level $100(1 - \alpha)\%$ is

$$\bar{x} - ks, \qquad \bar{x} + ks$$

where k is a tolerance interval factor found in Appendix Table XII. Values are given for $\gamma = 90\%$, 95%, and 99% and for 90%, 95%, and 99% confidence.

8-7 Tolerance and Prediction Intervals

EXAMPLE 8-10 Alloy Adhesion

Let's reconsider the tensile adhesion tests originally described in Example 8-5. The load at failure for $n = 22$ specimens was observed, and we found that $\bar{x} = 13.71$ and $s = 3.55$. We want to find a tolerance interval for the load at failure that includes 90% of the values in the population with 95% confidence. From Appendix Table XII the tolerance factor k for $n = 22$, $\gamma = 0.90$, and 95% confidence is $k = 2.264$. The desired tolerance interval is

$$(\bar{x} - ks, \ \bar{x} + ks) \quad \text{or} \quad [13.71 - (2.264)3.55, 13.71 + (2.264)3.55]$$

which reduces to (5.67, 21.74). We can be 95% confident that at least 90% of the values of load at failure for this particular alloy lie between 5.67 and 21.74 megapascals.

Simulation on Tolerance Intervals

9

Tests of Hypotheses for a Single Sample

9-3.4 Likelihood ratio test

Neyman-Pearson lemma

9-1 Hypothesis Testing

9-1.1 Statistical Hypotheses

Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a **comparative experiment,** in which the engineer is interested, for example, in comparing the mean of a population to a specified value.

Definition

A **statistical hypothesis** is a statement about the parameters of one or more populations.

9-1 Hypothesis Testing

9-1.1 Statistical Hypotheses

For example, suppose that we are interested in the burning rate of a solid propellant used to power aircrew escape systems.

- Now burning rate is a random variable that can be described by a probability distribution.

- Suppose that our interest focuses on the **mean** burning rate (a parameter of this distribution).

- Specifically, we are interested in deciding whether or not the mean burning rate is 50 centimeters per second.

9-1 Hypothesis Testing

9-1.1 Statistical Hypotheses

<u>Two-sided Alternative Hypothesis</u>

$H_0: \mu = 50$ centimeters per second   null hypothesis

$H_1: \mu \neq 50$ centimeters per second   alternative hypothesis

<u>One-sided Alternative Hypotheses</u>

$H_0: \mu = 50$ centimeters per second $H_0: \mu = 50$ centimeters per second

or

$H_1: \mu < 50$ centimeters per second $H_1: \mu > 50$ centimeters per second

9-1 Hypothesis Testing

9-1.1 Statistical Hypotheses

<u>Test of a Hypothesis</u>
- A procedure leading to a decision about a particular hypothesis

- Hypothesis-testing procedures rely on using the information in a **random sample from the population of interest**.

- If this information is *consistent* with the hypothesis, then we will conclude that the hypothesis is **true**; if this information is *inconsistent* with the hypothesis, we will conclude that the hypothesis is **false**.

9-1 Hypothesis Testing

9-1.2 Tests of Statistical Hypotheses

$H_0: \mu = 50$ centimeters per second
$H_1: \mu \neq 50$ centimeters per second

| Reject H_0 | Fail to Reject H_0 | Reject H_0 |
| $\mu \neq 50$ cm/s | $\mu = 50$ cm/s | $\mu \neq 50$ cm/s |

48.5   50   51.5 $\bar{x}$

Figure 9-1 Decision criteria for testing $H_0: \mu = 50$ centimeters per second versus $H_1: \mu \neq 50$ centimeters per second.

9-1 Hypothesis Testing

9-1.2 Tests of Statistical Hypotheses

Table 9-1   Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	no error	type II error
Reject H_0	type I error	no error

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error.**

Failing to reject the null hypothesis when it is false is defined as a **type II error.**

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

Sometimes the type I error probability is called the **significance level,** or the **α-error,** or the **size** of the test.

9-1 Hypothesis Testing

9-1.2 Tests of Statistical Hypotheses

• In the propellant burning rate example, a type I error will occur when

$$\bar{x} < 48.5 \text{ or } \bar{x} > 51.5$$

when the true mean burning rate is $\mu = 50$ centimeters per second.

• n=10.

• Suppose that the standard deviation of burning rate is $\sigma = 2.5$ centimeters per second and that the burning rate has a normal distribution, so the distribution of the sample mean is normal with mean $\mu = 50$ and standard deviation

$$\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{10}} = 0.79$$

• The probability of making a type I error (or the significance level of our test) is equal to the sum of the areas that have been shaded in the tails of the normal distribution in Fig. 9-2.

9-1 Hypothesis Testing

9-1.2 Tests of Statistical Hypotheses

$$\alpha = P(\overline{X} < 48.5 \text{ when } \mu = 50) + P(\overline{X} > 51.5 \text{ when } \mu = 50)$$

The z-values that correspond to the critical values 48.5 and 51.5 are

$$z_1 = \frac{48.5 - 50}{0.79} = -1.90 \quad \text{and} \quad z_2 = \frac{51.5 - 50}{0.79} = 1.90$$

Therefore

$$\alpha = P(Z < -1.90) + P(Z > 1.90) = 0.028717 + 0.028717 = 0.057434$$

9-1 Hypothesis Testing
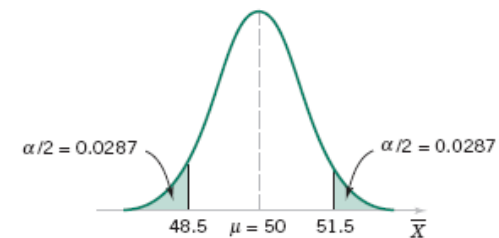
Figure 9-2   The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \qquad (9\text{-}3)$$

9-1 Hypothesis Testing

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \qquad (9\text{-}4)$$
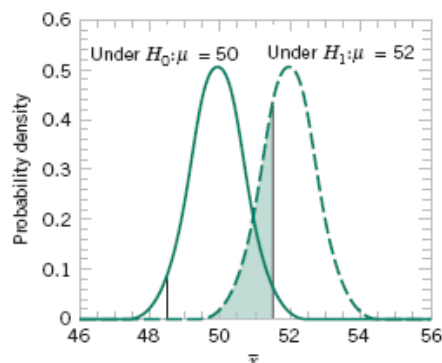
Figure 9-3 The probability of type II error when μ = 52 and n = 10.

9-1 Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 52)$$

The z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.79} = -4.43 \qquad \text{and} \qquad z_2 = \frac{51.5 - 52}{0.79} = -0.63$$

Therefore

$$\beta = P(-4.43 \leq Z \leq -0.63) = P(Z \leq -0.63) - P(Z \leq -4.43)$$
$$= 0.2643 - 0.0000 = 0.2643$$

9-1 Hypothesis Testing
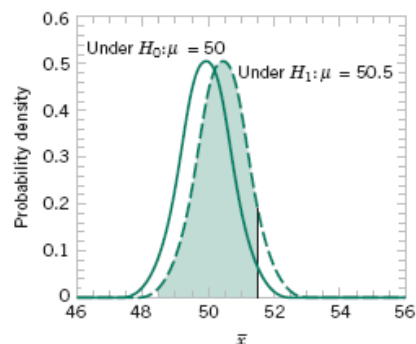
$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 50.5)$$

Figure 9-4 The probability of type II error when μ = 50.5 and n = 10.

9-1 Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 50.5)$$

As shown in Fig. 9-4, the z-values corresponding to 48.5 and 51.5 when $\mu = 50.5$ are

$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53 \qquad \text{and} \qquad z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$

Therefore

$$\beta = P(-2.53 \leq Z \leq 1.27) = P(Z \leq 1.27) - P(Z \leq -2.53)$$
$$= 0.8980 - 0.0057 = 0.8923$$

9-1 Hypothesis Testing
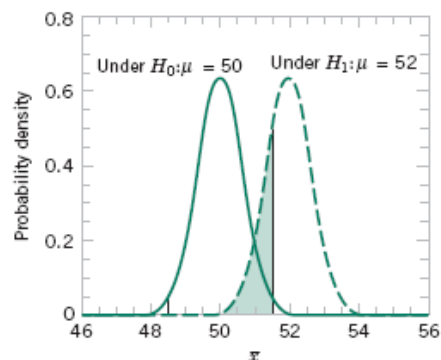
$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 52)$$

Figure 9-5 The probability of type II error when $\mu = 52$ and $n = 16$.

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 52)$$

When $n = 16$, the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$, and the z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.625} = -5.60 \quad \text{and} \quad z_2 = \frac{51.5 - 52}{0.625} = -0.80$$

Therefore

$$\beta = P(-5.60 \le Z \le -0.80) = P(Z \le -0.80) - P(Z \le -5.60)$$
$$= 0.2119 - 0.0000 = 0.2119$$

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \overline{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \overline{x} < 52$	10	0.0114	0.5000	0.9705
$48.5 < \overline{x} < 51.5$	16	0.0164	0.2119	0.9445
$48 < \overline{x} < 52$	16	0.0014	0.5000	0.9918

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \overline{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \overline{x} < 52$	10	0.0114	0.5000	0.9705
$48.81 < \overline{x} < 51.19$	16	0.0576	0.0966	0.8606
$48.42 < \overline{x} < 51.58$	16	0.0114	0.2515	0.9578

1. The size of the critical region, and consequently the probability of a type I error α, can always be reduced by appropriate selection of the critical values.

2. Type I and type II errors are related. A decrease in the probability of one type of error always results in an increase in the probability of the other, provided that the sample size n does not change.

3. An increase in sample size reduces β, provided that α is held constant.

4. When the null hypothesis is false, β increases as the true value of the parameter approaches the value hypothesized in the null hypothesis. The value of β decreases as the difference between the true mean and the hypothesized value increases.

9-1 Hypothesis Testing

Definition

> The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

- The power is computed as $1 - \beta$, and power can be interpreted as *the probability of correctly rejecting a false null hypothesis.* We often compare statistical tests by comparing their power properties.

- For example, consider the propellant burning rate problem when we are testing $H_0 : \mu = 50$ centimeters per second against $H_1 : \mu$ not equal 50 centimeters per second . Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

9-1 Hypothesis Testing

9-1.3 One-Sided and Two-Sided Hypotheses

Two-Sided Test:

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

One-Sided Tests:

$$H_0: \mu = \mu_0 \qquad \text{or} \qquad H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0 \qquad\qquad\qquad H_1: \mu < \mu_0$$

Rejecting H_0 is a strong conclusion.

9-1 Hypothesis Testing

Example 9-1

Consider the propellant burning rate problem. Suppose that if the burning rate is less than 50 centimeters per second, we wish to show this with a strong conclusion. The hypotheses should be stated as

$$H_0: \mu = 50 \text{ centimeters per second}$$
$$H_1: \mu < 50 \text{ centimeters per second}$$

Here the critical region lies in the lower tail of the distribution of $\overline{X}$. Since the rejection of H_0 is always a strong conclusion, this statement of the hypotheses will produce the desired outcome if H_0 is rejected. Notice that, although the null hypothesis is stated with an equal sign, it is understood to include any value of μ not specified by the alternative hypothesis. Therefore, failing to reject H_0 does not mean that $\mu = 50$ centimeters per second exactly, but only that we do not have strong evidence in support of H_1.

9-1 Hypothesis Testing

9-1.4 P-Values in Hypothesis Tests

P-value = P (test statistic will take on a value that is at least as extreme as the observed value when the null hypothesis H_0 is true)

Decision rule:

- If P-value > α , fail to reject H_0 at significance level α;

- If P-value < α , reject H_0 at significance level α.

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

9-1 Hypothesis Testing

9-1.4 P-Values in Hypothesis Tests

Consider the two-sided hypothesis test for burning rate

$$H_0: \mu = 50 \quad H_1: \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is $\bar{x} = 51.3$ centimeters per second. Figure 9-6 shows a critical region for this test with critical values at 51.3 and the symmetric value 48.7. The P-value of the test is the α associated with this critical region. Any smaller value for α expands the critical region and the test fails to reject the null hypothesis when $\bar{x} = 51.3$. The P-value is easy to compute after the test statistic is observed. In this example

$$
\begin{aligned}
P\text{-value} &= 1 - P(48.7 < \bar{X} < 51.3) \\
&= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right) \\
&= 1 - P(-2.08 < Z < 2.08) \\
&= 1 - 0.962 = 0.038
\end{aligned}
$$

9-1 Hypothesis Testing

Figure 9-6 P-value is area of shaded region when $\bar{x} = 51.3$.

9-1 Hypothesis Testing

9-1.5 Connection between Hypothesis Tests and Confidence Intervals

There is a close relationship between the test of a hypothesis about any parameter, say θ, and the confidence interval for θ. If $[l, u]$ is a $100(1 - \alpha)\%$ confidence interval for the parameter θ, the test of size α of the hypothesis

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is **not** in the $100(1 - \alpha)\%$ CI $[l, u]$. As an illustration, consider the escape system propellant problem with $\bar{x} = 51.3$, $\sigma = 2.5$, and $n = 16$. The null hypothesis $H_0: \mu = 50$ was rejected, using $\alpha = 0.05$. The 95% two-sided CI on μ can be calculated using Equation 8-7. This CI is $51.3 \pm 1.96(2.5/\sqrt{16})$ and this is $50.075 \leq \mu \leq 52.525$. Because the value $\mu_0 = 50$ is not included in this interval, the null hypothesis $H_0: \mu = 50$ is rejected.

9-1 Hypothesis Testing

9-1.6 General Procedure for Hypothesis Tests

1. From the problem context, identify the parameter of interest.

2. State the null hypothesis, H_0.

3. Specify an appropriate alternative hypothesis, H_1.

4. Choose a significance level, α.

5. Determine an appropriate test statistic.

6. State the rejection region for the statistic.

7. Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.

8. Decide whether or not H_0 should be rejected and report that in the problem context.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.1 Hypothesis Tests on the Mean

We wish to test:

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

The **test statistic** is:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{9-8}$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.1 Hypothesis Tests on the Mean

Reject H_0 if the observed value of the test statistic z_0 is either:

$$z_0 > z_{\alpha/2} \text{ or } z_0 < -z_{\alpha/2}$$

Fail to reject H_0 if

$$-z_{\alpha/2} < z_0 < z_{\alpha/2}$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known
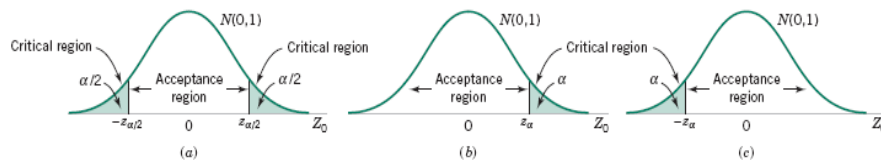
Figure 9-7 The distribution of Z_0 when $H_0: \mu = \mu_0$ is true, with critical region for (a) the two-sided alternative $H_1: \mu \neq \mu_0$, (b) the one-sided alternative $H_1: \mu > \mu_0$, and (c) the one-sided alternative $H_1: \mu < \mu_0$.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

Example 9-2

Aircrew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $n = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

9-2 Tests on the Mean of a Normal Distribution, Variance Known

Example 9-2

We may solve this problem by following the eight-step procedure outlined in Section 9-1.4. This results in

1. The parameter of interest is μ, the mean burning rate.
2. H_0: $\mu = 50$ centimeters per second
3. H_1: $\mu \neq 50$ centimeters per second
4. $\alpha = 0.05$
5. The test statistic is

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

6. Reject H_0 if $z_0 > 1.96$ or if $z_0 < -1.96$. Note that this results from step 4, where we specified $\alpha = 0.05$, and so the boundaries of the critical region are at $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.

7. Computations: Since $\bar{x} = 51.3$ and $\sigma = 2$,

$$z_0 = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

8. Conclusion: Since $z_0 = 3.25 > 1.96$, we reject H_0: $\mu = 50$ at the 0.05 level of significance. Stated more completely, we conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.1 Hypothesis Tests on the Mean

We may also develop procedures for testing hypotheses on the mean μ where the alternative hypothesis is one-sided. Suppose that we specify the hypotheses as

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0 \tag{9-11}$$

In defining the critical region for this test, we observe that a negative value of the test statistic Z_0 would never lead us to conclude that H_0: $\mu = \mu_0$ is false. Therefore, we would place the critical region in the **upper tail** of the standard normal distribution and reject H_0 if the computed value of z_0 is too large. That is, we would reject H_0 if

$$z_0 > z_\alpha \tag{9-12}$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.1 Hypothesis Tests on the Mean (Continued)

as shown in Figure 9-7(b). Similarly, to test

$$H_0: \mu = \mu_0$$
$$H_1: \mu < \mu_0 \tag{9-13}$$

we would calculate the test statistic Z_0 and reject H_0 if the value of z_0 is too small. That is, the critical region is in the **lower tail** of the standard normal distribution as shown in Figure 9-7(c), and we reject H_0 if

$$z_0 < -z_\alpha \tag{9-14}$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.1 Hypothesis Tests on the Mean (Continued)

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

Alternative hypothesis	Rejection criteria
$H_1: \mu \neq \mu_0$	$Z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: \mu > \mu_0$	$z_0 > z_\alpha$
$H_1: \mu < \mu_0$	$z_0 < -z_\alpha$

The notation on p. 307 includes n-1, which is wrong.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

P-Values in Hypothesis Tests

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

$$P = \begin{cases} 2[1 - \Phi(|z_0|)] & \text{for a two-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \\ 1 - \Phi(z_0) & \text{for a upper-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \\ \Phi(z_0) & \text{for a lower-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0 \end{cases} \qquad (9\text{-}15)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Finding the Probability of Type II Error β

Consider the two-sided hypothesis

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

Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The test statistic Z_0 is

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

Therefore, the distribution of Z_0 when H_1 is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \qquad (9\text{-}16)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Finding the Probability of Type II Error β

$\beta = P(\text{type II error}) = P(\text{failing to reject } H_0 \text{ when it is false})$

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \qquad (9\text{-}17)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Finding the Probability of Type II Error β
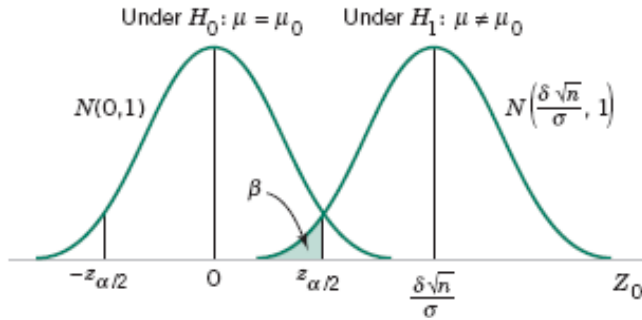
Figure 9-7 The distribution of Z_0 under H_0 and H_1

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Sample Size Formulas

For a two-sided alternative hypothesis:

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} \qquad \text{where} \qquad \delta = \mu - \mu_0 \qquad (9\text{-}19)$$

For a one-sided alternative hypothesis:

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\delta^2} \qquad \text{where} \qquad \delta = \mu - \mu_0 \qquad (9\text{-}20)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

Example 9-3

Consider the rocket propellant problem of Example 9-2. Suppose that the analyst wishes to design the test so that if the true mean burning rate differs from 50 centimeters per second by as much as 1 centimeter per second, the test will detect this (i.e., reject H_0: $\mu = 50$) with a high probability, say 0.90. Now, we note that $\sigma = 2$, $\delta = 51 - 50 = 1$, $\alpha = 0.05$, and $\beta = 0.10$. Since $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_{\beta} = z_{0.10} = 1.28$, the sample size required to detect this departure from H_0: $\mu = 50$ is found by Equation 9-19 as

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} = \frac{(1.96 + 1.28)^2 2^2}{(1)^2} \simeq 42$$

The approximation is good here, since $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) = \Phi(-1.96 - (1)\sqrt{42}/2) = \Phi(-5.20) \simeq 0$, which is small relative to β.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Using Operating Characteristic Curves

When performing sample size or type II error calculations, it is sometimes more convenient to use the **operating characteristic (OC) curves** in Appendix Charts VIa and VIb. These curves plot β as calculated from Equation 9-17 against a parameter d for various sample sizes n. Curves are provided for both $\alpha = 0.05$ and $\alpha = 0.01$. The parameter d is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} \qquad (9\text{-}21)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.2 Type II Error and Choice of Sample Size

Using Operating Characteristic Curves

so one set of operating characteristic curves can be used for all problems regardless of the values of μ_0 and σ. From examining the operating characteristic curves or Equation 9-17 and Fig. 9-7, we note that

1. The further the true value of the mean μ is from μ_0, the smaller the probability of type II error β for a given n and α. That is, we see that for a specified sample size and α, large differences in the mean are easier to detect than small ones.

2. For a given δ and α, the probability of type II error β decreases as n increases. That is, to detect a specified difference δ in the mean, we may make the test more powerful by increasing the sample size.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

Example 9-4

Consider the propellant problem in Example 9-2. Suppose that the analyst is concerned about the probability of type II error if the true mean burning rate is $\mu = 51$ centimeters per second. We may use the operating characteristic curves to find β. Note that $\delta = 51 - 50 = 1$, $n = 25$, $\sigma = 2$, and $\alpha = 0.05$. Then using Equation 9-21 gives

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} = \frac{1}{2}$$

and from Appendix Chart VIIa, with $n = 25$, we find that $\beta = 0.30$. That is, if the true mean burning rate is $\mu = 51$ centimeters per second, there is approximately a 30% chance that this will not be detected by the test with $n = 25$.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.3 Large Sample Test

We have developed the test procedure for the null hypothesis H_0: $\mu = \mu_0$ assuming that the population is normally distributed and that σ^2 is known. In many if not most practical situations σ^2 will be unknown. Furthermore, we may not be certain that the population is well modeled by a normal distribution. In these situations if n is large (say $n > 40$) the sample standard deviation s can be substituted for σ in the test procedures with little effect. Thus, while we have given a test for the mean of a normal distribution with known σ^2, it can be easily converted into a **large-sample test procedure for unknown σ^2** that is valid regardless of the form of the distribution of the population. This large-sample test relies on the central limit theorem just as the large-sample confidence interval on μ that was presented in the previous chapter did. Exact treatment of the case where the population is normal, σ^2 is unknown, and n is small involves use of the t distribution and will be deferred until Section 9-3.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.1 Hypothesis Tests on the Mean
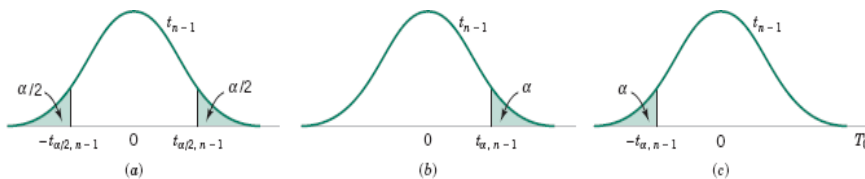
One-Sample t-Test

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic: $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

Alternative hypothesis	Rejection criteria
H_1: $\mu \neq \mu_0$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
H_1: $\mu > \mu_0$	$t_0 > t_{\alpha,n-1}$
H_1: $\mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.1 Hypothesis Tests on the Mean

Figure 9-9 The reference distribution for H_0: $\mu = \mu_0$ with critical region for (a) H_1: $\mu \neq \mu_0$, (b) H_1: $\mu > \mu_0$, and (c) H_1: $\mu < \mu_0$.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

Example 9-6

The increased availability of light materials with high strength has revolutionized the design and manufacture of golf clubs, particularly drivers. Clubs with hollow heads and very thin faces can result in much longer tee shots, especially for players of modest skills. This is due partly to the "spring-like effect" that the thin face imparts to the ball. Firing a golf ball at the head of the club and measuring the ratio of the outgoing velocity of the ball to the incoming velocity can quantify this spring-like effect. The ratio of velocities is called the coefficient of restitution of the club. An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of restitution measured. In the experiment the golf balls were fired from an air cannon so that the incoming velocity and spin rate of the ball could be precisely controlled. It is of interest to determine if there is evidence (with $\alpha = 0.05$) to support a claim that the mean coefficient of restitution exceeds 0.82. The observations follow:

0.8411	0.8191	0.8182	0.8125	0.8750
0.8580	0.8532	0.8483	0.8276	0.7983
0.8042	0.8730	0.8282	0.8359	0.8660

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

Example 9-6

The sample mean and sample standard deviation are $\bar{x} = 0.83725$ and $s = 0.02456$. The normal probability plot of the data in Fig. 9-9 supports the assumption that the coefficient of restitution is normally distributed. Since the objective of the experimenter is to demonstrate that the mean coefficient of restitution exceeds 0.82, a one-sided alternative hypothesis is appropriate.

The solution using the eight-step procedure for hypothesis testing is as follows:

1. The parameter of interest is the mean coefficient of restitution, μ.
2. H_0: $\mu = 0.82$
3. H_1: $\mu > 0.82$. We want to reject H_0 if the mean coefficient of restitution exceeds 0.82.
4. $\alpha = 0.05$
5. The test statistic is

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

6. Reject H_0 if $t_0 > t_{0.05,14} = 1.761$

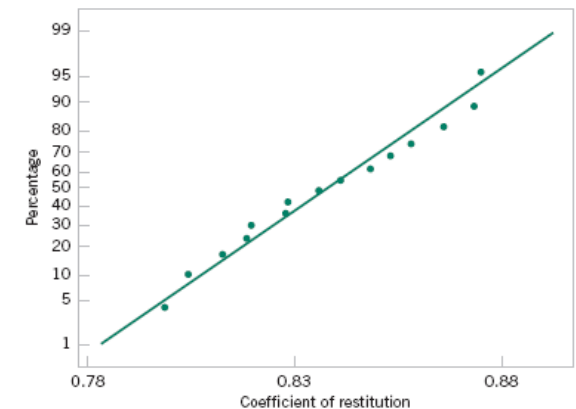## 9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

Example 9-6

Figure 9-10
Normal probability plot of the coefficient of restitution data from Example 9-6.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

Example 9-6

7. Computations: Since $\bar{x} = 0.83725$, $s = 0.02456$, $\mu_0 = 0.82$, and $n = 15$, we have

$$t_0 = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

8. Conclusions: Since $t_0 = 2.72 > 1.761$, we reject H_0 and conclude at the 0.05 level of significance that the mean coefficient of restitution exceeds 0.82.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.2 P-value for a t-Test

The P-value for a t-test is just the smallest level of significance at which the null hypothesis would be rejected.

To illustrate, consider the t-test based on 14 degrees of freedom in Example 9-6. The relevant critical values from Appendix Table IV are as follows:

Critical Value:	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
Tail Area:	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005

Notice that $t_0 = 2.72$ in Example 9-6, and that this is between two tabulated values, 2.624 and 2.977. Therefore, the P-value must be between 0.01 and 0.005. These are effectively the upper and lower bounds on the P-value.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.3 Type II Error and Choice of Sample Size

The type II error of the two-sided alternative (for example) would be

$$\beta = P\{-t_{\alpha/2,n-1} \le T_0 \le t_{\alpha/2,n-1} \mid \delta \ne 0\}$$
$$= P\{-t_{\alpha/2,n-1} \le T'_0 \le t_{\alpha/2,n-1}\}$$

where T'_0 denotes a noncentral t random variable.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

Example 9-7

Consider the golf club testing problem from Example 9-6. If the mean coefficient of restitution exceeds 0.82 by as much as 0.02, is the sample size $n = 15$ adequate to ensure that $H_0: \mu = 0.82$ will be rejected with probability at least 0.8?

To solve this problem, we will use the sample standard deviation $s = 0.02456$ to estimate σ. Then $d = |\delta|/\sigma = 0.02/0.02456 = 0.81$. By referring to the operating characteristic curves in Appendix Chart VIIg (for $\alpha = 0.05$) with $d = 0.81$ and $n = 15$, we find that $\beta = 0.10$, approximately. Thus, the probability of rejecting $H_0: \mu = 0.82$ if the true mean exceeds this by 0.02 is approximately $1 - \beta = 1 - 0.10 = 0.90$, and we conclude that a sample size of $n = 15$ is adequate to provide the desired sensitivity.

9-3.4 Likelihood Ratio Test (extra!)

Hypothesis testing is one of the most important techniques of statistical inference. Throughout this book we present many applications of hypothesis testing. While we have emphasized a heuristic development, many of these hypothesis-testing procedures can be developed using a general principle called the likelihood ratio principle. Tests developed by this method often turn out to be "best" test procedures in the sense that they minimize the type II error probability β among all tests that have the same type I error probability α.

The likelihood ratio principle is easy to illustrate. Suppose that the random variable X has a probability distribution that is described by an unknown parameter θ, say, $f(x, \theta)$. We wish to test the hypothesis H_0: θ is in Ω_0 versus H_1: θ is in Ω_1, where Ω_0 and Ω_1 are disjoint sets of values (such as H_0: $\mu \geq 0$ versus H_1: $\mu < 0$). Let $X_1, X_2, \ldots, X_n$ be the observations in a random sample. The joint distribution of these sample observations is

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \cdots \cdot f(x_n, \theta)$$

Recall from our discussion of maximum likelihood estimation in Chapter 7 that the likelihood function, say $L(\theta)$, is just this joint distribution considered as a function of the parameter θ. The likelihood ratio principle for test construction consists of the following steps:

9-3.4 Likelihood Ratio Test (extra!)

1. Find the largest value of the likelihood for any θ in Ω_0. This is done by finding the maximum likelihood estimator of θ restricted to values within Ω_0 and by substituting this value of θ back into the likelihood function. This results in a value of the likelihood function that we will call $L(\Omega_0)$.

2. Find the largest value of the likelihood for any θ in Ω_1. Call this the value of the likelihood function $L(\Omega_1)$.

3. Form the ratio

$$\lambda = \frac{L(\Omega_0)}{L(\Omega_1)}$$

This ratio λ is called the likelihood ratio test statistic.

The test procedure calls for rejecting the null hypothesis H_0 when the value of this ratio λ is small, say, whenever $\lambda < k$, where k is a constant. Thus, the likelihood ratio principle requires rejecting H_0 when $L(\Omega_1)$ is much larger than $L(\Omega_0)$, which would indicate that the sample data are more compatible with the alternative hypothesis H_1 than with the null hypothesis H_0. Usually, the constant k would be selected to give a specified value for α, the type I error probability.

9-3.4 Likelihood Ratio Test (extra!)

- Neyman-Pearson Lemma:

  Likelihood-ratio test is the **most powerful test** of a specified value α when testing two simple hypotheses.

- simple hypotheses

  H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$

The likelihood ratio principle is a very general procedure. Most of the tests presented in this book that utilize the t, chi-square, and F-distributions for testing means and variances of normal distributions are likelihood ratio tests. The principle can also be used in cases where the observations are dependent, or even in cases where their distributions are different.

9-3.4 Likelihood Ratio Test (extra!)

Suppose that we have a sample of n observations from a normal population with unknown mean μ and unknown variance σ^2, say, $X_1, X_2, \ldots, X_n$. We wish to test the hypothesis H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$. The likelihood function of the sample is

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^{n}(x_i - \mu)^2/(2\sigma^2)}$$

and the values of Ω_0 and Ω_1 are $\Omega_0 = \mu_0$ and $\Omega_1 = \{\mu: -\infty < \mu < \infty\}$, respectively. The values of μ and σ^2 that maximize L in Ω_1 are the usual maximum likelihood estimates for μ and σ^2:

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

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

Substituting these values in L, we have

$$L(\Omega_1) = \left[\frac{1}{(2\pi/n) \sum(x_i - \bar{x})^2} \right]^{n/2} e^{-(n/2)}$$

9-3.4 Likelihood Ratio Test (extra!)

To maximize L in Ω_0 we simply set $\mu = \mu_0$ and then find the value of σ^2 that maximizes L. This value is found to be

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

which gives

$$L(\Omega_0) = \left[\frac{1}{(2\pi/n) \Sigma(x_i - \mu_0)^2} \right]^{n/2} e^{-(n/2)}$$

The likelihood ratio is

$$\lambda = \frac{L(\Omega_0)}{L(\Omega_1)} = \left[\frac{\Sigma(x_i - \bar{x})^2}{\Sigma(x_i - \mu_0)^2} \right]^{n/2}$$

we may write the value of the likelihood ratio λ as

$$\lambda = \left\{ \frac{1}{1 + \left(\frac{1}{n-1}\right)\left[\frac{(\bar{x} - \mu_0)^2}{s^2/n}\right]} \right\}^n = \left\{ \frac{1}{1 + [t^2/(n-1)]} \right\}^{n/2}$$

It is easy to find the value for the constant k that would lead to rejection of the null hypothesis H_0. Since we reject H_0 if $\lambda < k$, this implies that small values of λ support the alternative hypothesis. Clearly, λ will be small when t^2 is large. So instead of specifying k we can specify a constant c and reject H_0: $\mu = \mu_0$ if $t^2 > c$. The critical values of t would be the extreme values, either positive or negative, and if we wish to control the type I error probability at α, the critical region in terms of t would be

$$t < -t_{\alpha/2, n-1} \quad \text{and} \quad t > t_{\alpha/2, n-1}$$

or, equivalently, we would reject H_0: $\mu = \mu_0$ if $t^2 > c = t_{\alpha/2, n-1}^2$. Therefore, the likelihood ratio test for H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$ is the familiar single-sample t-test.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.1 Hypothesis Test on the Variance

Suppose that we wish to test the hypothesis that the variance of a normal population σ^2 equals a specified value, say σ_0^2, or equivalently, that the standard deviation σ is equal to σ_0. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations from this population. To test

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 \neq \sigma_0^2 \quad\quad (9\text{-}26)$$

we will use the test statistic:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \quad\quad (9\text{-}27)$$

If the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ is true, the test statistic X_0^2 defined in Equation 9-27 follows the chi-square distribution with $n - 1$ degrees of freedom. This is the reference distribution for this test procedure. Therefore, we calculate χ_0^2, the value of the test statistic X_0^2, and the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ would be rejected if

$$\chi_0^2 > \chi_{\alpha/2, n-1}^2 \quad \text{or if} \quad \chi_0^2 < \chi_{1-\alpha/2, n-1}^2$$

where $\chi_{\alpha/2, n-1}^2$ and $\chi_{1-\alpha/2, n-1}^2$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n - 1$ degrees of freedom, respectively. Figure 9-10(a) shows the critical region.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.1 Hypothesis Test on the Variance

Figure 9-11 Reference distribution for the test of $H_0: \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1: \sigma^2 \neq \sigma_0^2$, (b) $H_1: \sigma^2 > \sigma_0^2$, and (c) $H_1: \sigma^2 < \sigma_0^2$.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.1 Hypothesis Test on the Variance

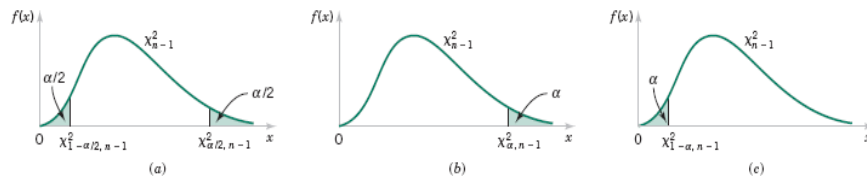The same test statistic is used for one-sided alternative hypotheses. For the one-sided hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 > \sigma_0^2 \qquad (9\text{-}28)$$

we would reject H_0 if $\chi_0^2 > \chi_{\alpha,n-1}^2$, whereas for the other one-sided hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 < \sigma_0^2 \qquad (9\text{-}29)$$

we would reject H_0 if $\chi_0^2 < \chi_{1-\alpha,n-1}^2$. The one-sided critical regions are shown in Figure 9-10(b) and (c).

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

Example 9-8

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

Using the eight-step procedure results in the following:

1. The parameter of interest is the population variance σ^2.
2. $H_0: \sigma^2 = 0.01$
3. $H_1: \sigma^2 > 0.01$
4. $\alpha = 0.05$
5. The test statistic is

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

Example 9-8

6. Reject H_0 if $\chi_0^2 > \chi_{0.05,19}^2 = 30.14$.
7. Computations:

$$\chi_0^2 = \frac{19(0.0153)}{0.01} = 29.07$$

8. Conclusions: Since $\chi_0^2 = 29.07 < \chi_{0.05,19}^2 = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.2 Type II Error and Choice of Sample Size

$$\lambda = \frac{\sigma}{\sigma_0}$$

Operating characteristic curves are provided in

- Charts VII(i) and VII(j) for the *two-sided* alternative

- Charts VII(k) and VII(l) for the *upper tail* alternative

- Charts VII(m) and VII(n) for the *lower tail* alternative

Example 9-9

Consider the bottle-filling problem from Example 9-8. If the variance of the filling process exceeds 0.01 (fluid ounces)^2, too many bottles will be underfilled. Thus, the hypothesized value of the standard deviation is $\sigma_0 = 0.10$. Suppose that if the true standard deviation of the filling process exceeds this value by 25%, we would like to detect this with probability at least 0.8. Is the sample size of $n = 20$ adequate?

To solve this problem, note that we require

$$\lambda = \frac{\sigma}{\sigma_0} = \frac{0.125}{0.10} = 1.25$$

This is the abscissa parameter for Chart VIIk. From this chart, with $n = 20$ and $\lambda = 1.25$, we find that $\beta \approx 0.6$. Therefore, there is only about a 40% chance that the null hypothesis will be rejected if the true standard deviation is really as large as $\sigma = 0.125$ fluid ounce.

To reduce the β-error, a larger sample size must be used. From the operating characteristic curve with $\beta = 0.20$ and $\lambda = 1.25$, we find that $n = 75$, approximately. Thus, if we want the test to perform as required above, the sample size must be at least 75 bottles.

9-5 Tests on a Population Proportion

9-5.1 Large-Sample Tests on a Proportion

Many engineering decision problems include hypothesis testing about p.

$$H_0: p = p_0$$
$$H_1: p \neq p_0$$

An appropriate **test statistic** is

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \qquad (9\text{-}32)$$

and reject $H_0: p = p_0$ if

$$z_0 > z_{\alpha/2} \quad \text{or} \quad z_0 < -z_{\alpha/2}$$

Example 9-10

A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?

We may solve this problem using the eight-step hypothesis-testing procedure as follows:

1. The parameter of interest is the process fraction defective p.
2. $H_0: p = 0.05$
3. $H_1: p < 0.05$
   This formulation of the problem will allow the manufacturer to make a strong claim about process capability if the null hypothesis $H_0: p = 0.05$ is rejected.
4. $\alpha = 0.05$

9-5 Tests on a Population Proportion

Example 9-10

5. The test statistic is (from Equation 9-32)

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}}$$

where $x = 4$, $n = 200$, and $p_0 = 0.05$.

6. Reject H_0: $p = 0.05$ if $z_0 < -z_{0.05} = -1.645$

7. Computations: The test statistic is

$$z_0 = \frac{4 - 200(0.05)}{\sqrt{200(0.05)(0.95)}} = -1.95$$

8. Conclusions: Since $z_0 = -1.95 < -z_{0.05} = -1.645$, we reject H_0 and conclude that the process fraction defective p is less than 0.05. The P-value for this value of the test statistic z_0 is $P = 0.0256$, which is less than $\alpha = 0.05$. We conclude that the process is capable.

9-5 Tests on a Population Proportion

Another form of the test statistic Z_0 is

$$Z_0 = \frac{X/n - p_0}{\sqrt{p_0(1 - p_0)/n}} \quad \text{or} \quad Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Think about: What are the distribution of Z_0 under H_0 and H_1?

9-5 Tests on a Population Proportion

9-5.2 Type II Error and Choice of Sample Size

For a two-sided alternative

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}34)$$

If the alternative is $p < p_0$

$$\beta = 1 - \Phi\left(\frac{p_0 - p - z_{\alpha}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}35)$$

If the alternative is $p > p_0$

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}36)$$

9-5 Tests on a Population Proportion

9-5.3 Type II Error and Choice of Sample Size

For a two-sided alternative

$$n = \left[\frac{z_{\alpha/2}\sqrt{p_0(1 - p_0)} + z_{\beta}\sqrt{p(1 - p)}}{p - p_0}\right]^2 \quad (9\text{-}37)$$

For a one-sided alternative

$$n = \left[\frac{z_{\alpha}\sqrt{p_0(1 - p_0)} + z_{\beta}\sqrt{p(1 - p)}}{p - p_0}\right]^2 \quad (9\text{-}38)$$

9-5 Tests on a Population Proportion

Example 9-11

Consider the semiconductor manufacturer from Example 9-10. Suppose that its process fall-out is really $p = 0.03$. What is the β-error for a test of process capability that uses $n = 200$ and $\alpha = 0.05$?

The β-error can be computed using Equation 9-35 as follows:

$$\beta = 1 - \Phi\left[\frac{0.05 - 0.03 - (1.645)\sqrt{0.05(0.95)/200}}{\sqrt{0.03(1 - 0.03)/200}}\right] = 1 - \Phi(-0.44) = 0.67$$

Thus, the probability is about 0.7 that the semiconductor manufacturer will fail to conclude that the process is capable if the true process fraction defective is $p = 0.03$ (3%). That is, the power of the test against this particular alternative is only about 0.3. This appears to be a large β-error (or small power), but the difference between $p = 0.05$ and $p = 0.03$ is fairly small, and the sample size $n = 200$ is not particularly large.

9-5 Tests on a Population Proportion

Example 9-11

Suppose that the semiconductor manufacturer was willing to accept a β-error as large as 0.10 if the true value of the process fraction defective was $p = 0.03$. If the manufacturer continues to use $\alpha = 0.05$, what sample size would be required?

The required sample size can be computed from Equation 9-38 as follows:

$$n = \left[\frac{1.645\sqrt{0.05(0.95)} + 1.28\sqrt{0.03(0.97)}}{0.03 - 0.05}\right]^2$$

$$\simeq 832$$

where we have used $p = 0.03$ in Equation 9-38. Note that $n = 832$ is a very large sample size. However, we are trying to detect a fairly small deviation from the null value $p_0 = 0.05$.

9-7 Testing for Goodness of Fit

- The test is based on the chi-square distribution.

- Assume there is a sample of size n from a population whose probability distribution is unknown.

- Arrange n observations in a frequency histogram.

- Let O_i be the observed frequency in the ith class interval.

- Let E_i be the expected frequency in the ith class interval.

The test statistic is

$$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (9\text{-}39)$$

which has approximately chi-square distribution with df=k-p-1.

9-7 Testing for Goodness of Fit

Example 9-12

A Poisson Distribution

The number of defects in printed circuit boards is hypothesized to follow a Poisson distribution. A random sample of $n = 60$ printed boards has been collected, and the following number of defects observed.

Number of Defects	Observed Frequency
0	32
1	15
2	9
3	4

9-7 Testing for Goodness of Fit

Example 9-12

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, $(32 \cdot 0 + 15 \cdot 1 + 9 \cdot 2 + 4 \cdot 3)/60 = 0.75$. From the Poisson distribution with parameter 0.75, we may compute p_i, the theoretical, hypothesized probability associated with the ith class interval. Since each class interval corresponds to a particular number of defects, we may find the p_i as follows:

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041$$

9-7 Testing for Goodness of Fit

Example 9-12

The expected frequencies are computed by multiplying the sample size $n = 60$ times the probabilities p_i. That is, $E_i = np_i$. The expected frequencies follow:

Number of Defects	Probability	Expected Frequency
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (or more)	0.041	2.46

9-7 Testing for Goodness of Fit

Example 9-12

Since the expected frequency in the last cell is less than 3, we combine the last two cells:

Number of Defects	Observed Frequency	Expected Frequency
0	32	28.32
1	15	21.24
2 (or more)	13	10.44

The chi-square test statistic in Equation 9-39 will have $k - p - 1 = 3 - 1 - 1 = 1$ degree of freedom, because the mean of the Poisson distribution was estimated from the data.

9-7 Testing for Goodness of Fit

Example 9-12

The eight-step hypothesis-testing procedure may now be applied, using $\alpha = 0.05$, as follows:

1. The variable of interest is the form of the distribution of defects in printed circuit boards.
2. H_0: The form of the distribution of defects is Poisson.
3. H_1: The form of the distribution of defects is not Poisson.
4. $\alpha = 0.05$
5. The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i}$$

9-7 Testing for Goodness of Fit

Example 9-12

6. Reject H_0 if $\chi_0^2 > \chi_{0.05,1}^2 = 3.84$.
7. Computations:

$$\chi_0^2 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} = 2.94$$

8. Conclusions: Since $\chi_0^2 = 2.94 < \chi_{0.05,1}^2 = 3.84$, we are unable to reject the null hypothesis that the distribution of defects in printed circuit boards is Poisson. The P-value for the test is $P = 0.0864$. (This value was computed using an HP-48 calculator.)

9-8 Contingency Table Tests

Many times, the *n* elements of a sample from a population may be classified according to two different criteria. It is then of interest to know whether the two methods of classification are statistically independent;

Table 9-2 An $r \times c$ Contingency Table

		Columns			
		1	2	...	c
Rows	1	O_{11}	O_{12}	...	O_{1c}
	2	O_{21}	O_{22}	...	O_{2c}
	⋮	⋮	⋮	⋮	⋮
	r	O_{r1}	O_{r2}	...	O_{rc}

9-8 Contingency Table Tests

We are interested in testing the hypothesis that the row-and-column methods of classification are independent. If we reject this hypothesis, we conclude there is some interaction between the two criteria of classification. The exact test procedures are difficult to obtain, but an approximate test statistic is valid for large n. Let p_{ij} be the probability that a randomly selected element falls in the ijth cell, given that the two classifications are independent. Then $p_{ij} = u_i v_j$, where u_i is the probability that a randomly selected element falls in row class i and v_j is the probability that a randomly selected element falls in column class j. Now, assuming independence, the estimators of u_i and v_j are

$$\hat{u}_i = \frac{1}{n} \sum_{j=1}^{c} O_{ij}$$

$$\hat{v}_j = \frac{1}{n} \sum_{i=1}^{r} O_{ij} \tag{9-40}$$

9-8 Contingency Table Tests

Therefore, the expected frequency of each cell is

$$E_{ij} = n\hat{u}_i\hat{v}_j = \frac{1}{n} \sum_{j=1}^{c} O_{ij} \sum_{i=1}^{r} O_{ij} \tag{9-41}$$

Then, for large n, the statistic

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{9-42}$$

has an approximate chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom if the null hypothesis is true. Therefore, we would reject the hypothesis of independence if the observed value of the test statistic χ_0^2 exceeded $\chi_{\alpha,(r-1)(c-1)}^2$.

9-8 Contingency Table Tests

Example 9-14

A company has to choose among three pension plans. Management wishes to know whether the preference for plans is independent of job classification and wants to use $\alpha = 0.05$. The opinions of a random sample of 500 employees are shown in Table 9-3.

To find the expected frequencies, we must first compute $\hat{u}_1 = (340/500) = 0.68$, $\hat{u}_2 = (160/500) = 0.32$, $\hat{v}_1 = (200/500) = 0.40$, $\hat{v}_2 = (200/500) = 0.40$, and $\hat{v}_3 = (100/500) = 0.20$. The expected frequencies may now be computed from Equation 9-41. For example, the expected number of salaried workers favoring pension plan 1 is

$$E_{11} = n\hat{u}_1\hat{v}_1 = 500(0.68)(0.40) = 136$$

The expected frequencies are shown in Table 9-4.

9-8 Contingency Table Tests

Example 9-14

Table 9-3   Observed Data for Example 9-14

Job Classification	Pension Plan			Totals
	1	2	3	
Salaried workers	160	140	40	340
Hourly workers	40	60	60	160
Totals	200	200	100	500

Table 9-4   Expected Frequencies for Example 9-14

Job Classification	Pension Plan			Totals
	1	2	3	
Salaried workers	136	136	68	340
Hourly workers	64	64	32	160
Totals	200	200	100	500

9-8 Contingency Table Tests

Example 9-14

The eight-step hypothesis-testing procedure may now be applied to this problem.

1. The variable of interest is employee preference among pension plans.
2. H_0: Preference is independent of salaried versus hourly job classification.
3. H_1: Preference is not independent of salaried versus hourly job classification.
4. $\alpha = 0.05$
5. The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

6. Since $r = 2$ and $c = 3$, the degrees of freedom for chi-square are $(r-1)(c-1) = (1)(2) = 2$, and we would reject H_0 if $\chi_0^2 > \chi_{0.05,2}^2 = 5.99$.

9-8 Contingency Table Tests

Example 9-14

7. Computations:

$$\chi_0^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(160 - 136)^2}{136} + \frac{(140 - 136)^2}{136} + \frac{(40 - 68)^2}{68} + \frac{(40 - 64)^2}{64}$$

$$+ \frac{(60 - 64)^2}{64} + \frac{(60 - 32)^2}{32} = 49.63$$

8. Conclusions: Since $\chi_0^2 = 49.63 > \chi_{0.05,2}^2 = 5.99$, we reject the hypothesis of independence and conclude that the preference for pension plans is not independent of job classification. The P-value for $\chi_0^2 = 49.63$ is $P = 1.671 \times 10^{-11}$. (This value was computed using a hand-held calculator.) Further analysis would be necessary to explore the nature of the association between these factors. It might be helpful to examine the table of observed minus expected frequencies.

10

Statistical Inference for Two Samples

CHAPTER OUTLINE

10-1 Introduction

The previous chapter presented hypothesis tests and confidence intervals for a single population parameter (the mean μ, the variance σ^2, or a proportion p). This chapter extends those results to the case of two independent populations.

The general situation is shown in Fig. 10-1. Population 1 has mean μ_1 and variance σ_1^2, while population 2 has mean μ_2 and variance σ_2^2. Inferences will be based on two random samples of sizes n1 and n2, respectively. That is, X11, X12, p , X_{1n_1} is a random sample of n1 observations from population 1, and X21, X22, p , X_{2n_2} is a random sample of n2 observations from population 2. Most of the practical applications of the procedures in this chapter arise in the context of **simple comparative experiments** in which the objective is to study the difference in the parameters of the two populations.

10-1 Introduction
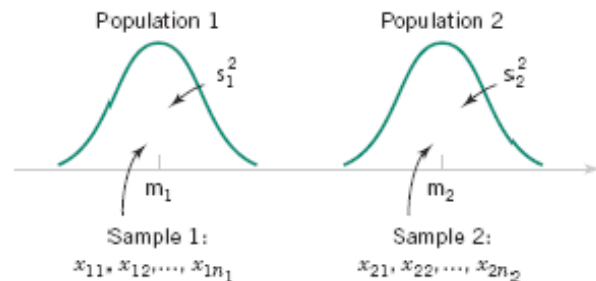
Figure 10-1 Two independent populations.

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

Assumptions

1. $X_{11}, X_{12}, \ldots, X_{1n_1}$ is a random sample from population 1.
2. $X_{21}, X_{22}, \ldots, X_{2n_2}$ 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.

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

$$V(\overline{X}_1 - \overline{X}_2) = V(\overline{X}_1) + V(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

The quantity

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \qquad (10\text{-}1)$$

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

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

10-2.1 Hypothesis Tests for a Difference in Means, Variances Known

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

Test statistic: $Z_0 = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ $\qquad (10\text{-}2)$

Alternative Hypotheses	Rejection Criterion
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$z_0 > z_\alpha$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$z_0 < -z_\alpha$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

Example 10-1

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 $\overline{x}_1 = 121$ minutes and $\overline{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

We apply the eight-step procedure to this problem as follows:

1. The quantity of interest is the difference in mean drying times, $\mu_1 - \mu_2$, and $\Delta_0 = 0$.
2. $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$.
3. $H_1: \mu_1 > \mu_2$. We want to reject H_0 if the new ingredient reduces mean drying time.
4. $\alpha = 0.05$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

Example 10-1

5. The test statistic is

$$z_0 = \frac{\overline{x}_1 - \overline{x}_2 - 0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where $\sigma_1^2 = \sigma_2^2 = (8)^2 = 64$ and $n_1 = n_2 = 10$.

6. Reject $H_0: \mu_1 = \mu_2$ if $z_0 > 1.645 = z_{0.05}$.
7. Computations: Since $\overline{x}_1 = 121$ minutes and $\overline{x}_2 = 112$ minutes, the test statistic is

$$z_0 = \frac{121 - 112}{\sqrt{\dfrac{(8)^2}{10} + \dfrac{(8)^2}{10}}} = 2.52$$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

8. Conclusion: Since $z_0 = 2.52 > 1.645$, 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 $\alpha \geq 0.0059$.

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

10-2.2 Type II Error and Choice of Sample Size

- Use of Operating Characteristic Curves
- Chart VII(a)-(d)
- Identical to 9-2.2 except

$$d = \frac{|\mu_1 - \mu_2 - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{|\Delta - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

10-2.3 Confidence Interval on a Difference in Means, Variances Known

Definition

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2 from two independent normal populations with known variances σ_1^2 and σ_2^2, respectively, a $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10\text{-}7)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

Example 10-4

Tensile strength tests were performed on two different grades of aluminum spars used in manufacturing the wing of a commercial transport aircraft. From past experience with the spar manufacturing process and the testing procedure, the standard deviations of tensile strengths are assumed to be known. The data obtained are as follows: $n_1 = 10$, $\bar{x}_1 = 87.6$, $\sigma_1 = 1$, $n_2 = 12$, $\bar{x}_2 = 74.5$, and $\sigma_2 = 1.5$. If μ_1 and μ_2 denote the true mean tensile strengths for the two grades of spars, we may find a 90% confidence interval on the difference in mean strength $\mu_1 - \mu_2$ as follows:

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$87.6 - 74.5 - 1.645\sqrt{\frac{(1)^2}{10} + \frac{(1.5)^3}{12}} \leq \mu_1 - \mu_2 \leq 87.6 - 74.5 + 1.645\sqrt{\frac{(1^2)}{10} + \frac{(1.5)^2}{12}}$$

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

Example 10-4

Therefore, the 90% confidence interval on the difference in mean tensile strength (in kilograms per square millimeter) is

$$12.22 \le \mu_1 - \mu_2 \le 13.98 \text{ (in kilograms per square millimeter)}$$

Notice that the confidence interval does not include zero, implying that the mean strength of aluminum grade 1 (μ_1) exceeds the mean strength of aluminum grade 2 (μ_2). In fact, we can state that we are 90% confident that the mean tensile strength of aluminum grade 1 exceeds that of aluminum grade 2 by between 12.22 and 13.98 kilograms per square millimeter.

10-2 Inference for a Difference in Means of Two Normal Distributions, Variances Known

One-Sided Confidence Bounds

Upper Confidence Bound

$$\mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (10\text{-}9)$$

Lower Confidence Bound

$$\bar{x}_1 - \bar{x}_2 - z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \qquad (10\text{-}10)$$

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

10-3.1 Hypotheses Tests for a Difference in Means, Variances Unknown

We wish to test:

$$H_0: \mu_1 - \mu_2 = \Delta_0$$
$$H_1: \mu_1 - \mu_2 \ne \Delta_0$$

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

Case 2: $\quad \sigma_1^2 \ne \sigma_2^2$

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

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

The pooled estimator of σ^2:

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

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (10\text{-}12)$$

The pooled estimator is an **unbiased** estimator of σ^2

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

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

Given the assumptions of this section, the quantity

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad (10\text{-}13)$$

has a t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Definition: The Two-Sample or Pooled t-Test*

Null hypothesis:	$H_0: \mu_1 - \mu_2 = \Delta_0$
Test statistic:	$T_0 = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ (10-14)

Alternative Hypothesis	Rejection Criterion
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$t_0 > t_{\alpha/2, n_1+n_2-2}$ or $t_0 < -t_{\alpha/2, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$t_0 > t_{\alpha, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$t_0 < -t_{\alpha, n_1+n_2-2}$

Example 10-5

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. A test is run in the pilot plant and results in the data shown in Table 10-1. Is there any difference between the mean yields? Use $\alpha = 0.05$, and assume equal variances.

Table 10-1   Catalyst Yield Data, Example 10-5

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
	$\overline{x}_1 = 92.255$	$\overline{x}_2 = 92.733$
	$s_1 = 2.39$	$s_2 = 2.98$

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown
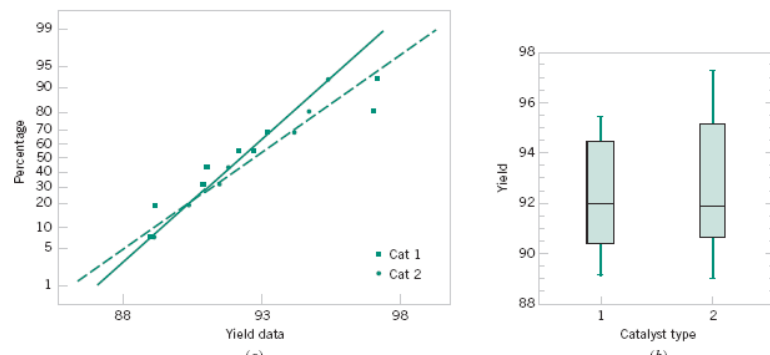
Figure 10-2 Normal probability plot and comparative box plot for the catalyst yield data in Example 10-5. (a) Normal probability plot, (b) Box plots.

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-5

The solution using the eight-step hypothesis-testing procedure is as follows:

1. The parameters of interest are μ_1 and μ_2, the mean process yield using catalysts 1 and 2, respectively, and we want to know if $\mu_1 - \mu_2 = 0$.

2. H_0: $\mu_1 - \mu_2 = 0$, or H_0: $\mu_1 = \mu_2$

3. H_1: $\mu_1 \neq \mu_2$

4. $\alpha = 0.05$

5. The test statistic is

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

6. Reject H_0 if $t_0 > t_{0.025,14} = 2.145$ or if $t_0 < -t_{0.025,14} = -2.145$.

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-5

7. Computations: From Table 10-1 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

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

8. Conclusions: Since $-2.145 < t_0 = -0.35 < 2.145$, 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.

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

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

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad (10\text{-}15)$$

is distributed approximately as t with degrees of freedom given by

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} \qquad (10\text{-}16)$$

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown
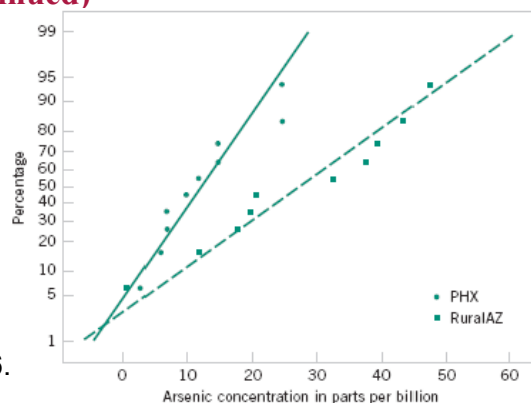
Example 10-6

Arsenic concentration in public drinking water supplies is a potential health risk. An article in the *Arizona Republic* (Sunday, May 27, 2001) reported drinking water arsenic concentrations in parts per billion (ppb) for 10 metropolitan Phoenix communities and 10 communities in rural Arizona. The data follow:

Metro Phoenix ($\bar{x}_1 = 12.5$, $s_1 = 7.63$)	Rural Arizona ($\bar{x}_2 = 27.5$, $s_2 = 15.3$)
Phoenix, 3	Rimrock, 48
Chandler, 7	Goodyear, 44
Gilbert, 25	New River, 40
Glendale, 10	Apachie Junction, 38
Mesa, 15	Buckeye, 33
Paradise Valley, 6	Nogales, 21
Peoria, 12	Black Canyon City, 20
Scottsdale, 25	Sedona, 12
Tempe, 15	Payson, 1
Sun City, 7	Casa Grande, 18

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-6 (Continued)

We wish to determine it there is any difference in mean arsenic concentrations between metropolitan Phoenix communities and communities in rural Arizona. Figure 10-3 shows a normal probability plot for the two samples of arsenic concentration. The assumption of normality appears quite reasonable, but since the slopes of the two straight lines are very different, it is unlikely that the population variances are the same.

Applying the eight-step procedure gives the following:

1. The parameters of interest are the mean arsenic concentrations for the two geographic regions, say, μ_1 and μ_2, and we are interested in determining whether $\mu_1 - \mu_2 = 0$.
2. $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$
3. $H_1: \mu_1 \neq \mu_2$
4. $\alpha = 0.05$ (say)

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-6 (Continued)

Figure 10-3 Normal probability plot of the arsenic concentration data from Example 10-6.

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-6 (Continued)

5. The test statistic is

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

6. The degrees of freedom on t_0^* are found from Equation 10-16 as

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{\left[\dfrac{(7.63)^2}{10} + \dfrac{(15.3)^2}{10}\right]^2}{\dfrac{[(7.63)^2/10]^2}{9} + \dfrac{[(15.3)^2/10]^2}{9}} = 13.2 \approx 13$$

Therefore, using $\alpha = 0.05$, we would reject $H_0: \mu_1 = \mu_2$ if $t_0^* > t_{0.025,13} = 2.160$ or if $t_0^* < -t_{0.025,13} = -2.160$

10-3 Inference for a Difference in Means of Two Normal Distributions, Variances Unknown

Example 10-6 (Continued)

7. Computations: Using the sample data we find

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{12.5 - 27.5}{\sqrt{\frac{(7.63)^2}{10} + \frac{(15.3)^2}{10}}} = -2.77$$

8. Conclusions: Because $t_0^* = -2.77 < t_{0.025,13} = -2.160$, we reject the null hypothesis. Therefore, there is evidence to conclude that mean arsenic concentration in the drinking water in rural Arizona is different from the mean arsenic concentration in metropolitan Phoenix drinking water. Furthermore, the mean arsenic concentration is higher in rural Arizona communities. The P-value for this test is approximately $P = 0.016$.

10-3.3 Confidence Interval on the Difference in Means, Variance 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 sample 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

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (10\text{-}19)$$

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.

10-3.3 Confidence Interval on the Difference in Means, Variance 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, an approximate $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (10\text{-}20)$$

where v is given by Equation 10-16 and $t_{\alpha/2,v}$ is the upper $\alpha/2$ percentage point of the t distribution with v degrees of freedom.

10-4 Paired t-Test

• A special case of the two-sample t-tests of Section 10-3 occurs when the observations on the two populations of interest are collected in **pairs**.

• Each pair of observations, say (X_{1j}, X_{2j}), is taken under homogeneous conditions, but these conditions may change from one pair to another.

• The test procedure consists of analyzing the **differences** between two observations from each pair.

10-4 Paired t-Test

The Paired t-Test

Null hypothesis: $H_0: \mu_D = \Delta_0$

Test statistic: $T_0 = \dfrac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$ 　　　　　　　(10-22)

Alternative Hypothesis	Rejection Region
$H_1: \mu_D \neq \Delta_0$	$t_0 > t_{\alpha/2,\,n-1}$　or　$t_0 < -t_{\alpha/2,\,n-1}$
$H_1: \mu_D > \Delta_0$	$t_0 > t_{\alpha,\,n-1}$
$H_1: \mu_D < \Delta_0$	$t_0 < -t_{\alpha,\,n-1}$

In Equation 10-22, $\overline{D}$ is the sample average of the n differences $D_1, D_2, \ldots, D_n$, and S_D is the sample standard deviation of these differences.

10-4 Paired t-Test

Example 10-9

An article in the *Journal of Strain Analysis* (1983, Vol. 18, No. 2) compares several methods for predicting the shear strength for steel plate girders. Data for two of these methods, the Karlsruhe and Lehigh procedures, when applied to nine specific girders, are shown in Table 10-2. We wish to determine whether there is any difference (on the average) between the two methods.

Table 10-2　Strength Predictions for Nine Steel Plate Girders (Predicted Load/Observed Load)

Girder	Karlsruhe Method	Lehigh Method	Difference d_j
S1/1	1.186	1.061	0.119
S2/1	1.151	0.992	0.159
S3/1	1.322	1.063	0.259
S4/1	1.339	1.062	0.277
S5/1	1.200	1.065	0.138
S2/1	1.402	1.178	0.224
S2/2	1.365	1.037	0.328
S2/3	1.537	1.086	0.451
S2/4	1.559	1.052	0.507

10-4 Paired t-Test

Example 10-9

The eight-step procedure is applied as follows:

1. The parameter of interest is the difference in mean shear strength between the two methods, say, $\mu_D = \mu_1 - \mu_2 = 0$.

2. $H_0: \mu_D = 0$

3. $H_1: \mu_D \neq 0$

4. $\alpha = 0.05$

5. The test statistic is

$$t_0 = \frac{\overline{d}}{s_D/\sqrt{n}}$$

6. Reject H_0 if $t_0 > t_{0.025,8} = 2.306$ or if $t_0 < -t_{0.025,8} = -2.306$.

10-4 Paired t-Test

Example 10-9

7. Computations: The sample average and standard deviation of the differences d_j are $\overline{d} = 0.2736$ and $s_D = 0.1356$, so the test statistic is

$$t_0 = \frac{\overline{d}}{s_D/\sqrt{n}} = \frac{0.2736}{0.1356/\sqrt{9}} = 6.05$$

8. Conclusions: Since $t_0 = 6.05 > 2.306$, we conclude that the strength prediction methods yield different results. Specifically, the data indicate that the Karlsruhe method produces, on the average, higher strength predictions than does the Lehigh method. The P-value for $t_0 = 6.05$ is $P = 0.0002$, so the test statistic is well into the critical region.

10-4 Paired *t*-Test

Paired Versus Unpaired Comparisons

$$T_0 = \frac{\bar{D} - \Delta_0}{S_D / \sqrt{n}}$$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{S_p \sqrt{\frac{1}{n} + \frac{1}{n}}}$$

So how do we decide to conduct the experiment? Should we pair the observations or not? Although there is no general answer to this question, we can give some guidelines based on the above discussion.

1. If the experimental units are relatively homogeneous (small σ) and the correlation within pairs is small, the gain in precision attributable to pairing will be offset by the loss of degrees of freedom, so an independent-sample experiment should be used.

2. If the experimental units are relatively heterogeneous (large σ) and there is large positive correlation within pairs, the paired experiment should be used. Typically, this case occurs when the experimental units are the *same* for both treatments; as in Example 10-9, the same girders were used to test the two methods.

10-4 Paired *t*-Test

A Confidence Interval for μ_D

If $\bar{d}$ and s_D are the sample mean and standard deviation of the difference of n random pairs of normally distributed measurements, a $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{\alpha/2,n-1}s_D/\sqrt{n} \leq \mu_D \leq \bar{d} + t_{\alpha/2,n-1}s_D/\sqrt{n} \qquad (10\text{-}23)$$

where $t_{\alpha/2,n-1}$ is the upper $\alpha/2\%$ point of the t-distribution with $n - 1$ degrees of freedom.

11

Simple Linear
Regression and
Correlation

10-5.1 The F distribution

11-1 Empirical Models

• Many problems in engineering and science involve exploring the relationships between two or more variables.

• **Regression analysis** is a statistical technique that is very useful for these types of problems.

• For example, in a chemical process, suppose that the yield of the product is related to the process-operating temperature.

• Regression analysis can be used to build a model to predict yield at a given temperature level.

11-1 Empirical Models

Table 11-1   Oxygen and Hydrocarbon Levels

Observation Number	Hydrocarbon Level $x(\%)$	Purity $y(\%)$
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52
13	0.98	90.56
14	1.01	89.54
15	1.11	89.85
16	1.20	90.39
17	1.26	93.25
18	1.32	93.41
19	1.43	94.98
20	0.95	87.33
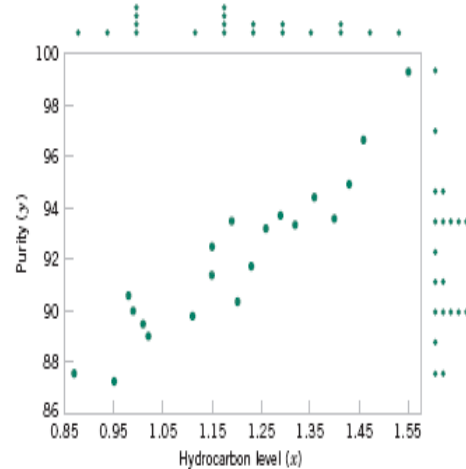
11-1 Empirical Models

Based on the scatter diagram, it is probably reasonable to assume that the mean of the random variable Y is related to x by the following straight-line relationship:

$$E(Y|x) = \mu_{Y|x} = \beta_0 + \beta_1 x$$

where the slope and intercept of the line are called **regression coefficients**.

The **simple linear regression model** is given by

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where ϵ is the random error term.

11-1 Empirical Models

We think of the regression model as an empirical model.

Suppose that the mean and variance of ϵ are 0 and σ^2, respectively, then

$$E(Y|x) = E(\beta_0 + \beta_1 x + \epsilon) = \beta_0 + \beta_1 x + E(\epsilon) = \beta_0 + \beta_1 x$$

The variance of Y given x is

$$V(Y|x) = V(\beta_0 + \beta_1 x + \epsilon) = V(\beta_0 + \beta_1 x) + V(\epsilon) = 0 + \sigma^2 = \sigma^2$$

11-1 Empirical Models

• The true regression model is a line of mean values:
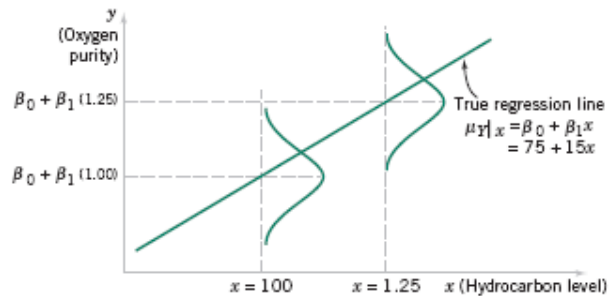
$$\mu_{Y|x} = \beta_0 + \beta_1 x$$

where β_1 can be interpreted as the change in the mean of Y for a unit change in x.

• Also, the variability of Y at a particular value of x is determined by the error variance, σ^2.

• This implies there is a distribution of Y-values at each x and that the variance of this distribution is the same at each x.

11-1 Empirical Models

Figure 11-2 The distribution of *Y* for a given value of *x* for the oxygen purity-hydrocarbon data.

11-2 Simple Linear Regression

• The case of **simple linear regression** considers a single **regressor** or **predictor** *x* and a **dependent** or **response variable** *Y*.

• The expected value of *Y* at each level of *x* is a random variable:

$$E(Y|x) = \beta_0 + \beta_1 x$$

• We assume that each observation, *Y*, can be described by the model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

11-2 Simple Linear Regression

• Suppose that we have *n* pairs of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Figure 11-3
Deviations of the data from the estimated regression model.

11-2 Simple Linear Regression

• The **method of least squares** is used to estimate the parameters, β_0 and β_1 by minimizing the sum of the squares of the vertical deviations in Figure 11-3.
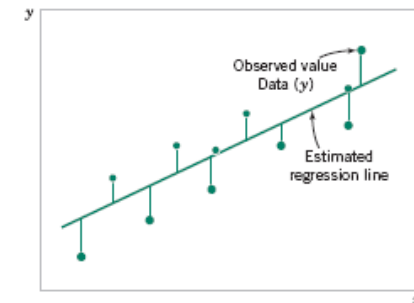
Figure 11-3
Deviations of the data from the estimated regression model.

11-2 Simple Linear Regression

• Using Equation 11-2, the *n* observations in the sample can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

• The sum of the squares of the deviations of the observations from the true regression line is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

11-2 Simple Linear Regression

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

The least squares estimators of β_0 and β_1, say, $\hat{\beta}_0$ and $\hat{\beta}_1$, must satisfy

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\left. \frac{\partial L}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

11-2 Simple Linear Regression

Simplifying these two equations yields

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \qquad (11\text{-}6)$$

Equations 11-6 are called the **least squares normal equations.** The solution to the normal equations results in the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

11-2 Simple Linear Regression

The **least squares estimates** of the intercept and slope in the simple linear regression model are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (11\text{-}7)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \dfrac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (11\text{-}8)$$

where $\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$.

The **fitted** or **estimated regression line** is therefore

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{11-9}$$

Note that each pair of observations satisfies the relationship

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \qquad i = 1, 2, \ldots, n$$

where $e_i = y_i - \hat{y}_i$ is called the **residual.** The residual describes the error in the fit of the model to the ith observation y_i. Later in this chapter we will use the residuals to provide information about the adequacy of the fitted model.

Notation

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

Example 11-1

We will fit a simple linear regression model to the oxygen purity data in Table 11-1. The following quantities may be computed:

$$n = 20 \quad \sum_{i=1}^{20} x_i = 23.92 \quad \sum_{i=1}^{20} y_i = 1{,}843.21 \quad \bar{x} = 1.1960 \quad \bar{y} = 92.1605$$

$$\sum_{i=1}^{20} y_i^2 = 170{,}044.5321 \quad \sum_{i=1}^{20} x_i^2 = 29.2892 \quad \sum_{i=1}^{20} x_i y_i = 2{,}214.6566$$

$$S_{xx} = \sum_{i=1}^{20} x_i^2 - \frac{\left(\sum_{i=1}^{20} x_i\right)^2}{20} = 29.2892 - \frac{(23.92)^2}{20} = 0.68088$$

and

$$S_{xy} = \sum_{i=1}^{20} x_i y_i - \frac{\left(\sum_{i=1}^{20} x_i\right)\left(\sum_{i=1}^{20} y_i\right)}{20} = 2{,}214.6566 - \frac{(23.92)(1{,}843.21)}{20} = 10.17744$$

Example 11-1

Therefore, the least squares estimates of the slope and intercept are

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{10.17744}{0.68088} = 14.94748$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 92.1605 - (14.94748)1.196 = 74.28331$$

The fitted simple linear regression model (with the coefficients reported to three decimal places) is

$$\hat{y} = 74.283 + 14.947x$$

This model is plotted in Fig. 11-4, along with the sample data.

11-2 Simple Linear Regression

Example 11-1

Figure 11-4 Scatter plot of oxygen purity y versus hydrocarbon level x and regression model $\hat{y} = 74.20 + 14.97x$.
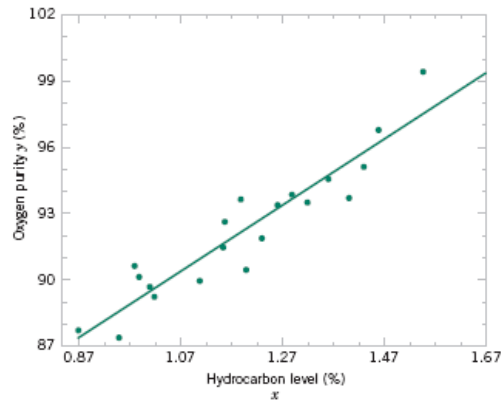
Table 11-2   Minitab Output for the Oxygen Purity Data in Example 11-1

Regression Analysis

The regression equation is

Purity = 74.3 + 14.9 HC Level

Predictor	Coef	SE Coef	T	P
Constant	74.283 ◄ $\hat{\beta}_0$	1.593	46.62	0.000
HC Level	14.947 ◄ $\hat{\beta}_1$	1.317	11.35	0.000

S = 1.087   R-Sq = 87.7%   R-Sq (adj) = 87.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	152.13	152.13	128.86	0.000
Residual Error	18	21.25 ◄ SS_E	1.18 ◄ $\hat{\sigma}^2$		
Total	19	173.38			

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	89.231	0.354	(88.486, 89.975)	(86.830, 91.632)

Values of Predictors for New Observations

New Obs	HC Level
1	1.00

11-2 Simple Linear Regression

Estimating σ^2

The error sum of squares is

$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

It can be shown that the expected value of the error sum of squares is $E(SS_E) = (n-2)\sigma^2$.

11-2 Simple Linear Regression

Estimating σ^2

An **unbiased estimator** of σ^2 is

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \qquad (11\text{-}13)$$

where SS_E can be easily computed using

$$SS_E = SS_T - \hat{\beta}_1 S_{xy} \qquad (11\text{-}14)$$

where $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = S_{yy}$

11-3 Properties of the Least Squares Estimators

• Slope Properties

$$E(\hat{\beta}_1) = \beta_1 \qquad V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

• Intercept Properties

$$E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad V(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]$$

11-4 Hypothesis Tests in Simple Linear Regression

11-4.1 Use of t-Tests

Suppose we wish to test

$$H_0: \beta_1 = \beta_{1,0}$$
$$H_1: \beta_1 \neq \beta_{1,0}$$

An appropriate test statistic would be

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2/S_{xx}}}$$

11-4 Hypothesis Tests in Simple Linear Regression

Assumptions:

To test hypotheses about the slope and intercept of the regression model, we must make the additional assumption that the error component in the model, ε, is normally distributed.

Thus, the complete assumptions are that the errors are normally and independently distributed with mean zero and variance σ^2, abbreviated NID(0, σ^2).

11-4 Hypothesis Tests in Simple Linear Regression

11-4.1 Use of t-Tests

The test statistic could also be written as:

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

We would reject the null hypothesis if

$$|t_0| > t_{\alpha/2, n-2}$$

11-4.1 Use of *t*-Tests

Suppose we wish to test

$$H_0: \beta_0 = \beta_{0,0}$$
$$H_1: \beta_0 \neq \beta_{0,0}$$

An appropriate test statistic would be

$$T_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} = \frac{\hat{\beta}_0 - \beta_{0,0}}{se(\hat{\beta}_0)}$$

We would reject the null hypothesis if

$$|t_0| > t_{\alpha/2, n-2}$$

11-4 Hypothesis Tests in Simple Linear Regression

11-4.1 Use of *t*-Tests

An important special case of the hypotheses of Equation 11-18 is

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

These hypotheses relate to the **significance of regression**.

Failure to reject H_0 is equivalent to concluding that there is no linear relationship between x and Y.

11-4 Hypothesis Tests in Simple Linear Regression
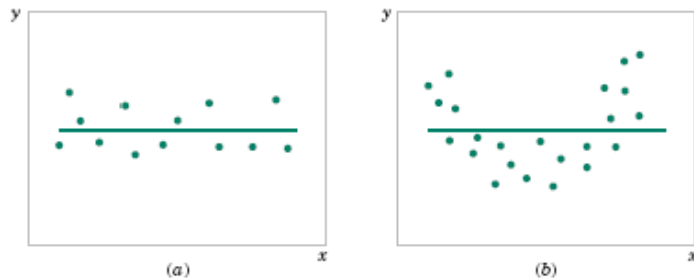
Figure 11-5 The hypothesis $H_0: \beta_1 = 0$ is not rejected.

11-4 Hypothesis Tests in Simple Linear Regression
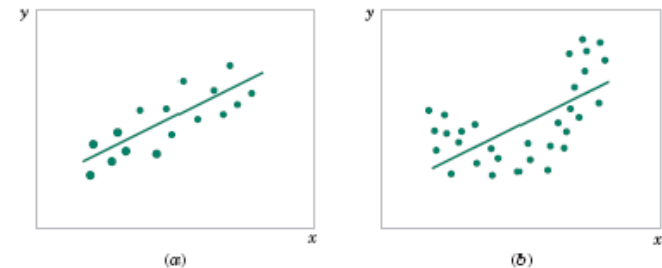
Figure 11-6 The hypothesis $H_0: \beta_1 = 0$ is rejected.

11-4 Hypothesis Tests in Simple Linear Regression

Example 11-2

We will test for significance of regression using the model for the oxygen purity data from Example 11-1. The hypotheses are

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

and we will use $\alpha = 0.01$. From Example 11-1 and Table 11-2 we have

$$\hat{\beta}_1 = 14.97 \quad n = 20, \quad S_{xx} = 0.68088, \quad \hat{\sigma}^2 = 1.18$$

so the t-statistic in Equation 10-20 becomes

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{14.947}{\sqrt{1.18/0.68088}} = 11.35$$

Since the reference value of t is $t_{0.005,18} = 2.88$, the value of the test statistic is very far into the critical region, implying that $H_0: \beta_1 = 0$ should be rejected. The P-value for this test is $P \simeq 1.23 \times 10^{-9}$. This was obtained manually with a calculator.

R commands and outputs

```
> dat=read.table("table11-1.txt", h=T)
> g=lm(y~x, dat)
> summary(g)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   74.283      1.593   46.62  < 2e-16 ***
x             14.947      1.317   11.35 1.23e-09 ***

Residual standard error: 1.087 on 18 degrees of freedom
Multiple R-Squared: 0.8774,   Adjusted R-squared: 0.8706
F-statistic: 128.9 on 1 and 18 DF,  p-value: 1.227e-09

> anova(g)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value      Pr(>F)
x          1 152.127 152.127  128.86 1.227e-09 ***
Residuals 18  21.250   1.181
```

10-5.1 The *F* Distribution

Let W and Y be independent chi-square random variables with u and v degrees of freedom, respectively. Then the ratio

$$F = \frac{W/u}{Y/v} \qquad (10\text{-}26)$$

has the probability density function

$$f(x) = \frac{\Gamma\left(\dfrac{u+v}{2}\right)\left(\dfrac{u}{v}\right)^{u/2} x^{(u/2)-1}}{\Gamma\left(\dfrac{u}{2}\right)\Gamma\left(\dfrac{v}{2}\right)\left[\left(\dfrac{u}{v}\right)x + 1\right]^{(u+v)/2}}, \qquad 0 < x < \infty \qquad (10\text{-}27)$$

and is said to follow the F distribution with u degrees of freedom in the numerator and v degrees of freedom in the denominator. It is usually abbreviated as $F_{u,v}$.
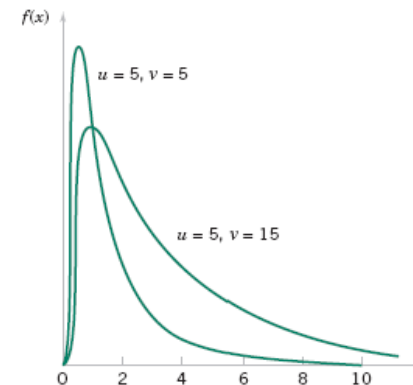
10-5.1 The *F* Distribution

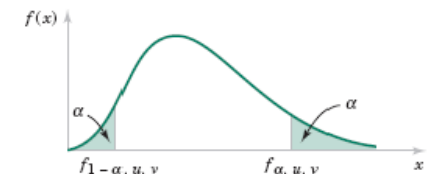Figure 10-4 Probability density functions of two F distributions.

Figure 10-5 Upper and lower percentage points of the F distribution.

10-5.1 The *F* Distribution

The lower-tail percentage points $f_{1-\alpha,u,v}$ can be found as follows.

$$f_{1-\alpha,u,v} = \frac{1}{f_{\alpha,v,u}} \qquad (10\text{-}28)$$

11-4 Hypothesis Tests in Simple Linear Regression

11-4.2 Analysis of Variance Approach to Test Significance of Regression

The analysis of variance identity is

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (11\text{-}24)$$

Symbolically,

$$SS_T = SS_R + SS_E \qquad (11\text{-}25)$$

11-4 Hypothesis Tests in Simple Linear Regression

11-4.2 Analysis of Variance Approach to Test Significance of Regression

If the null hypothesis, $H_0: \beta_1 = 0$ is true, the statistic

$$F_0 = \frac{SS_R/1}{SS_E/(n-2)} = \frac{MS_R}{MS_E} \qquad (11\text{-}26)$$

follows the $F_{1,n-2}$ distribution and we would reject if $f_0 > f_{\alpha,1,n-2}$.

11-4 Hypothesis Tests in Simple Linear Regression

11-4.2 Analysis of Variance Approach to Test Significance of Regression

The quantities, MS_R and MS_E are called **mean squares**.

Analysis of variance table:

Table 11-3 Analysis of Variance for Testing Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_E
Error	$SS_E = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_E	
Total	SS_T	$n-1$		

Note that $MS_E = \hat{\sigma}^2$.

11-4 Hypothesis Tests in Simple Linear Regression

Example 11-3

We will use the analysis of variance approach to test for significance of regression using the oxygen purity data model from Example 11-1. Recall that $SS_T = 173.38$, $\hat{\beta}_1 = 14.947$, $S_{xy} = 10.17744$, and $n = 20$. The regression sum of squares is

$$SS_R = \hat{\beta}_1 S_{xy} = (14.947)10.17744 = 152.13$$

and the error sum of squares is

$$SS_E = SS_T - SS_R = 173.38 - 152.13 = 21.25$$

The analysis of variance for testing $H_0: \beta_1 = 0$ is summarized in the Minitab output in Table 11-2. The test statistic is $f_0 = MS_R/MS_E = 152.13/1.18 = 128.86$, for which we find that the P-value is $P \simeq 1.23 \times 10^{-9}$, so we conclude that β_1 is not zero.

There are frequently minor differences in terminology among computer packages. For example, sometimes the regression sum of squares is called the "model" sum of squares, and the error sum of squares is called the "residual" sum of squares.

11-4 Hypothesis Tests in Simple Linear Regression

Note that the analysis of variance procedure for testing for significance of regression is equivalent to the t-test in Section 11-5.1. That is, either procedure will lead to the same conclusions. This is easy to demonstrate by starting with the t-test statistic in Equation 11-19 with $\beta_{1,0} = 0$, say

$$T_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \qquad (11\text{-}27)$$

Squaring both sides of Equation 11-27 and using the fact that $\hat{\sigma}^2 = MS_E$ results in

$$T_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{MS_E} = \frac{\hat{\beta}_1 S_{xY}}{MS_E} = \frac{MS_R}{MS_E} \qquad (11\text{-}28)$$

Note that T_0^2 in Equation 11-28 is identical to F_0 in Equation 11-26 It is true, in general, that the square of a t random variable with v degrees of freedom is an F random variable, with one and v degrees of freedom in the numerator and denominator, respectively. Thus, the test using T_0 is equivalent to the test based on F_0. Note, however, that the t-test is somewhat more flexible in that it would allow testing against a one-sided alternative hypothesis, while the F-test is restricted to a two-sided alternative.

11-5 Confidence Intervals

11-5.1 Confidence Intervals on the Slope and Intercept

Definition

Under the assumption that the observations are normally and independently distributed, a $100(1 - \alpha)\%$ confidence interval on the slope β_1 in simple linear regression is

$$\hat{\beta}_1 - t_{\alpha/2, n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \qquad (11\text{-}29)$$

Similarly, a $100(1 - \alpha)\%$ confidence interval on the intercept β_0 is

$$\hat{\beta}_0 - t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$$
$$\leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} \qquad (11\text{-}30)$$

11-6 Confidence Intervals

Example 11-4

We will find a 95% confidence interval on the slope of the regression line using the data in Example 11-1. Recall that $\hat{\beta}_1 = 14.947$, $S_{xx} = 0.68088$, and $\hat{\sigma}^2 = 1.18$ (see Table 11-2). Then, from Equation 10-31 we find

$$\hat{\beta}_1 - t_{0.025,18}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{0.025,18}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

or

$$14.947 - 2.101\sqrt{\frac{1.18}{0.68088}} \leq \beta_1 \leq 14.947 + 2.101\sqrt{\frac{1.18}{0.68088}}$$

This simplifies to

$$12.197 \leq \beta_1 \leq 17.697$$

11-5 Confidence Intervals

11-5.2 Confidence Interval on the Mean Response

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Definition

A $100(1 - \alpha)\%$ **confidence interval** about the **mean response** at the value of $x = x_0$, say $\mu_{Y|x_0}$, is given by

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]}$$

$$\leq \mu_{Y|x_0} \leq \hat{\mu}_{Y|x_0} + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \qquad (11\text{-}31)$$

where $\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$ is computed from the fitted regression model.

11-5 Confidence Intervals

Example 11-5

We will construct a 95% confidence interval about the mean response for the data in Example 11-1. The fitted model is $\hat{\mu}_{Y|x_0} = 74.283 + 14.947x_0$, and the 95% confidence interval on $\mu_{Y|x_0}$ is found from Equation 11-31 as

$$\hat{\mu}_{Y|x_0} \pm 2.101\sqrt{1.18\left[\frac{1}{20} + \frac{(x_0 - 1.1960)^2}{0.68088}\right]}$$

Suppose that we are interested in predicting mean oxygen purity when $x_0 = 1.00\%$. Then

$$\hat{\mu}_{Y|x_{1.00}} = 74.283 + 14.947(1.00) = 89.23$$

and the 95% confidence interval is

$$\left\{89.23 \pm 2.101\sqrt{1.18\left[\frac{1}{20} + \frac{(1.00 - 1.1960)^2}{0.68088}\right]}\right\}$$

11-5 Confidence Intervals

or

$$89.23 \pm 0.75$$

Therefore, the 95% confidence interval on $\mu_{Y|1.00}$ is

$$88.48 \leq \mu_{Y|1.00} \leq 89.98$$

Minitab will also perform these calculations. Refer to Table 11-2. The predicted value of y at $x = 1.00$ is shown along with the 95% CI on the mean of y at this level of x.

By repeating these calculations for several different values for x_0 we can obtain confidence limits for each corresponding value of $\mu_{Y|x_0}$. Figure 11-7 displays the scatter diagram with the fitted model and the corresponding 95% confidence limits plotted as the upper and lower lines. The 95% confidence level applies only to the interval obtained at one value of x and not to the entire set of x-levels. Notice that the width of the confidence interval on $\mu_{Y|x_0}$ increases as $|x_0 - \bar{x}|$ increases.

11-5 Confidence Intervals

Example 11-5

Figure 11-7
Scatter diagram of oxygen purity data from Example 11-1 with fitted regression line and 95 percent confidence limits on $\mu_{Y|x0}$.
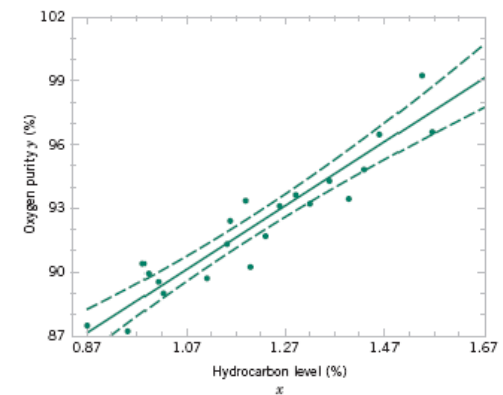
11-6 Prediction of New Observations

If x_0 is the value of the regressor variable of interest,

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

is the point estimator of the new or future value of the response, Y_0.

11-6 Prediction of New Observations

A $100(1 - \alpha)\%$ **prediction interval** on a **future observation** Y_0 at the value x_0 is given by

$$\hat{y}_0 - t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]}$$

$$\leq Y_0 \leq \hat{y}_0 + t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \qquad (11\text{-}33)$$

The value $\hat{y}_0$ is computed from the regression model $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

11-6 Prediction of New Observations

Example 11-6

To illustrate the construction of a prediction interval, suppose we use the data in Example 11-1 and find a 95% prediction interval on the next observation of oxygen purity at $x_0 = 1.00\%$. Using Equation 11-33 and recalling from Example 11-5 that $\hat{y}_0 = 89.23$, we find that the prediction interval is

$$89.23 - 2.101\sqrt{1.18\left[1 + \frac{1}{20} + \frac{(1.00 - 1.1960)^2}{0.68088}\right]}$$

$$\leq Y_0 \leq 89.23 + 2.101\sqrt{1.18\left[1 + \frac{1}{20} + \frac{(1.00 - 1.1960)^2}{0.68088}\right]}$$

11-6 Prediction of New Observations

Example 11-6

which simplifies to

$$86.83 \leq y_0 \leq 91.63$$

Minitab will also calculate prediction intervals. Refer to the output in Table 11-2. The 95% PI on the future observation at $x_0 = 1.00$ is shown in the display.

By repeating the foregoing calculations at different levels of x_0, we may obtain the 95% prediction intervals shown graphically as the lower and upper lines about the fitted regression model in Fig. 11-8. Notice that this graph also shows the 95% confidence limits on $\mu_{Y|x_0}$ calculated in Example 11-5. It illustrates that the prediction limits are always wider than the confidence limits.

11-6 Prediction of New Observations

Example 11-6

Figure 11-8 Scatter diagram of oxygen purity data from Example 11-1 with fitted regression line, 95% prediction limits (outer lines), and 95% confidence limits on $\mu_{Y|x0}$.
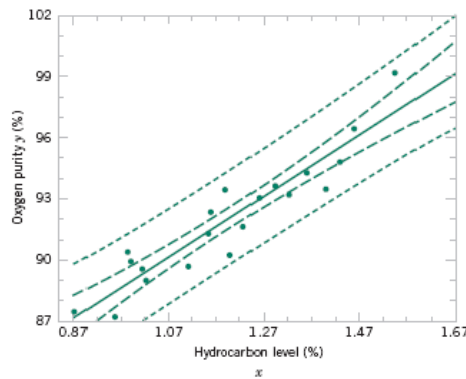
11-7 Adequacy of the Regression Model

- Fitting a regression model requires several **assumptions.**

    1. Errors are uncorrelated random variables with mean zero;

    2. Errors have constant variance; and,

    3. Errors be normally distributed.

- The analyst should always consider the validity of these assumptions to be doubtful and conduct analyses to examine the adequacy of the model

11-7 Adequacy of the Regression Model

11-7.1 Residual Analysis

- The **residuals** from a regression model are $e_i = y_i - \hat{y}_i$, where y_i is an actual observation and $\hat{y}_i$ is the corresponding fitted value from the regression model.

- Analysis of the residuals is frequently helpful in checking the assumption that the errors are approximately normally distributed with constant variance, and in determining whether additional terms in the model would be useful.

11-7 Adequacy of the Regression Model

11-7.1 Residual Analysis

Figure 11-9 Patterns for residual plots. (a) satisfactory, (b) funnel, (c) double bow, (d) nonlinear.
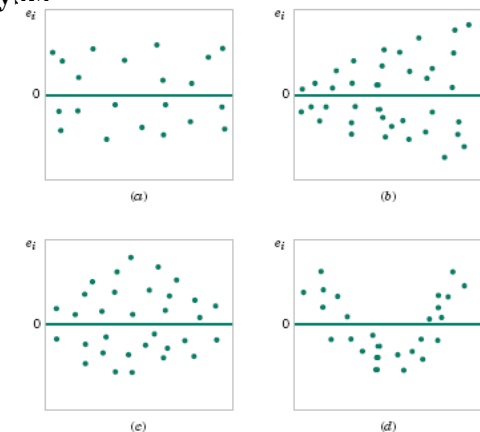
[Adapted from Montgomery, Peck, and Vining (2001).]

11-7 Adequacy of the Regression Model

Example 11-7

The regression model for the oxygen purity data in Example 11-1 is $\hat{y} = 74.283 + 14.947x$. Table 11-4 presents the observed and predicted values of y at each value of x from this data set, along with the corresponding residual. These values were computed using Minitab and show the number of decimal places typical of computer output. A normal probability plot of the residuals is shown in Fig. 11-10. Since the residuals fall approximately along a straight line in the figure, we conclude that there is no severe departure from normality. The residuals are also plotted against the predicted value $\hat{y}_i$ in Fig. 11-11 and against the hydrocarbon levels x_i in Fig. 11-12. These plots do not indicate any serious model inadequacies.
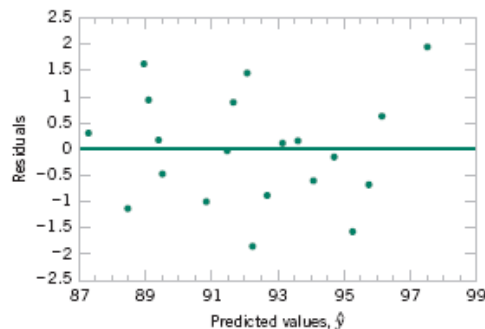
Table 11-4   Oxygen Purity Data from Example 11-1, Predicted Values, and Residuals

	Hydrocarbon Level, x	Oxygen Purity, y	Predicted Value, ŷ	Residual e = y − ŷ		Hydrocarbon Level, x	Oxygen Purity, y	Predicted Value, ŷ	Residual e = y − ŷ
1	0.99	90.01	89.069009	0.940991	11	1.19	93.54	92.063189	1.476811
2	1.02	89.05	89.518136	−0.468136	12	1.15	92.52	91.614062	0.905938
3	1.15	91.43	91.464353	−0.034353	13	0.98	90.56	88.919300	1.640700
4	1.29	93.74	93.560279	0.179721	14	1.01	89.54	89.368427	0.171573
5	1.46	96.73	96.105332	0.624668	15	1.11	89.85	90.865517	−1.015517
6	1.36	94.45	94.608242	−0.158242	16	1.20	90.39	92.212898	−1.822898
7	0.87	87.59	87.272501	0.317499	17	1.26	93.25	93.111152	0.138848
8	1.23	91.77	92.662025	−0.892025	18	1.32	93.41	94.009406	−0.599406
9	1.55	99.42	97.452713	1.967287	19	1.43	94.98	95.656205	−0.676205
10	1.40	93.65	95.207078	−1.557078	20	0.95	87.33	88.470173	−1.140173

11-7 Adequacy of the Regression Model

Example 11-7

Figure 11-10 Normal probability plot of residuals, Example 11-7.

11-7 Adequacy of the Regression Model

Example 11-7

Figure 11-11 Plot of residuals versus predicted oxygen purity, $\hat{y}$, Example 11-7.
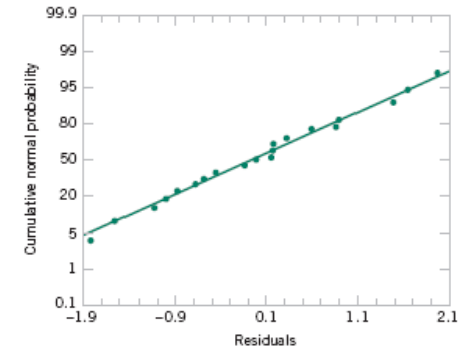
11-7 Adequacy of the Regression Model

11-7.2 Coefficient of Determination (R^2)

• The quantity

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

is called the **coefficient of determination** and is often used to judge the adequacy of a regression model.

• $0 \le R^2 \le 1$;

• We often refer (loosely) to R^2 as the amount of variability in the data explained or accounted for by the regression model.

11-7 Adequacy of the Regression Model

11-7.2 Coefficient of Determination (R^2)

• For the oxygen purity regression model,

$$R^2 = SS_R/SS_T$$

$$= 152.13/173.38$$

$$= 0.877$$

• Thus, the model accounts for 87.7% of the variability in the data.

11-9 Transformation and Logistic Regression

We occasionally find that the straight-line regression model $Y = \beta_0 + \beta_1 x + \epsilon$ is inappropriate because the true regression function is nonlinear. Sometimes nonlinearity is visually determined from the scatter diagram, and sometimes, because of prior experience or underlying theory, we know in advance that the model is nonlinear. Occasionally, a scatter diagram will exhibit an apparent nonlinear relationship between Y and x. In some of these situations, a nonlinear function can be expressed as a straight line by using a suitable transformation. Such nonlinear models are called **intrinsically linear.**

11-9 Transformation and Logistic Regression

Example 11-9

A research engineer is investigating the use of a windmill to generate electricity and has collected data on the DC output from this windmill and the corresponding wind velocity. The data are plotted in Figure 11-14 and listed in Table 11-5.

Observation Number, i	Wind Velocity (mph), x_i	DC Output, y_i
1	5.00	1.582
2	6.00	1.822
3	3.40	1.057
4	2.70	0.500
5	10.00	2.236
6	9.70	2.386
7	9.55	2.294
8	3.05	0.558
9	8.15	2.166
10	6.20	1.866
11	2.90	0.653
12	6.35	1.930
13	4.60	1.562
14	5.80	1.737
15	7.40	2.088
16	3.60	1.137
17	7.85	2.179
18	8.80	2.112
19	7.00	1.800
20	5.45	1.501
21	9.10	2.303
22	10.20	2.310
23	4.10	1.194
24	3.95	1.144
25	2.45	0.123

Table 11-5 Observed Values y_i and Regressor Variable x_i for Example 11-9.

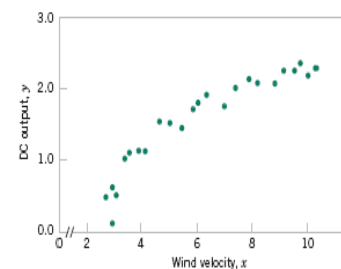# 11-9 Transformation and Logistic Regression

Figure 11-14 Plot of DC output y versus wind velocity x for the windmill data.
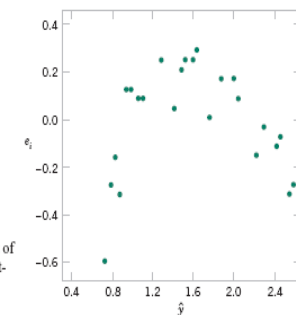

Figure 11-15 Plot of residuals e_i versus fitted values $\hat{y}_i$ for the windmill data.
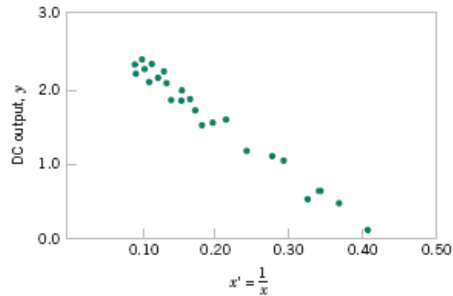
11-9 Transformation and Logistic Regression

Figure 11-16 Plot of DC output versus $x' = 1/x$ for the windmill data.

Figure 11-16 is a scatter diagram with the transformed variable $x' = 1/x$. This plot appears linear, indicating that the reciprocal transformation is appropriate. The fitted regression model is

$$\hat{y} = 2.9789 - 6.9345x'$$

The summary statistics for this model are $R^2 = 0.9800$, $MS_E = \hat{\sigma}^2 = 0.0089$, and $F_0 = 1128.43$ (the P value is <0.0001).
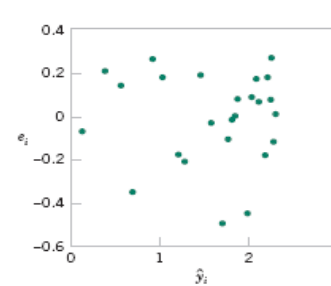
Figure 11-17 Plot of residuals versus fitted values $\hat{y}_i$ for the transformed model for the windmill data.
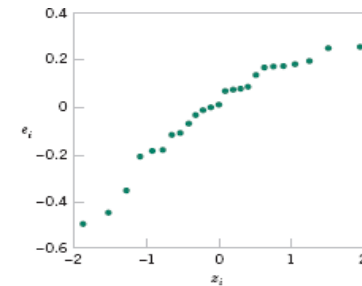
Figure 11-18 Normal probability plot of the residuals for the transformed model for the windmill data.

A plot of the residuals from the transformed model versus $\hat{y}$ is shown in Figure 11-17. This plot does not reveal any serious problem with inequality of variance. The normal probability plot, shown in Figure 11-18, gives a mild indication that the errors come from a distribution with heavier tails than the normal (notice the slight upward and downward curve at the extremes). This normal probability plot has the z-score value plotted on the horizontal axis. Since there is no strong signal of model inadequacy, we conclude that the transformed model is satisfactory.

12

Multiple Linear Regression

CHAPTER OUTLINE

12-1 Multiple Linear Regression Models

12-1.1 Introduction

• Many applications of regression analysis involve situations in which there are more than one regressor variable.

• A regression model that contains more than one regressor variable is called a **multiple regression model**.

12-1 Multiple Linear Regression Models

12-1.1 Introduction

- For example, suppose that the effective life of a cutting tool depends on the cutting speed and the tool angle. A possible multiple regression model could be

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where

Y – tool life

x_1 – cutting speed

x_2 – tool angle

12-1 Multiple Linear Regression Models
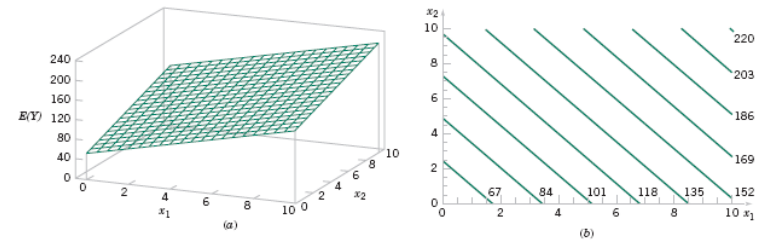
12-1.1 Introduction

Figure 12-1 (a) The regression plane for the model $E(Y)$ = 50 + 10x$_1$ + 7x$_2$. (b) The contour plot

12-1 Multiple Linear Regression Models

12-1.1 Introduction

In general, the **dependent variable** or **response** Y may be related to k **independent** or **regressor variables.** The model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{12-2}$$

is called a multiple linear regression model with k regressor variables. The parameters $\beta_j, j = 0,$ 1, ... , k, are called the regression coefficients. This model describes a hyperplane in the k-dimensional space of the regressor variables $\{x_j\}$. The parameter β_j represents the expected change in response Y per unit change in x_j when all the remaining regressors x_i $(i \neq j)$ are held constant.

12-1 Multiple Linear Regression Models

12-1.1 Introduction

Figure 12-2 (a) Three-dimensional plot of the regression model $E(Y)$ = 50 + 10x$_1$ + 7x$_2$ + 5x$_1$x$_2$.

(b) The contour plot

12-1 Multiple Linear Regression Models
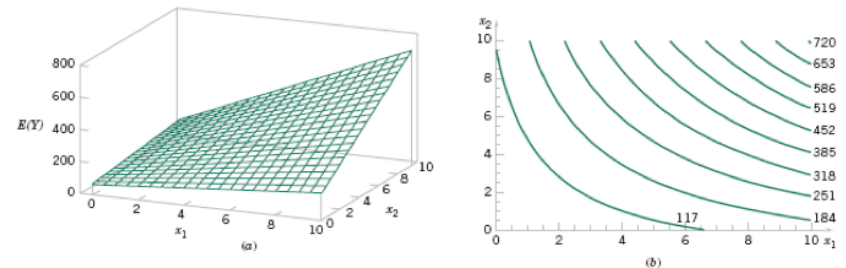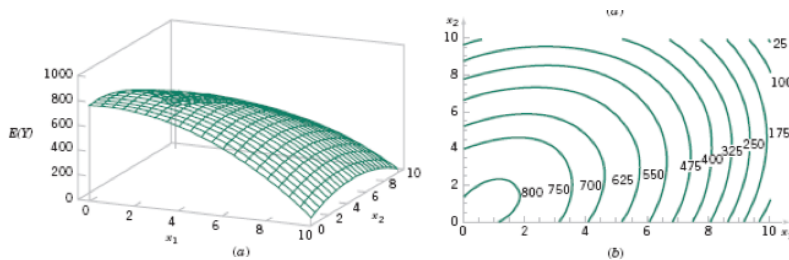
12-1.1 Introduction

Figure 12-3 (a) 3-D plot of the regression model

$E(Y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$.

(b) The contour plot

12-1 Multiple Linear Regression Models

12-1.2 Least Squares Estimation of the Parameters

The **method of least squares** may be used to estimate the regression coefficients in the multiple regression model, Equation 12-2. Suppose that $n > k$ observations are available, and let x_{ij} denote the ith observation or level of variable x_j. The observations are

$$(x_{i1}, x_{i2}, \dots, x_{ik}, y_i), \quad i = 1, 2, \dots, n \quad \text{and} \quad n > k$$

It is customary to present the data for multiple regression in a table such as Table 12-1.

Table 12-1 Data for Multiple Linear Regression

y	x_1	x_2	$\cdots$	x_k
y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

12-1 Multiple Linear Regression Models

12-1.2 Least Squares Estimation of the Parameters

• The **least squares function** is given by

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2$$

• The **least squares estimates** must satisfy

$$\frac{\partial L}{\partial \beta_0}\bigg|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0$$

and

$$\frac{\partial L}{\partial \beta_j}\bigg|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) x_{ij} = 0 \quad j = 1, 2, \dots, k$$

12-1 Multiple Linear Regression Models

12-1.2 Least Squares Estimation of the Parameters

• The **least squares normal Equations** are

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} = \sum_{i=1}^{n} x_{i1} y_i$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{ik} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{ik} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 = \sum_{i=1}^{n} x_{ik} y_i$$

• The solution to the normal Equations are the **least squares estimators** of the regression coefficients.

12-1 Multiple Linear Regression Models

12-1.3 Matrix Approach to Multiple Linear Regression

Suppose the model relating the regressors to the response is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \qquad i = 1, 2, \ldots, n$$

In matrix notation this model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

12-1 Multiple Linear Regression Models

12-1.3 Matrix Approach to Multiple Linear Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$$
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad
\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \qquad
\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad
\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

12-1 Multiple Linear Regression Models

12-1.3 Matrix Approach to Multiple Linear Regression

We wish to find the vector of least squares estimators that minimizes:

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

The resulting least squares estimate is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (12\text{-}13)$$

12-1 Multiple Linear Regression Models

12-1.3 Matrix Approach to Multiple Linear Regression

The fitted regression model is

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \qquad i = 1, 2, \ldots, n \qquad (12\text{-}14)$$

In matrix notation, the fitted model is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

The difference between the observation y_i and the fitted value $\hat{y}_i$ is a **residual,** say, $e_i = y_i - \hat{y}_i$. The $(n \times 1)$ vector of residuals is denoted by

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \qquad (12\text{-}15)$$

Example 12-2

Table 12-2 Wire Bond Data for Example 11-1

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2	Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50	14	11.66	2	360
2	24.45	8	110	15	21.65	4	205
3	31.75	11	120	16	17.89	4	400
4	35.00	10	550	17	69.00	20	600
5	25.02	8	295	18	10.30	1	585
6	16.86	4	200	19	34.93	10	540
7	14.38	2	375	20	46.59	15	250
8	9.60	2	52	21	44.88	15	290
9	24.35	9	100	22	54.12	16	510
10	27.50	8	300	23	56.63	17	590
11	17.08	4	412	24	22.13	6	100
12	37.00	11	400	25	21.15	5	400
13	41.95	12	500				

Figure 12-4 Matrix of scatter plots (from Minitab) for the wire bond pull strength data in Table 12-2.

Example 12-2

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ 1 & 11 & 120 \\ 1 & 10 & 550 \\ 1 & 8 & 295 \\ 1 & 4 & 200 \\ 1 & 2 & 375 \\ 1 & 2 & 52 \\ 1 & 9 & 100 \\ 1 & 8 & 300 \\ 1 & 4 & 412 \\ 1 & 11 & 400 \\ 1 & 12 & 500 \\ 1 & 2 & 360 \\ 1 & 4 & 205 \\ 1 & 4 & 400 \\ 1 & 20 & 600 \\ 1 & 1 & 585 \\ 1 & 10 & 540 \\ 1 & 15 & 250 \\ 1 & 15 & 290 \\ 1 & 16 & 510 \\ 1 & 17 & 590 \\ 1 & 6 & 100 \\ 1 & 5 & 400 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 9.95 \\ 24.45 \\ 31.75 \\ 35.00 \\ 25.02 \\ 16.86 \\ 14.38 \\ 9.60 \\ 24.35 \\ 27.50 \\ 17.08 \\ 37.00 \\ 41.95 \\ 11.66 \\ 21.65 \\ 17.89 \\ 69.00 \\ 10.30 \\ 34.93 \\ 46.59 \\ 44.88 \\ 54.12 \\ 56.63 \\ 22.13 \\ 21.15 \end{bmatrix}$$

Example 12-2

The $\mathbf{X'X}$ matrix is

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 8 & \cdots & 5 \\ 50 & 110 & \cdots & 400 \end{bmatrix} \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ \vdots & \vdots & \vdots \\ 1 & 5 & 400 \end{bmatrix} = \begin{bmatrix} 25 & 206 & 8{,}294 \\ 206 & 2{,}396 & 77{,}177 \\ 8{,}294 & 77{,}177 & 3{,}531{,}848 \end{bmatrix}$$

and the $\mathbf{X'y}$ vector is

$$\mathbf{X'y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 8 & \cdots & 5 \\ 50 & 110 & \cdots & 400 \end{bmatrix} \begin{bmatrix} 9.95 \\ 24.45 \\ \vdots \\ 21.15 \end{bmatrix} = \begin{bmatrix} 725.82 \\ 8{,}008.37 \\ 274{,}811.31 \end{bmatrix}$$

The least squares estimates are found from Equation 12-13 as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

12-1 Multiple Linear Regression Models

Example 12-2

or

$$
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 206 & 8{,}294 \\ 206 & 2{,}396 & 77{,}177 \\ 8{,}294 & 77{,}177 & 3{,}531{,}848 \end{bmatrix}^{-1} \begin{bmatrix} 725.82 \\ 8{,}008.37 \\ 274{,}811.31 \end{bmatrix}
$$

$$
= \begin{bmatrix} 0.214653 & -0.007491 & -0.000340 \\ -0.007491 & 0.001671 & -0.000019 \\ -0.000340 & -0.000019 & +0.0000015 \end{bmatrix} \begin{bmatrix} 725.82 \\ 8{,}008.47 \\ 274{,}811.31 \end{bmatrix} = \begin{bmatrix} 2.26379143 \\ 2.74426964 \\ 0.01252781 \end{bmatrix}
$$

Therefore, the fitted regression model with the regression coefficients rounded to five decimal places is

$$\hat{y} = 2.26379 + 2.74427x_1 + 0.01253x_2$$

This is identical to the results obtained in Example 12-1.

12-1 Multiple Linear Regression Models

Example 12-2

This regression model can be used to predict values of pull strength for various values of wire length (x_1) and die height (x_2). We can also obtain the **fitted values** $\hat{y}_i$ by substituting each observation (x_{i1}, x_{i2}), $i = 1, 2, \ldots, n$, into the equation. For example, the first observation has $x_{11} = 2$ and $x_{12} = 50$, and the fitted value is

$$
\begin{aligned}
\hat{y}_1 &= 2.26379 + 2.74427x_{11} + 0.01253x_{12} \\
&= 2.26379 + 2.74427(2) + 0.01253(50) \\
&= 8.38
\end{aligned}
$$

The corresponding observed value is $y_1 = 9.95$. The *residual* corresponding to the first observation is

$$
\begin{aligned}
e_1 &= y_1 - \hat{y}_1 \\
&= 9.95 - 8.38 \\
&= 1.57
\end{aligned}
$$

Table 12-3 displays all 25 fitted values $\hat{y}_i$ and the corresponding residuals. The fitted values and residuals are calculated to the same accuracy as the original data.

12-1 Multiple Linear Regression Models

Example 12-2

Table 12-3   Observations, Fitted Values, and Residuals for Example 12-2

Observation Number	y_t	$\hat{y}_t$	$e_t = y_t - \hat{y}_t$	Observation Number	y_t	$\hat{y}_t$	$e_t = y_t - \hat{y}_t$
1	9.95	8.38	1.57	14	11.66	12.26	-0.60
2	24.45	25.60	-1.15	15	21.65	15.81	5.84
3	31.75	33.95	-2.20	16	17.89	18.25	-0.36
4	35.00	36.60	-1.60	17	69.00	64.67	4.33
5	25.02	27.91	-2.89	18	10.30	12.34	-2.04
6	16.86	15.75	1.11	19	34.93	36.47	-1.54
7	14.38	12.45	1.93	20	46.59	46.56	-0.03
8	9.60	8.40	1.20	21	44.88	47.06	-2.18
9	24.35	28.21	-3.86	22	54.12	52.56	1.56
10	27.50	27.98	-0.48	23	56.63	56.31	0.32
11	17.08	18.40	-1.32	24	22.13	19.98	2.15
12	37.00	37.46	-0.46	25	21.15	21.00	0.15
13	41.95	41.46	0.49				

Table 12-4   Minitab Multiple Regression Output for the Wire Bond Pull Strength Data

Regression Analysis: Strength versus Length, Height

The regression equation is
Strength = 2.26 + 2.74 Length + 0.0125 Height

Predictor		Coef	SE Coef	T	P	VIF
Constant	$\hat{\beta}_0 \rightarrow$ 2.264		1.060	2.14	0.044	
Length	$\hat{\beta}_1 \rightarrow$ 2.74427		0.09352	29.34	0.000	1.2
Height	$\hat{\beta}_2 \rightarrow$ 0.012528		0.002798	4.48	0.000	1.2

S = 2.288   R-Sq = 98.1%   R-Sq (adj) = 97.9%
PRESS = 156.163   R-Sq (pred) = 97.44%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5990.8	2995.4	572.17	0.000
Residual Error	22	115.2	5.2 $\leftarrow \hat{\sigma}^2$		
Total	24	6105.9			

Source	DF	Seq SS
Length	1	5885.9
Height	1	104.9

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	27.663	0.482	(26.663, 28.663)	(22.814, 32.512)

Values of Predictors for New Observations

New Obs	Length	Height
1	8.00	275

12-1 Multiple Linear Regression Models

Estimating σ^2

An unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n - p} = \frac{SS_E}{n - p} \qquad (12\text{-}16)$$

12-1 Multiple Linear Regression Models

12-1.4 Properties of the Least Squares Estimators

Individual variances and covariances:

$$V(\hat{\beta}_j) = \sigma^2 C_{jj}, \qquad j = 0, 1, 2$$

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 C_{ij}, \qquad i \neq j$$

In general,

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X'X})^{-1} = \sigma^2 \mathbf{C}$$

12-1 Multiple Linear Regression Models

12-1.4 Properties of the Least Squares Estimators

Unbiased estimators:

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X'X})^{-1}\mathbf{X'Y}] \\
&= E[(\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})] \\
&= E[(\mathbf{X'X})^{-1}\mathbf{X'X}\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\epsilon}] \\
&= \boldsymbol{\beta}
\end{aligned}$$

Covariance Matrix:

$$\mathbf{C} = (\mathbf{X'X})^{-1} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

12-2 Hypothesis Tests in Multiple Linear Regression

12-2.1 Test for Significance of Regression

The appropriate hypotheses are

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1: \beta_j \neq 0 \quad \text{for at least one } j \qquad (12\text{-}17)$$

The test statistic is

$$F_0 = \frac{SS_R/k}{SS_E/(n - p)} = \frac{MS_R}{MS_E} \qquad (12\text{-}18)$$

12-2 Hypothesis Tests in Multiple Linear Regression

12-2.1 Test for Significance of Regression

Table 12-9 Analysis of Variance for Testing Significance of Regression in Multiple Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	SS_R	k	MS_R	MS_R/MS_E
Error or residual	SS_E	$n - p$	MS_E	
Total	SS_T	$n - 1$		

12-2 Hypothesis Tests in Multiple Linear Regression

Example 12-3

Table 12-10 Test for Significance of Regression for Example 12-3

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Regression	5990.7712	2	2995.3856	572.17	1.08E-19
Error or residual	115.1735	22	5.2352		
Total	6105.9447	24			

The analysis of variance is shown in Table 12-10. To test H_0: $\beta_1 = \beta_2 = 0$, we calculate the statistic

$$f_0 = \frac{MS_R}{MS_E} = \frac{2995.3856}{5.2352} = 572.17$$

Since $f_0 > f_{0.05,2,22} = 3.44$ (or since the P-value is considerably smaller than $\alpha = 0.05$), we reject the null hypothesis and conclude that pull strength is linearly related to either wire length or die height, or both. However, we note that this does not necessarily imply that the

12-2 Hypothesis Tests in Multiple Linear Regression

R^2 and Adjusted R^2

The **coefficient of multiple determination**

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

- For the wire bond pull strength data, we find that $R^2 = SS_R/SS_T = 5990.7712/6105.9447 = 0.9811$.

- Thus, the model accounts for about 98% of the variability in the pull strength response.

12-2 Hypothesis Tests in Multiple Linear Regression

R^2 and Adjusted R^2

The **adjusted R^2** is

$$R_{adj}^2 = 1 - \frac{SS_E/(n - p)}{SS_T/(n - 1)} \tag{12-22}$$

- The adjusted R^2 statistic penalizes the analyst for adding terms to the model.

- It can help guard against overfitting (including regressors that are not really useful)

12-2 Hypothesis Tests in Multiple Linear Regression

12-2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients

$$H_0: \beta_j = \beta_{j0}$$
$$H_1: \beta_j \neq \beta_{j0} \qquad (12\text{-}23)$$

The test statistic for this hypothesis is

$$T_0 = \frac{\beta_{j0} - \beta_j}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{se(\beta_{j0})} \qquad (12\text{-}24)$$

- Reject H_0 if $|t_0| > t_{\alpha/2,\,n-p}$.

- This is called a **partial** or marginal test

12-2 Hypothesis Tests in Multiple Linear Regression

Example 12-4

The P-value for $t_0 = 4.4767$ is $P = 0.0002$, so with $\alpha = 0.05$ we would reject the null hypothesis. Note that this test measures the marginal or partial contribution of x_2 given that x_1 is in the model. That is, the t-test measures the contribution of adding the variable x_2 = die height to a model that already contains x_1 = wire length. Table 12-4 shows the value of the t-test computed by Minitab. The Minitab t-test statistic is reported to two decimal places. Note that the computer produces a t-test for each regression coefficient in the model. These t-tests indicate that both regressors contribute to the model.

12-2 Hypothesis Tests in Multiple Linear Regression

Example 12-4

Consider the wire bond pull strength data, and suppose that we want to test the hypothesis that the regression coefficient for x_2 (die height) is zero. The hypotheses are

$$H_0: \beta_2 = 0$$
$$H_1: \beta_2 \neq 0$$

The main diagonal element of the $(\mathbf{X'X})^{-1}$ matrix corresponding to $\hat{\beta}_2$ is $C_{22} = 0.0000015$, so the t-statistic in Equation 12-24 is

$$t_0 = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{0.01253}{\sqrt{(5.2352)(0.0000015)}} = 4.4767$$

Note that we have used the estimate of σ^2 reported to four decimal places in Table 12-10. Since $t_{0.025,22} = 2.074$, we reject $H_0: \beta_2 = 0$ and conclude that the variable x_2 (die height) contributes significantly to the model. We could also have used a P-value to draw conclusions.

R commands and outputs

```
> dat=read.table("http://www.stat.ucla.edu/~hqxu/stat105/
data/table12_2.txt", h=T)

> g=lm(Strength~Length+Height, dat)

> summary(g)

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.263791    1.060066    2.136 0.044099 *
Length      2.744270    0.093524   29.343  < 2e-16 ***
Height      0.012528    0.002798    4.477 0.000188 ***

Residual standard error: 2.288 on 22 degrees of freedom

Multiple R-Squared: 0.9811,     Adjusted R-squared: 0.9794

F-statistic: 572.2 on 2 and 22 DF,  p-value: < 2.2e-16
```

13

Design and Analysis
of Single-Factor
Experiments: The
Analysis of Variance

CHAPTER OUTLINE

13-1 Designing Engineering Experiments

Every experiment involves a sequence of activities:

1. Conjecture – the original hypothesis that motivates the experiment.

2. Experiment – the test performed to investigate the conjecture.

3. Analysis – the statistical analysis of the data from the experiment.

4. Conclusion – what has been learned about the original conjecture from the experiment. Often the experiment will lead to a revised conjecture, and a new experiment, and so forth.

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level, using a pilot plant. All 24 specimens are tested on a laboratory tensile tester, in random order. The data from this experiment are shown in Table 13-1.

Table 13-1 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

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

- The levels of the factor are sometimes called **treatments**.

- Each treatment has six observations or **replicates**.

- The runs are run in **random** order.

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example
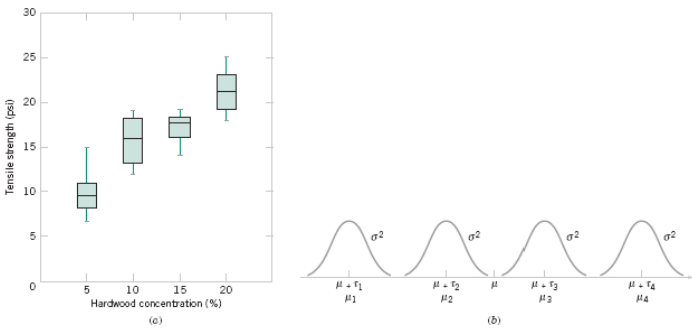
Figure 13-1 (a) Box plots of hardwood concentration data. (b) Display of the model in Equation 13-1 for the completely randomized single-factor experiment

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Suppose there are a different levels of a single factor that we wish to compare. The levels are sometimes called **treatments**.

Table 13-2   Typical Data for a Single-Factor Experiment

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_1.$	$\bar{y}_1.$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_2.$	$\bar{y}_2.$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_a.$	$\bar{y}_a.$
					$y..$	$\bar{y}..$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

We may describe the observations in Table 13-2 by the linear statistical model:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \quad (13\text{-}1)$$

The model could be written as

$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Fixed-effects Model

The treatment effects are usually defined as deviations from the overall mean so that:

$$\sum_{i=1}^{a} \tau_i = 0$$

Also,
$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i\cdot} = y_{i\cdot}/n \qquad i = 1, 2, \ldots, a$$

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N$$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

We wish to test the hypotheses:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \quad \text{for at least one } i$$

The analysis of variance partitions the total variability into two parts.

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

The **sum of squares identity** is

$$\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = n \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 \qquad (13\text{-}5)$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_E \qquad (13\text{-}6)$$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

The expected value of the treatment sum of squares is

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

and the expected value of the error sum of squares is

$$E(SS_E) = a(n - 1)\sigma^2$$

The ratio $MS_{\text{Treatments}} = SS_{\text{Treatments}}/(a - 1)$ is called the **mean square for treatments**.

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

The appropriate test statistic is

$$F_0 = \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \qquad (13\text{-}7)$$

We would reject H_0 if $f_0 > f_{\alpha, a-1, a(n-1)}$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

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

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \qquad (13\text{-}8)$$

and

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \qquad (13\text{-}9)$$

The error sum of squares is obtained by subtraction as

$$SS_E = SS_T - SS_{\text{Treatments}} \qquad (13\text{-}10)$$

where $N = na$ is the total number of observations.

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Analysis of Variance Table

Table 13-3   The Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model

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

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

Consider the paper tensile strength experiment described in Section 13-2.1. 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

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$
$$H_1: \tau_i \neq 0 \text{ for at least one } i.$$

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

We will use $\alpha = 0.01$. The sums of squares for the analysis of variance are computed from Equations 13-8, 13-9, and 13-10 as follows:

$$SS_T = \sum_{i=1}^{4} \sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$= (7)^2 + (8)^2 + \cdots + (20)^2 - \frac{(383)^2}{24} = 512.96$$

$$SS_{Treatments} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

$$= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24} = 382.79$$

$$SS_E = SS_T - SS_{Treatments}$$
$$= 512.96 - 382.79 = 130.17$$

Table 13-5   Minitab Analysis of Variance Output for Example 13-1

One-Way ANOVA: Strength versus CONC
Analysis of Variance for Strength

Source	DF	SS	MS	F	P
Conc	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	(—*—)

—+———+———+———+— +-
10.0   15.0   20.0   25.0

Pooled StDev =   2.551

Fisher's pairwise comparisons

Family error rate = 0.192
Individual error rate = 0.0500

Critical value = 2.086

Intervals for (column level mean) − (row level mean)

	5	10	15
10	−8.739		
	−2.594		
15	−10.072	−4.406	
	−3.928	1.739	
20	−14.239	−8.572	−7.239
	−8.094	−2.428	−1.094

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

The ANOVA is summarized in Table 13-4. Since $f_{0.01,3,20} = 4.94$, we reject H_0 and conclude that hardwood concentration in the pulp significantly affects the mean strength of the paper. We can also find a P-value for this test statistic as follows:

$$P = P(F_{3,20} > 19.60) \simeq 3.59 \times 10^{-6}$$

Since $P \simeq 3.59 \times 10^{-6}$ is considerably smaller than $\alpha = 0.01$, we have strong evidence to conclude that H_0 is not true.

Table 13-4   ANOVA for the Tensile Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Hardwood concentration	382.79	3	127.60	19.60	3.59 E-6
Error	130.17	20	6.51		
Total	512.96	23			

13-2 The Completely Randomized Single-Factor Experiment

A $100(1 - \alpha)$ percent confidence interval on the mean of the ith treatment μ_i is

$$\bar{y}_{i.} - t_{\alpha/2,a(n-1)} \sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i.} + t_{\alpha/2,a(n-1)} \sqrt{\frac{MS_E}{n}} \qquad (13\text{-}11)$$

The 95% CI on the mean of the 20% hardwood is

$$\left[\bar{y}_{4.} \pm t_{0.025,20} \sqrt{MS_E / n} \right]$$

$$\left[21.167 \pm (2.086) \sqrt{6.51 / 6} \right]$$

$$19.00 \text{ psi} \le \mu_4 \le 23.34 \text{ psi}$$

13-2 The Completely Randomized Single-Factor Experiment

A $100(1 - \alpha)$ percent confidence interval on the difference in two treatment means $\mu_i - \mu_j$ is

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2, a(n-1)} \sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + t_{\alpha/2, a(n-1)} \sqrt{\frac{2MS_E}{n}}$$

$$(13\text{-}12)$$

For the hardwood concentration example,

A 95% CI on the difference in means $\mu_3 - \mu_2$ is

$$\left[\bar{y}_{3\cdot} - \bar{y}_{2\cdot} \pm t_{0.025,20} \sqrt{2MS_E / n} \right]$$

$$\left[17.00 - 15.67 \pm (2.086) \sqrt{2(6.51) / 6} \right]$$

$$-1.74 \leq \mu_3 - \mu_2 \leq 4.40$$

13-2 The Completely Randomized Single-Factor Experiment

An Unbalanced Experiment

The sums of squares computing formulas for the ANOVA with unequal sample sizes n_i in each treatment are

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \qquad (13\text{-}13)$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{N} \qquad (13\text{-}14)$$

and

$$SS_E = SS_T - SS_{\text{Treatments}} \qquad (13\text{-}15)$$

13-2 The Completely Randomized Single-Factor Experiment

13-2.3 Multiple Comparisons Following the ANOVA

The least significant difference (LSD) is

$$LSD = t_{\alpha/2, a(n-1)} \sqrt{\frac{2MS_E}{n}} \qquad (13\text{-}16)$$

If the sample sizes are different in each treatment:

$$LSD = t_{\alpha/2, N-a} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

We will apply the Fisher LSD method to the hardwood concentration experiment. There are $a = 4$ means, $n = 6$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$. The treatment means are

$$\bar{y}_{1\cdot} = 10.00 \text{ psi}$$
$$\bar{y}_{2\cdot} = 15.67 \text{ psi}$$
$$\bar{y}_{3\cdot} = 17.00 \text{ psi}$$
$$\bar{y}_{4\cdot} = 21.17 \text{ psi}$$

The value of LSD is $LSD = t_{0.025,20} \sqrt{2MS_E/n} = 2.086\sqrt{2(6.51)/6} = 3.07$. Therefore, any pair of treatment averages that differs by more than 3.07 implies that the corresponding pair of treatment means are different.

13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

The comparisons among the observed treatment averages are as follows:

$$4 \text{ vs. } 1 = 21.17 - 10.00 = 11.17 > 3.07$$
$$4 \text{ vs. } 2 = 21.17 - 15.67 = 5.50 > 3.07$$
$$4 \text{ vs. } 3 = 21.17 - 17.00 = 4.17 > 3.07$$
$$3 \text{ vs. } 1 = 17.00 - 10.00 = 7.00 > 3.07$$
$$3 \text{ vs. } 2 = 17.00 - 15.67 = 1.33 < 3.07$$
$$2 \text{ vs. } 1 = 15.67 - 10.00 = 5.67 > 3.07$$

From this analysis, we see that there are significant differences between all pairs of means except 2 and 3. This implies that 10 and 15% hardwood concentration produce approximately the same tensile strength and that all other concentration levels tested produce different tensile strengths. It is often helpful to draw a graph of the treatment means, such as in Fig. 13-2, with the means that are *not* different underlined. This graph clearly reveals the results of the experiment and shows that 20% hardwood produces the maximum tensile strength.

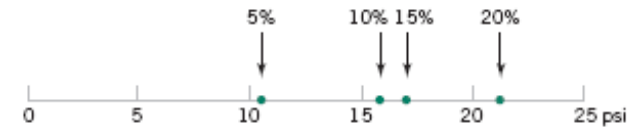# 13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

Figure 13-2 Results of Fisher's LSD method in Example 13-2.

Figure 13-2 Results of Fisher's LSD method in Example 13-2

13-2 The Completely Randomized Single-Factor Experiment
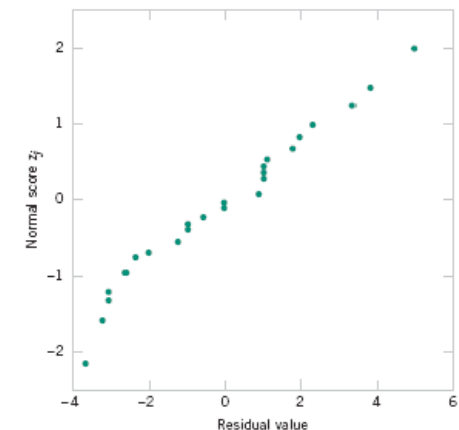
13-2.4 Residual Analysis and Model Checking

Table 13-6 Residuals for the Tensile Strength 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

13-2 The Completely Randomized Single-Factor Experiment

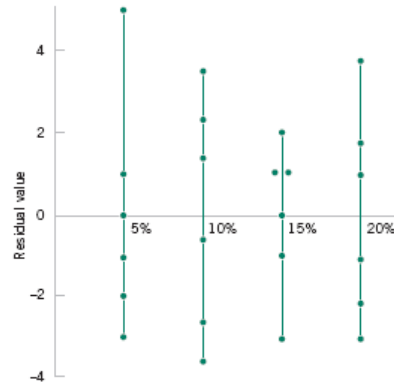## 13-2.4 Residual Analysis and Model Checking

Figure 13-4 Normal probability plot of residuals from the hardwood concentration experiment.

13-2 The Completely Randomized Single-Factor Experiment
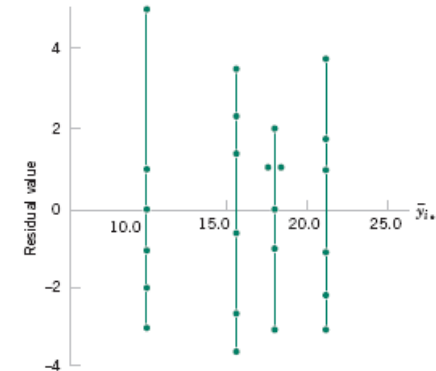
13-2.4 Residual Analysis and Model Checking

Figure 13-5 Plot of residuals versus factor levels (hardwood concentration).

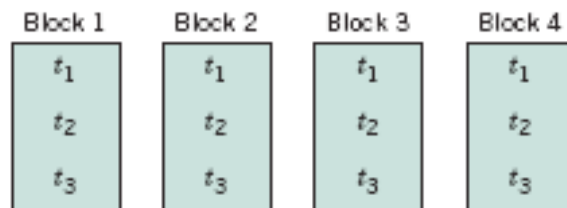# 13-2 The Completely Randomized Single-Factor Experiment

Figure 13-6 Plot of residuals versus $\bar{y}_i$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The **randomized block design** is an extension of the paired t-test to situations where the factor of interest has more than two levels.

Figure 13-9 A randomized complete block design.

13-4 Randomized Complete Block Designs

For example, consider the situation of Example 10-9, where two different methods were used to predict the shear strength of steel plate girders. Say we use four girders as the experimental units.

Table 13-9   A Randomized Complete Block Design

Treatments (Method)	Block (Girder)			
	1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

General procedure for a randomized complete block design:

Table 13-10   A Randomized Complete Block Design with a Treatments and b Blocks

Treatments	Blocks				Totals	Averages
	1	2	...	b		
1	y_{11}	y_{12}	...	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	...	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	...	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	...	$y_{\cdot b}$	$y_{\cdot\cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The appropriate linear statistical model:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases}$$

We assume

- treatments and blocks are initially fixed effects

- blocks do not interact

- $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

We are interested in testing:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \text{ at least one } i$$

The **sum of squares identity** for the randomized complete block design is

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{..})^2 + a \sum_{j=1}^{b} (\bar{y}_{\cdot j} - \bar{y}_{..})^2$$
$$+ \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{\cdot j} - \bar{y}_{i\cdot} + \bar{y}_{..})^2 \qquad (13\text{-}27)$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The mean squares are:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a - 1}$$

$$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b - 1}$$

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

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The expected values of these mean squares are:

$$E(MS_{Treatments}) = \sigma^2 + \frac{b \sum_{i=1}^{a} \tau_i^2}{a-1}$$

$$E(MS_{Blocks}) = \sigma^2 + \frac{a \sum_{j=1}^{b} \beta_j^2}{b-1}$$

$$E(MS_E) = \sigma^2$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

Table 13-11 ANOVA for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{Treatments}$	$a-1$	$\dfrac{SS_{Treatments}}{a-1}$	$\dfrac{MS_{Treatments}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\dfrac{SS_{Blocks}}{b-1}$	
Error	SS_E (by subtraction)	$(a-1)(b-1)$	$\dfrac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$ab-1$		

13-4 Randomized Complete Block Designs

Example 13-5

An experiment was performed to determine the effect of four different chemicals on the strength of a fabric. These chemicals are used as part of the permanent press finishing process. Five fabric samples were selected, and a randomized complete block design was run by testing each chemical type once in random order on each fabric sample. The data are shown in Table 13-12. We will test for differences in means using an ANOVA with $\alpha = 0.01$.

Table 13-12 Fabric Strength Data—Randomized Complete Block Design

Chemical Type	Fabric Sample					Treatment Totals $y_{i\cdot}$	Treatment Averages $\bar{y}_{i\cdot}$
	1	2	3	4	5		
1	1.3	1.6	0.5	1.2	1.1	5.7	1.14
2	2.2	2.4	0.4	2.0	1.8	8.8	1.76
3	1.8	1.7	0.6	1.5	1.3	6.9	1.38
4	3.9	4.4	2.0	4.1	3.4	17.8	3.56
Block totals $y_{\cdot j}$	9.2	10.1	3.5	8.8	7.6	39.2($y_{\cdot\cdot}$)	
Block averages $\bar{y}_{\cdot j}$	2.30	2.53	0.88	2.20	1.90		1.96($\bar{y}_{\cdot\cdot}$)

13-4 Randomized Complete Block Designs

Example 13-5

Table 13-13 Analysis of Variance for the Randomized Complete Block Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Chemical types (treatments)	18.04	3	6.01	75.13	4.79 E-8
Fabric samples (blocks)	6.69	4	1.67		
Error	0.96	12	0.08		
Total	25.69	19			

The ANOVA is summarized in Table 13-13. Since $f_0 = 75.13 > f_{0.01,3,12} = 5.95$ (the P-value is 4.79×10^{-8}), we conclude that there is a significant difference in the chemical types so far as their effect on strength is concerned.

13-4 Randomized Complete Block Designs

Minitab Output for Example 13-5

Table 13-14 Minitab Analysis of Variance for the Randomized Complete Block Design in Example 13-5

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values				
Chemical	fixed	4	1	2	3	4	
Fabric S	fixed	5	1	2	3	4	5

Analysis of Variance for strength

Source	DF	SS	MS	F	P
Chemical	3	18.0440	6.0147	75.89	0.000
Fabric S	4	6.6930	1.6733	21.11	0.000
Error	12	0.9510	0.0792		
Total	19	25.6880			

F-test with denominator: Error
Denominator MS = 0.079250 with 12 degrees of freedom

Numerator	DF	MS	F	P
Chemical	3	6.015	75.89	0.000
Fabric S	4	1.673	21.11	0.000

13-4 Randomized Complete Block Designs

13-4.2 Multiple Comparisons

Fisher's Least Significant Difference for Example 13-5

$$\bar{y}_{1.}=1.14 \quad \bar{y}_{2.}=1.76 \quad \bar{y}_{3.}=1.38 \quad \bar{y}_{4.}=3.56$$

$$LSD = t_{0.025,12}\sqrt{\frac{2MS_E}{b}} = 2.179\sqrt{\frac{2(0.08)}{5}} = 0.39$$
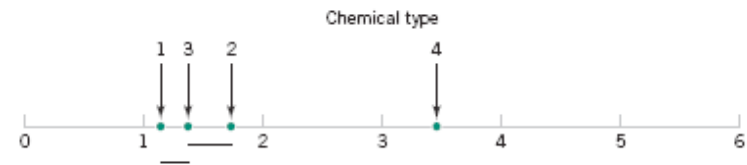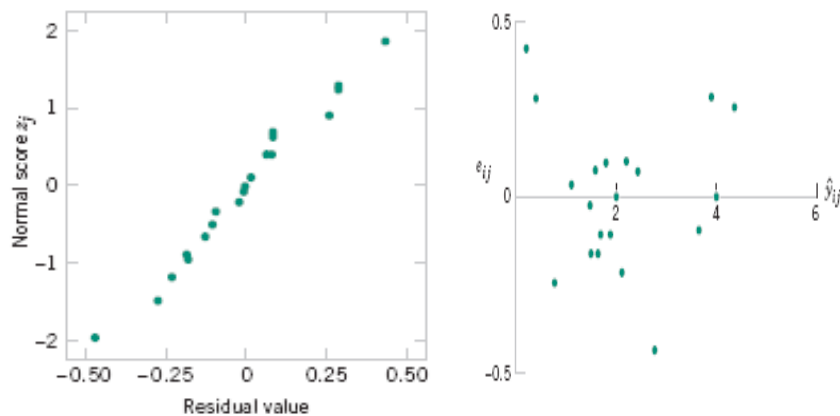

Figure 13-10 Results of Fisher's LSD method.
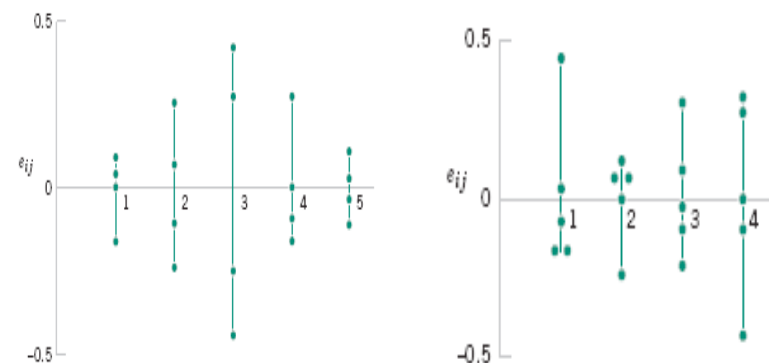
13-4 Randomized Complete Block Designs

13-4.3 Residual Analysis and Model Checking

(a) Normal prob. plot of residuals

(b) Residuals versus $\hat{y}_{ij}$

13-4 Randomized Complete Block Designs

(a) Residuals by block.

(b) Residuals by treatment

14 Design of Experiments with Several Factors

14-1 Introduction

- An **experiment** is a test or series of tests.

- The **design** of an experiment plays a major role in the eventual solution of the problem.

- In a **factorial experimental design**, experimental trials (or runs) are performed at all combinations of the factor levels.

- The **analysis of variance** (ANOVA) will be used as one of the primary tools for statistical data analysis.

14-2 Factorial Experiments

Definition

By a **factorial experiment** we mean that in each complete trial or replicate of the experiment all possible combinations of the levels of the factors are investigated.

Table 14-1 A Factorial Experiment with Two Factors

Factor A	Factor B	
	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	40

Table 14-2 A Factorial Experiment with Interaction

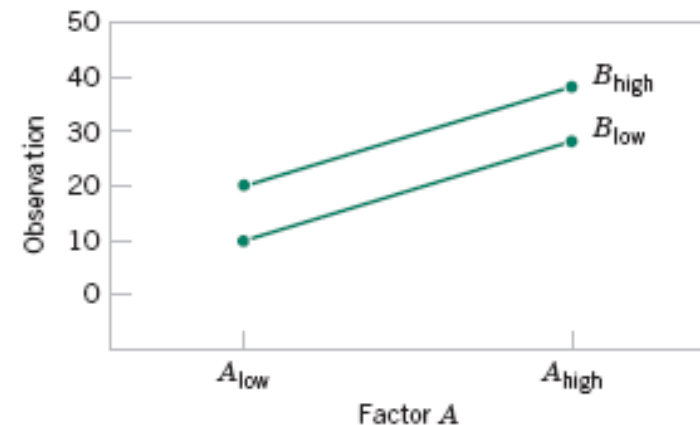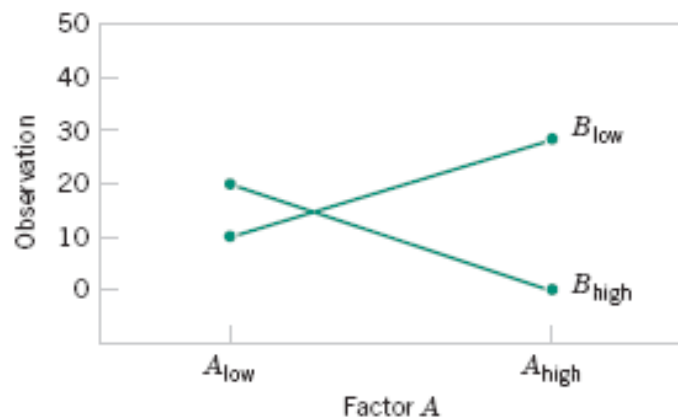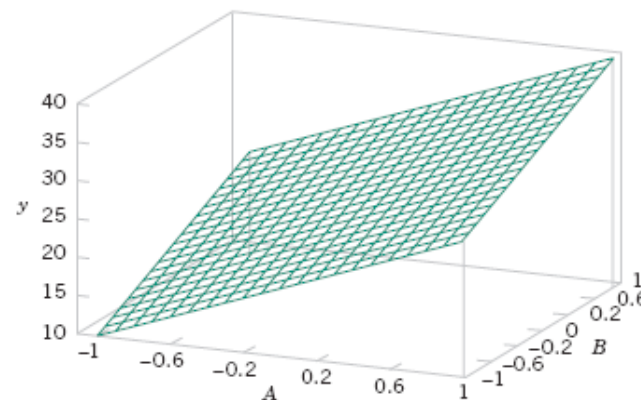| Factor A	Factor B	
	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	0

14-2 Factorial Experiments

Figure 14-3 Factorial Experiment, no interaction.
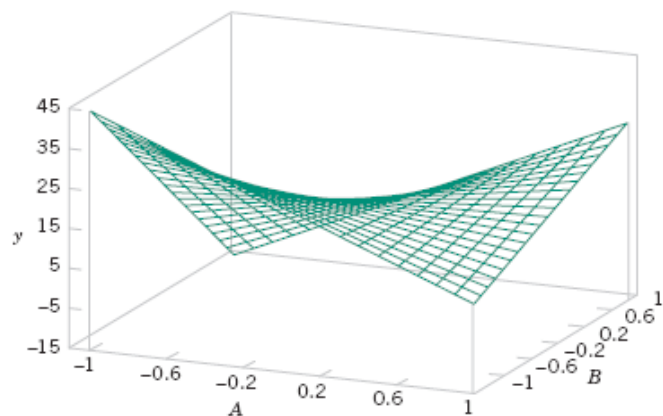
14-2 Factorial Experiments

Figure 14-4 Factorial Experiment, with interaction.
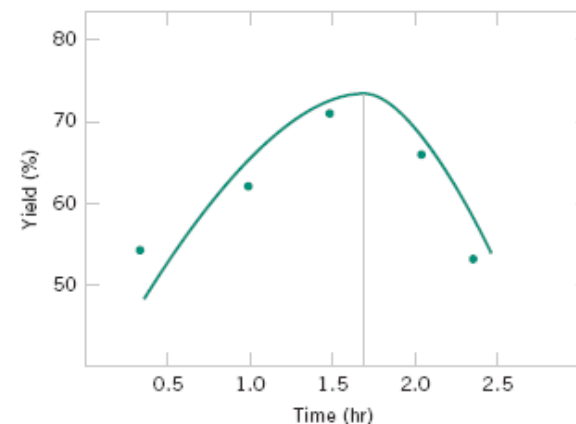
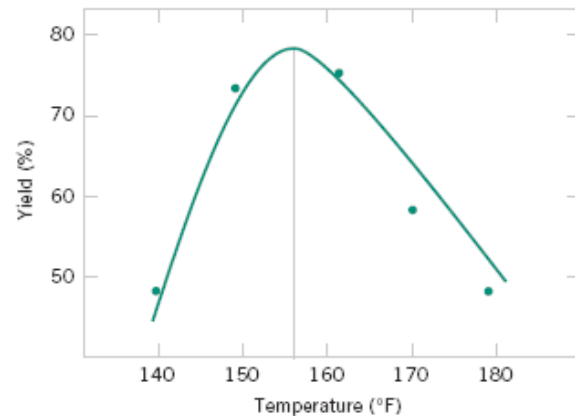**Figure 14-5** Three-dimensional surface plot of the data from Table 14-1, showing main effects of the two factors A and B.

Figure 14-6 Three-dimensional surface plot of the data from Table 14-2, showing main effects of the A and B interaction.

Figure 14-7 Yield versus reaction time with temperature constant at 155° F.

14-2 Factorial Experiments

Figure 14-8 Yield versus temperature with reaction time constant at 1.7 hours.
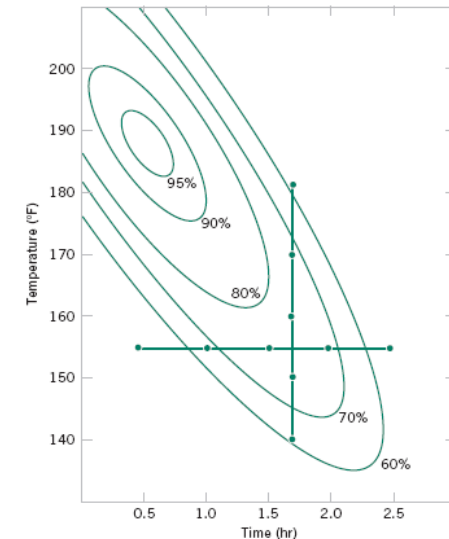
Figure 14-9 Optimization experiment using the one-factor-at-a-time method.

14-3 Two-Factor Factorial Experiments

Table 14-3 Data Arrangement for a Two-Factor Factorial Design

The observations may be described by the linear statistical model:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

$$y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn} \qquad i = 1, 2, \ldots, a$$

$$y_{.j.} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{an} \qquad j = 1, 2, \ldots, b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \qquad \begin{array}{l} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{array}$$

$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{...} = \frac{y_{...}}{abn}$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

The hypotheses that we will test are as follows:

1. $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$ (no main effect of factor A)
   $H_1:$ at least one $\tau_i \neq 0$

2. $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ (no main effect of factor B)
   $H_1:$ at least one $\beta_j \neq 0$

3. $H_0: (\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab} = 0$ (no interaction)
   $H_1:$ at least one $(\tau\beta)_{ij} \neq 0$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

The sum of squares identity for a two-factor ANOVA is

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{...})^2 = bn \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$+ \, an \sum_{j=1}^{b} (\bar{y}_{.j.} - \bar{y}_{...})^2$$

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

$$+ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij.})^2 \qquad (14\text{-}3)$$

or symbolically,

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \qquad (14\text{-}4)$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

To test $H_0: \tau_i = 0$ use the ratio

$$F_0 = \frac{MS_A}{MS_E}$$

To test $H_0: \beta_j = 0$ use the ratio

$$F_0 = \frac{MS_B}{MS_E}$$

To test $H_0: (\tau\beta)_{ij} = 0$ use the ratio

$$F_0 = \frac{MS_{AB}}{MS_E}$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Definition

Computing formulas for the sums of squares in a two-factor analysis of variance.

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn} \qquad (14\text{-}5)$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn} \qquad (14\text{-}6)$$

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn} \qquad (14\text{-}7)$$

$$SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_A - SS_B \qquad (14\text{-}8)$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B \qquad (14\text{-}9)$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Table 14-4 ANOVA Table for a Two-Factor Factorial, Fixed-Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \dfrac{SS_A}{a-1}$	$\dfrac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \dfrac{SS_B}{b-1}$	$\dfrac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \dfrac{SS_{AB}}{(a-1)(b-1)}$	$\dfrac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \dfrac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

Aircraft primer paints are applied to aluminum surfaces by two methods: dipping and spraying. The purpose of the primer is to improve paint adhesion, and some parts can be primed using either application method. The process engineering group responsible for this operation is interested in learning whether three different primers differ in their adhesion properties. A factorial experiment was performed to investigate the effect of paint primer type and application method on paint adhesion. For each combination of primer type and application method, three specimens were painted, then a finish paint was applied, and the adhesion force was measured. The data from the experiment are shown in Table 14-5. The circled numbers in the cells are the cell totals $y_{ij.}$. The sums of squares required to perform the ANOVA are computed as follows:

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

Table 14-5 Adhesion Force Data for Example 14-1

Primer Type	Dipping		Spraying		$y_{i..}$
1	4.0, 4.5, 4.3	12.8	5.4, 4.9, 5.6	15.9	28.7
2	5.6, 4.9, 5.4	15.9	5.8, 6.1, 6.3	18.2	34.1
3	3.8, 3.7, 4.0	11.5	5.5, 5.0, 5.0	15.5	27.0
$y_{.j.}$	40.2		49.6		89.8 = $y_{...}$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$= (4.0)^2 + (4.5)^2 + \cdots + (5.0)^2 - \frac{(89.8)^2}{18} = 10.72$$

$$SS_{types} = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$

$$= \frac{(28.7)^2 + (34.1)^2 + (27.0)^2}{6} - \frac{(89.8)^2}{18} = 4.58$$

$$SS_{methods} = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$= \frac{(40.2)^2 + (49.6)^2}{9} - \frac{(89.8)^2}{18} = 4.91$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

$$SS_{interaction} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_{types} - SS_{methods}$$

$$= \frac{(12.8)^2 + (15.9)^2 + (11.5)^2 + (15.9)^2 + (18.2)^2 + (15.5)^2}{3}$$

$$- \frac{(89.8)^2}{18} - 4.58 - 4.91 = 0.24$$

and

$$SS_E = SS_T - SS_{types} - SS_{methods} - SS_{interaction}$$

$$= 10.72 - 4.58 - 4.91 - 0.24 = 0.99$$

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

The ANOVA is summarized in Table 14-6. The experimenter has decided to use $\alpha = 0.05$. Since $f_{0.05,2,12} = 3.89$ and $f_{0.05,1,12} = 4.75$, we conclude that the main effects of primer type and application method affect adhesion force. Furthermore, since $1.5 < f_{0.05,2,12}$, there is no indication of interaction between these factors. The last column of Table 14-6 shows the P-value for each F-ratio. Notice that the P-values for the two test statistics for the main effects are considerably less than 0.05, while the P-value for the test statistic for the interaction is greater than 0.05.

A graph of the cell adhesion force averages $\{\bar{y}_{ij.}\}$ versus levels of primer type for each application method is shown in Fig. 14-8. The no-interaction conclusion is obvious in this graph, because the two lines are nearly parallel. Furthermore, since a large response indicates greater adhesion force, we conclude that spraying is the best application method and that primer type 2 is most effective.

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

Table 14-6   ANOVA for Example 14-1

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
Primer types	4.58	2	2.29	28.63	$2.7 \times$ E-5
Application methods	4.91	1	4.91	61.38	$4.7 \times$ E-7
Interaction	0.24	2	0.12	1.50	0.2621
Error	0.99	12	0.08		
Total	10.72	17			

14-3 Two-Factor Factorial Experiments

14-3.1 Statistical Analysis of the Fixed-Effects Model

Example 14-1

R commands and outputs
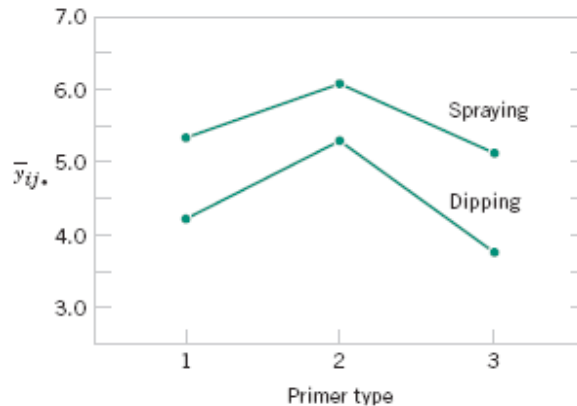
```
Example 14-1: enter data by row
> Adhesion=c(4.0, 4.5, 4.3, 5.4, 4.9, 5.6,  5.6, 4.9, 5.4, 5.8, 6.1, 6.3,  3.8,
  3.7, 4.0, 5.5, 5.0, 5.0)
> Primer=c(1,1,1,1,1,1, 2,2,2,2,2,2, 3,3,3,3,3,3)
> Method=c(1,1,1,2,2,2, 1,1,1,2,2,2, 1,1,1,2,2,2)  # 1=Dipping, 2=Spraying
> g=lm(Adhesion ~ as.factor(Primer) * as.factor(Method))
> anova(g)
Response: Adhesion
                                    Df Sum Sq Mean Sq F value      Pr(>F)
as.factor(Primer)                    2 4.5811  2.2906 27.8581 3.097e-05
as.factor(Method)                    1 4.9089  4.9089 59.7027 5.357e-06
as.factor(Primer):as.factor(Method)  2 0.2411  0.1206  1.4662    0.2693
Residuals                           12 0.9867  0.0822

> interaction.plot(Primer, Method, Adhesion)

See ch14.R for more commands
```

14-3 Two-Factor Factorial Experiments

14-3.2 Model Adequacy Checking
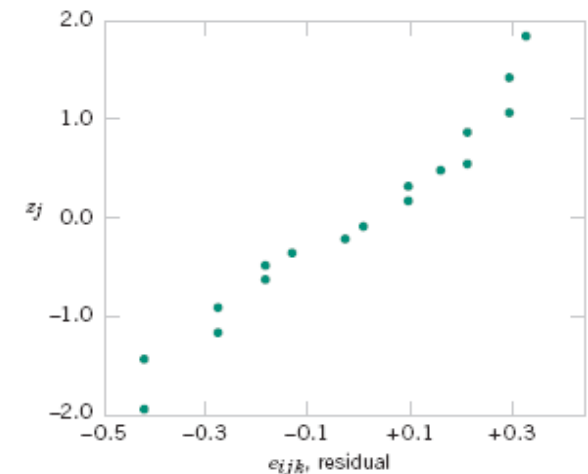
Table 14-8   Residuals for the Aircraft Primer Experiment in Example 14-1

Primer Type	Application Method	
	Dipping	**Spraying**
1	−0.27, 0.23, 0.03	0.10, −0.40, 0.30
2	0.30, −0.40, 0.10	−0.27, 0.03, 0.23
3	−0.03, −0.13, 0.17	0.33, −0.17, −0.17

14-3 Two-Factor Factorial Experiments

14-3.2 Model Adequacy Checking

14-3 Two-Factor Factorial Experiments

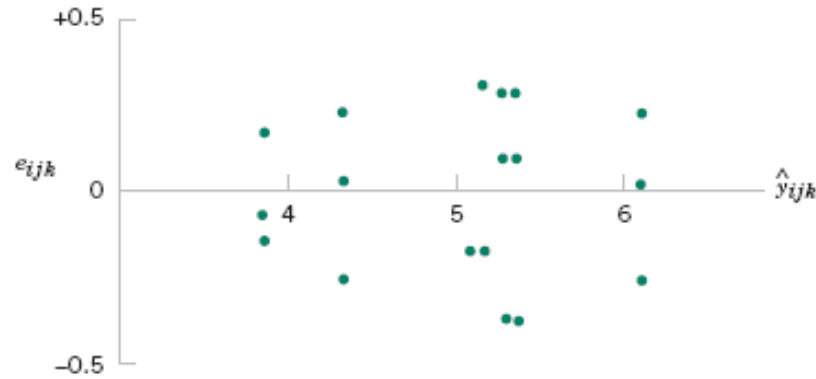## 14-3.2 Model Adequacy Checking

Figure 14-14 Plot of residuals versus predicted values.

14-4 General Factorial Experiments

Model for a **three-factor factorial experiment**

$$Y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk}$$

$$+ (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, c \\ l = 1, 2, \ldots, n \end{cases}$$

Table 14-9 Analysis of Variance Table for the Three-Factor Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Squares	F_0
A	SS_A	$a - 1$	MS_A	$\sigma^2 + \dfrac{bcn\Sigma\tau_i^2}{a-1}$	$\dfrac{MS_A}{MS_E}$
B	SS_B	$b - 1$	MS_B	$\sigma^2 + \dfrac{acn\Sigma\beta_j^2}{b-1}$	$\dfrac{MS_B}{MS_E}$
C	SS_C	$c - 1$	MS_C	$\sigma^2 + \dfrac{abn\Sigma\gamma_k^2}{c-1}$	$\dfrac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \dfrac{cn\Sigma\Sigma(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\dfrac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \dfrac{bn\Sigma\Sigma(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\dfrac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \dfrac{an\Sigma\Sigma(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\dfrac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \dfrac{n\Sigma\Sigma\Sigma(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\dfrac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

14-4 General Factorial Experiments

Example 14-2

A mechanical engineer is studying the surface roughness of a part produced in a metal-cutting operation. Three factors, feed rate (A), depth of cut (B), and tool angle (C), are of interest. All three factors have been assigned two levels, and two replicates of a factorial design are run. The coded data are shown in Table 14-10.

Table 14-10 Coded Surface Roughness Data for Example 14-2

Feed Rate (A)	Depth of Cut (B)				$y_{i\ldots}$
	0.025 inch		0.040 inch		
	Tool Angle (C)		Tool Angle (C)		
	15°	25°	15°	25°	
20 inches per minute	9	11	9	10	
	7	10	11	8	75
30 inches per minute	10	10	12	16	
	12	13	15	14	102

R commands and outputs

```
Example 14-2: enter data by row
> Roughness=c(9,11,9,10, 7,10,11,8, 10,10,12,16, 12,13,15,14)
> Feed=c(1,1,1,1, 1,1,1,1, 2,2,2,2, 2,2,2,2)
> Depth=c(1,1,2,2, 1,1,2,2, 1,1,2,2, 1,1,2,2)
> Angle=c(1,2,1,2, 1,2,1,2, 1,2,1,2, 1,2,1,2)
> g=lm(Roughness ~ Feed*Depth*Angle)
> anova(g)
Response: Roughness
                 Df Sum Sq Mean Sq F value   Pr(>F)
Feed              1 45.562  45.562 18.6923 0.002534 **
Depth             1 10.562  10.562  4.3333 0.070931 .
Angle             1  3.062   3.062  1.2564 0.294849
Feed:Depth        1  7.562   7.562  3.1026 0.116197
Feed:Angle        1  0.062   0.062  0.0256 0.876749
Depth:Angle       1  1.562   1.562  0.6410 0.446463
Feed:Depth:Angle  1  5.062   5.062  2.0769 0.187512
Residuals         8 19.500   2.438

> par(mfrow=c(1,3)) #
> interaction.plot(Feed, Depth, Roughness)
> interaction.plot(Feed, Angle, Roughness)
> interaction.plot(Angle, Depth, Roughness)
```

14-4 General Factorial Experiments

Example 14-2

The F-ratios for all three main effects and the interactions are formed by dividing the mean square for the effect of interest by the error mean square. Since the experimenter has selected $\alpha = 0.05$, the critical value for each of these F-ratios is $f_{0.05,1,8} = 5.32$. Alternately, we could use the P-value approach. The P-values for all the test statistics are shown in the last column of Table 14-11. Inspection of these P-values is revealing. There is a strong main effect of feed rate, since the F-ratio is well into the critical region. However, there is some indication of an effect due to the depth of cut, since $P = 0.0710$ is not much greater than $\alpha = 0.05$. The next largest effect is the AB or feed rate × depth of cut interaction. Most likely, both feed rate and depth of cut are important process variables.

15

Nonparametric
Statistics

CHAPTER OUTLINE

Bootstrap Method

15-1 Introduction

- Most of the hypothesis-testing and confidence interval procedures discussed in previous chapters are based on the assumption that we are working with random samples from normal populations.

- These procedures are often called **parametric methods**

- In this chapter, **nonparametric** and **distribution free** methods will be discussed.

- We usually make no assumptions about the distribution of the underlying population.

15-2 Sign Test

15-2.1 Description of the Test

• The **sign test** is used to test hypotheses about the **median** of a continuous distribution.

$$H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} < \tilde{\mu}_0$$

• Let R^+ represent the number of differences

$$X_i - \tilde{\mu}_0$$

that are positive.

• What is the sampling distribution of R^+ under H_0?

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

$$H_0: \tilde{\mu} = \tilde{\mu}_0$$
$$H_1: \tilde{\mu} < \tilde{\mu}_0$$

The appropriate P-value is

$$P = P\left(R^+ \le r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

$$H_0: \tilde{\mu} = \tilde{\mu}_0$$
$$H_1: \tilde{\mu} > \tilde{\mu}_0$$

The appropriate P-value is

$$P = P\left(R^+ \ge r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

$$H_0: \tilde{\mu} = \tilde{\mu}_0$$
$$H_1: \tilde{\mu} \ne \tilde{\mu}_0$$

If $r^+ < n/2$, then the appropriate P-value is

$$P = 2P\left(R^+ \le r^+ \text{ when } p = \frac{1}{2}\right)$$

If $r^+ > n/2$, then the appropriate P-value is

$$P = 2P\left(R^+ \ge r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

Example 15-1

Montgomery, Peck, and Vining (2001) report on a study in which a rocket motor is formed by binding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two propellant types is an important characteristic. The results of testing 20 randomly selected motors are shown in Table 15-1. We would like to test the hypothesis that the median shear strength is 2000 psi, using $\alpha = 0.05$.

This problem can be solved using the eight-step hypothesis-testing procedure introduced in Chapter 9:

1. The parameter of interest is the median of the distribution of propellant shear strength.
2. H_0: $\tilde{\mu} = 2000$ psi
3. H_1: $\tilde{\mu} \neq 2000$ psi
4. $\alpha = 0.05$
5. The test statistic is the observed number of plus differences in Table 15-1, or $r^+ = 14$.
6. We will reject H_0 if the P-value corresponding to $r^+ = 14$ is less than or equal to $\alpha = 0.05$.

Example 15-1

Table 15-1 Propellant Shear Strength Data

Observation i	Shear Strength x_i	Differences $x_i - 2000$	Sign
1	2158.70	+158.70	+
2	1678.15	−321.85	−
3	2316.00	+316.00	+
4	2061.30	+61.30	+
5	2207.50	+207.50	+
6	1708.30	−291.70	−
7	1784.70	−215.30	−
8	2575.10	+575.10	+
9	2357.90	+357.90	+
10	2256.70	+256.70	+
11	2165.20	+165.20	+
12	2399.55	+399.55	+
13	1779.80	−220.20	−
14	2336.75	+336.75	+
15	1765.30	−234.70	−
16	2053.50	+53.50	+
17	2414.40	+414.40	+
18	2200.50	+200.50	+
19	2654.20	+654.20	+
20	1753.70	−246.30	−

15-2 Sign Test

Example 15-1

7. Computations: Since $r^+ = 14$ is greater than $n/2 = 20/2 = 10$, we calculate the P-value from

$$P = 2P\left(R^+ \geq 14 \text{ when } p = \frac{1}{2}\right)$$

$$= 2 \sum_{r=14}^{20} \binom{20}{r}(0.5)^r(0.5)^{20-r}$$

$$= 0.1153$$

8. Conclusions: Since $P = 0.1153$ is not less than $\alpha = 0.05$, we cannot reject the null hypothesis that the median shear strength is 2000 psi. Another way to say this is that the observed number of plus signs $r^+ = 14$ was not large or small enough to indicate that median shear strength is different from 2000 psi at the $\alpha = 0.05$ level of significance.

15-2 Sign Test

15-2.2 Sign Test for Paired Samples

The sign test can also be applied to paired observations drawn from continuous populations. Let (X_{1j}, X_{2j}), $j = 1, 2, \ldots, n$ be a collection of paired observations from two continuous populations, and let

$$D_j = X_{1j} - X_{2j} \quad j = 1, 2, \ldots, n$$

be the paired differences. We wish to test the hypothesis that the two populations have a common median, that is, that $\tilde{\mu}_1 = \tilde{\mu}_2$. This is equivalent to testing that the median of the differences $\tilde{\mu}_D = 0$. This can be done by applying the sign test to the n observed differences d_j, as illustrated in the following example.
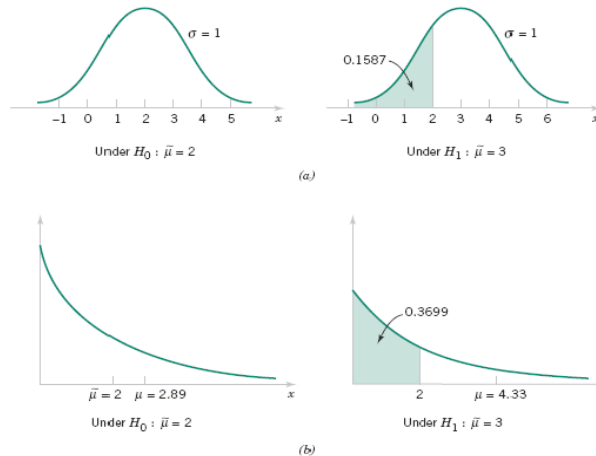
See Example 15-3.

15-2 Sign Test

15-2.3 Type II Error for the Sign Test

• Depends on both the true population distribution and alternative value!

Figure 15-1
Calculation of β for the sign test. (a) Normal distributions. (b) Exponential distributions

15-3 Wilcoxon Signed-Rank Test

• The **Wilcoxon signed-rank test** applies to the case of symmetric continuous distributions.

• Under this assumption, the mean equals the median.

• The null hypothesis is $H_0: \mu = \mu_0$

15-3 Wilcoxon Signed-Rank Test

• 15-3.1 Description of the Test

• Assume that $X_1, X_2, \ldots, X_n$ is a random sample from a continuous and symmetric distribution with mean (and median) μ.

Procedure:

• Compute the differences $X_i - \mu_0$, $i = 1, 2, \ldots, n$.
• Rank the absolute differences $|X_i - \mu_0|$, $i = 1, 2, \ldots, n$ in ascending order.
• Give the ranks the signs of their corresponding differences.
• Let W^+ be the sum of the positive ranks and W^- be the absolute value of the sum of the negative ranks.
• Let $W = \min(W^+, W^-)$.

15-3 Wilcoxon Signed-Rank Test

Decision rules:

Appendix Table IX contains critical values of W, say w_α^*

If the alternative is $H_1: \mu \neq \mu_0$, reject $H_0: \mu = \mu_0$ if $w \leq w_\alpha^*$

If the alternative is $H_1: \mu > \mu_0$, reject $H_0: \mu = \mu_0$ if $w^- \leq w_\alpha^*$

If the alternative is $H_1: \mu < \mu_0$, reject $H_0: \mu = \mu_0$ if $w^+ \leq w_\alpha^*$

Appendix Table IX provides significance levels of $\alpha = 0.10$, $\alpha = 0.05$, $\alpha = 0.02$, $\alpha = 0.01$ for the two-sided test.

The significance levels for one-sided tests provided in Appendix Table IX are $\alpha = 0.05, 0.025, 0.01$, and 0.005.

15-3 Wilcoxon Signed-Rank Test

Example 15-4

We will illustrate the Wilcoxon signed-rank test by applying it to the propellant shear strength data from Table 15-1. Assume that the underlying distribution is a continuous symmetric distribution. The eight-step procedure is applied as follows:

1. The parameter of interest is the mean (or median) of the distribution of propellant shear strength.
2. H_0: $\mu = 2000$ psi
3. H_1: $\mu \neq 2000$ psi
4. $\alpha = 0.05$
5. The test statistic is

$$w = \min(w^+, w^-)$$

6. We will reject H_0 if $w \leq w^*_{0.05} = 52$ from Appendix Table VIII.

Example 15-4

7. Computations: The signed ranks from Table 15-1 are shown in the following table:

Observation	Difference $x_i - 2000$	Signed Rank
16	+53.50	+1
4	+61.30	+2
1	+158.70	+3
11	+165.20	+4
18	+200.50	+5
5	+207.50	+6
7	−215.30	−7
13	−220.20	−8
15	−234.70	−9
20	−246.30	−10
10	+256.70	+11
6	−291.70	−12
3	+316.00	+13
2	−321.85	−14
14	+336.75	+15
9	+357.90	+16
12	+399.55	+17
17	+414.40	+18
8	+575.10	+19
19	+654.20	+20

15-3 Wilcoxon Signed-Rank Test

Example 15-4

The sum of the positive ranks is $w^+ = (1 + 2 + 3 + 4 + 5 + 6 + 11 + 13 + 15 + 16 + 17 + 18 + 19 + 20) = 150$, and the sum of the absolute values of the negative ranks is $w^- = (7 + 8 + 9 + 10 + 12 + 14) = 60$. Therefore,

$$w = \min(150, 60) = 60$$

8. Conclusions: Since $w = 60$ is not less than or equal to the critical value $w_{0.05} = 52$, we cannot reject the null hypothesis that the mean (or median, since the population is assumed to be symmetric) shear strength is 2000 psi.

15-3 Wilcoxon Signed-Rank Test

15-3.2 Large-Sample Approximation

Therefore, a test of H_0: $\mu = \mu_0$ can be based on the statistic

$$Z_0 = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \qquad (15\text{-}6)$$

Z_0 is approximately standard normal when n is large.

15-4 Wilcoxon Rank-Sum Test

Suppose that we have two independent continuous populations X_1 and X_2 with means μ_1 and μ_2. Assume that the distributions of X_1 and X_2 have the same shape and spread and differ only (possibly) in their locations. The Wilcoxon rank-sum test can be used to test the hypothesis H_0: $\mu_1 = \mu_2$. This procedure is sometimes called the Mann-Whitney test, although the Mann-Whitney test statistic is usually expressed in a different form.

15-4.1 Description of the Test

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes $n_1 \leq n_2$ from the continuous populations X_1 and X_2 described earlier. We wish to test the hypotheses

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

15-4.1 Description of the Test

The test procedure is as follows. Arrange all $n_1 + n_2$ observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical), use the mean of the ranks that would have been assigned if the observations differed.

Let W_1 be the sum of the ranks in the smaller sample (1), and define W_2 to be the sum of the ranks in the other sample. Then,

$$W_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - W_1 \qquad (15\text{-}7)$$

Now if the sample means do not differ, we will expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we will conclude that the means are not equal.

Example 15-6

The mean axial stress in tensile members used in an aircraft structure is being studied. Two alloys are being investigated. Alloy 1 is a traditional material, and alloy 2 is a new aluminum-lithium alloy that is much lighter than the standard material. Ten specimens of each alloy type are tested, and the axial stress is measured. The sample data are assembled in Table 15-3. Using $\alpha = 0.05$, we wish to test the hypothesis that the means of the two stress distributions are identical.

TABLE 15-3

Axial Stress for Two Aluminum-Lithium Alloys

Alloy 1		Alloy 2	
3238 psi	3254 psi	3261 psi	3248 psi
3195	3229	3187	3215
3246	3225	3209	3226
3190	3217	3212	3240
3204	3241	3258	3234

Example 15-6

We will apply the eight-step hypothesis-testing procedure to this problem:

1. The parameters of interest are the means of the two distributions of axial stress.
2. H_0: $\mu_1 = \mu_2$
3. H_1: $\mu_1 \neq \mu_2$
4. $\alpha = 0.05$
5. We will use the Wilcoxon rank-sum test statistic in Equation 15-7,

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

6. Since $\alpha = 0.05$ and $n_1 = n_2 = 10$, Appendix Table IX gives the critical value as $w_{0.05} = 78$. If either w_1 or w_2 is less than or equal to $w_{0.05} = 78$, we will reject H_0: $\mu_1 = \mu_2$.

Example 15-6

7. Computations: The data from Table 15-3 are analyzed in ascending order and ranked as follows:

Alloy Number	Axial Stress	Rank
2	3187 psi	1
1	3190	2
1	3195	3
1	3204	4
2	3209	5
2	3212	6
2	3215	7
1	3217	8
1	3225	9
2	3226	10
1	3229	11
2	3234	12
1	3238	13
2	3240	14
1	3241	15
1	3246	16
2	3248	17
1	3254	18
2	3258	19
2	3261	20

15-5 Nonparametric Methods in the Analysis of Variance

The single-factor analysis of variance model for comparing a population means is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n_i \end{cases}$$

The hypothesis of interest is

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$$

The Kruskal-Wallis test (w/o assumption of normality)

• Basic idea: Use **ranks** instead of actual numbers

15-4 Wilcoxon Rank-Sum Test

Example 15-6

The sum of the ranks for alloy 1 is

$$w_1 = 2 + 3 + 4 + 8 + 9 + 11 + 13 + 15 + 16 + 18 = 99$$

and for alloy 2

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 = \frac{(10 + 10)(10 + 10 + 1)}{2} - 99 = 111$$

8. Conclusions: Since neither w_1 nor w_2 is less than or equal to $w_{0.05} = 78$, we cannot reject the null hypothesis that both alloys exhibit the same mean axial stress.

Parametric vs. Nonparametric Tests

• When the normality assumption is correct, t-test or F-test is more powerful.
  – Wilcoxon signed-rank or rank-sum test is approximately 95% as efficient as the t-test in large samples.
• On the other hand, regardless of the form of the distributions, nonparametric tests may be more powerful.
  – Wilcoxon signed-rank or rank-sum test will always be at least 86% as efficient.
• The efficiency of the Wilcoxon test relative to the t-test is usually high if the underlying distribution has heavier tails than the normal
  – because the behavior of the t-test is very dependent on the sample mean, which is quite unstable in heavy-tailed distributions.