Probability & Statistics
for Engineers & Scientists

NINTH EDITION

WALPOLE | MYERS | MYERS | YE

Probability & Statistics
for Engineers & Scientists

Probability & Statistics for Engineers & Scientists

NINTH EDITION

Ronald E. Walpole
Roanoke College

Raymond H. Myers
Virginia Tech

Sharon L. Myers
Radford University

Keying Ye
University of Texas at San Antonio

Prentice Hall
Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: *Deirdre Lynch*
Acquisitions Editor: *Christopher Cummings*
Executive Content Editor: *Christine O'Brien*
Associate Editor: *Christina Lepre*
Senior Managing Editor: *Karen Wernholm*
Senior Production Project Manager: *Tracy Patruno*
Design Manager: *Andrea Nix*
Cover Designer: *Heather Scott*
Digital Assets Manager: *Marianne Groth*
Associate Media Producer: *Vicki Dreyfus*
Marketing Manager: *Alex Gay*
Marketing Assistant: *Kathleen DeChavez*
Senior Author Support/Technology Specialist: *Joe Vetere*
Rights and Permissions Advisor: *Michael Joyce*
Senior Manufacturing Buyer: *Carol Melville*
Production Coordination: *Lifland et al. Bookmakers*
Composition: *Keying Ye*
Cover photo: *Marjory Dressler/Dressler Photo-Graphics*

Library of Congress Cataloging-in-Publication Data

Probability & statistics for engineers & scientists/Ronald E. Walpole ... [et al.] — 9th ed.
p. cm.
ISBN 978-0-321-62911-1
1. Engineering—Statistical methods. 2. Probabilities. I. Walpole, Ronald E.

TA340.P738 2011
519.02'462–dc22

2010004857

1 2 3 4 5 6 7 8 9 10—EB—14 13 12 11 10

Prentice Hall
is an imprint of
PEARSON

www.pearsonhighered.com

ISBN 10: 0-321-62911-6
ISBN 13: 978-0-321-62911-1

This book is dedicated to

Billy and Julie
R.H.M. and S.L.M.
Limin, Carolyn and Emily
K.Y.

Contents

3 Random Variables and Probability Distributions 81

4 Mathematical Expectation 111

5 Some Discrete Probability Distributions 143

6 Some Continuous Probability Distributions 171

7 Functions of Random Variables (Optional) 211

8 Fundamental Sampling Distributions and Data Descriptions 225

9 One- and Two-Sample Estimation Problems 265

10 One- and Two-Sample Tests of Hypotheses 319

11 Simple Linear Regression and Correlation 389

12 Multiple Linear Regression and Certain Nonlinear Regression Models 443

13 One-Factor Experiments: General 507

14 Factorial Experiments (Two or More Factors) 561

15 2^k Factorial Experiments and Fractions 597

16 Nonparametric Statistics 655

Chapter 11

Simple Linear Regression and Correlation

11.1 Introduction to Linear Regression

Often, in practice, one is called upon to solve problems involving sets of variables when it is known that there exists some inherent relationship among the variables. For example, in an industrial situation it may be known that the tar content in the outlet stream in a chemical process is related to the inlet temperature. It may be of interest to develop a method of prediction, that is, a procedure for estimating the tar content for various levels of the inlet temperature from experimental information. Now, of course, it is highly likely that for many example runs in which the inlet temperature is the same, say 130°C, the outlet tar content will not be the same. This is much like what happens when we study several automobiles with the same engine volume. They will not all have the same gas mileage. Houses in the same part of the country that have the same square footage of living space will not all be sold for the same price. Tar content, gas mileage (mpg), and the price of houses (in thousands of dollars) are natural **dependent variables**, or responses, in these three scenarios. Inlet temperature, engine volume (cubic feet), and square feet of living space are, respectively, natural **independent variables**, or **regressors**. A reasonable form of a relationship between the **response $\boldsymbol{Y}$** and the regressor x is the linear relationship

$$Y = \beta_0 + \beta_1 x,$$

where, of course, β_0 is the **intercept** and β_1 is the **slope**. The relationship is illustrated in Figure 11.1.

If the relationship is exact, then it is a **deterministic** relationship between two scientific variables and there is no random or probabilistic component to it. However, in the examples listed above, as well as in countless other scientific and engineering phenomena, the relationship is not deterministic (i.e., a given x does not always give the same value for Y). As a result, important problems here are probabilistic in nature since the relationship above cannot be viewed as being exact. The concept of **regression analysis** deals with finding the best relationship

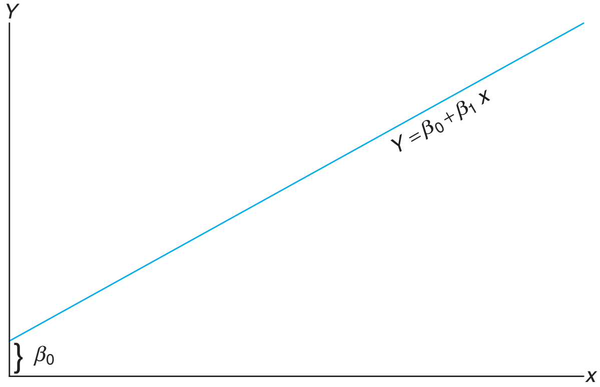


Figure 11.1: A linear relationship; β_0: intercept; β_1: slope.

between Y and x, quantifying the strength of that relationship, and using methods that allow for prediction of the response values given values of the regressor x.

In many applications, there will be more than one regressor (i.e., more than one independent variable **that helps to explain Y**). For example, in the case where the response is the price of a house, one would expect the age of the house to contribute to the explanation of the price, so in this case the multiple regression structure might be written

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

where Y is price, x_1 is square footage, and x_2 is age in years. In the next chapter, we will consider problems with multiple regressors. The resulting analysis is termed **multiple regression**, while the analysis of the single regressor case is called **simple regression**. As a second illustration of multiple regression, a chemical engineer may be concerned with the amount of hydrogen lost from samples of a particular metal when the material is placed in storage. In this case, there may be two inputs, storage time x_1 in hours and storage temperature x_2 in degrees centigrade. The response would then be hydrogen loss Y in parts per million.

In this chapter, we deal with the topic of **simple linear regression**, treating only the case of a single regressor variable in which the relationship between y and x is linear. For the case of more than one regressor variable, the reader is referred to Chapter 12. Denote a random sample of size n by the set $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$. If additional samples were taken using exactly the same values of x, we should expect the y values to vary. Hence, the value y_i in the ordered pair (x_i, y_i) is a value of some random variable Y_i.

11.2 The Simple Linear Regression (SLR) Model

We have already confined the terminology *regression analysis* to situations in which relationships among variables are not deterministic (i.e., not exact). In other words, there must be a **random component** to the equation that relates the variables.

This random component takes into account considerations that are not being measured or, in fact, are not understood by the scientists or engineers. Indeed, in most applications of regression, the linear equation, say $Y = \beta_0 + \beta_1 x$, is an approximation that is a simplification of something unknown and much more complicated. For example, in our illustration involving the response Y= tar content and x = inlet temperature, $Y = \beta_0 + \beta_1 x$ is likely a reasonable approximation that may be operative within a confined range on x. More often than not, the models that are simplifications of more complicated and unknown structures are linear in nature (i.e., linear in the **parameters** β_0 and β_1 or, in the case of the model involving the price, size, and age of the house, linear in the **parameters** β_0, β_1, and β_2). These linear structures are simple and empirical in nature and are thus called **empirical models**.

An analysis of the relationship between Y and x requires the statement of a **statistical model**. A model is often used by a statistician as a representation of an **ideal** that essentially defines how we perceive that the data were generated by the system in question. The model must include the set $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$ of data involving n pairs of (x, y) values. One must bear in mind that the value y_i depends on x_i via a linear structure that also has the random component involved. The basis for the use of a statistical model relates to how the random variable Y moves with x and the random component. The model also includes what is assumed about the statistical properties of the random component. The statistical model for simple linear regression is given below. The response Y is related to the independent variable x through the equation

Simple Linear Regression Model

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

In the above, β_0 and β_1 are unknown intercept and slope parameters, respectively, and ϵ is a random variable that is assumed to be distributed with $E(\epsilon) = 0$ and $\text{Var}(\epsilon) = \sigma^2$. The quantity σ^2 is often called the error variance or residual variance.

From the model above, several things become apparent. The quantity Y is a random variable since ϵ is random. The value x of the regressor variable is not random and, in fact, is measured with negligible error. The quantity ϵ, often called a **random error** or **random disturbance**, has constant variance. This portion of the assumptions is often called the **homogeneous variance assumption**. The presence of this random error, ϵ, keeps the model from becoming simply a deterministic equation. Now, the fact that $E(\epsilon) = 0$ implies that at a specific x the y-values are distributed around the **true**, or population, **regression line** $y = \beta_0 + \beta_1 x$. If the model is well chosen (i.e., there are no additional important regressors and the linear approximation is good within the ranges of the data), then positive and negative errors around the true regression are reasonable. We must keep in mind that in practice β_0 and β_1 are not known and must be estimated from data. In addition, the model described above is conceptual in nature. As a result, we never observe the actual ϵ values in practice and thus we can never draw the true regression line (but we assume it is there). We can only draw an estimated line. Figure 11.2 depicts the nature of hypothetical (x, y) data scattered around a true regression line for a case in which only $n = 5$ observations are available. Let us emphasize that what we see in Figure 11.2 is not the line that is used by the scientist or engineer. Rather, the picture merely describes what the assumptions mean! The regression that the user has at his or her disposal will now be described.

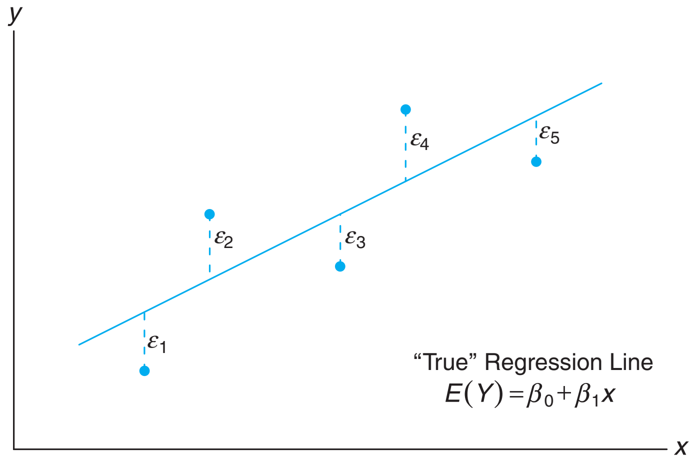


Figure 11.2: Hypothetical (x, y) data scattered around the true regression line for $n = 5$.

The Fitted Regression Line

An important aspect of regression analysis is, very simply, to estimate the parameters β_0 and β_1 (i.e., estimate the so-called **regression coefficients**). The method of estimation will be discussed in the next section. Suppose we denote the estimates b_0 for β_0 and b_1 for β_1. Then the estimated or **fitted regression** line is given by

$$\hat{y} = b_0 + b_1 x,$$

where $\hat{y}$ is the predicted or fitted value. Obviously, the fitted line is an estimate of the true regression line. We expect that the fitted line should be closer to the true regression line when a large amount of data are available. In the following example, we illustrate the fitted line for a real-life pollution study.

One of the more challenging problems confronting the water pollution control field is presented by the tanning industry. Tannery wastes are chemically complex. They are characterized by high values of chemical oxygen demand, volatile solids, and other pollution measures. Consider the experimental data in Table 11.1, which were obtained from 33 samples of chemically treated waste in a study conducted at Virginia Tech. Readings on x, the percent reduction in total solids, and y, the percent reduction in chemical oxygen demand, were recorded.

The data of Table 11.1 are plotted in a **scatter diagram** in Figure 11.3. From an inspection of this scatter diagram, it can be seen that the points closely follow a straight line, indicating that the assumption of linearity between the two variables appears to be reasonable.

Table 11.1: Measures of Reduction in Solids and Oxygen Demand

Solids Reduction, x (%)	**Oxygen Demand Reduction, y (%)**	**Solids Reduction, x (%)**	**Oxygen Demand Reduction, y (%)**
3	5	36	34
7	11	37	36
11	21	38	38
15	16	39	37
18	16	39	36
27	28	39	45
29	27	40	39
30	25	41	41
30	35	42	40
31	30	42	44
31	40	43	37
32	32	44	44
33	34	45	46
33	32	46	46
34	34	47	49
36	37	50	51
36	38		

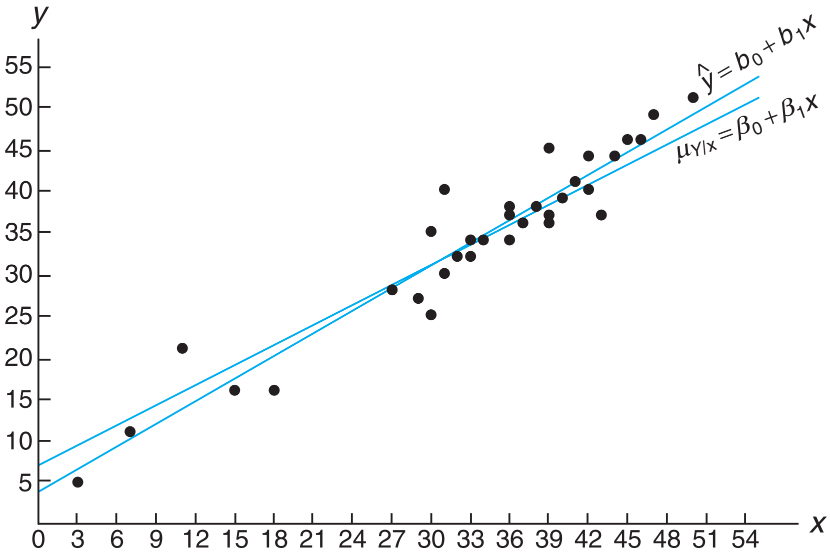


Figure 11.3: Scatter diagram with regression lines.

The fitted regression line and a *hypothetical true regression line* are shown on the scatter diagram of Figure 11.3. This example will be revisited as we move on to the method of estimation, discussed in Section 11.3.

Another Look at the Model Assumptions

It may be instructive to revisit the simple linear regression model presented previously and discuss in a graphical sense how it relates to the so-called true regression. Let us expand on Figure 11.2 by illustrating not merely where the ϵ_i fall on a graph but also what the implication is of the normality assumption on the ϵ_i.

Suppose we have a simple linear regression with $n = 6$ evenly spaced values of x and a single y-value at each x. Consider the graph in Figure 11.4. This illustration should give the reader a clear representation of the model and the assumptions involved. The line in the graph is the true regression line. The points plotted are actual (y, x) points which are scattered about the line. Each point is on its own normal distribution with the center of the distribution (i.e., the mean of y) falling on the line. This is certainly expected since $E(Y) = \beta_0 + \beta_1 x$. As a result, the true regression line **goes through the means of the response**, and the actual observations are on the distribution around the means. Note also that all distributions have the same variance, which we referred to as σ^2. Of course, the deviation between an individual y and the point on the line will be its individual ϵ value. This is clear since

$$y_i - E(Y_i) = y_i - (\beta_0 + \beta_1 x_i) = \epsilon_i.$$

Thus, at a given x, Y and the corresponding ϵ both have variance σ^2.

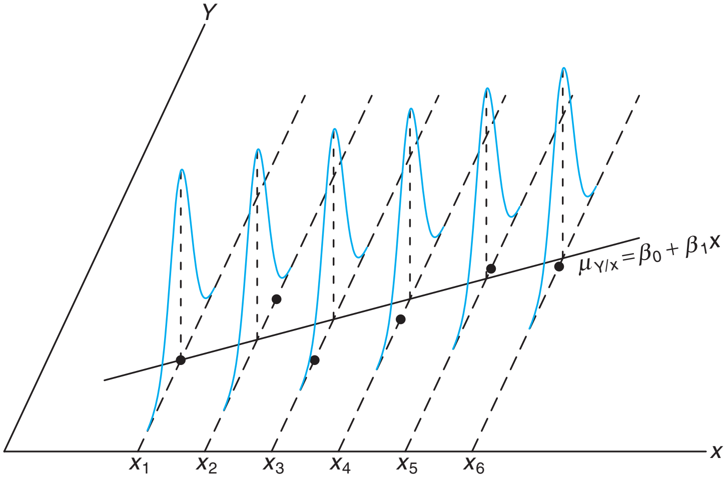


Figure 11.4: Individual observations around true regression line.

Note also that we have written the true regression line here as $\mu_{Y|x} = \beta_0 + \beta_1 x$ in order to reaffirm that the line goes through the mean of the Y random variable.

11.3 Least Squares and the Fitted Model

In this section, we discuss the method of fitting an estimated regression line to the data. This is tantamount to the determination of estimates b_0 for β_0 and b_1

for β_1. This of course allows for the computation of predicted values from the fitted line $\hat{y} = b_0 + b_1 x$ and other types of analyses and diagnostic information that will ascertain the strength of the relationship and the adequacy of the fitted model. Before we discuss the method of least squares estimation, it is important to introduce the concept of a **residual**. A residual is essentially an error in the fit of the model $\hat{y} = b_0 + b_1 x$.

Residual: Error in Fit

Given a set of regression data $\{(x_i, y_i); i = 1, 2, \ldots, n\}$ and a fitted model, $\hat{y}_i = b_0 + b_1 x_i$, the ith residual e_i is given by

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \ldots, n.$$

Obviously, if a set of n residuals is large, then the fit of the model is not good. Small residuals are a sign of a good fit. Another interesting relationship which is useful at times is the following:

$$y_i = b_0 + b_1 x_i + e_i.$$

The use of the above equation should result in clarification of the distinction between the residuals, e_i, and the conceptual model errors, ϵ_i. One must bear in mind that whereas the ϵ_i are not observed, the e_i not only are observed but also play an important role in the total analysis.

Figure 11.5 depicts the line fit to this set of data, namely $\hat{y} = b_0 + b_1 x$, and the line reflecting the model $\mu_{Y|x} = \beta_0 + \beta_1 x$. Now, of course, β_0 and β_1 are unknown parameters. The fitted line is an estimate of the line produced by the statistical model. Keep in mind that the line $\mu_{Y|x} = \beta_0 + \beta_1 x$ is not known.

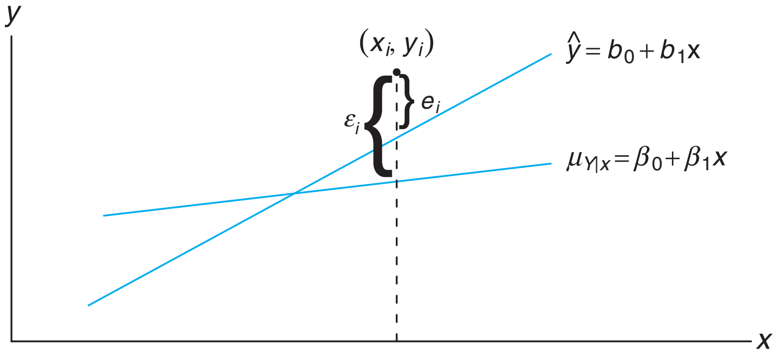


Figure 11.5: Comparing ϵ_i with the residual, e_i.

The Method of Least Squares

We shall find b_0 and b_1, the estimates of β_0 and β_1, so that the sum of the squares of the residuals is a minimum. The residual sum of squares is often called the sum of squares of the errors about the regression line and is denoted by SSE. This minimization procedure for estimating the parameters is called the **method of least squares**. Hence, we shall find a and b so as to minimize

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2.$$

Differentiating SSE with respect to b_0 and b_1, we have

$$\frac{\partial(SSE)}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i), \quad \frac{\partial(SSE)}{\partial b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i.$$

Setting the partial derivatives equal to zero and rearranging the terms, we obtain the equations (called the **normal equations**)

$$nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \quad b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i,$$

which may be solved simultaneously to yield computing formulas for b_0 and b_1.

Estimating the Regression Coefficients

Given the sample $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$, the least squares estimates b_0 and b_1 of the regression coefficients β_0 and β_1 are computed from the formulas

$$b_1 = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \text{ and}$$

$$b_0 = \frac{\sum\limits_{i=1}^{n} y_i - b_1 \sum\limits_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}.$$

The calculations of b_0 and b_1, using the data of Table 11.1, are illustrated by the following example.

Example 11.1: Estimate the regression line for the pollution data of Table 11.1.

Solution:

$$\sum_{i=1}^{33} x_i = 1104, \quad \sum_{i=1}^{33} y_i = 1124, \quad \sum_{i=1}^{33} x_i y_i = 41{,}355, \quad \sum_{i=1}^{33} x_i^2 = 41{,}086$$

Therefore,

$$b_1 = \frac{(33)(41{,}355) - (1104)(1124)}{(33)(41{,}086) - (1104)^2} = 0.903643 \text{ and}$$

$$b_0 = \frac{1124 - (0.903643)(1104)}{33} = 3.829633.$$

Thus, the estimated regression line is given by

$$\hat{y} = 3.8296 + 0.9036x.$$

Using the regression line of Example 11.1, we would predict a 31% reduction in the chemical oxygen demand when the reduction in the total solids is 30%. The

31% reduction in the chemical oxygen demand may be interpreted as an estimate of the population mean $\mu_{Y|30}$ or as an estimate of a new observation when the reduction in total solids is 30%. Such estimates, however, are subject to error. Even if the experiment were controlled so that the reduction in total solids was 30%, it is unlikely that we would measure a reduction in the chemical oxygen demand exactly equal to 31%. In fact, the original data recorded in Table 11.1 show that measurements of 25% and 35% were recorded for the reduction in oxygen demand when the reduction in total solids was kept at 30%.

What Is Good about Least Squares?

It should be noted that the least squares criterion is designed to provide a fitted line that results in a "closeness" between the line and the plotted points. There are many ways of measuring closeness. For example, one may wish to determine b_0 and b_1 for which $\sum\limits_{i=1}^{n} |y_i - \hat{y}_i|$ is minimized or for which $\sum\limits_{i=1}^{n} |y_i - \hat{y}_i|^{1.5}$ is minimized. These are both viable and reasonable methods. Note that both of these, as well as the least squares procedure, result in forcing residuals to be "small" in some sense. One should remember that the residuals are the empirical counterpart to the ϵ values. Figure 11.6 illustrates a set of residuals. One should note that the fitted line has predicted values as points on the line and hence the residuals are vertical deviations from points to the line. As a result, the least squares procedure produces a line that **minimizes the sum of squares of vertical deviations** from the points to the line.

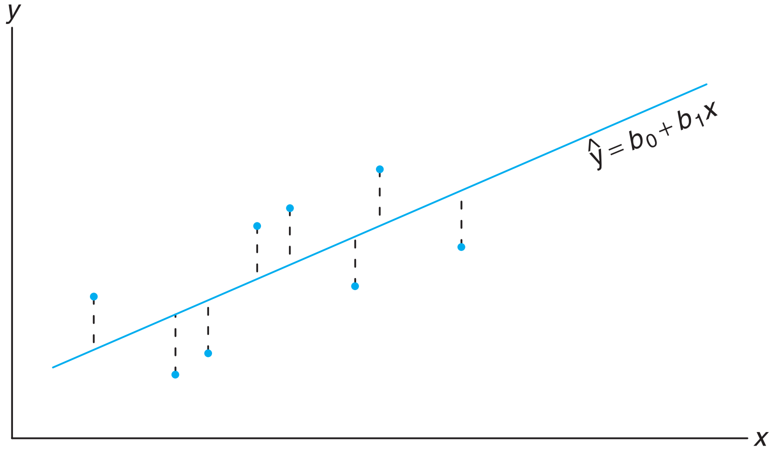


Figure 11.6: Residuals as vertical deviations.

Exercises

11.1 A study was conducted at Virginia Tech to determine if certain static arm-strength measures have an influence on the "dynamic lift" characteristics of an individual. Twenty-five individuals were subjected to strength tests and then were asked to perform a weight-lifting test in which weight was dynamically lifted overhead. The data are given here.

Individual	**Arm Strength, x**	**Dynamic Lift, y**
1	17.3	71.7
2	19.3	48.3
3	19.5	88.3
4	19.7	75.0
5	22.9	91.7
6	23.1	100.0
7	26.4	73.3
8	26.8	65.0
9	27.6	75.0
10	28.1	88.3
11	28.2	68.3
12	28.7	96.7
13	29.0	76.7
14	29.6	78.3
15	29.9	60.0
16	29.9	71.7
17	30.3	85.0
18	31.3	85.0
19	36.0	88.3
20	39.5	100.0
21	40.4	100.0
22	44.3	100.0
23	44.6	91.7
24	50.4	100.0
25	55.9	71.7

(a) Estimate β_0 and β_1 for the linear regression curve $\mu_{Y|x} = \beta_0 + \beta_1 x$.

(b) Find a point estimate of $\mu_{Y|30}$.

(c) Plot the residuals versus the x's (arm strength). Comment.

11.2 The grades of a class of 9 students on a midterm report (x) and on the final examination (y) are as follows:

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

(a) Estimate the linear regression line.

(b) Estimate the final examination grade of a student who received a grade of 85 on the midterm report.

11.3 The amounts of a chemical compound y that dissolved in 100 grams of water at various temperatures x were recorded as follows:

x (°C)	**y (grams)**		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

(a) Find the equation of the regression line.

(b) Graph the line on a scatter diagram.

(c) Estimate the amount of chemical that will dissolve in 100 grams of water at 50°C.

11.4 The following data were collected to determine the relationship between pressure and the corresponding scale reading for the purpose of calibration.

Pressure, x (lb/sq in.)	**Scale Reading, y**
10	13
10	18
10	16
10	15
10	20
50	86
50	90
50	88
50	88
50	92

(a) Find the equation of the regression line.

(b) The purpose of calibration in this application is to estimate pressure from an observed scale reading. Estimate the pressure for a scale reading of 54 using $\hat{x} = (54 - b_0)/b_1$.

11.5 A study was made on the amount of converted sugar in a certain process at various temperatures. The data were coded and recorded as follows:

Temperature, x	**Converted Sugar, y**
1.0	8.1
1.1	7.8
1.2	8.5
1.3	9.8
1.4	9.5
1.5	8.9
1.6	8.6
1.7	10.2
1.8	9.3
1.9	9.2
2.0	10.5

(a) Estimate the linear regression line.

(b) Estimate the mean amount of converted sugar produced when the coded temperature is 1.75.

(c) Plot the residuals versus temperature. Comment.

11.6 In a certain type of metal test specimen, the normal stress on a specimen is known to be functionally related to the shear resistance. The following is a set of coded experimental data on the two variables:

Normal Stress, x	**Shear Resistance, y**
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5
26.9	21.7
27.4	21.4
22.6	25.8
25.6	24.9

(a) Estimate the regression line $\mu_{Y|x} = \beta_0 + \beta_1 x$.

(b) Estimate the shear resistance for a normal stress of 24.5.

11.7 The following is a portion of a classic data set called the "pilot plot data" in *Fitting Equations to Data* by Daniel and Wood, published in 1971. The response y is the acid content of material produced by titration, whereas the regressor x is the organic acid content produced by extraction and weighing.

y	x	y	x
76	123	70	109
62	55	37	48
66	100	82	138
58	75	88	164
88	159	43	28

(a) Plot the data; does it appear that a simple linear regression will be a suitable model?

(b) Fit a simple linear regression; estimate a slope and intercept.

(c) Graph the regression line on the plot in (a).

11.8 A mathematics placement test is given to all entering freshmen at a small college. A student who receives a grade below 35 is denied admission to the regular mathematics course and placed in a remedial class. The placement test scores and the final grades for 20 students who took the regular course were recorded.

(a) Plot a scatter diagram.

(b) Find the equation of the regression line to predict course grades from placement test scores.

(c) Graph the line on the scatter diagram.

(d) If 60 is the minimum passing grade, below which placement test score should students in the future be denied admission to this course?

Placement Test	**Course Grade**
50	53
35	41
35	61
40	56
55	68
65	36
35	11
60	70
90	79
35	59
90	54
80	91
60	48
60	71
60	71
40	47
55	53
50	68
65	57
50	79

11.9 A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.

Advertising Costs (\$)	**Sales (\$)**
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

(a) Plot a scatter diagram.

(b) Find the equation of the regression line to predict weekly sales from advertising expenditures.

(c) Estimate the weekly sales when advertising costs are \$35.

(d) Plot the residuals versus advertising costs. Comment.

11.10 The following data are the selling prices z of a certain make and model of used car w years old. Fit a curve of the form $\mu_{z|w} = \gamma\delta^w$ by means of the nonlinear sample regression equation $\hat{z} = cd^w$. [*Hint*: Write $\ln \hat{z} = \ln c + (\ln d)w = b_0 + b_1 w$.]

w (years)	**z (dollars)**	**w (years)**	**z (dollars)**
1	6350	3	5395
2	5695	5	4985
2	5750	5	4895

11.11 The thrust of an engine (y) is a function of exhaust temperature (x) in °F when other important variables are held constant. Consider the following data.

y	x	y	x
4300	1760	4010	1665
4650	1652	3810	1550
3200	1485	4500	1700
3150	1390	3008	1270
4950	1820		

(a) Plot the data.

(b) Fit a simple linear regression to the data and plot the line through the data.

11.12 A study was done to study the effect of ambient temperature x on the electric power consumed by a chemical plant y. Other factors were held constant, and the data were collected from an experimental pilot plant.

y (BTU)	x (°F)	y (BTU)	x (°F)
250	27	265	31
285	45	298	60
320	72	267	34
295	58	321	74

(a) Plot the data.

(b) Estimate the slope and intercept in a simple linear regression model.

(c) Predict power consumption for an ambient temperature of 65°F.

11.13 A study of the amount of rainfall and the quantity of air pollution removed produced the following data:

Daily Rainfall, x (0.01 cm)	Particulate Removed, y (μg/m^3)
4.3	126
4.5	121
5.9	116
5.6	118
6.1	114
5.2	118
3.8	132
2.1	141
7.5	108

(a) Find the equation of the regression line to predict the particulate removed from the amount of daily rainfall.

(b) Estimate the amount of particulate removed when the daily rainfall is $x = 4.8$ units.

11.14 A professor in the School of Business in a university polled a dozen colleagues about the number of professional meetings they attended in the past five years (x) and the number of papers they submitted to refereed journals (y) during the same period. The summary data are given as follows:

$$n = 12, \quad \bar{x} = 4, \quad \bar{y} = 12,$$
$$\sum_{i=1}^{n} x_i^2 = 232, \quad \sum_{i=1}^{n} x_i y_i = 318.$$

Fit a simple linear regression model between x and y by finding out the estimates of intercept and slope. Comment on whether attending more professional meetings would result in publishing more papers.

11.4 Properties of the Least Squares Estimators

In addition to the assumptions that the error term in the model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

is a random variable with mean 0 and constant variance σ^2, suppose that we make the further assumption that $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent from run to run in the experiment. This provides a foundation for finding the means and variances for the estimators of β_0 and β_1.

It is important to remember that our values of b_0 and b_1, based on a given sample of n observations, are only estimates of true parameters β_0 and β_1. If the experiment is repeated over and over again, each time using the same fixed values of x, the resulting estimates of β_0 and β_1 will most likely differ from experiment to experiment. These different estimates may be viewed as values assumed by the random variables B_0 and B_1, while b_0 and b_1 are specific realizations.

Since the values of x remain fixed, the values of B_0 and B_1 depend on the variations in the values of y or, more precisely, on the values of the random variables,

$Y_1, Y_2, \dots, Y_n$. The distributional assumptions imply that the Y_i, $i = 1, 2, \dots, n$, are also independently distributed, with mean $\mu_{Y|x_i} = \beta_0 + \beta_1 x_i$ and equal variances σ^2; that is,

$$\sigma^2_{Y|x_i} = \sigma^2 \quad \text{for} \quad i = 1, 2, \dots, n.$$

Mean and Variance of Estimators

In what follows, we show that the estimator B_1 is unbiased for β_1 and demonstrate the variances of both B_0 and B_1. This will begin a series of developments that lead to hypothesis testing and confidence interval estimation on the intercept and slope.

Since the estimator

$$B_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

is of the form $\sum_{i=1}^{n} c_i Y_i$, where

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad i = 1, 2, \dots, n,$$

we may conclude from Theorem 7.11 that B_1 has a $n(\mu_{B_1}, \sigma_{B_1})$ distribution with

$$\mu_{B_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 \text{ and } \sigma^2_{B_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma^2_{Y_i}}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

It can also be shown (Review Exercise 11.60 on page 438) that the random variable B_0 is normally distributed with

$$\text{mean } \mu_{B_0} = \beta_0 \text{ and variance } \sigma^2_{B_0} = \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\sigma^2.$$

From the foregoing results, it is apparent that the **least squares estimators for β_0 and β_1 are both unbiased estimators**.

Partition of Total Variability and Estimation of σ^2

To draw inferences on β_0 and β_1, it becomes necessary to arrive at an estimate of the parameter σ^2 appearing in the two preceding variance formulas for B_0 and B_1. The parameter σ^2, the model error variance, reflects random variation or experimental error variation around the regression line. In much of what follows, it is advantageous to use the notation

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

Now we may write the error sum of squares as follows:

$$\begin{aligned} SSE &= \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n}[(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2 \\ &= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2b_1 \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + b_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} = S_{yy} - b_1 S_{xy}, \end{aligned}$$

the final step following from the fact that $b_1 = S_{xy}/S_{xx}$.

Theorem 11.1: An unbiased estimate of σ^2 is

$$s^2 = \frac{SSE}{n-2} = \sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n-2} = \frac{S_{yy} - b_1 S_{xy}}{n-2}.$$

The proof of Theorem 11.1 is left as an exercise (see Review Exercise 11.59).

The Estimator of σ^2 as a Mean Squared Error

One should observe the result of Theorem 11.1 in order to gain some intuition about the estimator of σ^2. The parameter σ^2 measures variance or squared deviations between Y values and their mean given by $\mu_{Y|x}$ (i.e., squared deviations between Y and $\beta_0 + \beta_1 x$). Of course, $\beta_0 + \beta_1 x$ is estimated by $\hat{y} = b_0 + b_1 x$. Thus, it would make sense that the variance σ^2 is best depicted as a squared deviation of the typical observation y_i from the estimated mean, $\hat{y}_i$, which is the corresponding point on the fitted line. Thus, $(y_i - \hat{y}_i)^2$ values reveal the appropriate variance, much like the way $(y_i - \bar{y})^2$ values measure variance when one is sampling in a nonregression scenario. In other words, $\bar{y}$ estimates the mean in the latter simple situation, whereas $\hat{y}_i$ estimates the mean of y_i in a regression structure. Now, what about the divisor $n-2$? In future sections, we shall note that these are the degrees of freedom associated with the estimator s^2 of σ^2. Whereas in the standard normal i.i.d. scenario, one degree of freedom is subtracted from n in the denominator and a reasonable explanation is that one parameter is estimated, namely the mean μ by, say, $\bar{y}$, but in the regression problem, **two parameters are estimated**, namely β_0 and β_1 by b_0 and b_1. Thus, the important parameter σ^2, estimated by

$$s^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-2),$$

is called a **mean squared error**, depicting a type of mean (division by $n-2$) of the squared residuals.

11.5 Inferences Concerning the Regression Coefficients

Aside from merely estimating the linear relationship between x and Y for purposes of prediction, the experimenter may also be interested in drawing certain inferences about the slope and intercept. In order to allow for the testing of hypotheses and the construction of confidence intervals on β_0 and β_1, one must be willing to make the further assumption that each ϵ_i, $i = 1, 2, \ldots, n$, is normally distributed. This assumption implies that $Y_1, Y_2, \ldots, Y_n$ are also normally distributed, each with probability distribution $n(y_i; \beta_0 + \beta_1 x_i, \sigma)$.

From Section 11.4 we know that B_1 follows a normal distribution. It turns out that under the normality assumption, a result very much analogous to that given in Theorem 8.4 allows us to conclude that $(n-2)S^2/\sigma^2$ is a chi-squared variable with $n-2$ degrees of freedom, independent of the random variable B_1. Theorem 8.5 then assures us that the statistic

$$T = \frac{(B_1 - \beta_1)/(\sigma/\sqrt{S_{xx}})}{S/\sigma} = \frac{B_1 - \beta_1}{S/\sqrt{S_{xx}}}$$

has a t-distribution with $n-2$ degrees of freedom. The statistic T can be used to construct a $100(1-\alpha)\%$ confidence interval for the coefficient β_1.

Confidence Interval for β_1

A $100(1-\alpha)\%$ confidence interval for the parameter β_1 in the regression line $\mu_{Y|x} = \beta_0 + \beta_1 x$ is

$$b_1 - t_{\alpha/2}\frac{s}{\sqrt{S_{xx}}} < \beta_1 < b_1 + t_{\alpha/2}\frac{s}{\sqrt{S_{xx}}},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n-2$ degrees of freedom.

Example 11.2: Find a 95% confidence interval for β_1 in the regression line $\mu_{Y|x} = \beta_0 + \beta_1 x$, based on the pollution data of Table 11.1.

***Solution*:** From the results given in Example 11.1 we find that $S_{xx} = 4152.18$ and $S_{xy} = 3752.09$. In addition, we find that $S_{yy} = 3713.88$. Recall that $b_1 = 0.903643$. Hence,

$$s^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{3713.88 - (0.903643)(3752.09)}{31} = 10.4299.$$

Therefore, taking the square root, we obtain $s = 3.2295$. Using Table A.4, we find $t_{0.025} \approx 2.045$ for 31 degrees of freedom. Therefore, a 95% confidence interval for β_1 is

$$0.903643 - \frac{(2.045)(3.2295)}{\sqrt{4152.18}} < \beta < 0.903643 + \frac{(2.045)(3.2295)}{\sqrt{4152.18}},$$

which simplifies to

$$0.8012 < \beta_1 < 1.0061.$$

■

Hypothesis Testing on the Slope

To test the null hypothesis H_0 that $\beta_1 = \beta_{10}$ against a suitable alternative, we again use the t-distribution with $n-2$ degrees of freedom to establish a critical region and then base our decision on the value of

$$t = \frac{b_1 - \beta_{10}}{s/\sqrt{S_{xx}}}.$$

The method is illustrated by the following example.

Example 11.3: Using the estimated value $b_1 = 0.903643$ of Example 11.1, test the hypothesis that $\beta_1 = 1.0$ against the alternative that $\beta_1 < 1.0$.

Solution: The hypotheses are H_0: $\beta_1 = 1.0$ and H_1: $\beta_1 < 1.0$. So

$$t = \frac{0.903643 - 1.0}{3.2295/\sqrt{4152.18}} = -1.92,$$

with $n - 2 = 31$ degrees of freedom ($P \approx 0.03$).

Decision: The t-value is significant at the 0.03 level, suggesting strong evidence that $\beta_1 < 1.0$.

One important t-test on the slope is the test of the hypothesis

$$H_0\colon\ \beta_1 = 0 \text{ versus } H_1\colon\ \beta_1 \neq 0.$$

When the null hypothesis is not rejected, the conclusion is that there is no significant linear relationship between $E(y)$ and the independent variable x. The plot of the data for Example 11.1 would suggest that a linear relationship exists. However, in some applications in which σ^2 is large and thus considerable "noise" is present in the data, a plot, while useful, may not produce clear information for the researcher. Rejection of H_0 above implies that a significant linear regression exists.

Figure 11.7 displays a *MINITAB* printout showing the t-test for

$$H_0\colon\ \beta_1 = 0 \text{ versus } H_1\colon\ \beta_1 \neq 0,$$

for the data of Example 11.1. Note the regression coefficient (Coef), standard error (SE Coef), t-value (T), and P-value (P). The null hypothesis is rejected. Clearly, there is a significant linear relationship between mean chemical oxygen demand reduction and solids reduction. Note that the t-statistic is computed as

$$t = \frac{\text{coefficient}}{\text{standard error}} = \frac{b_1}{s/\sqrt{S_{xx}}}.$$

The failure to reject H_0: $\beta_1 = 0$ suggests that there is no linear relationship between Y and x. Figure 11.8 is an illustration of the implication of this result. It may mean that changing x has little impact on changes in Y, as seen in (a). However, it may also indicate that the true relationship is nonlinear, as indicated by (b).

When H_0: $\beta_1 = 0$ is rejected, there is an implication that the linear term in x residing in the model explains a significant portion of variability in Y. The two

```
Regression Analysis: COD versus Per_Red
The regression equation is COD = 3.83 + 0.904 Per_Red

Predictor      Coef   SE Coef      T      P
Constant      3.830     1.768   2.17  0.038
Per_Red     0.90364   0.05012  18.03  0.000

S = 3.22954   R-Sq = 91.3%   R-Sq(adj) = 91.0%
Analysis of Variance
Source          DF      SS      MS       F      P
Regression       1  3390.6  3390.6  325.08  0.000
Residual Error  31   323.3    10.4
Total           32  3713.9
```

Figure 11.7: *MINITAB* printout for t-test for data of Example 11.1.

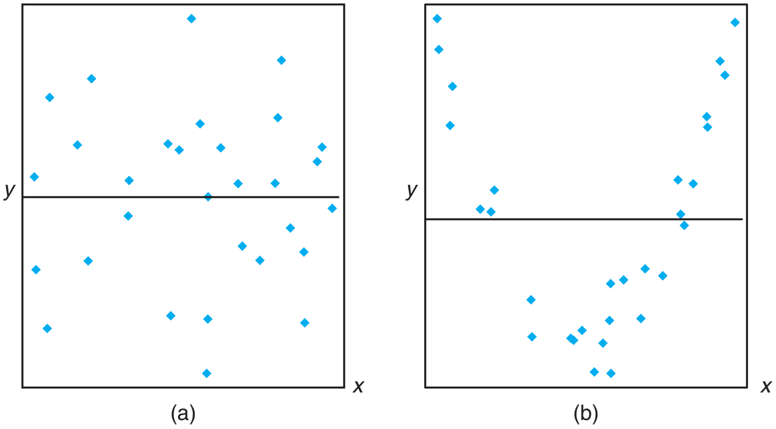


Figure 11.8: The hypothesis H_0: $\beta_1 = 0$ is not rejected.

plots in Figure 11.9 illustrate possible scenarios. As depicted in (a) of the figure, rejection of H_0 may suggest that the relationship is, indeed, linear. As indicated in (b), it may suggest that while the model does contain a linear effect, a better representation may be found by including a polynomial (perhaps quadratic) term (i.e., terms that supplement the linear term).

Statistical Inference on the Intercept

Confidence intervals and hypothesis testing on the coefficient β_0 may be established from the fact that B_0 is also normally distributed. It is not difficult to show that

$$T = \frac{B_0 - \beta_0}{S\sqrt{\sum_{i=1}^{n} x_i^2/(nS_{xx})}}$$

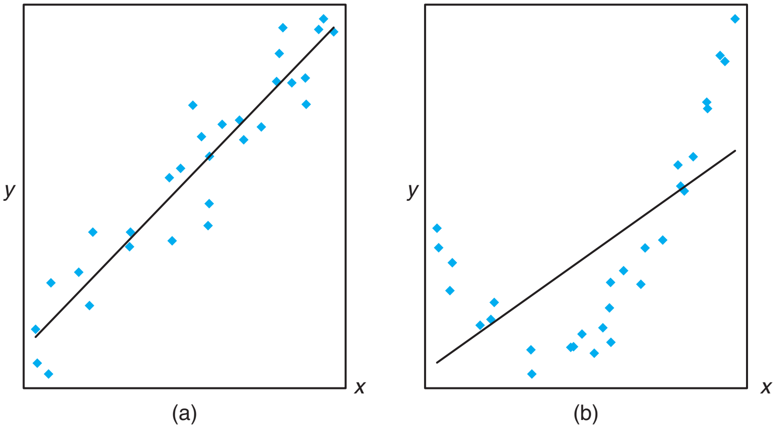


Figure 11.9: The hypothesis H_0: $\beta_1 = 0$ is rejected.

has a t-distribution with $n - 2$ degrees of freedom from which we may construct a $100(1 - \alpha)\%$ confidence interval for α.

Confidence Interval for β_0

A $100(1 - \alpha)\%$ confidence interval for the parameter β_0 in the regression line $\mu_{Y|x} = \beta_0 + \beta_1 x$ is

$$b_0 - t_{\alpha/2}\frac{s}{\sqrt{nS_{xx}}}\sqrt{\sum_{i=1}^{n} x_i^2} < \beta_0 < b_0 + t_{\alpha/2}\frac{s}{\sqrt{nS_{xx}}}\sqrt{\sum_{i=1}^{n} x_i^2},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n - 2$ degrees of freedom.

Example 11.4: Find a 95% confidence interval for β_0 in the regression line $\mu_{Y|x} = \beta_0 + \beta_1 x$, based on the data of Table 11.1.

Solution: In Examples 11.1 and 11.2, we found that

$$S_{xx} = 4152.18 \quad \text{and} \quad s = 3.2295.$$

From Example 11.1 we had

$$\sum_{i=1}^{n} x_i^2 = 41{,}086 \quad \text{and} \quad b_0 = 3.829633.$$

Using Table A.4, we find $t_{0.025} \approx 2.045$ for 31 degrees of freedom. Therefore, a 95% confidence interval for β_0 is

$$3.829633 - \frac{(2.045)(3.2295)\sqrt{41{,}086}}{\sqrt{(33)(4152.18)}} < \beta_0 < 3.829633 + \frac{(2.045)(3.2295)\sqrt{41{,}086}}{\sqrt{(33)(4152.18)}},$$

which simplifies to $0.2132 < \beta_0 < 7.4461$. ▟

To test the null hypothesis H_0 that $\beta_0 = \beta_{00}$ against a suitable alternative, we can use the t-distribution with $n-2$ degrees of freedom to establish a critical region and then base our decision on the value of

$$t = \frac{b_0 - \beta_{00}}{s\sqrt{\sum_{i=1}^{n} x_i^2/(nS_{xx})}}.$$

Example 11.5: Using the estimated value $b_0 = 3.829633$ of Example 11.1, test the hypothesis that $\beta_0 = 0$ at the 0.05 level of significance against the alternative that $\beta_0 \neq 0$.

***Solution*:** The hypotheses are H_0: $\beta_0 = 0$ and H_1: $\beta_0 \neq 0$. So

$$t = \frac{3.829633 - 0}{3.2295\sqrt{41{,}086/[(33)(4152.18)]}} = 2.17,$$

with 31 degrees of freedom. Thus, $P = P$-value ≈ 0.038 and we conclude that $\beta_0 \neq 0$. Note that this is merely Coef/StDev, as we see in the *MINITAB* printout in Figure 11.7. The SE Coef is the standard error of the estimated intercept. ■

A Measure of Quality of Fit: Coefficient of Determination

Note in Figure 11.7 that an item denoted by R-Sq is given with a value of 91.3%. This quantity, R^2, is called the **coefficient of determination**. This quantity is a measure of the **proportion of variability explained by the fitted model**. In Section 11.8, we shall introduce the notion of an analysis-of-variance approach to hypothesis testing in regression. The analysis-of-variance approach makes use of the error sum of squares $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ and the **total corrected sum of squares** $SST = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$. The latter represents the variation in the response values that *ideally* would be explained by the model. The SSE value is the variation due to error, or **variation unexplained**. Clearly, if $SSE = 0$, all variation is explained. The quantity that represents variation explained is $SST - SSE$. The R^2 is

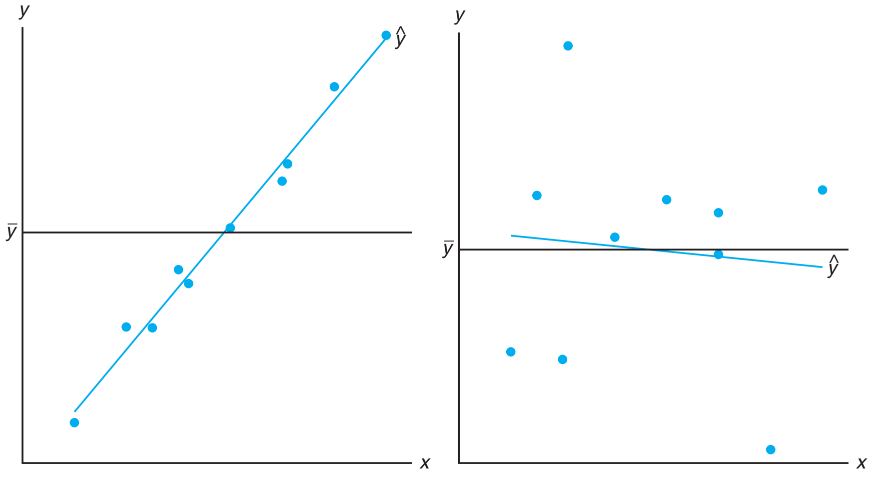
$$\text{Coeff. of determination:} \quad R^2 = 1 - \frac{SSE}{SST}.$$

Note that if the fit is perfect, *all residuals are zero*, and thus $R^2 = 1.0$. But if SSE is only slightly smaller than SST, $R^2 \approx 0$. Note from the printout in Figure 11.7 that the coefficient of determination suggests that the model fit to the data explains 91.3% of the variability observed in the response, the reduction in chemical oxygen demand.

Figure 11.10 provides an illustration of a good fit ($R^2 \approx 1.0$) in plot (a) and a poor fit ($R^2 \approx 0$) in plot (b).

Pitfalls in the Use of R^2

Analysts quote values of R^2 quite often, perhaps due to its simplicity. However, there are pitfalls in its interpretation. The reliability of R^2 is a function of the



(a) $R^2 \approx 1.0$ (b) $R^2 \approx 0$

Figure 11.10: Plots depicting a very good fit and a poor fit.

size of the regression data set and the type of application. Clearly, $0 \leq R^2 \leq 1$ and the upper bound is achieved when the fit to the data is perfect (i.e., all of the residuals are zero). What is an acceptable value for R^2? This is a difficult question to answer. A chemist, charged with doing a linear calibration of a high-precision piece of equipment, certainly expects to experience a very high R^2-value (perhaps exceeding 0.99), while a behavioral scientist, dealing in data impacted by variability in human behavior, may feel fortunate to experience an R^2 as large as 0.70. An experienced model fitter senses when a value is large enough, given the situation confronted. Clearly, some scientific phenomena lend themselves to modeling with more precision than others.

The R^2 criterion is dangerous to use for comparing *competing models* for the same data set. Adding additional terms to the model (e.g., an additional regressor) decreases SSE and thus increases R^2 (or at least does not decrease it). This implies that R^2 can be made artificially high by an unwise practice of **overfitting** (i.e., the inclusion of too many model terms). Thus, the inevitable increase in R^2 enjoyed by adding an additional term does not imply the additional term was needed. In fact, the simple model may be superior for predicting response values. The role of overfitting and its influence on prediction capability will be discussed at length in Chapter 12 as we visit the notion of models involving **more than a single regressor**. Suffice it to say at this point that one *should not subscribe to a model selection process that solely involves the consideration of* R^2.

11.6 Prediction

There are several reasons for building a linear regression. One, of course, is to predict response values at one or more values of the independent variable. In this section, the focus is on errors associated with prediction.

The equation $\hat{y} = b_0 + b_1 x$ may be used to predict or estimate the **mean response** $\mu_{Y|x_0}$ at $x = x_0$, where x_0 is not necessarily one of the prechosen values, or it may be used to predict a single value y_0 of the variable Y_0, when $x = x_0$. We would expect the error of prediction to be higher in the case of a single predicted value than in the case where a mean is predicted. This, then, will affect the width of our intervals for the values being predicted.

Suppose that the experimenter wishes to construct a confidence interval for $\mu_{Y|x_0}$. We shall use the point estimator $\hat{Y}_0 = B_0 + B_1 x_0$ to estimate $\mu_{Y|x_0} = \beta_0 + \beta_1 x$. It can be shown that the sampling distribution of $\hat{Y}_0$ is normal with mean

$$\mu_{Y|x_0} = E(\hat{Y}_0) = E(B_0 + B_1 x_0) = \beta_0 + \beta_1 x_0 = \mu_{Y|x_0}$$

and variance

$$\sigma^2_{\hat{Y}_0} = \sigma^2_{B_0 + B_1 x_0} = \sigma^2_{\bar{Y} + B_1(x_0 - \bar{x})} = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right],$$

the latter following from the fact that $\mathrm{Cov}(\bar{Y}, B_1) = 0$ (see Review Exercise 11.61 on page 438). Thus, a $100(1-\alpha)\%$ confidence interval on the mean response $\mu_{Y|x_0}$ can now be constructed from the statistic

$$T = \frac{\hat{Y}_0 - \mu_{Y|x_0}}{S\sqrt{1/n + (x_0 - \bar{x})^2/S_{xx}}},$$

which has a t-distribution with $n-2$ degrees of freedom.

Confidence Interval for $\mu_{Y|x_0}$

A $100(1-\alpha)\%$ confidence interval for the mean response $\mu_{Y|x_0}$ is

$$\hat{y}_0 - t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n-2$ degrees of freedom.

Example 11.6: Using the data of Table 11.1, construct 95% confidence limits for the mean response $\mu_{Y|x_0}$.

Solution: From the regression equation we find for $x_0 = 20\%$ solids reduction, say,

$$\hat{y}_0 = 3.829633 + (0.903643)(20) = 21.9025.$$

In addition, $\bar{x} = 33.4545$, $S_{xx} = 4152.18$, $s = 3.2295$, and $t_{0.025} \approx 2.045$ for 31 degrees of freedom. Therefore, a 95% confidence interval for $\mu_{Y|20}$ is

$$21.9025 - (2.045)(3.2295)\sqrt{\frac{1}{33} + \frac{(20 - 33.4545)^2}{4152.18}} < \mu_{Y|20}$$
$$< 21.9025 + (2.045)(3.2295)\sqrt{\frac{1}{33} + \frac{(20 - 33.4545)^2}{4152.18}},$$

or simply $20.1071 < \mu_{Y|20} < 23.6979$. ■

Repeating the previous calculations for each of several different values of x_0, one can obtain the corresponding confidence limits on each $\mu_{Y|x_0}$. Figure 11.11 displays the data points, the estimated regression line, and the upper and lower confidence limits on the mean of $Y|x$.

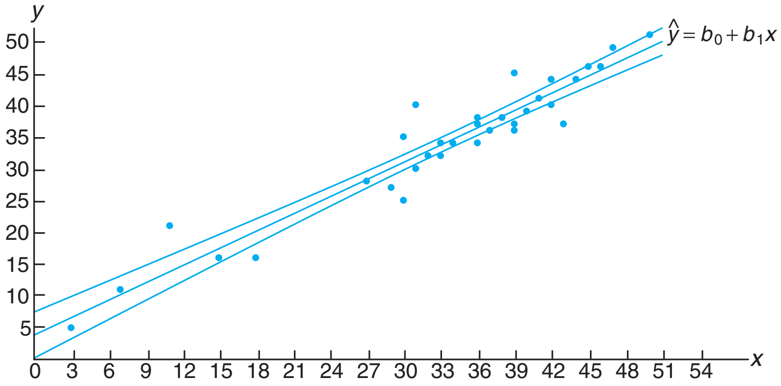


Figure 11.11: Confidence limits for the mean value of $Y|x$.

In Example 11.6, we are 95% confident that the population mean reduction in chemical oxygen demand is between 20.1071% and 23.6979% when solid reduction is 20%.

Prediction Interval

Another type of interval that is often misinterpreted and confused with that given for $\mu_{Y|x}$ is the prediction interval for a future observed response. Actually in many instances, the prediction interval is more relevant to the scientist or engineer than the confidence interval on the mean. In the tar content and inlet temperature example cited in Section 11.1, there would certainly be interest not only in estimating the mean tar content at a specific temperature but also in constructing an interval that reflects the error in predicting a future observed amount of tar content at the given temperature.

To obtain a **prediction interval** for any single value y_0 of the variable Y_0, it is necessary to estimate the variance of the differences between the ordinates $\hat{y}_0$, obtained from the computed regression lines in repeated sampling when $x = x_0$, and the corresponding true ordinate y_0. We can think of the difference $\hat{y}_0 - y_0$ as a value of the random variable $\hat{Y}_0 - Y_0$, whose sampling distribution can be shown to be normal with mean

$$\mu_{\hat{Y}_0 - Y_0} = E(\hat{Y}_0 - Y_0) = E[B_0 + B_1x_0 - (\beta_0 + \beta_1x_0 + \epsilon_0)] = 0$$

and variance

$$\sigma^2_{\hat{Y}_0 - Y_0} = \sigma^2_{B_0 + B_1x_0 - \epsilon_0} = \sigma^2_{\bar{Y} + B_1(x_0 - \bar{x}) - \epsilon_0} = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right].$$

Thus, a $100(1-\alpha)\%$ prediction interval for a single predicted value y_0 can be constructed from the statistic

$$T = \frac{\hat{Y}_0 - Y_0}{S\sqrt{1 + 1/n + (x_0 - \bar{x})^2/S_{xx}}},$$

which has a t-distribution with $n-2$ degrees of freedom.

Prediction Interval for y_0

A $100(1-\alpha)\%$ prediction interval for a single response y_0 is given by

$$\hat{y}_0 - t_{\alpha/2}s\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}} < y_0 < \hat{y}_0 + t_{\alpha/2}s\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n-2$ degrees of freedom.

Clearly, there is a distinction between the concept of a confidence interval and the prediction interval described previously. The interpretation of the confidence interval is identical to that described for all confidence intervals on population parameters discussed throughout the book. Indeed, $\mu_{Y|x_0}$ is a population parameter. The computed prediction interval, however, represents an interval that has a probability equal to $1-\alpha$ of containing not a parameter but a future value y_0 of the random variable Y_0.

Example 11.7: Using the data of Table 11.1, construct a 95% prediction interval for y_0 when $x_0 = 20\%$.

***Solution*:** We have $n = 33$, $x_0 = 20$, $\bar{x} = 33.4545$, $\hat{y}_0 = 21.9025$, $S_{xx} = 4152.18$, $s = 3.2295$, and $t_{0.025} \approx 2.045$ for 31 degrees of freedom. Therefore, a 95% prediction interval for y_0 is

$$21.9025 - (2.045)(3.2295)\sqrt{1 + \frac{1}{33} + \frac{(20-33.4545)^2}{4152.18}} < y_0$$
$$< 21.9025 + (2.045)(3.2295)\sqrt{1 + \frac{1}{33} + \frac{(20-33.4545)^2}{4152.18}},$$

which simplifies to $15.0585 < y_0 < 28.7464$. ┘

Figure 11.12 shows another plot of the chemical oxygen demand reduction data, with both the confidence interval on the mean response and the prediction interval on an individual response plotted. The plot reflects a much tighter interval around the regression line in the case of the mean response.

Exercises

11.15 With reference to Exercise 11.1 on page 398,

(a) evaluate s^2;

(b) test the hypothesis that $\beta_1 = 0$ against the alternative that $\beta_1 \neq 0$ at the 0.05 level of significance and interpret the resulting decision.

11.16 With reference to Exercise 11.2 on page 398,

(a) evaluate s^2;

(b) construct a 95% confidence interval for β_0;

(c) construct a 95% confidence interval for β_1.

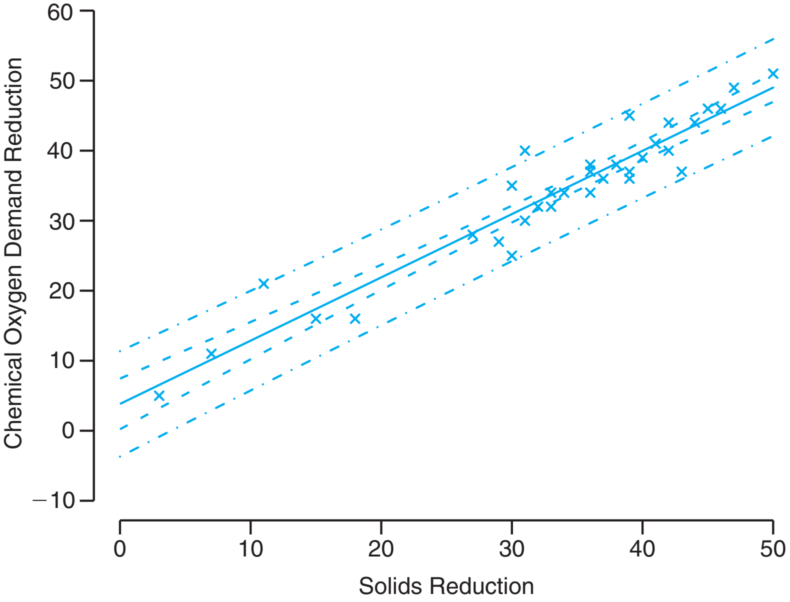


Figure 11.12: Confidence and prediction intervals for the chemical oxygen demand reduction data; inside bands indicate the confidence limits for the mean responses and outside bands indicate the prediction limits for the future responses.

11.17 With reference to Exercise 11.5 on page 398,

(a) evaluate s^2;

(b) construct a 95% confidence interval for β_0;

(c) construct a 95% confidence interval for β_1.

11.18 With reference to Exercise 11.6 on page 399,

(a) evaluate s^2;

(b) construct a 99% confidence interval for β_0;

(c) construct a 99% confidence interval for β_1.

11.19 With reference to Exercise 11.3 on page 398,

(a) evaluate s^2;

(b) construct a 99% confidence interval for β_0;

(c) construct a 99% confidence interval for β_1.

11.20 Test the hypothesis that $\beta_0 = 10$ in Exercise 11.8 on page 399 against the alternative that $\beta_0 < 10$. Use a 0.05 level of significance.

11.21 Test the hypothesis that $\beta_1 = 6$ in Exercise 11.9 on page 399 against the alternative that $\beta_1 < 6$. Use a 0.025 level of significance.

11.22 Using the value of s^2 found in Exercise 11.16(a), construct a 95% confidence interval for $\mu_{Y|85}$ in Exercise 11.2 on page 398.

11.23 With reference to Exercise 11.6 on page 399, use the value of s^2 found in Exercise 11.18(a) to compute

(a) a 95% confidence interval for the mean shear resistance when $x = 24.5$;

(b) a 95% prediction interval for a single predicted value of the shear resistance when $x = 24.5$.

11.24 Using the value of s^2 found in Exercise 11.17(a), graph the regression line and the 95% confidence bands for the mean response $\mu_{Y|x}$ for the data of Exercise 11.5 on page 398.

11.25 Using the value of s^2 found in Exercise 11.17(a), construct a 95% confidence interval for the amount of converted sugar corresponding to $x = 1.6$ in Exercise 11.5 on page 398.

11.26 With reference to Exercise 11.3 on page 398, use the value of s^2 found in Exercise 11.19(a) to compute

(a) a 99% confidence interval for the average amount of chemical that will dissolve in 100 grams of water at 50°C;

(b) a 99% prediction interval for the amount of chemical that will dissolve in 100 grams of water at 50°C.

11.27 Consider the regression of mileage for certain automobiles, measured in miles per gallon (mpg) on their weight in pounds (wt). The data are from *Consumer Reports* (April 1997). Part of the *SAS* output from the procedure is shown in Figure 11.13.

(a) Estimate the mileage for a vehicle weighing 4000 pounds.

(b) Suppose that Honda engineers claim that, on average, the Civic (or any other model weighing 2440 pounds) gets more than 30 mpg. Based on the results of the regression analysis, would you believe that claim? Why or why not?

(c) The design engineers for the Lexus ES300 targeted 18 mpg as being ideal for this model (or any other model weighing 3390 pounds), although it is expected that some variation will be experienced. Is it likely that this target value is realistic? Discuss.

11.28 There are important applications in which, due to known scientific constraints, the regression line **must go through the origin** (i.e., the intercept must be zero). In other words, the model should read

$$Y_i = \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

and only a simple parameter requires estimation. The model is often called the **regression through the origin model**.

(a) Show that the least squares estimator of the slope is

$$b_1 = \left(\sum_{i=1}^{n} x_i y_i\right) \Bigg/ \left(\sum_{i=1}^{n} x_i^2\right).$$

(b) Show that $\sigma^2_{B_1} = \sigma^2 \Bigg/ \left(\sum\limits_{i=1}^{n} x_i^2\right)$.

(c) Show that b_1 in part (a) is an unbiased estimator for β_1. That is, show $E(B_1) = \beta_1$.

11.29 Use the data set

y	x
7	2
50	15
100	30
40	10
70	20

(a) Plot the data.

(b) Fit a regression line through the origin.

(c) Plot the regression line on the graph with the data.

(d) Give a general formula (in terms of the y_i and the slope b_1) for the estimator of σ^2.

(e) Give a formula for $\text{Var}(\hat{y}_i)$, $i = 1, 2, \ldots, n$, for this case.

(f) Plot 95% confidence limits for the mean response on the graph around the regression line.

11.30 For the data in Exercise 11.29, find a 95% prediction interval at $x = 25$.

```
                   Root MSE              1.48794    R-Square      0.9509
                   Dependent Mean       21.50000    Adj R-Sq      0.9447
                                  Parameter Estimates
                                Parameter       Standard
            Variable      DF     Estimate          Error    t Value    Pr > |t|
            Intercept      1     44.78018        1.92919      23.21      <.0001
            WT             1     -0.00686     0.00055133     -12.44      <.0001
MODEL           WT      MPG     Predict      LMean      UMean      Lpred      Upred     Residual
GMC           4520       15     13.7720    11.9752    15.5688     9.8988    17.6451      1.22804
Geo           2065       29     30.6138    28.6063    32.6213    26.6385    34.5891     -1.61381
Honda         2440       31     28.0412    26.4143    29.6681    24.2439    31.8386      2.95877
Hyundai       2290       28     29.0703    27.2967    30.8438    25.2078    32.9327     -1.07026
Infiniti      3195       23     22.8618    21.7478    23.9758    19.2543    26.4693      0.13825
Isuzu         3480       21     20.9066    19.8160    21.9972    17.3062    24.5069      0.09341
Jeep          4090       15     16.7219    15.3213    18.1224    13.0158    20.4279     -1.72185
Land          4535       13     13.6691    11.8570    15.4811     9.7888    17.5493     -0.66905
Lexus         3390       22     21.5240    20.4390    22.6091    17.9253    25.1227      0.47599
Lincoln       3930       18     17.8195    16.5379    19.1011    14.1568    21.4822      0.18051
```

Figure 11.13: *SAS* printout for Exercise 11.27.

11.7 Choice of a Regression Model

Much of what has been presented thus far on regression involving a single independent variable depends on the assumption that the model chosen is correct, the presumption that $\mu_{Y|x}$ is related to x linearly in the parameters. Certainly, one cannot expect the prediction of the response to be good if there are several independent variables, not considered in the model, that are affecting the response and are varying in the system. In addition, the prediction will certainly be inadequate if the true structure relating $\mu_{Y|x}$ to x is extremely nonlinear in the range of the variables considered.

Often the simple linear regression model is used even though it is known that the model is something other than linear or that the true structure is unknown. This approach is often sound, particularly when the range of x is narrow. Thus, the model used becomes an approximating function that one hopes is an adequate representation of the true picture in the region of interest. One should note, however, the effect of an inadequate model on the results presented thus far. For example, if the true model, unknown to the experimenter, is linear in more than one x, say

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

then the ordinary least squares estimate $b_1 = S_{xy}/S_{xx}$, calculated by only considering x_1 in the experiment, is, under general circumstances, a biased estimate of the coefficient β_1, the bias being a function of the additional coefficient β_2 (see Review Exercise 11.65 on page 438). Also, the estimate s^2 for σ^2 is biased due to the additional variable.

11.8 Analysis-of-Variance Approach

Often the problem of analyzing the quality of the estimated regression line is handled by an **analysis-of-variance** (ANOVA) approach: a procedure whereby the total variation in the dependent variable is subdivided into meaningful components that are then observed and treated in a systematic fashion. The analysis of variance, discussed in Chapter 13, is a powerful resource that is used for many applications.

Suppose that we have n experimental data points in the usual form (x_i, y_i) and that the regression line is estimated. In our estimation of σ^2 in Section 11.4, we established the identity

$$S_{yy} = b_1 S_{xy} + SSE.$$

An alternative and perhaps more informative formulation is

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

We have achieved a partitioning of the **total corrected sum of squares of y** into two components that should reflect particular meaning to the experimenter. We shall indicate this partitioning symbolically as

$$SST = SSR + SSE.$$

The first component on the right, SSR, is called the **regression sum of squares**, and it reflects the amount of variation in the y-values **explained by the model**, in this case the postulated straight line. The second component is the familiar error sum of squares, which reflects variation about the regression line.

Suppose that we are interested in testing the hypothesis

$$H_0\colon\ \beta_1 = 0 \text{ versus } H_1\colon\ \beta_1 \neq 0,$$

where the null hypothesis says essentially that the model is $\mu_{Y|x} = \beta_0$. That is, the variation in Y results from chance or random fluctuations which are independent of the values of x. This condition is reflected in Figure 11.10(b). Under the conditions of this null hypothesis, it can be shown that SSR/σ^2 and SSE/σ^2 are values of independent chi-squared variables with 1 and $n-2$ degrees of freedom, respectively, and then by Theorem 7.12 it follows that SST/σ^2 is also a value of a chi-squared variable with $n-1$ degrees of freedom. To test the hypothesis above, we compute

$$f = \frac{SSR/1}{SSE/(n-2)} = \frac{SSR}{s^2}$$

and reject H_0 at the α-level of significance when $f > f_\alpha(1, n-2)$.

The computations are usually summarized by means of an **analysis-of-variance table**, as in Table 11.2. It is customary to refer to the various sums of squares divided by their respective degrees of freedom as the **mean squares**.

Table 11.2: Analysis of Variance for Testing $\beta_1 = 0$

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Regression	SSR	1	SSR	$\frac{SSR}{s^2}$
Error	SSE	$n-2$	$s^2 = \frac{SSE}{n-2}$	
Total	SST	$n-1$		

When the null hypothesis is rejected, that is, when the computed F-statistic exceeds the critical value $f_\alpha(1, n-2)$, we conclude that **there is a significant amount of variation in the response accounted for by the postulated model, the straight-line function**. If the F-statistic is in the fail to reject region, we conclude that the data did not reflect sufficient evidence to support the model postulated.

In Section 11.5, a procedure was given whereby the statistic

$$T = \frac{B_1 - \beta_{10}}{S/\sqrt{S_{xx}}}$$

is used to test the hypothesis

$$H_0\colon\ \beta_1 = \beta_{10} \text{ versus } H_1\colon\ \beta_1 \neq \beta_{10},$$

where T follows the t-distribution with $n-2$ degrees of freedom. The hypothesis is rejected if $|t| > t_{\alpha/2}$ for an α-level of significance. It is interesting to note that in the special case in which we are testing

$$H_0\colon\ \beta_1 = 0 \text{ versus } H_1\colon\ \beta_1 \neq 0,$$

the value of our T-statistic becomes

$$t = \frac{b_1}{s/\sqrt{S_{xx}}},$$

and the hypothesis under consideration is identical to that being tested in Table 11.2. Namely, the null hypothesis states that the variation in the response is due merely to chance. The analysis of variance uses the F-distribution rather than the t-distribution. For the two-sided alternative, the two approaches are identical. This we can see by writing

$$t^2 = \frac{b_1^2 S_{xx}}{s^2} = \frac{b_1 S_{xy}}{s^2} = \frac{SSR}{s^2},$$

which is identical to the f-value used in the analysis of variance. The basic relationship between the t-distribution with v degrees of freedom and the F-distribution with 1 and v degrees of freedom is

$$t^2 = f(1, v).$$

Of course, the t-test allows for testing against a one-sided alternative while the F-test is restricted to testing against a two-sided alternative.

Annotated Computer Printout for Simple Linear Regression

Consider again the chemical oxygen demand reduction data of Table 11.1. Figures 11.14 and 11.15 show more complete annotated computer printouts. Again we illustrate it with *MINITAB* software. The t-ratio column indicates tests for null hypotheses of zero values on the parameter. The term “Fit” denotes $\hat{y}$-values, often called **fitted values**. The term “SE Fit” is used in computing confidence intervals on mean response. The item R^2 is computed as $(SSR/SST) \times 100$ and signifies the proportion of variation in y explained by the straight-line regression. Also shown are confidence intervals on the mean response and prediction intervals on a new observation.

11.9 Test for Linearity of Regression: Data with Repeated Observations

In certain kinds of experimental situations, the researcher has the capability of obtaining repeated observations on the response for each value of x. Although it is not necessary to have these repetitions in order to estimate β_0 and β_1, nevertheless repetitions enable the experimenter to obtain quantitative information concerning the appropriateness of the model. In fact, if repeated observations are generated, the experimenter can make a significance test to aid in determining whether or not the model is adequate.

```
 The regression equation is COD = 3.83 + 0.904 Per_Red
 Predictor    Coef  SE Coef      T      P
  Constant   3.830    1.768   2.17  0.038
   Per_Red 0.90364  0.05012  18.03  0.000
S = 3.22954   R-Sq = 91.3%   R-Sq(adj) = 91.0%
             Analysis of Variance
 Source         DF      SS      MS       F    P
 Regression      1  3390.6  3390.6  325.08  0.000
 Residual Error 31   323.3    10.4
 Total          32  3713.9

Obs  Per_Red     COD     Fit  SE Fit  Residual  St Resid
  1      3.0   5.000   6.541   1.627    -1.541     -0.55
  2     36.0  34.000  36.361   0.576    -2.361     -0.74
  3      7.0  11.000  10.155   1.440     0.845      0.29
  4     37.0  36.000  37.264   0.590    -1.264     -0.40
  5     11.0  21.000  13.770   1.258     7.230      2.43
  6     38.0  38.000  38.168   0.607    -0.168     -0.05
  7     15.0  16.000  17.384   1.082    -1.384     -0.45
  8     39.0  37.000  39.072   0.627    -2.072     -0.65
  9     18.0  16.000  20.095   0.957    -4.095     -1.33
 10     39.0  36.000  39.072   0.627    -3.072     -0.97
 11     27.0  28.000  28.228   0.649    -0.228     -0.07
 12     39.0  45.000  39.072   0.627     5.928      1.87
 13     29.0  27.000  30.035   0.605    -3.035     -0.96
 14     40.0  39.000  39.975   0.651    -0.975     -0.31
 15     30.0  25.000  30.939   0.588    -5.939     -1.87
 16     41.0  41.000  40.879   0.678     0.121      0.04
 17     30.0  35.000  30.939   0.588     4.061      1.28
 18     42.0  40.000  41.783   0.707    -1.783     -0.57
 19     31.0  30.000  31.843   0.575    -1.843     -0.58
 20     42.0  44.000  41.783   0.707     2.217      0.70
 21     31.0  40.000  31.843   0.575     8.157      2.57
 22     43.0  37.000  42.686   0.738    -5.686     -1.81
 23     32.0  32.000  32.746   0.567    -0.746     -0.23
 24     44.0  44.000  43.590   0.772     0.410      0.13
 25     33.0  34.000  33.650   0.563     0.350      0.11
 26     45.0  46.000  44.494   0.807     1.506      0.48
 27     33.0  32.000  33.650   0.563    -1.650     -0.52
 28     46.0  46.000  45.397   0.843     0.603      0.19
 29     34.0  34.000  34.554   0.563    -0.554     -0.17
 30     47.0  49.000  46.301   0.881     2.699      0.87
 31     36.0  37.000  36.361   0.576     0.639      0.20
 32     50.0  51.000  49.012   1.002     1.988      0.65
 33     36.0  38.000  36.361   0.576     1.639      0.52
```

Figure 11.14: *MINITAB* printout of simple linear regression for chemical oxygen demand reduction data; part I.

Let us select a random sample of n observations using k distinct values of x, say $x_1, x_2, \ldots, x_n$, such that the sample contains n_1 observed values of the random variable Y_1 corresponding to x_1, n_2 observed values of Y_2 corresponding to $x_2, \ldots$, n_k observed values of Y_k corresponding to x_k. Of necessity, $n = \sum\limits_{i=1}^{k} n_i$.

```
Obs     Fit  SE Fit        95% CI            95% PI
  1   6.541   1.627  ( 3.223,  9.858)  (-0.834, 13.916)
  2  36.361   0.576  (35.185, 37.537)  (29.670, 43.052)
  3  10.155   1.440  ( 7.218, 13.092)  ( 2.943, 17.367)
  4  37.264   0.590  (36.062, 38.467)  (30.569, 43.960)
  5  13.770   1.258  (11.204, 16.335)  ( 6.701, 20.838)
  6  38.168   0.607  (36.931, 39.405)  (31.466, 44.870)
  7  17.384   1.082  (15.177, 19.592)  (10.438, 24.331)
  8  39.072   0.627  (37.793, 40.351)  (32.362, 45.781)
  9  20.095   0.957  (18.143, 22.047)  (13.225, 26.965)
 10  39.072   0.627  (37.793, 40.351)  (32.362, 45.781)
 11  28.228   0.649  (26.905, 29.551)  (21.510, 34.946)
 12  39.072   0.627  (37.793, 40.351)  (32.362, 45.781)
 13  30.035   0.605  (28.802, 31.269)  (23.334, 36.737)
 14  39.975   0.651  (38.648, 41.303)  (33.256, 46.694)
 15  30.939   0.588  (29.739, 32.139)  (24.244, 37.634)
 16  40.879   0.678  (39.497, 42.261)  (34.149, 47.609)
 17  30.939   0.588  (29.739, 32.139)  (24.244, 37.634)
 18  41.783   0.707  (40.341, 43.224)  (35.040, 48.525)
 19  31.843   0.575  (30.669, 33.016)  (25.152, 38.533)
 20  41.783   0.707  (40.341, 43.224)  (35.040, 48.525)
 21  31.843   0.575  (30.669, 33.016)  (25.152, 38.533)
 22  42.686   0.738  (41.181, 44.192)  (35.930, 49.443)
 23  32.746   0.567  (31.590, 33.902)  (26.059, 39.434)
 24  43.590   0.772  (42.016, 45.164)  (36.818, 50.362)
 25  33.650   0.563  (32.502, 34.797)  (26.964, 40.336)
 26  44.494   0.807  (42.848, 46.139)  (37.704, 51.283)
 27  33.650   0.563  (32.502, 34.797)  (26.964, 40.336)
 28  45.397   0.843  (43.677, 47.117)  (38.590, 52.205)
 29  34.554   0.563  (33.406, 35.701)  (27.868, 41.239)
 30  46.301   0.881  (44.503, 48.099)  (39.473, 53.128)
 31  36.361   0.576  (35.185, 37.537)  (29.670, 43.052)
 32  49.012   1.002  (46.969, 51.055)  (42.115, 55.908)
 33  36.361   0.576  (35.185, 37.537)  (29.670, 43.052)
```

Figure 11.15: *MINITAB* printout of simple linear regression for chemical oxygen demand reduction data; part II.

We define

$$\begin{aligned} y_{ij} &= \text{the } j\text{th value of the random variable } Y_i,\\ y_{i.} &= T_{i.} = \sum_{j=1}^{n_i} y_{ij},\\ \bar{y}_{i.} &= \frac{T_{i.}}{n_i}. \end{aligned}$$

Hence, if $n_4 = 3$ measurements of Y were made corresponding to $x = x_4$, we would indicate these observations by y_{41}, y_{42}, and y_{43}. Then

$$T_{i.} = y_{41} + y_{42} + y_{43}.$$

Concept of Lack of Fit

The error sum of squares consists of two parts: the amount due to the variation between the values of Y within given values of x and a component that is normally called the **lack-of-fit** contribution. The first component reflects mere random variation, or **pure experimental error**, while the second component is a measure of the systematic variation brought about by higher-order terms. In our case, these are terms in x other than the linear, or first-order, contribution. Note that in choosing a linear model we are essentially assuming that this second component does not exist and hence our error sum of squares is completely due to random errors. If this should be the case, then $s^2 = SSE/(n-2)$ is an unbiased estimate of σ^2. However, if the model does not adequately fit the data, then the error sum of squares is inflated and produces a biased estimate of σ^2. Whether or not the model fits the data, an unbiased estimate of σ^2 can always be obtained when we have repeated observations simply by computing

$$s_i^2 = \frac{\sum\limits_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}{n_i - 1}, \quad i = 1, 2, \ldots, k,$$

for each of the k distinct values of x and then pooling these variances to get

$$s^2 = \frac{\sum\limits_{i=1}^{k}(n_i - 1)s_i^2}{n-k} = \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}{n-k}.$$

The numerator of s^2 is a **measure of the pure experimental error**. A computational procedure for separating the error sum of squares into the two components representing pure error and lack of fit is as follows:

Computation of Lack-of-Fit Sum of Squares

1. Compute the pure error sum of squares

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

This sum of squares has $n - k$ degrees of freedom associated with it, and the resulting mean square is our unbiased estimate s^2 of σ^2.

2. Subtract the pure error sum of squares from the error sum of squares SSE, thereby obtaining the sum of squares due to lack of fit. The degrees of freedom for lack of fit are obtained by simply subtracting $(n-2) - (n-k) = k - 2$.

The computations required for testing hypotheses in a regression problem with repeated measurements on the response may be summarized as shown in Table 11.3.

Figures 11.16 and 11.17 display the sample points for the "correct model" and "incorrect model" situations. In Figure 11.16, where the $\mu_{Y|x}$ fall on a straight line, there is no lack of fit when a linear model is assumed, so the sample variation around the regression line is a pure error resulting from the variation that occurs among repeated observations. In Figure 11.17, where the $\mu_{Y|x}$ clearly do not fall on a straight line, the lack of fit from erroneously choosing a linear model accounts for a large portion of the variation around the regression line, supplementing the pure error.

Table 11.3: Analysis of Variance for Testing Linearity of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	SSR	1	SSR	$\frac{SSR}{s^2}$
Error	SSE	$n-2$		
Lack of fit	$SSE - SSE$ (pure)	$k-2$	$\frac{SSE-SSE(\text{pure})}{k-2}$	$\frac{SSE-SSE(\text{pure})}{s^2(k-2)}$
Pure error	SSE (pure)	$n-k$	$s^2 = \frac{SSE(\text{pure})}{n-k}$	
Total	SST	$n-1$		

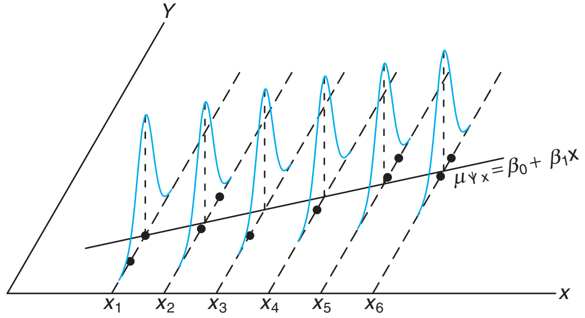


Figure 11.16: Correct linear model with no lack-of-fit component.

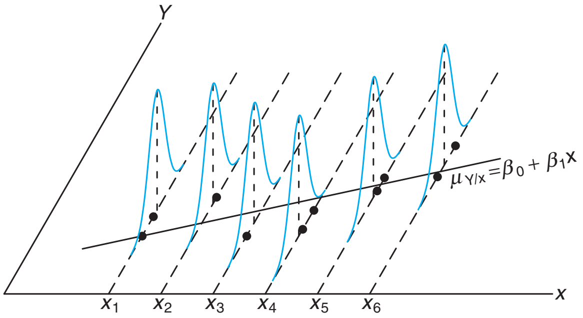


Figure 11.17: Incorrect linear model with lack-of-fit component.

What Is the Importance in Detecting Lack of Fit?

The concept of lack of fit is extremely important in applications of regression analysis. In fact, the need to construct or design an experiment that will account for lack of fit becomes more critical as the problem and the underlying mechanism involved become more complicated. Surely, one cannot always be certain that his or her postulated structure, in this case the linear regression model, is correct or even an adequate representation. The following example shows how the error sum of squares is partitioned into the two components representing pure error and lack of fit. The adequacy of the model is tested at the α-level of significance by comparing the lack-of-fit mean square divided by s^2 with $f_\alpha(k-2, n-k)$.

Example 11.8: Observations of the yield of a chemical reaction taken at various temperatures were recorded in Table 11.4. Estimate the linear model $\mu_{Y|x} = \beta_0 + \beta_1 x$ and test for lack of fit.

***Solution*:** Results of the computations are shown in Table 11.5.

Conclusion: The partitioning of the total variation in this manner reveals a significant variation accounted for by the linear model and an insignificant amount of variation due to lack of fit. Thus, the experimental data do not seem to suggest the need to consider terms higher than first order in the model, and the null hypothesis is not rejected.

Table 11.4: Data for Example 11.8

y (%)	x (°C)	y (%)	x (°C)
77.4	150	88.9	250
76.7	150	89.2	250
78.2	150	89.7	250
84.1	200	94.8	300
84.5	200	94.7	300
83.7	200	95.9	300

Table 11.5: Analysis of Variance on Yield-Temperature Data

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**	***P*-Values**
Regression	509.2507	1	509.2507	1531.58	<0.0001
Error	3.8660	10			
Lack of fit	1.2060	2	0.6030	1.81	0.2241
Pure error	2.6600	8	0.3325		
Total	513.1167	11			

Annotated Computer Printout for Test for Lack of Fit

Figure 11.18 is an annotated computer printout showing analysis of the data of Example 11.8 with *SAS*. Note the "LOF" with 2 degrees of freedom, representing the quadratic and cubic contribution to the model, and the *P*-value of 0.22, suggesting that the linear (first-order) model is adequate.

```
Dependent Variable: yield
                                              Sum of
 Source                   DF         Squares      Mean Square     F Value     Pr > F
 Model                     3     510.4566667      170.1522222      511.74     <.0001
 Error                     8       2.6600000        0.3325000
 Corrected Total          11     513.1166667
              R-Square       Coeff Var       Root MSE     yield Mean
              0.994816        0.666751       0.576628       86.48333
 Source                   DF       Type I SS      Mean Square     F Value     Pr > F
 temperature               1     509.2506667      509.2506667     1531.58     <.0001
 LOF                       2       1.2060000        0.6030000        1.81     0.2241
```

Figure 11.18: *SAS* printout, showing analysis of data of Example 11.8.

Exercises

11.31 Test for linearity of regression in Exercise 11.3 on page 398. Use a 0.05 level of significance. Comment.

11.32 Test for linearity of regression in Exercise 11.8 on page 399. Comment.

11.33 Suppose we have a linear equation through the origin (Exercise 11.28) $\mu_{Y|x} = \beta x$.

(a) Estimate the regression line passing through the origin for the following data:

x	0.5	1.5	3.2	4.2	5.1	6.5
y	1.3	3.4	6.7	8.0	10.0	13.2

(b) Suppose it is not known whether the true regression should pass through the origin. Estimate the linear model $\mu_{Y|x} = \beta_0 + \beta_1 x$ and test the hypothesis that $\beta_0 = 0$, at the 0.10 level of significance, against the alternative that $\beta_0 \neq 0$.

11.34 Use an analysis-of-variance approach to test the hypothesis that $\beta_1 = 0$ against the alternative hypothesis $\beta_1 \neq 0$ in Exercise 11.5 on page 398 at the 0.05 level of significance.

11.35 The following data are a result of an investigation as to the effect of reaction temperature x on percent conversion of a chemical process y. (See Myers, Montgomery and Anderson-Cook, 2009.) Fit a simple linear regression, and use a lack-of-fit test to determine if the model is adequate. Discuss.

Observation	**Temperature (°C), x**	**Conversion (%), y**
1	200	43
2	250	78
3	200	69
4	250	73
5	189.65	48
6	260.35	78
7	225	65
8	225	74
9	225	76
10	225	79
11	225	83
12	225	81

11.36 Transistor gain between emitter and collector in an integrated circuit device (hFE) is related to two variables (Myers, Montgomery and Anderson-Cook, 2009) that can be controlled at the deposition process, emitter drive-in time (x_1, in minutes) and emitter dose (x_2, in ions × 10^{14}). Fourteen samples were observed following deposition, and the resulting data are shown in the table below. We will consider linear regression models using gain as the response and emitter drive-in time or emitter dose as the regressor variable.

Obs.	**x_1 (drive-in time, min)**	**x_2 (dose, ions $\times 10^{14}$)**	**y (gain, or hFE)**
1	195	4.00	1004
2	255	4.00	1636
3	195	4.60	852
4	255	4.60	1506
5	255	4.20	1272
6	255	4.10	1270
7	255	4.60	1269
8	195	4.30	903
9	255	4.30	1555
10	255	4.00	1260
11	255	4.70	1146
12	255	4.30	1276
13	255	4.72	1225
14	340	4.30	1321

(a) Determine if emitter drive-in time influences gain in a linear relationship. That is, test H_0: $\beta_1 = 0$, where β_1 is the slope of the regressor variable.

(b) Do a lack-of-fit test to determine if the linear relationship is adequate. Draw conclusions.

(c) Determine if emitter dose influences gain in a linear relationship. Which regressor variable is the better predictor of gain?

11.37 Organophosphate (OP) compounds are used as pesticides. However, it is important to study their effect on species that are exposed to them. In the laboratory study *Some Effects of Organophosphate Pesticides on Wildlife Species*, by the Department of Fisheries and Wildlife at Virginia Tech, an experiment was conducted in which different dosages of a particular OP pesticide were administered to 5 groups of 5 mice (*peromysius leucopus*). The 25 mice were females of similar age and condition. One group received no chemical. The basic response y was a measure of activity in the brain. It was postulated that brain activity would decrease with an increase in OP dosage. The data are as follows:

Animal	**Dose, x (mg/kg body weight)**	**Activity, y (moles/liter/min)**
1	0.0	10.9
2	0.0	10.6
3	0.0	10.8
4	0.0	9.8
5	0.0	9.0
6	2.3	11.0
7	2.3	11.3
8	2.3	9.9
9	2.3	9.2
10	2.3	10.1
11	4.6	10.6
12	4.6	10.4
13	4.6	8.8
14	4.6	11.1
15	4.6	8.4
16	9.2	9.7
17	9.2	7.8
18	9.2	9.0
19	9.2	8.2
20	9.2	2.3
21	18.4	2.9
22	18.4	2.2
23	18.4	3.4
24	18.4	5.4
25	18.4	8.2

(a) Using the model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, 25,$$

find the least squares estimates of β_0 and β_1.

(b) Construct an analysis-of-variance table in which the lack of fit and pure error have been separated.

Determine if the lack of fit is significant at the 0.05 level. Interpret the results.

11.38 Heat treating is often used to carburize metal parts such as gears. The thickness of the carburized layer is considered an important feature of the gear, and it contributes to the overall reliability of the part. Because of the critical nature of this feature, a lab test is performed on each furnace load. The test is a destructive one, where an actual part is cross sectioned and soaked in a chemical for a period of time. This test involves running a carbon analysis on the surface of both the gear pitch (top of the gear tooth) and the gear root (between the gear teeth). The data below are the results of the pitch carbon-analysis test for 19 parts.

Soak Time	Pitch	Soak Time	Pitch
0.58	0.013	1.17	0.021
0.66	0.016	1.17	0.019
0.66	0.015	1.17	0.021
0.66	0.016	1.20	0.025
0.66	0.015	2.00	0.025
0.66	0.016	2.00	0.026
1.00	0.014	2.20	0.024
1.17	0.021	2.20	0.025
1.17	0.018	2.20	0.024
1.17	0.019		

(a) Fit a simple linear regression relating the pitch carbon analysis y against soak time. Test H_0: $\beta_1 = 0$.

(b) If the hypothesis in part (a) is rejected, determine if the linear model is adequate.

11.39 A regression model is desired relating temperature and the proportion of impurities passing through solid helium. Temperature is listed in degrees centigrade. The data are as follows:

Temperature (°C)	Proportion of Impurities
−260.5	0.425
−255.7	0.224
−264.6	0.453
−265.0	0.475
−270.0	0.705
−272.0	0.860
−272.5	0.935
−272.6	0.961
−272.8	0.979
−272.9	0.990

(a) Fit a linear regression model.

(b) Does it appear that the proportion of impurities passing through helium increases as the temperature approaches −273 degrees centigrade?

(c) Find R^2.

(d) Based on the information above, does the linear model seem appropriate? What additional information would you need to better answer that question?

11.40 It is of interest to study the effect of population size in various cities in the United States on ozone concentrations. The data consist of the 1999 population in millions and the amount of ozone present per hour in ppb (parts per billion). The data are as follows.

Ozone (ppb/hour), y	Population, x
126	0.6
135	4.9
124	0.2
128	0.5
130	1.1
128	0.1
126	1.1
128	2.3
128	0.6
129	2.3

(a) Fit the linear regression model relating ozone concentration to population. Test H_0: $\beta_1 = 0$ using the ANOVA approach.

(b) Do a test for lack of fit. Is the linear model appropriate based on the results of your test?

(c) Test the hypothesis of part (a) using the pure mean square error in the F-test. Do the results change? Comment on the advantage of each test.

11.41 Evaluating nitrogen deposition from the atmosphere is a major role of the National Atmospheric Deposition Program (NADP), a partnership of many agencies. NADP is studying atmospheric deposition and its effect on agricultural crops, forest surface waters, and other resources. Nitrogen oxides may affect the ozone in the atmosphere and the amount of pure nitrogen in the air we breathe. The data are as follows:

Year	Nitrogen Oxide
1978	0.73
1979	2.55
1980	2.90
1981	3.83
1982	2.53
1983	2.77
1984	3.93
1985	2.03
1986	4.39
1987	3.04
1988	3.41
1989	5.07
1990	3.95
1991	3.14
1992	3.44
1993	3.63
1994	4.50
1995	3.95
1996	5.24
1997	3.30
1998	4.36
1999	3.33

(a) Plot the data.

(b) Fit a linear regression model and find R^2.

(c) What can you say about the trend in nitrogen oxide across time?

11.42 For a particular variety of plant, researchers wanted to develop a formula for predicting the quantity of seeds (in grams) as a function of the density of plants. They conducted a study with four levels of the factor x, the number of plants per plot. Four replications were used for each level of x. The data are shown as follows:

Plants per Plot, x	**Quantity of Seeds, y (grams)**			
10	12.6	11.0	12.1	10.9
20	15.3	16.1	14.9	15.6
30	17.9	18.3	18.6	17.8
40	19.2	19.6	18.9	20.0

Is a simple linear regression model adequate for analyzing this data set?

11.10 Data Plots and Transformations

In this chapter, we deal with building regression models where there is one independent, or regressor, variable. In addition, we are assuming, through model formulation, that both x and y enter the model in a *linear fashion*. Often it is advisable to work with an alternative model in which either x or y (or both) enters in a nonlinear way. A **transformation** of the data may be indicated because of theoretical considerations inherent in the scientific study, or a simple plotting of the data may suggest the need to *reexpress* the variables in the model. The need to perform a transformation is rather simple to diagnose in the case of simple linear regression because two-dimensional plots give a true pictorial display of how each variable enters the model.

A model in which x or y is transformed should not be viewed as a *nonlinear regression model*. We normally refer to a regression model as linear when it is **linear in the parameters**. In other words, suppose the complexion of the data or other scientific information suggests that we should **regress y* against x***, where each is a transformation on the natural variables x and y. Then the model of the form

$$y_i^* = \beta_0 + \beta_1 x_i^* + \epsilon_i$$

is a linear model since it is linear in the parameters β_0 and β_1. The material given in Sections 11.2 through 11.9 remains intact, with y_i^* and x_i^* replacing y_i and x_i. A simple and useful example is the log-log model

$$\log y_i = \beta_0 + \beta_1 \log x_i + \epsilon_i.$$

Although this model is not linear in x and y, it is linear in the parameters and is thus treated as a linear model. On the other hand, an example of a truly nonlinear model is

$$y_i = \beta_0 + \beta_1 x^{\beta_2} + \epsilon_i,$$

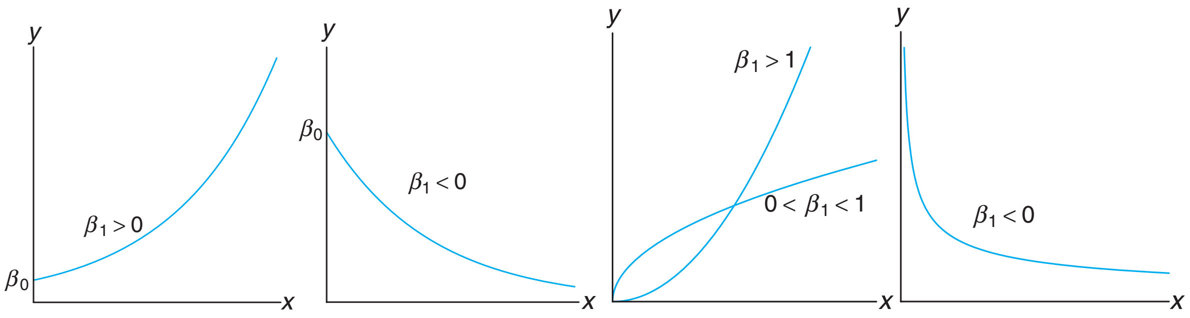
where the parameter β_2 (as well as β_0 and β_1) is to be estimated. The model is not linear in β_2.

Transformations that may enhance the fit and predictability of a model are many in number. For a thorough discussion of transformations, the reader is referred to Myers (1990, see the Bibliography). We choose here to indicate a few of them and show the appearance of the graphs that serve as a diagnostic tool. Consider Table 11.6. Several functions are given describing relationships between y and x that can produce a *linear regression* through the transformation indicated.

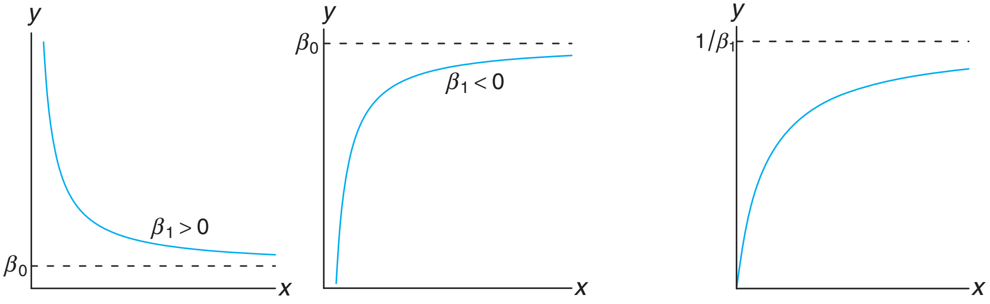
In addition, for the sake of completeness the reader is given the dependent and independent variables to use in the resulting *simple linear regression*. Figure 11.19 depicts functions listed in Table 11.6. These serve as a guide for the analyst in choosing a transformation from the observation of the plot of y against x.

Table 11.6: Some Useful Transformations to Linearize

Functional Form Relating y to x	**Proper Transformation**	**Form of Simple Linear Regression**
Exponential: $y = \beta_0 e^{\beta_1 x}$	$y^* = \ln y$	Regress y^* against x
Power: $y = \beta_0 x^{\beta_1}$	$y^* = \log y; \quad x^* = \log x$	Regress y^* against x^*
Reciprocal: $y = \beta_0 + \beta_1 \left(\frac{1}{x}\right)$	$x^* = \frac{1}{x}$	Regress y against x^*
Hyperbolic: $y = \frac{x}{\beta_0 + \beta_1 x}$	$y^* = \frac{1}{y}; \quad x^* = \frac{1}{x}$	Regress y^* against x^*



(a) Exponential function (b) Power function



(c) Reciprocal function (d) Hyperbolic function

Figure 11.19: Diagrams depicting functions listed in Table 11.6.

What Are the Implications of a Transformed Model?

The foregoing is intended as an aid for the analyst when it is apparent that a transformation will provide an improvement. However, before we provide an example, two important points should be made. The first one revolves around the formal writing of the model when the data are transformed. Quite often the analyst does not think about this. He or she merely performs the transformation without any concern about the model form *before* and *after* the transformation. The exponential model serves as a good illustration. The model in the natural (untransformed) variables that produces an *additive error model* in the transformed variables is given by

$$y_i = \beta_0 e^{\beta_1 x_i} \cdot \epsilon_i,$$

which is a *multiplicative error model.* Clearly, taking logs produces

$$\ln y_i = \ln \beta_0 + \beta_1 x_i + \ln \epsilon_i.$$

As a result, it is on $\ln \epsilon_i$ that the basic assumptions are made. The purpose of this presentation is merely to remind the reader that one should not view a transformation as merely an algebraic manipulation with an error added. Often a model in the transformed variables that has a proper *additive error structure* is a result of a model in the natural variables with a different type of error structure.

The second important point deals with the notion of measures of improvement. Obvious measures of comparison are, of course, R^2 and the residual mean square, s^2. (Other measures of performance used to compare competing models are given in Chapter 12.) Now, if the response y is not transformed, then clearly s^2 and R^2 can be used in measuring the utility of the transformation. The residuals will be in the same units for both the transformed and the untransformed models. But when y is transformed, performance criteria for the transformed model should be based on values of the residuals in the metric of the untransformed response so that comparisons that are made are proper. The example that follows provides an illustration.

Example 11.9: The pressure P of a gas corresponding to various volumes V is recorded, and the data are given in Table 11.7.

Table 11.7: Data for Example 11.9

V (cm^3)	50	60	70	90	100
P (kg/cm^2)	64.7	51.3	40.5	25.9	7.8

The ideal gas law is given by the functional form $PV^\gamma = C$, where γ and C are constants. Estimate the constants C and γ.

Solution: Let us take natural logs of both sides of the model

$$P_i V^\gamma = C \cdot \epsilon_i, \quad i = 1, 2, 3, 4, 5.$$

As a result, a linear model can be written

$$\ln P_i = \ln C - \gamma \ln V_i + \epsilon_i^*, \quad i = 1, 2, 3, 4, 5,$$

where $\epsilon_i^* = \ln \epsilon_i$. The following represents results of the simple linear regression:

$$\text{Intercept: } \widehat{\ln C} = 14.7589, \ \widehat{C} = 2{,}568{,}862.88, \ \text{Slope: } \hat{\gamma} = 2.65347221.$$

The following represents information taken from the regression analysis.

P_i	V_i	$\ln P_i$	$\ln V_i$	$\widehat{\ln P_i}$	$\widehat{P_i}$	$e_i = P_i - \widehat{P_i}$
64.7	50	4.16976	3.91202	4.37853	79.7	−15.0
51.3	60	3.93769	4.09434	3.89474	49.1	2.2
40.5	70	3.70130	4.24850	3.48571	32.6	7.9
25.9	90	3.25424	4.49981	2.81885	16.8	9.1
7.8	100	2.05412	4.60517	2.53921	12.7	−4.9

It is instructive to plot the data and the regression equation. Figure 11.20 shows a plot of the data in the untransformed pressure and volume and the curve representing the regression equation. ▟

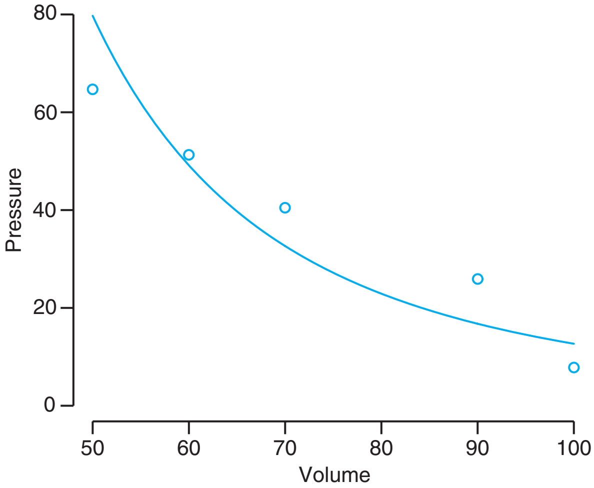


Figure 11.20: Pressure and volume data and fitted regression.

Diagnostic Plots of Residuals: Graphical Detection of Violation of Assumptions

Plots of the raw data can be extremely helpful in determining the nature of the model that should be fit to the data when there is a single independent variable. We have attempted to illustrate this in the foregoing. Detection of proper model form is, however, not the only benefit gained from diagnostic plotting. As in much of the material associated with significance testing in Chapter 10, plotting methods can illustrate and detect violation of assumptions. The reader should recall that much of what is illustrated in this chapter requires assumptions made on the model errors, the ϵ_i. In fact, we assume that the ϵ_i are independent $N(0, \sigma)$ random variables. Now, of course, the ϵ_i are not observed. However, the $e_i = y_i - \hat{y}_i$, the *residuals*, are the error in the fit of the regression line and thus serve to mimic the ϵ_i. Thus, the general complexion of these residuals can often highlight difficulties. Ideally, of course, the plot of the residuals is as depicted in Figure 11.21. That is, they should truly show random fluctuations around a value of zero.

Nonhomogeneous Variance

Homogeneous variance is an important assumption made in regression analysis. Violations can often be detected through the appearance of the residual plot. Increasing error variance with an increase in the regressor variable is a common condition in scientific data. Large error variance produces large residuals, and hence a residual plot like the one in Figure 11.22 is a signal of nonhomogeneous variance. More discussion regarding these residual plots and information regard-

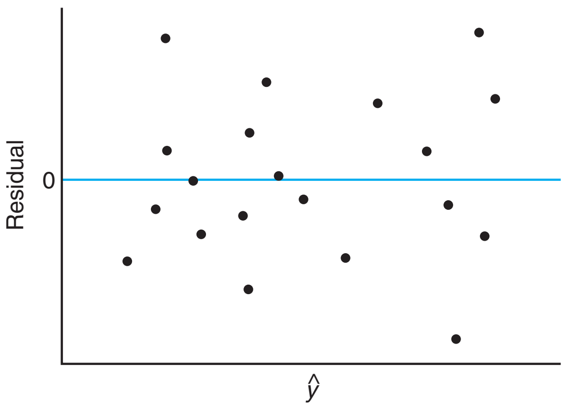


Figure 11.21: Ideal residual plot.

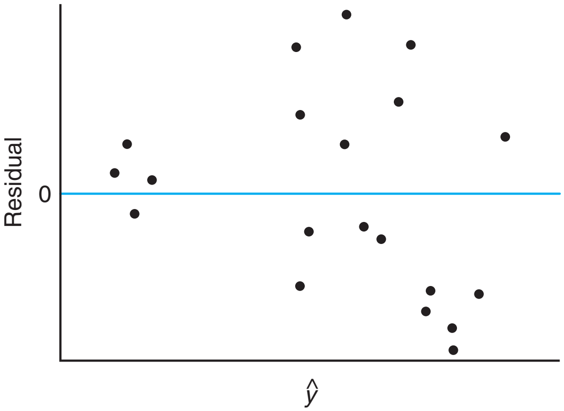


Figure 11.22: Residual plot depicting heterogeneous error variance.

ing different types of residuals appears in Chapter 12, where we deal with multiple linear regression.

Normal Probability Plotting

The assumption that the model errors are normal is made when the data analyst deals in either hypothesis testing or confidence interval estimation. Again, the numerical counterpart to the ϵ_i, namely the residuals, are subjects of diagnostic plotting to detect any extreme violations. In Chapter 8, we introduced normal quantile-quantile plots and briefly discussed normal probability plots. These plots on residuals are illustrated in the case study introduced in the next section.

11.11 Simple Linear Regression Case Study

In the manufacture of commercial wood products, it is important to estimate the relationship between the density of a wood product and its stiffness. A relatively new type of particleboard is being considered that can be formed with considerably more ease than the accepted commercial product. It is necessary to know at what density the stiffness is comparable to that of the well-known, well-documented commercial product. A study was done by Terrance E. Conners, *Investigation of Certain Mechanical Properties of a Wood-Foam Composite* (M.S. Thesis, Department of Forestry and Wildlife Management, University of Massachusetts). Thirty particleboards were produced at densities ranging from roughly 8 to 26 pounds per cubic foot, and the stiffness was measured in pounds per square inch. Table 11.8 shows the data.

It is necessary for the data analyst to focus on an appropriate fit to the data and use inferential methods discussed in this chapter. Hypothesis testing on the slope of the regression, as well as confidence or prediction interval estimation, may well be appropriate. We begin by demonstrating a simple scatter plot of the raw data with a simple linear regression superimposed. Figure 11.23 shows this plot.

The simple linear regression fit to the data produced the fitted model

$$\hat{y} = -25{,}433.739 + 3884.976x \quad (R^2 = 0.7975),$$

Table 11.8: Density and Stiffness for 30 Particleboards

Density, x	**Stiffness, y**	**Density, x**	**Stiffness, y**
9.50	14,814.00	8.40	17,502.00
9.80	14,007.00	11.00	19,443.00
8.30	7573.00	9.90	14,191.00
8.60	9714.00	6.40	8076.00
7.00	5304.00	8.20	10,728.00
17.40	43,243.00	15.00	25,319.00
15.20	28,028.00	16.40	41,792.00
16.70	49,499.00	15.40	25,312.00
15.00	26,222.00	14.50	22,148.00
14.80	26,751.00	13.60	18,036.00
25.60	96,305.00	23.40	104,170.00
24.40	72,594.00	23.30	49,512.00
19.50	32,207.00	21.20	48,218.00
22.80	70,453.00	21.70	47,661.00
19.80	38,138.00	21.30	53,045.00

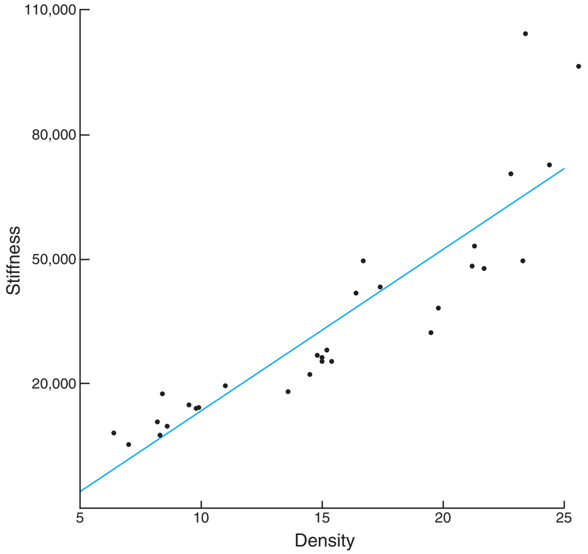


Figure 11.23: Scatter plot of the wood density data.

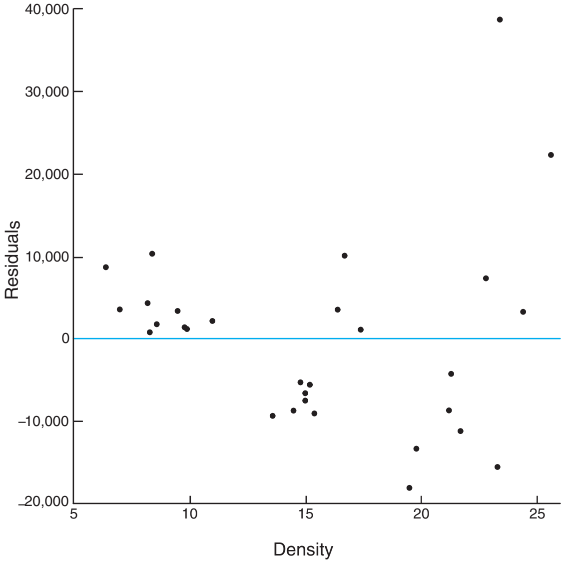


Figure 11.24: Residual plot for the wood density data.

and the residuals were computed. Figure 11.24 shows the residuals plotted against the measurements of density. This is hardly an ideal or healthy set of residuals. They do not show a random scatter around a value of zero. In fact, clusters of positive and negative values suggest that a curvilinear trend in the data should be investigated.

To gain some type of idea regarding the normal error assumption, a normal probability plot of the residuals was generated. This is the type of plot discussed in

Section 8.8 in which the horizontal axis represents the empirical normal distribution function on a scale that produces a straight-line plot when plotted against the residuals. Figure 11.25 shows the normal probability plot of the residuals. The normal probability plot does not reflect the straight-line appearance that one would like to see. This is another symptom of a faulty, perhaps overly simplistic choice of a regression model.

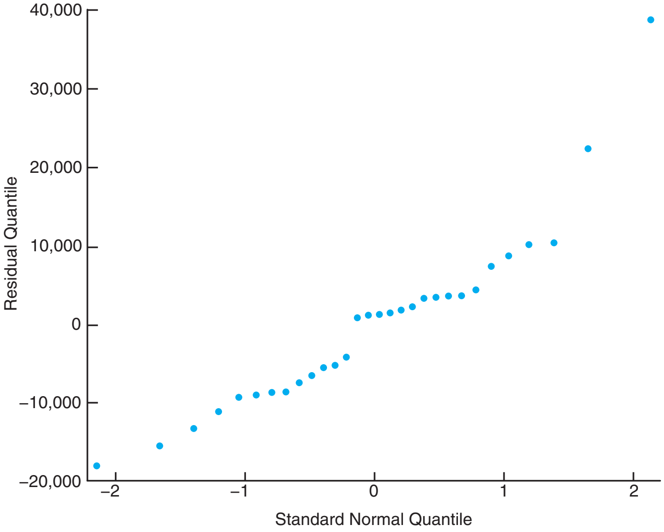


Figure 11.25: Normal probability plot of residuals for wood density data.

Both types of residual plots and, indeed, the scatter plot itself suggest here that a somewhat complicated model would be appropriate. One possible approach is to use a natural log transformation. In other words, one might choose to regress $\ln y$ against x. This produces the regression

$$\widehat{\ln y} = 8.257 + 0.125x \quad (R^2 = 0.9016).$$

To gain some insight into whether the transformed model is more appropriate, consider Figures 11.26 and 11.27, which reveal plots of the residuals in stiffness [i.e., y_i-antilog $(\widehat{\ln y})$] against density. Figure 11.26 appears to be closer to a random pattern around zero, while Figure 11.27 is certainly closer to a straight line. This in addition to the higher R^2-value would suggest that the transformed model is more appropriate.

11.12 Correlation

Up to this point we have assumed that the independent regressor variable x is a physical or scientific variable but not a random variable. In fact, in this context, x is often called a **mathematical variable**, which, in the sampling process, is measured with negligible error. In many applications of regression techniques, it is more realistic to assume that both X and Y are random variables and the measurements $\{(x_i, y_i); i = 1, 2, \dots, n\}$ are observations from a population having

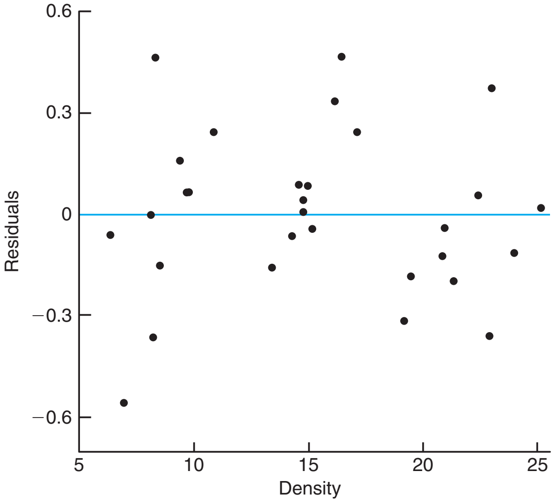


Figure 11.26: Residual plot using the log transformation for the wood density data.

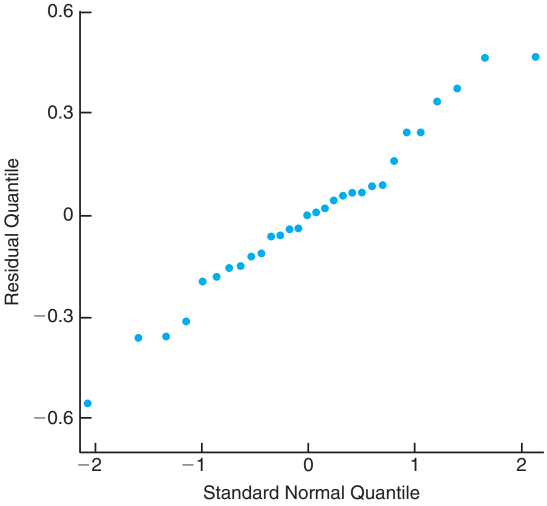


Figure 11.27: Normal probability plot of residuals using the log transformation for the wood density data.

the joint density function $f(x, y)$. We shall consider the problem of measuring the relationship between the two variables X and Y. For example, if X and Y represent the length and circumference of a particular kind of bone in the adult body, we might conduct an anthropological study to determine whether large values of X are associated with large values of Y, and vice versa.

On the other hand, if X represents the age of a used automobile and Y represents the retail book value of the automobile, we would expect large values of X to correspond to small values of Y and small values of X to correspond to large values of Y. **Correlation analysis** attempts to measure the strength of such relationships between two variables by means of a single number called a **correlation coefficient**.

In theory, it is often assumed that the conditional distribution $f(y|x)$ of Y, for fixed values of X, is normal with mean $\mu_{Y|x} = \beta_0 + \beta_1 x$ and variance $\sigma^2_{Y|x} = \sigma^2$ and that X is likewise normally distributed with mean μ and variance σ^2_x. The joint density of X and Y is then

$$\begin{aligned} f(x, y) &= n(y|x; \beta_0 + \beta_1 x, \sigma) n(x; \mu_X, \sigma_X) \\ &= \frac{1}{2\pi\sigma_x\sigma} \exp\left\{-\frac{1}{2}\left[\left(\frac{y - \beta_0 - \beta_1 x}{\sigma}\right)^2 + \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right]\right\}, \end{aligned}$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

Let us write the random variable Y in the form

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where X is now a random variable independent of the random error ϵ. Since the mean of the random error ϵ is zero, it follows that

$$\mu_Y = \beta_0 + \beta_1 \mu_X \quad \text{and} \quad \sigma_Y^2 = \sigma^2 + \beta_1^2 \sigma_X^2.$$

Substituting for α and σ^2 into the preceding expression for $f(x, y)$, we obtain the **bivariate normal distribution**

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\},$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, where

$$\rho^2 = 1 - \frac{\sigma^2}{\sigma_Y^2} = \beta_1^2 \frac{\sigma_X^2}{\sigma_Y^2}.$$

The constant ρ (rho) is called the **population correlation coefficient** and plays a major role in many bivariate data analysis problems. It is important for the reader to understand the physical interpretation of this correlation coefficient and the distinction between correlation and regression. The term *regression* still has meaning here. In fact, the straight line given by $\mu_{Y|x} = \beta_0 + \beta_1 x$ is still called the regression line as before, and the estimates of β_0 and β_1 are identical to those given in Section 11.3. The value of ρ is 0 when $\beta_1 = 0$, which results when there essentially is no linear regression; that is, the regression line is horizontal and any knowledge of X is useless in predicting Y. Since $\sigma_Y^2 \geq \sigma^2$, we must have $\rho^2 \leq 1$ and hence $-1 \leq \rho \leq 1$. Values of $\rho = \pm 1$ only occur when $\sigma^2 = 0$, in which case we have a perfect linear relationship between the two variables. Thus, a value of ρ equal to $+1$ implies a perfect linear relationship with a positive slope, while a value of ρ equal to -1 results from a perfect linear relationship with a negative slope. It might be said, then, that sample estimates of ρ close to unity in magnitude imply good correlation, or **linear association**, between X and Y, whereas values near zero indicate little or no correlation.

To obtain a sample estimate of ρ, recall from Section 11.4 that the error sum of squares is

$$SSE = S_{yy} - b_1 S_{xy}.$$

Dividing both sides of this equation by S_{yy} and replacing S_{xy} by $b_1 S_{xx}$, we obtain the relation

$$b_1^2 \frac{S_{xx}}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}.$$

The value of $b_1^2 S_{xx}/S_{yy}$ is zero when $b_1 = 0$, which will occur when the sample points show no linear relationship. Since $S_{yy} \geq SSE$, we conclude that $b_1^2 S_{xx}/S_{xy}$ must be between 0 and l. Consequently, $b_1\sqrt{S_{xx}/S_{yy}}$ must range from -1 to $+1$, negative values corresponding to lines with negative slopes and positive values to lines with positive slopes. A value of -1 or $+1$ will occur when $SSE = 0$, but this is the case where all sample points lie in a straight line. Hence, a perfect linear relationship appears in the sample data when $b_1\sqrt{S_{xx}/S_{yy}} = \pm 1$. Clearly, the quantity $b_1\sqrt{S_{xx}/S_{yy}}$, which we shall henceforth designate as r, can be used as an estimate of the population correlation coefficient ρ. It is customary to refer to the estimate r as the **Pearson product-moment correlation coefficient** or simply the **sample correlation coefficient**.

Correlation Coefficient

The measure ρ of linear association between two variables X and Y is estimated by the **sample correlation coefficient** r, where

$$r = b_1\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}.$$

For values of r between -1 and $+1$ we must be careful in our interpretation. For example, values of r equal to 0.3 and 0.6 only mean that we have two positive correlations, one somewhat stronger than the other. It is wrong to conclude that $r = 0.6$ indicates a linear relationship twice as good as that indicated by the value $r = 0.3$. On the other hand, if we write

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{SSR}{S_{yy}},$$

then r^2, which is usually referred to as the **sample coefficient of determination**, represents the proportion of the variation of S_{yy} explained by the regression of Y on x, namely SSR. That is, r^2 expresses the proportion of the total variation in the values of the variable Y that can be accounted for or explained by a linear relationship with the values of the random variable X. Thus, a correlation of 0.6 means that 0.36, or 36%, of the total variation of the values of Y in our sample is accounted for by a linear relationship with values of X.

Example 11.10: It is important that scientific researchers in the area of forest products be able to study correlation among the anatomy and mechanical properties of trees. For the study *Quantitative Anatomical Characteristics of Plantation Grown Loblolly Pine (Pinus Taeda L.) and Cottonwood (Populus deltoides Bart. Ex Marsh.) and Their Relationships to Mechanical Properties*, conducted by the Department of Forestry and Forest Products at Virginia Tech, 29 loblolly pines were randomly selected for investigation. Table 11.9 shows the resulting data on the specific gravity in grams/cm^3 and the modulus of rupture in kilopascals (kPa). Compute and interpret the sample correlation coefficient.

Table 11.9: Data on 29 Loblolly Pines for Example 11.10

Specific Gravity, x (g/cm^3)	**Modulus of Rupture, y (kPa)**	**Specific Gravity, x (g/cm^3)**	**Modulus of Rupture, y (kPa)**
0.414	29,186	0.581	85,156
0.383	29,266	0.557	69,571
0.399	26,215	0.550	84,160
0.402	30,162	0.531	73,466
0.442	38,867	0.550	78,610
0.422	37,831	0.556	67,657
0.466	44,576	0.523	74,017
0.500	46,097	0.602	87,291
0.514	59,698	0.569	86,836
0.530	67,705	0.544	82,540
0.569	66,088	0.557	81,699
0.558	78,486	0.530	82,096
0.577	89,869	0.547	75,657
0.572	77,369	0.585	80,490
0.548	67,095		

Solution: From the data we find that

$$S_{xx} = 0.11273, \quad S_{yy} = 11{,}807{,}324{,}805, \quad S_{xy} = 34{,}422.27572.$$

Therefore,

$$r = \frac{34{,}422.27572}{\sqrt{(0.11273)(11{,}807{,}324{,}805)}} = 0.9435.$$

A correlation coefficient of 0.9435 indicates a good linear relationship between X and Y. Since $r^2 = 0.8902$, we can say that approximately 89% of the variation in the values of Y is accounted for by a linear relationship with X.

A test of the special hypothesis $\rho = 0$ versus an appropriate alternative is equivalent to testing $\beta_1 = 0$ for the simple linear regression model, and therefore the procedures of Section 11.8 using either the t-distribution with $n-2$ degrees of freedom or the F-distribution with 1 and $n-2$ degrees of freedom are applicable. However, if one wishes to avoid the analysis-of-variance procedure and compute only the sample correlation coefficient, it can be verified (see Review Exercise 11.66 on page 438) that the t-value

$$t = \frac{b_1}{s/\sqrt{S_{xx}}}$$

can also be written as

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

which, as before, is a value of the statistic T having a t-distribution with $n-2$ degrees of freedom.

Example 11.11: For the data of Example 11.10, test the hypothesis that there is no linear association among the variables.

Solution:
1. H_0: $\rho = 0$.
2. H_1: $\rho \neq 0$.
3. $\alpha = 0.05$.
4. Critical region: $t < -2.052$ or $t > 2.052$.
5. Computations: $t = \frac{0.9435\sqrt{27}}{\sqrt{1-0.9435^2}} = 14.79$, $P < 0.0001$.
6. Decision: Reject the hypothesis of no linear association.

A test of the more general hypothesis $\rho = \rho_0$ against a suitable alternative is easily conducted from the sample information. If X and Y follow the bivariate normal distribution, the quantity

$$\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

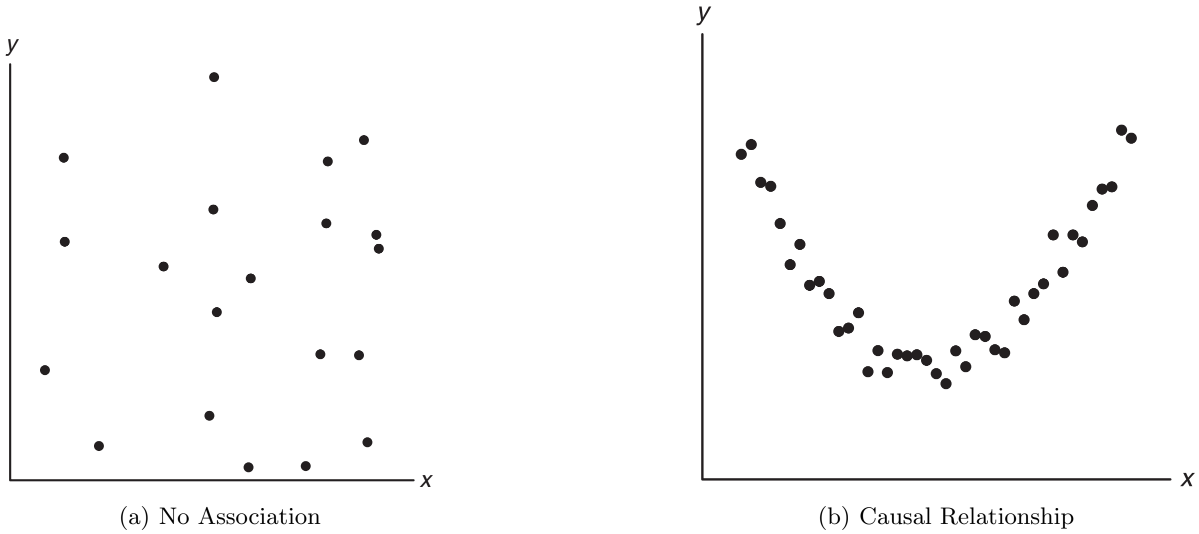
is the value of a random variable that follows approximately the normal distribution with mean $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ and variance $1/(n-3)$. Thus, the test procedure is to compute

$$z = \frac{\sqrt{n-3}}{2}\left[\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right] = \frac{\sqrt{n-3}}{2}\ln\left[\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)}\right]$$

and compare it with the critical points of the standard normal distribution.

Example 11.12: For the data of Example 11.10, test the null hypothesis that $\rho = 0.9$ against the alternative that $\rho > 0.9$. Use a 0.05 level of significance.

Solution:
1. H_0: $\rho = 0.9$.
2. H_1: $\rho > 0.9$.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$.



(a) No Association (b) Causal Relationship

Figure 11.28: Scatter diagram showing zero correlation.

5. Computations:

$$z = \frac{\sqrt{26}}{2} \ln \left[\frac{(1 + 0.9435)(0.1)}{(1 - 0.9435)(1.9)} \right] = 1.51, \quad P = 0.0655.$$

6. Decision: There is certainly some evidence that the correlation coefficient does not exceed 0.9.

It should be pointed out that in correlation studies, as in linear regression problems, the results obtained are only as good as the model that is assumed. In the correlation techniques studied here, a bivariate normal density is assumed for the variables X and Y, with the mean value of Y at each x-value being linearly related to x. To observe the suitability of the linearity assumption, a preliminary plotting of the experimental data is often helpful. A value of the sample correlation coefficient close to zero will result from data that display a strictly random effect as in Figure 11.28(a), thus implying little or no causal relationship. It is important to remember that the correlation coefficient between two variables is a measure of their linear relationship and that a value of $r = 0$ implies *a lack of linearity and not a lack of association.* Hence, if a strong quadratic relationship exists between X and Y, as indicated in Figure 11.28(b), we can still obtain a zero correlation indicating a nonlinear relationship.

Exercises

11.43 Compute and interpret the correlation coefficient for the following grades of 6 students selected at random:

Mathematics grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

11.44 With reference to Exercise 11.1 on page 398, assume that x and y are random variables with a bivariate normal distribution.

(a) Calculate r.

(b) Test the hypothesis that $\rho = 0$ against the alternative that $\rho \neq 0$ at the 0.05 level of significance.

11.45 With reference to Exercise 11.13 on page 400, assume a bivariate normal distribution for x and y.

(a) Calculate r.

(b) Test the null hypothesis that $\rho = -0.5$ against the alternative that $\rho < -0.5$ at the 0.025 level of significance.

(c) Determine the percentage of the variation in the amount of particulate removed that is due to changes in the daily amount of rainfall.

11.46 Test the hypothesis that $\rho = 0$ in Exercise 11.43 against the alternative that $\rho \neq 0$. Use a 0.05 level of significance.

11.47 The following data were obtained in a study of the relationship between the weight and chest size of infants at birth.

Weight (kg)	**Chest Size (cm)**
2.75	29.5
2.15	26.3
4.41	32.2
5.52	36.5
3.21	27.2
4.32	27.7
2.31	28.3
4.30	30.3
3.71	28.7

(a) Calculate r.

(b) Test the null hypothesis that $\rho = 0$ against the alternative that $\rho > 0$ at the 0.01 level of significance.

(c) What percentage of the variation in infant chest sizes is explained by difference in weight?

Review Exercises

11.48 With reference to Exercise 11.8 on page 399, construct

(a) a 95% confidence interval for the average course grade of students who make a 35 on the placement test;

(b) a 95% prediction interval for the course grade of a student who made a 35 on the placement test.

11.49 The Statistics Consulting Center at Virginia Tech analyzed data on normal woodchucks for the Department of Veterinary Medicine. The variables of interest were body weight in grams and heart weight in grams. It was desired to develop a linear regression equation in order to determine if there is a significant linear relationship between heart weight and total body weight. Use heart weight as the independent variable and body weight as the dependent variable and fit a simple linear regression using the following data. In addition, test the hypothesis H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$. Draw conclusions.

Body Weight (grams)	**Heart Weight (grams)**
4050	11.2
2465	12.4
3120	10.5
5700	13.2
2595	9.8
3640	11.0
2050	10.8
4235	10.4
2935	12.2
4975	11.2
3690	10.8
2800	14.2
2775	12.2
2170	10.0
2370	12.3
2055	12.5
2025	11.8
2645	16.0
2675	13.8

11.50 The amounts of solids removed from a particular material when exposed to drying periods of different lengths are as shown.

x **(hours)**	y **(grams)**	
4.4	13.1	14.2
4.5	9.0	11.5
4.8	10.4	11.5
5.5	13.8	14.8
5.7	12.7	15.1
5.9	9.9	12.7
6.3	13.8	16.5
6.9	16.4	15.7
7.5	17.6	16.9
7.8	18.3	17.2

(a) Estimate the linear regression line.

(b) Test at the 0.05 level of significance whether the linear model is adequate.

11.51 With reference to Exercise 11.9 on page 399, construct

(a) a 95% confidence interval for the average weekly sales when $45 is spent on advertising;

(b) a 95% prediction interval for the weekly sales when $45 is spent on advertising.

11.52 An experiment was designed for the Department of Materials Engineering at Virginia Tech to study hydrogen embrittlement properties based on electrolytic hydrogen pressure measurements. The solution used was 0.1 *N* NaOH, and the material was a certain type of stainless steel. The cathodic charging current density was controlled and varied at four levels. The effective hydrogen pressure was observed as the response. The data follow.

Run	Charging Current Density, x (mA/cm^2)	Effective Hydrogen Pressure, y (atm)
1	0.5	86.1
2	0.5	92.1
3	0.5	64.7
4	0.5	74.7
5	1.5	223.6
6	1.5	202.1
7	1.5	132.9
8	2.5	413.5
9	2.5	231.5
10	2.5	466.7
11	2.5	365.3
12	3.5	493.7
13	3.5	382.3
14	3.5	447.2
15	3.5	563.8

(a) Run a simple linear regression of y against x.

(b) Compute the pure error sum of squares and make a test for lack of fit.

(c) Does the information in part (b) indicate a need for a model in x beyond a first-order regression? Explain.

11.53 The following data represent the chemistry grades for a random sample of 12 freshmen at a certain college along with their scores on an intelligence test administered while they were still seniors in high school.

Student	Test Score, x	Chemistry Grade, y
1	65	85
2	50	74
3	55	76
4	65	90
5	55	85
6	70	87
7	65	94
8	70	98
9	55	81
10	70	91
11	50	76
12	55	74

(a) Compute and interpret the sample correlation coefficient.

(b) State necessary assumptions on random variables.

(c) Test the hypothesis that $\rho = 0.5$ against the alternative that $\rho > 0.5$. Use a P-value in the conclusion.

11.54 The business section of the *Washington Times* in March of 1997 listed 21 different used computers and printers and their sale prices. Also listed was the average hover bid. Partial results from regression analysis using *SAS* software are shown in Figure 11.29 on page 439.

(a) Explain the difference between the confidence interval on the mean and the prediction interval.

(b) Explain why the standard errors of prediction vary from observation to observation.

(c) Which observation has the lowest standard error of prediction? Why?

11.55 Consider the vehicle data from *Consumer Reports* in Figure 11.30 on page 440. Weight is in tons, mileage in miles per gallon, and drive ratio is also indicated. A regression model was fitted relating weight x to mileage y. A partial *SAS* printout in Figure 11.30 on page 440 shows some of the results of that regression analysis, and Figure 11.31 on page 441 gives a plot of the residuals and weight for each vehicle.

(a) From the analysis and the residual plot, does it appear that an improved model might be found by using a transformation? Explain.

(b) Fit the model by replacing weight with log weight. Comment on the results.

(c) Fit a model by replacing mpg with gallons per 100 miles traveled, as mileage is often reported in other countries. Which of the three models is preferable? Explain.

11.56 Observations on the yield of a chemical reaction taken at various temperatures were recorded as follows:

x (°C)	y (%)	x (°C)	y (%)
150	75.4	150	77.7
150	81.2	200	84.4
200	85.5	200	85.7
250	89.0	250	89.4
250	90.5	300	94.8
300	96.7	300	95.3

(a) Plot the data.

(b) Does it appear from the plot as if the relationship is linear?

(c) Fit a simple linear regression and test for lack of fit.

(d) Draw conclusions based on your result in (c).

11.57 Physical fitness testing is an important aspect of athletic training. A common measure of the magnitude of cardiovascular fitness is the maximum volume of oxygen uptake during strenuous exercise. A study was conducted on 24 middle-aged men to determine the influence on oxygen uptake of the time required to complete a two-mile run. Oxygen uptake was measured with standard laboratory methods as the subjects performed on a treadmill. The work was published in "Maximal Oxygen Intake Prediction in Young and Middle Aged Males," *Journal of Sports Medicine* **9**, 1969, 17–22. The data are as follows:

Subject	***y*, Maximum Volume of O_2**	***x*, Time in Seconds**
1	42.33	918
2	53.10	805
3	42.08	892
4	50.06	962
5	42.45	968
6	42.46	907
7	47.82	770
8	49.92	743
9	36.23	1045
10	49.66	810
11	41.49	927
12	46.17	813
13	46.18	858
14	43.21	860
15	51.81	760
16	53.28	747
17	53.29	743
18	47.18	803
19	56.91	683
20	47.80	844
21	48.65	755
22	53.67	700
23	60.62	748
24	56.73	775

(a) Estimate the parameters in a simple linear regression model.

(b) Does the time it takes to run two miles have a significant influence on maximum oxygen uptake? Use H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$.

(c) Plot the residuals on a graph against x and comment on the appropriateness of the simple linear model.

11.58 Suppose a scientist postulates a model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, n,$$

and β_0 is a **known value**, not necessarily zero.

(a) What is the appropriate least squares estimator of β_1? Justify your answer.

(b) What is the variance of the slope estimator?

11.59 For the simple linear regression model, prove that $E(s^2) = \sigma^2$.

11.60 Assuming that the ϵ_i are independent and normally distributed with zero means and common variance σ^2, show that B_0, the least squares estimator of β_0 in $\mu_{Y|x} = \beta_0 + \beta_1 x$, is normally distributed with mean β_0 and variance

$$\sigma_{B_0}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2} \sigma^2.$$

11.61 For a simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where the ϵ_i are independent and normally distributed with zero means and equal variances σ^2, show that $\bar{Y}$ and

$$B_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

have zero covariance.

11.62 Show, in the case of a least squares fit to the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

that $\sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0.$

11.63 Consider the situation of Review Exercise 11.62 but suppose $n = 2$ (i.e., only two data points are available). Give an argument that the least squares regression line will result in $(y_1 - \hat{y}_1) = (y_2 - \hat{y}_2) = 0$. Also show that for this case $R^2 = 1.0$.

11.64 In Review Exercise 11.62, the student was required to show that $\sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$ for a standard simple linear regression model. Does the same hold for a model with zero intercept? Show why or why not.

11.65 Suppose that an experimenter postulates a model of the type

$$Y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

when in fact an additional variable, say x_2, also contributes linearly to the response. The true model is then given by

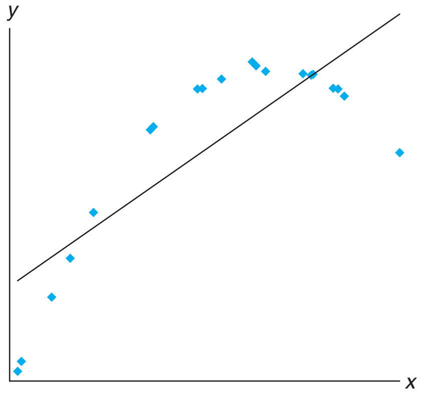
$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

Compute the expected value of the estimator

$$B_1 = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1) Y_i}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2}.$$

11.66 Show the necessary steps in converting the equation $r = \frac{b_1}{s/\sqrt{S_{xx}}}$ to the equivalent form $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

11.67 Consider the fictitious set of data shown below, where the line through the data is the fitted simple linear regression line. Sketch a residual plot.



11.68 **Project**: This project can be done in groups or as individuals. Each group or person must find a set of data, preferably but not restricted to their field of study. The data need to fit the regression framework with a regression variable x and a response variable y. Carefully make the assignment as to which variable is x and which y. It may be necessary to consult a journal or periodical from your field if you do not have other research data available.

(a) Plot y versus x. Comment on the relationship as seen from the plot.

(b) Fit an appropriate regression model from the data. Use simple linear regression or fit a polynomial model to the data. Comment on measures of quality.

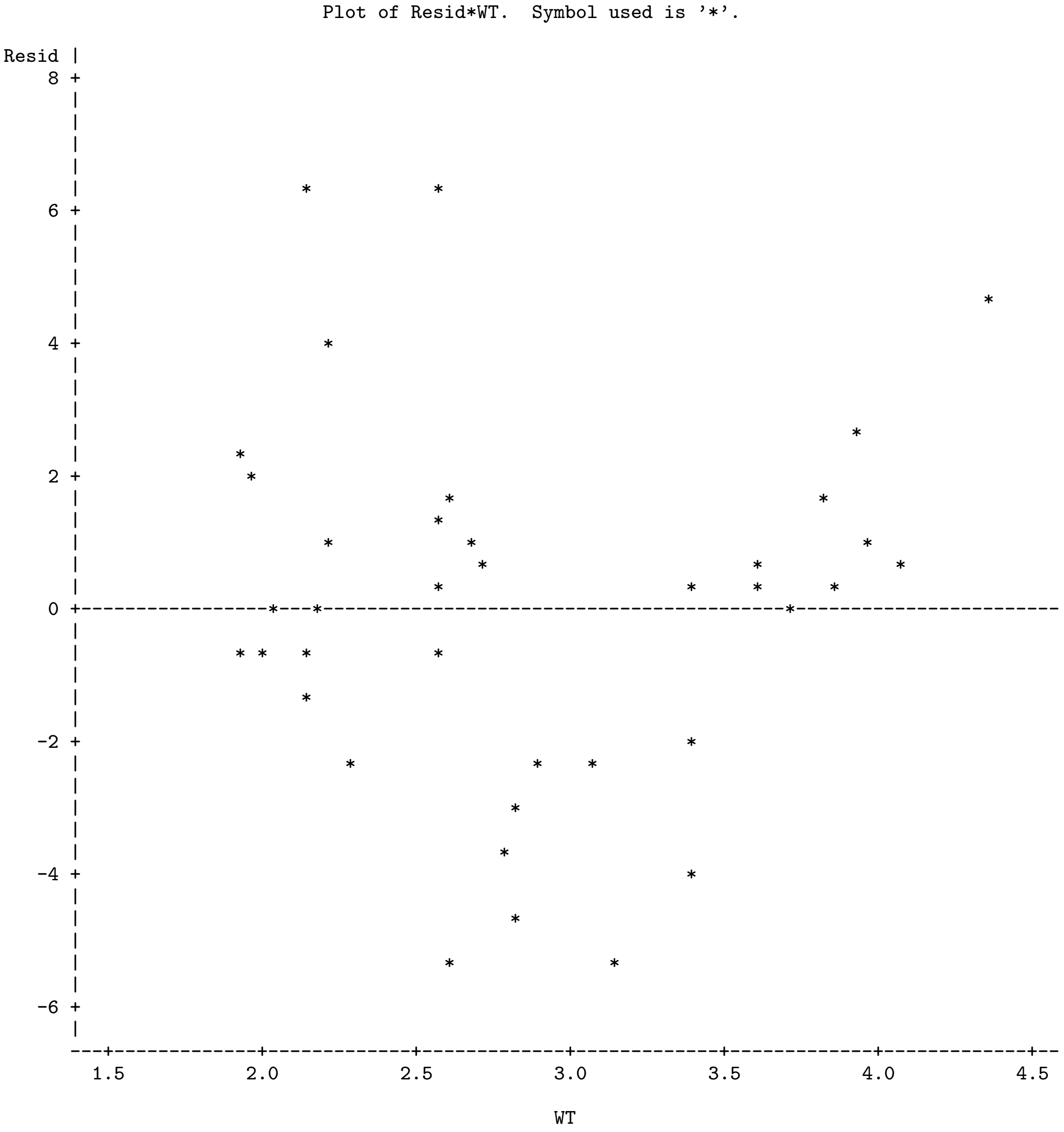
(c) Plot residuals as illustrated in the text. Check possible violation of assumptions. Show graphically a plot of confidence intervals on a mean response plotted against x. Comment.

```
     R-Square     Coeff Var      Root MSE    Price Mean
     0.967472      7.923338      70.83841      894.0476
                                  Standard
Parameter         Estimate           Error    t Value    Pr > |t|
Intercept      59.93749137     38.34195754       1.56      0.1345
Buyer           1.04731316      0.04405635      23.77      <.0001
                                      Predict Std Err Lower 95% Upper 95% Lower 95% Upper 95%
 product                  Buyer Price   Value Predict      Mean      Mean   Predict   Predict
 IBM PS/1 486/66 420MB      325   375  400.31 25.8906    346.12    454.50    242.46    558.17
 IBM ThinkPad 500           450   625  531.23 21.7232    485.76    576.70    376.15    686.31
 IBM Think-Dad 755CX       1700  1850 1840.37 42.7041   1750.99   1929.75   1667.25   2013.49
 AST Pentium 90 540MB       800   875  897.79 15.4590    865.43    930.14    746.03   1049.54
 Dell Pentium 75 1GB        650   700  740.69 16.7503    705.63    775.75    588.34    893.05
 Gateway 486/75 320MB       700   750  793.06 16.0314    759.50    826.61    641.04    945.07
 Clone 586/133 1GB          500   600  583.59 20.2363    541.24    625.95    429.40    737.79
 Compaq Contura 4/25 120MB  450   600  531.23 21.7232    485.76    576.70    376.15    686.31
 Compaq Deskpro P90 1.2GB   800   850  897.79 15.4590    865.43    930.14    746.03   1049.54
 Micron P75 810MB           800   675  897.79 15.4590    865.43    930.14    746.03   1049.54
 Micron P100 1.2GB          900   975 1002.52 16.1176    968.78   1036.25    850.46   1154.58
 Mac Quadra 840AV 500MB     450   575  531.23 21.7232    485.76    576.70    376.15    686.31
 Mac Performer 6116 700MB   700   775  793.06 16.0314    759.50    826.61    641.04    945.07
 PowerBook 540c 320MB      1400  1500 1526.18 30.7579   1461.80   1590.55   1364.54   1687.82
 PowerBook 5300 500MB      1350  1575 1473.81 28.8747   1413.37   1534.25   1313.70   1633.92
 Power Mac 7500/100 1GB    1150  1325 1264.35 21.9454   1218.42   1310.28   1109.13   1419.57
 NEC Versa 486 340MB        800   900  897.79 15.4590    865.43    930.14    746.03   1049.54
 Toshiba 1960CS 320MB       700   825  793.06 16.0314    759.50    826.61    641.04    945.07
 Toshiba 4800VCT 500MB     1000  1150 1107.25 17.8715   1069.85   1144.66    954.34   1260.16
 HP Laser jet III           350   475  426.50 25.0157    374.14    478.86    269.26    583.74
 Apple Laser Writer Pro 63  750   800  845.42 15.5930    812.79    878.06    693.61    997.24
```

Figure 11.29: *SAS* printout, showing partial analysis of data of Review Exercise 11.54.

```
 Obs    Model                          WT      MPG    DR_RATIO
   1    Buick Estate Wagon          4.360    16.9      2.73
   2    Ford Country Squire Wagon   4.054    15.5      2.26
   3    Chevy Ma libu Wagon         3.605    19.2      2.56
   4    Chrysler LeBaron Wagon      3.940    18.5      2.45
   5    Chevette                    2.155    30.0      3.70
   6    Toyota Corona               2.560    27.5      3.05
   7    Datsun 510                  2.300    27.2      3.54
   8    Dodge Omni                  2.230    30.9      3.37
   9    Audi 5000                   2.830    20.3      3.90
  10    Volvo 240 CL                3.140    17.0      3.50
  11    Saab 99 GLE                 2.795    21.6      3.77
  12    Peugeot 694 SL              3.410    16.2      3.58
  13    Buick Century Special       3.380    20.6      2.73
  14    Mercury Zephyr              3.070    20.8      3.08
  15    Dodge Aspen                 3.620    18.6      2.71
  16    AMC Concord D/L             3.410    18.1      2.73
  17    Chevy Caprice Classic       3.840    17.0      2.41
  18    Ford LTP                    3.725    17.6      2.26
  19    Mercury Grand Marquis       3.955    16.5      2.26
  20    Dodge St Regis              3.830    18.2      2.45
  21    Ford Mustang 4              2.585    26.5      3.08
  22    Ford Mustang Ghia           2.910    21.9      3.08
  23    Macda GLC                   1.975    34.1      3.73
  24    Dodge Colt                  1.915    35.1      2.97
  25    AMC Spirit                  2.670    27.4      3.08
  26    VW Scirocco                 1.990    31.5      3.78
  27    Honda Accord LX             2.135    29.5      3.05
  28    Buick Skylark               2.570    28.4      2.53
  29    Chevy Citation              2.595    28.8      2.69
  30    Olds Omega                  2.700    26.8      2.84
  31    Pontiac Phoenix             2.556    33.5      2.69
  32    Plymouth Horizon            2.200    34.2      3.37
  33    Datsun 210                  2.020    31.8      3.70
  34    Fiat Strada                 2.130    37.3      3.10
  35    VW Dasher                   2.190    30.5      3.70
  36    Datsun 810                  2.815    22.0      3.70
  37    BMW 320i                    2.600    21.5      3.64
  38    VW Rabbit                   1.925    31.9      3.78
    R-Square     Coeff Var      Root MSE      MPG Mean
    0.817244      11.46010      2.837580      24.76053
                                  Standard
Parameter         Estimate           Error    t Value    Pr > |t|
Intercept      48.67928080      1.94053995      25.09      <.0001
WT             -8.36243141      0.65908398     -12.69      <.0001
```

Figure 11.30: *SAS* printout, showing partial analysis of data of Review Exercise 11.55.



```
                             Plot of Resid*WT.  Symbol used is '*'.

Resid |
    8 +
      |
      |
      |
      |
      |                    *           *
    6 +
      |
      |
      |
      |                                                                                  *
      |
    4 +                      *
      |
      |
      |
      |                                                                      *
      |              *
    2 +               *
      |                                 *                                 *
      |                                *
      |                      *            *                                   *
      |                                    *                        *            *
      |                                *                      *     *      *
    0 +-----------------*---*------------------------------------------*------------------------
      |
      |              * *   *           *
      |
      |                    *
      |
   -2 +                                                       *
      |                        *                *    *
      |
      |                                       *
      |
      |                                      *
   -4 +                                                       *
      |
      |                                       *
      |
      |                                 *              *
      |
   -6 +
      |
      ---+-------------+-------------+-------------+-------------+-------------+-------------+--
        1.5           2.0           2.5           3.0           3.5           4.0           4.5

                                                  WT
```

Figure 11.31: *SAS* printout, showing residual plot of Review Exercise 11.55.

11.13 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

Anytime one is considering the use of simple linear regression, a plot of the data is not only recommended but essential. A plot of the ordinary residuals and a normal probability plot of these residuals are always edifying. In addition, we introduce and illustrate an additional type of residual in Chapter 12 that is in a standardized form. All of these plots are designed to detect violation of assumptions.

The use of t-statistics for tests on regression coefficients is reasonably robust to the normality assumption. The homogeneous variance assumption is crucial, and residual plots are designed to detect a violation.

The material in this chapter is used heavily in Chapters 12 and 15. All of the information involving the method of least squares in the development of regression models carries over into Chapter 12. The difference is that Chapter 12 deals with the scientific conditions in which there is more than a single x variable, i.e., more than one regression variable. However, material in the current chapter that deals with regression diagnostics, types of residual plots, measures of model quality, and so on, applies and will carry over. The student will realize that more complications occur in Chapter 12 because the problems in multiple regression models often involve the backdrop of questions regarding how the various regression variables enter the model and even issues of which variables should remain in the model. Certainly Chapter 15 heavily involves the use of regression modeling, but we will preview the connection in the summary at the end of Chapter 12.

Chapter 12

Multiple Linear Regression and Certain Nonlinear Regression Models

---

12.1 Introduction

In most research problems where regression analysis is applied, more than one independent variable is needed in the regression model. The complexity of most scientific mechanisms is such that in order to be able to predict an important response, a **multiple regression model** is needed. When this model is linear in the coefficients, it is called a **multiple linear regression model**. For the case of k independent variables $x_1, x_2, \ldots, x_k$, the mean of $Y|x_1, x_2, \ldots, x_k$ is given by the multiple linear regression model

$$\mu_{Y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k,$$

and the estimated response is obtained from the sample regression equation

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_k x_k,$$

where each regression coefficient β_i is estimated by b_i from the sample data using the method of least squares. As in the case of a single independent variable, the multiple linear regression model can often be an adequate representation of a more complicated structure within certain ranges of the independent variables.

Similar least squares techniques can also be applied for estimating the coefficients when the linear model involves, say, powers and products of the independent variables. For example, when $k = 1$, the experimenter may believe that the means $\mu_{Y|x}$ do not fall on a straight line but are more appropriately described by the **polynomial regression model**

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r,$$

and the estimated response is obtained from the polynomial regression equation

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \cdots + b_r x^r.$$

Confusion arises occasionally when we speak of a polynomial model as a linear model. However, statisticians normally refer to a linear model as one in which the parameters occur linearly, regardless of how the independent variables enter the model. An example of a nonlinear model is the **exponential relationship**

$$\mu_{Y|x} = \alpha\beta^x,$$

whose response is estimated by the regression equation

$$\hat{y} = ab^x.$$

There are many phenomena in science and engineering that are inherently nonlinear in nature, and when the true structure is known, an attempt should certainly be made to fit the actual model. The literature on estimation by least squares of nonlinear models is voluminous. The nonlinear models discussed in this chapter deal with nonideal conditions in which the analyst is certain that the response and hence the response model error are not normally distributed but, rather, have a binomial or Poisson distribution. These situations do occur extensively in practice.

A student who wants a more general account of nonlinear regression should consult *Classical and Modern Regression with Applications* by Myers (1990; see the Bibliography).

12.2 Estimating the Coefficients

In this section, we obtain the least squares estimators of the parameters $\beta_0, \beta_1, \ldots, \beta_k$ by fitting the multiple linear regression model

$$\mu_{Y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

to the data points

$$\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i); \quad i = 1, 2, \ldots, n \text{ and } n > k\},$$

where y_i is the observed response to the values $x_{1i}, x_{2i}, \ldots, x_{ki}$ of the k independent variables $x_1, x_2, \ldots, x_k$. Each observation $(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i)$ is assumed to satisfy the following equation.

Multiple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

or

$$y_i = \hat{y}_i + e_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + e_i,$$

where ϵ_i and e_i are the random error and residual, respectively, associated with the response y_i and fitted value $\hat{y}_i$.

As in the case of simple linear regression, it is assumed that the ϵ_i are independent and identically distributed with mean 0 and common variance σ^2.

In using the concept of least squares to arrive at estimates $b_0, b_1, \ldots, b_k$, we minimize the expression

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \cdots - b_k x_{ki})^2.$$

Differentiating SSE in turn with respect to $b_0, b_1, \ldots, b_k$ and equating to zero, we generate the set of $k + 1$ **normal equations for multiple linear regression**.

Normal Estimation Equations for Multiple Linear Regression

$$\begin{aligned}
nb_0 + b_1\sum_{i=1}^{n} x_{1i} \quad &+ b_2\sum_{i=1}^{n} x_{2i} \quad + \cdots + b_k\sum_{i=1}^{n} x_{ki} \quad = \sum_{i=1}^{n} y_i \\
b_0\sum_{i=1}^{n} x_{1i} + b_1\sum_{i=1}^{n} x_{1i}^2 \quad &+ b_2\sum_{i=1}^{n} x_{1i}x_{2i} + \cdots + b_k\sum_{i=1}^{n} x_{1i}x_{ki} = \sum_{i=1}^{n} x_{1i}y_i \\
\vdots \qquad\quad \vdots \qquad\qquad &\qquad\quad \vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots \\
b_0\sum_{i=1}^{n} x_{ki} + b_1\sum_{i=1}^{n} x_{ki}x_{1i} &+ b_2\sum_{i=1}^{n} x_{ki}x_{2i} + \cdots + b_k\sum_{i=1}^{n} x_{ki}^2 \quad = \sum_{i=1}^{n} x_{ki}y_i
\end{aligned}$$

These equations can be solved for $b_0, b_1, b_2, \ldots, b_k$ by any appropriate method for solving systems of linear equations. Most statistical software can be used to obtain numerical solutions of the above equations.

Example 12.1: A study was done on a diesel-powered light-duty pickup truck to see if humidity, air temperature, and barometric pressure influence emission of nitrous oxide (in ppm). Emission measurements were taken at different times, with varying experimental conditions. The data are given in Table 12.2. The model is

$$\mu_{Y|x_1,x_2,x_3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

or, equivalently,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \quad i = 1, 2, \ldots, 20.$$

Fit this multiple linear regression model to the given data and then estimate the amount of nitrous oxide emitted for the conditions where humidity is 50%, temperature is 76°F, and barometric pressure is 29.30.

Table 12.1: Data for Example 12.1

Nitrous Oxide, y	**Humidity, x_1**	**Temp., x_2**	**Pressure, x_3**	**Nitrous Oxide, y**	**Humidity, x_1**	**Temp., x_2**	**Pressure, x_3**
0.90	72.4	76.3	29.18	1.07	23.2	76.8	29.38
0.91	41.6	70.3	29.35	0.94	47.4	86.6	29.35
0.96	34.3	77.1	29.24	1.10	31.5	76.9	29.63
0.89	35.1	68.0	29.27	1.10	10.6	86.3	29.56
1.00	10.7	79.0	29.78	1.10	11.2	86.0	29.48
1.10	12.9	67.4	29.39	0.91	73.3	76.3	29.40
1.15	8.3	66.8	29.69	0.87	75.4	77.9	29.28
1.03	20.1	76.9	29.48	0.78	96.6	78.7	29.29
0.77	72.2	77.7	29.09	0.82	107.4	86.8	29.03
1.07	24.0	67.7	29.60	0.95	54.9	70.9	29.37

Source: Charles T. Hare, "Light-Duty Diesel Emission Correction Factors for Ambient Conditions," EPA-600/2-77-116. U.S. Environmental Protection Agency.

***Solution*:** The solution of the set of estimating equations yields the unique estimates

$$b_0 = -3.507778, \; b_1 = -0.002625, \; b_2 = 0.000799, \; b_3 = 0.154155.$$

Therefore, the regression equation is

$$\hat{y} = -3.507778 - 0.002625\,x_1 + 0.000799\,x_2 + 0.154155\,x_3.$$

For 50% humidity, a temperature of 76°F, and a barometric pressure of 29.30, the estimated amount of nitrous oxide emitted is

$$\begin{aligned}\hat{y} &= -3.507778 - 0.002625(50.0) + 0.000799(76.0) + 0.1541553(29.30)\\ &= 0.9384 \text{ ppm.}\end{aligned}$$

Polynomial Regression

Now suppose that we wish to fit the polynomial equation

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r$$

to the n pairs of observations $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$. Each observation, y_i, satisfies the equation

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_r x_i^r + \epsilon_i$$

or

$$y_i = \hat{y}_i + e_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_r x_i^r + e_i,$$

where r is the degree of the polynomial and ϵ_i and e_i are again the random error and residual associated with the response y_i and fitted value $\hat{y}_i$, respectively. Here, the number of pairs, n, must be at least as large as $r+1$, the number of parameters to be estimated.

Notice that the polynomial model can be considered a special case of the more general multiple linear regression model, where we set $x_1 = x, x_2 = x^2, \ldots, x_r = x^r$. The normal equations assume the same form as those given on page 445. They are then solved for $b_0, b_1, b_2, \ldots, b_r$.

Example 12.2: Given the data

x	0	1	2	3	4	5	6	7	8	9
y	9.1	7.3	3.2	4.6	4.8	2.9	5.7	7.1	8.8	10.2

fit a regression curve of the form $\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$ and then estimate $\mu_{Y|2}$.

***Solution*:** From the data given, we find that

$$\begin{aligned}10b_0 + 45\,b_1 + 285\,b_2 &= 63.7,\\ 45b_0 + 285b_1 + 2025\,b_2 &= 307.3,\\ 285b_0 + 2025\,b_1 + 15{,}333b_2 &= 2153.3.\end{aligned}$$

Solving these normal equations, we obtain

$$b_0 = 8.698, \quad b_1 = -2.341, \quad b_2 = 0.288.$$

Therefore,

$$\hat{y} = 8.698 - 2.341x + 0.288x^2.$$

When $x = 2$, our estimate of $\mu_{Y|2}$ is

$$\hat{y} = 8.698 - (2.341)(2) + (0.288)(2^2) = 5.168.$$

Example 12.3: The data in Table 12.2 represent the percent of impurities that resulted for various temperatures and sterilizing times during a reaction associated with the manufacturing of a certain beverage. Estimate the regression coefficients in the polynomial model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \epsilon_i,$$

for $i = 1, 2, \ldots, 18$.

Table 12.2: Data for Example 12.3

Sterilizing Time, x_2 (min)	**Temperature, x_1 (°C)**		
	75	**100**	**125**
15	14.05	10.55	7.55
	14.93	9.48	6.59
20	16.56	13.63	9.23
	15.85	11.75	8.78
25	22.41	18.55	15.93
	21.66	17.98	16.44

***Solution*:** Using the normal equations, we obtain

$$b_0 = 56.4411, \quad b_1 = -0.36190, \quad b_2 = -2.75299,$$
$$b_{11} = 0.00081, \quad b_{22} = 0.08173, \quad b_{12} = 0.00314,$$

and our estimated regression equation is

$$\hat{y} = 56.4411 - 0.36190x_1 - 2.75299x_2 + 0.00081x_1^2 + 0.08173x_2^2 + 0.00314x_1x_2.$$

Many of the principles and procedures associated with the estimation of polynomial regression functions fall into the category of **response surface methodology**, a collection of techniques that have been used quite successfully by scientists and engineers in many fields. The x_i^2 are called **pure quadratic terms**, and the $x_i x_j$ $(i \neq j)$ are called **interaction terms**. Such problems as selecting a proper experimental design, particularly in cases where a large number of variables are in the model, and choosing optimum operating conditions for $x_1, x_2, \ldots, x_k$ are often approached through the use of these methods. For an extensive exposure, the reader is referred to *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* by Myers, Montgomery, and Anderson-Cook (2009; see the Bibliography).

12.3 Linear Regression Model Using Matrices

In fitting a multiple linear regression model, particularly when the number of variables exceeds two, a knowledge of matrix theory can facilitate the mathematical manipulations considerably. Suppose that the experimenter has k independent variables $x_1, x_2, \ldots, x_k$ and n observations $y_1, y_2, \ldots, y_n$, each of which can be expressed by the equation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i.$$

This model essentially represents n equations describing how the response values are generated in the scientific process. Using matrix notation, we can write the following equation:

General Linear Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Then the least squares method for estimation of β, illustrated in Section 12.2, involves finding $\mathbf{b}$ for which

$$SSE = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

is minimized. This minimization process involves solving for $\mathbf{b}$ in the equation

$$\frac{\partial}{\partial \mathbf{b}}(SSE) = \mathbf{0}.$$

We will not present the details regarding solution of the equations above. The result reduces to the solution of $\mathbf{b}$ in

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

Notice the nature of the $\mathbf{X}$ matrix. Apart from the initial element, the ith row represents the x-values that give rise to the response y_i. Writing

$$\mathbf{A} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{ki}x_{1i} & \sum_{i=1}^{n} x_{ki}x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{bmatrix}$$

and

$$\mathbf{g} = \mathbf{X}'\mathbf{y} = \begin{bmatrix} g_0 = \sum_{i=1}^{n} y_i \\ g_1 = \sum_{i=1}^{n} x_{1i}y_i \\ \vdots \\ g_k = \sum_{i=1}^{n} x_{ki}y_i \end{bmatrix}$$

allows the normal equations to be put in the matrix form

$$\mathbf{A}\mathbf{b} = \mathbf{g}.$$

If the matrix $\mathbf{A}$ is nonsingular, we can write the solution for the regression coefficients as

$$\mathbf{b} = \mathbf{A}^{-1}\mathbf{g} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Thus, we can obtain the prediction equation or regression equation by solving a set of $k+1$ equations in a like number of unknowns. This involves the inversion of the $k+1$ by $k+1$ matrix $\mathbf{X}'\mathbf{X}$. Techniques for inverting this matrix are explained in most textbooks on elementary determinants and matrices. Of course, there are many high-speed computer packages available for multiple regression problems, packages that not only print out estimates of the regression coefficients but also provide other information relevant to making inferences concerning the regression equation.

Example 12.4: The percent survival rate of sperm in a certain type of animal semen, after storage, was measured at various combinations of concentrations of three materials used to increase chance of survival. The data are given in Table 12.3. Estimate the multiple linear regression model for the given data.

Table 12.3: Data for Example 12.4

y (% survival)	**x_1 (weight %)**	**x_2 (weight %)**	**x_3 (weight %)**
25.5	1.74	5.30	10.80
31.2	6.32	5.42	9.40
25.9	6.22	8.41	7.20
38.4	10.52	4.63	8.50
18.4	1.19	11.60	9.40
26.7	1.22	5.85	9.90
26.4	4.10	6.62	8.00
25.9	6.32	8.72	9.10
32.0	4.08	4.42	8.70
25.2	4.15	7.60	9.20
39.7	10.15	4.83	9.40
35.7	1.72	3.12	7.60
26.5	1.70	5.30	8.20

Solution: The least squares estimating equations, $(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}$, are

$$\begin{bmatrix} 13.0 & 59.43 & 81.82 & 115.40 \\ 59.43 & 394.7255 & 360.6621 & 522.0780 \\ 81.82 & 360.6621 & 576.7264 & 728.3100 \\ 115.40 & 522.0780 & 728.3100 & 1035.9600 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 377.5 \\ 1877.567 \\ 2246.661 \\ 3337.780 \end{bmatrix}.$$

From a computer readout we obtain the elements of the inverse matrix

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 8.0648 & -0.0826 & -0.0942 & -0.7905 \\ -0.0826 & 0.0085 & 0.0017 & 0.0037 \\ -0.0942 & 0.0017 & 0.0166 & -0.0021 \\ -0.7905 & 0.0037 & -0.0021 & 0.0886 \end{bmatrix},$$

and then, using the relation $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, the estimated regression coefficients are obtained as

$$b_0 = 39.1574,\ b_1 = 1.0161,\ b_2 = -1.8616,\ b_3 = -0.3433.$$

Hence, our estimated regression equation is

$$\hat{y} = 39.1574 + 1.0161x_1 - 1.8616x_2 - 0.3433x_3.$$

Exercises

12.1 A set of experimental runs was made to determine a way of predicting cooking time y at various values of oven width x_1 and flue temperature x_2. The coded data were recorded as follows:

y	x_1	x_2
6.40	1.32	1.15
15.05	2.69	3.40
18.75	3.56	4.10
30.25	4.41	8.75
44.85	5.35	14.82
48.94	6.20	15.15
51.55	7.12	15.32
61.50	8.87	18.18
100.44	9.80	35.19
111.42	10.65	40.40

Estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

12.2 In *Applied Spectroscopy*, the infrared reflectance spectra properties of a viscous liquid used in the electronics industry as a lubricant were studied. The designed experiment consisted of the effect of band frequency x_1 and film thickness x_2 on optical density y using a Perkin-Elmer Model 621 infrared spectrometer. (*Source*: Pacansky, J., England, C. D., and Wattman, R., 1986.)

y	x_1	x_2
0.231	740	1.10
0.107	740	0.62
0.053	740	0.31
0.129	805	1.10
0.069	805	0.62
0.030	805	0.31
1.005	980	1.10
0.559	980	0.62
0.321	980	0.31
2.948	1235	1.10
1.633	1235	0.62
0.934	1235	0.31

Estimate the multiple linear regression equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

12.3 Suppose in Review Exercise 11.53 on page 437 that we were also given the number of class periods missed by the 12 students taking the chemistry course. The complete data are shown.

Student	**Chemistry Grade, y**	**Test Score, x_1**	**Classes Missed, x_2**
1	85	65	1
2	74	50	7
3	76	55	5
4	90	65	2
5	85	55	6
6	87	70	3
7	94	65	2
8	98	70	5
9	81	55	4
10	91	70	3
11	76	50	1
12	74	55	4

(a) Fit a multiple linear regression equation of the form $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$.

(b) Estimate the chemistry grade for a student who has an intelligence test score of 60 and missed 4 classes.

12.4 An experiment was conducted to determine if the weight of an animal can be predicted after a given period of time on the basis of the initial weight of the animal and the amount of feed that was eaten. The following data, measured in kilograms, were recorded:

Final Weight, y	**Initial Weight, x_1**	**Feed Weight, x_2**
95	42	272
77	33	226
80	33	259
100	45	292
97	39	311
70	36	183
50	32	173
80	41	236
92	40	230
84	38	235

(a) Fit a multiple regression equation of the form

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

(b) Predict the final weight of an animal having an initial weight of 35 kilograms that is given 250 kilograms of feed.

12.5 The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature x_1, the number of days in the month x_2, the average product purity x_3, and the tons of product produced x_4. The past year's historical data are available and are presented in the following table.

y	x_1	x_2	x_3	x_4
240	25	24	91	100
236	31	21	90	95
290	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

(a) Fit a multiple linear regression model using the above data set.

(b) Predict power consumption for a month in which $x_1 = 75°$F, $x_2 = 24$ days, $x_3 = 90\%$, and $x_4 = 98$ tons.

12.6 An experiment was conducted on a new model of a particular make of automobile to determine the stopping distance at various speeds. The following data were recorded.

Speed, v (km/hr)	35	50	65	80	95	110
Stopping Distance, d (m)	16	26	41	62	88	119

(a) Fit a multiple regression curve of the form $\mu_{D|v} = \beta_0 + \beta_1 v + \beta_2 v^2$.

(b) Estimate the stopping distance when the car is traveling at 70 kilometers per hour.

12.7 An experiment was conducted in order to determine if cerebral blood flow in human beings can be predicted from arterial oxygen tension (millimeters of mercury). Fifteen patients participated in the study, and the following data were collected:

Blood Flow, y	**Arterial Oxygen Tension, x**
84.33	603.40
87.80	582.50
82.20	556.20
78.21	594.60
78.44	558.90
80.01	575.20
83.53	580.10
79.46	451.20
75.22	404.00
76.58	484.00
77.90	452.40
78.80	448.40
80.67	334.80
86.60	320.30
78.20	350.30

Estimate the quadratic regression equation

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2.$$

12.8 The following is a set of coded experimental data on the compressive strength of a particular alloy at various values of the concentration of some additive:

Concentration, x	**Compressive Strength, y**		
10.0	25.2	27.3	28.7
15.0	29.8	31.1	27.8
20.0	31.2	32.6	29.7
25.0	31.7	30.1	32.3
30.0	29.4	30.8	32.8

(a) Estimate the quadratic regression equation $\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$.

(b) Test for lack of fit of the model.

12.9 (a) Fit a multiple regression equation of the form $\mu_{Y|x} = \beta_0 + \beta_1 x_1 + \beta_2 x^2$ to the data of Example 11.8 on page 420.

(b) Estimate the yield of the chemical reaction for a temperature of 225°C.

12.10 The following data are given:

x	0	1	2	3	4	5	6
y	1	4	5	3	2	3	4

(a) Fit the cubic model $\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

(b) Predict Y when $x = 2$.

12.11 An experiment was conducted to study the size of squid eaten by sharks and tuna. The regressor variables are characteristics of the beaks of the squid. The data are given as follows:

x_1	x_2	x_3	x_4	x_5	y
1.31	1.07	0.44	0.75	0.35	1.95
1.55	1.49	0.53	0.90	0.47	2.90
0.99	0.84	0.34	0.57	0.32	0.72
0.99	0.83	0.34	0.54	0.27	0.81
1.01	0.90	0.36	0.64	0.30	1.09
1.09	0.93	0.42	0.61	0.31	1.22
1.08	0.90	0.40	0.51	0.31	1.02
1.27	1.08	0.44	0.77	0.34	1.93
0.99	0.85	0.36	0.56	0.29	0.64
1.34	1.13	0.45	0.77	0.37	2.08
1.30	1.10	0.45	0.76	0.38	1.98
1.33	1.10	0.48	0.77	0.38	1.90
1.86	1.47	0.60	1.01	0.65	8.56
1.58	1.34	0.52	0.95	0.50	4.49
1.97	1.59	0.67	1.20	0.59	8.49
1.80	1.56	0.66	1.02	0.59	6.17
1.75	1.58	0.63	1.09	0.59	7.54
1.72	1.43	0.64	1.02	0.63	6.36
1.68	1.57	0.72	0.96	0.68	7.63
1.75	1.59	0.68	1.08	0.62	7.78
2.19	1.86	0.75	1.24	0.72	10.15
1.73	1.67	0.64	1.14	0.55	6.88

In the study, the regressor variables and response considered are

$x_1 =$ rostral length, in inches,
$x_2 =$ wing length, in inches,
$x_3 =$ rostral to notch length, in inches,
$x_4 =$ notch to wing length, in inches,
$x_5 =$ width, in inches,
$y =$ weight, in pounds.

Estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2,x_3,x_4,x_5} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5.$$

12.12 The following data reflect information from 17 U.S. Naval hospitals at various sites around the world. The regressors are workload variables, that is, items that result in the need for personnel in a hospital. A brief description of the variables is as follows:

$y =$ monthly labor-hours,
$x_1 =$ average daily patient load,
$x_2 =$ monthly X-ray exposures,
$x_3 =$ monthly occupied bed-days,
$x_4 =$ eligible population in the area/1000,
$x_5 =$ average length of patient's stay, in days.

Site	x_1	x_2	x_3	x_4	x_5	y
1	15.57	2463	472.92	18.0	4.45	566.52
2	44.02	2048	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	6505	568.33	36.7	3.90	1003.62
5	49.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11,520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.15	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20,106	3655.08	180.5	6.15	3503.93
11	96.00	13,313	2912.00	60.9	5.88	3571.59
12	131.42	10,771	3921.00	103.7	4.88	3741.40
13	127.21	15,543	3865.67	126.8	5.50	4026.52
14	252.90	36,194	7684.10	157.7	7.00	10,343.81
15	409.20	34,703	12,446.33	169.4	10.75	11,732.17
16	463.70	39,204	14,098.40	331.4	7.05	15,414.94
17	510.22	86,533	15,524.00	371.6	6.35	18,854.45

The goal here is to produce an empirical equation that will estimate (or predict) personnel needs for Naval hospitals. Estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2,x_3,x_4,x_5} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5.$$

12.13 A study was performed on a type of bearing to find the relationship of amount of wear y to $x_1 =$ oil viscosity and $x_2 =$ load. The following data were obtained. (From *Response Surface Methodology*, Myers, Montgomery, and Anderson-Cook, 2009.)

y	x_1	x_2
193	1.6	851
172	22.0	1058
113	33.0	1357
230	15.5	816
91	43.0	1201
125	40.0	1115

(a) Estimate the unknown parameters of the multiple linear regression equation

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

(b) Predict wear when oil viscosity is 20 and load is 1200.

12.14 Eleven student teachers took part in an evaluation program designed to measure teacher effectiveness and determine what factors are important. The response measure was a quantitative evaluation of the teacher. The regressor variables were scores on four standardized tests given to each teacher. The data are as follows:

y	x_1	x_2	x_3	x_4
410	69	125	59.00	55.66
569	57	131	31.75	63.97
425	77	141	80.50	45.32
344	81	122	75.00	46.67
324	0	141	49.00	41.21
505	53	152	49.35	43.83
235	77	141	60.75	41.61
501	76	132	41.25	64.57
400	65	157	50.75	42.41
584	97	166	32.25	57.95
434	76	141	54.50	57.90

Estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2,x_3,x_4} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4.$$

12.15 The personnel department of a certain industrial firm used 12 subjects in a study to determine the relationship between job performance rating (y) and scores on four tests. The data are as follows:

y	x_1	x_2	x_3	x_4
11.2	56.5	71.0	38.5	43.0
14.5	59.5	72.5	38.2	44.8
17.2	69.2	76.0	42.5	49.0
17.8	74.5	79.5	43.4	56.3
19.3	81.2	84.0	47.5	60.2
24.5	88.0	86.2	47.4	62.0
21.2	78.2	80.5	44.5	58.1
16.9	69.0	72.0	41.8	48.1
14.8	58.1	68.0	42.1	46.0
20.0	80.5	85.0	48.1	60.3
13.2	58.3	71.0	37.5	47.1
22.5	84.0	87.2	51.0	65.2

Estimate the regression coefficients in the model

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4.$$

12.16 An engineer at a semiconductor company wants to model the relationship between the gain or hFE of a device (y) and three parameters: emitter-RS (x_1), base-RS (x_2), and emitter-to-base-RS (x_3). The data are shown below:

x_1, **Emitter-RS**	x_2, **Base-RS**	x_3, **E-B-RS**	y, **hFE**
14.62	226.0	7.000	128.40
15.63	220.0	3.375	52.62
14.62	217.4	6.375	113.90
15.00	220.0	6.000	98.01
14.50	226.5	7.625	139.90
15.25	224.1	6.000	102.60
16.12	220.5	3.375	48.14
15.13	223.5	6.125	109.60
15.50	217.6	5.000	82.68
15.13	228.5	6.625	112.60
15.50	230.2	5.750	97.52
16.12	226.5	3.750	59.06
15.13	226.6	6.125	111.80
15.63	225.6	5.375	89.09
15.38	234.0	8.875	171.90
15.50	230.0	4.000	66.80
14.25	224.3	8.000	157.10
14.50	240.5	10.870	208.40
14.62	223.7	7.375	133.40

(cont.)

(Data from Myers, Montgomery, and Anderson-Cook, 2009.)

(a) Fit a multiple linear regression to the data.

(b) Predict hFE when $x_1 = 14$, $x_2 = 220$, and $x_3 = 5$.

12.4 Properties of the Least Squares Estimators

The means and variances of the estimators $b_0, b_1, \ldots, b_k$ are readily obtained under certain assumptions on the random errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_k$ that are identical to those made in the case of simple linear regression. When we assume these errors to be independent, each with mean 0 and variance σ^2, it can be shown that $b_0, b_1, \ldots, b_k$ are, respectively, unbiased estimators of the regression coefficients $\beta_0, \beta_1, \ldots, \beta_k$. In addition, the variances of the b's are obtained through the elements of the inverse of the $\mathbf{A}$ matrix. Note that the off-diagonal elements of $\mathbf{A} = \mathbf{X}'\mathbf{X}$ represent sums of products of elements in the columns of $\mathbf{X}$, while the diagonal elements of $\mathbf{A}$ represent sums of squares of elements in the columns of $\mathbf{X}$. The inverse matrix, $\mathbf{A}^{-1}$, apart from the multiplier σ^2, represents the **variance-covariance matrix** of the estimated regression coefficients. That is, the elements of the matrix $\mathbf{A}^{-1}\sigma^2$ display the variances of $b_0, b_1, \ldots, b_k$ on the main diagonal and covariances on the off-diagonal. For example, in a $k = 2$ multiple linear regression problem, we might write

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix}$$

with the elements below the main diagonal determined through the symmetry of the matrix. Then we can write

$$\begin{aligned} \sigma_{b_i}^2 &= c_{ii}\sigma^2, \qquad\qquad i = 0, 1, 2, \\ \sigma_{b_i b_j} &= \text{Cov}(b_i, b_j) = c_{ij}\sigma^2, \;\; i \neq j. \end{aligned}$$

Of course, the estimates of the variances and hence the standard errors of these estimators are obtained by replacing σ^2 with the appropriate estimate obtained through experimental data. An unbiased estimate of σ^2 is once again defined in terms of the error sum of squares, which is computed using the formula established in Theorem 12.1. In the theorem, we are making the assumptions on the ϵ_i described above.

Theorem 12.1: For the linear regression equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

an unbiased estimate of σ^2 is given by the error or residual mean square

$$s^2 = \frac{SSE}{n-k-1}, \quad \text{where} \quad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

We can see that Theorem 12.1 represents a generalization of Theorem 11.1 for the simple linear regression case. The proof is left for the reader. As in the simpler linear regression case, the estimate s^2 is a measure of the variation in the prediction errors or residuals. Other important inferences regarding the fitted regression equation, based on the values of the individual residuals $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$, are discussed in Sections 12.10 and 12.11.

The error and regression sums of squares take on the same form and play the same role as in the simple linear regression case. In fact, the sum-of-squares identity

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

continues to hold, and we retain our previous notation, namely

$$SST = SSR + SSE,$$

with

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{total sum of squares}$$

and

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \text{regression sum of squares}.$$

There are k degrees of freedom associated with SSR, and, as always, SST has $n - 1$ degrees of freedom. Therefore, after subtraction, SSE has $n - k - 1$ degrees of freedom. Thus, our estimate of σ^2 is again given by the error sum of squares divided by its degrees of freedom. All three of these sums of squares will appear on the printouts of most multiple regression computer packages. Note that the condition $n > k$ in Section 12.2 guarantees that the degrees of freedom of SSE cannot be negative.

Analysis of Variance in Multiple Regression

The partition of the total sum of squares into its components, the regression and error sums of squares, plays an important role. An **analysis of variance** can be conducted to shed light on the quality of the regression equation. A useful hypothesis that determines if a significant amount of variation is explained by the model is

$$H_0\colon\ \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0.$$

The analysis of variance involves an F-test via a table given as follows:

Source	**Sum of Squares**	**Degrees of Freedom**	**Mean Squares**	$\boldsymbol{F}$
Regression	SSR	k	$MSR = \frac{SSR}{k}$	$f = \frac{MSR}{MSE}$
Error	SSE	$n-(k+1)$	$MSE = \frac{SSE}{n-(k+1)}$	
Total	SST	$n-1$		

This test is an **upper-tailed test**. Rejection of H_0 implies that the **regression equation differs from a constant**. That is, at least one regressor variable is important. More discussion of the use of analysis of variance appears in subsequent sections.

Further utility of the mean square error (or residual mean square) lies in its use in hypothesis testing and confidence interval estimation, which is discussed in Section 12.5. In addition, the mean square error plays an important role in situations where the scientist is searching for the best from a set of competing models. Many model-building criteria involve the statistic s^2. Criteria for comparing competing models are discussed in Section 12.11.

12.5 Inferences in Multiple Linear Regression

A knowledge of the distributions of the individual coefficient estimators enables the experimenter to construct confidence intervals for the coefficients and to test hypotheses about them. Recall from Section 12.4 that the b_j $(j = 0, 1, 2, \ldots, k)$ are normally distributed with mean β_j and variance $c_{jj}\sigma^2$. Thus, we can use the statistic

$$t = \frac{b_j - \beta_{j0}}{s\sqrt{c_{jj}}}$$

with $n - k - 1$ degrees of freedom to test hypotheses and construct confidence intervals on β_j. For example, if we wish to test

$$H_0\colon\ \beta_j = \beta_{j0},$$
$$H_1\colon\ \beta_j \neq \beta_{j0},$$

we compute the above t-statistic and do not reject H_0 if $-t_{\alpha/2} < t < t_{\alpha/2}$, where $t_{\alpha/2}$ has $n - k - 1$ degrees of freedom.

Example 12.5: For the model of Example 12.4, test the hypothesis that $\beta_2 = -2.5$ at the 0.05 level of significance against the alternative that $\beta_2 > -2.5$.

Solution:

$$H_0\colon\ \beta_2 = -2.5,$$
$$H_1\colon\ \beta_2 > -2.5.$$

Computations:

$$t = \frac{b_2 - \beta_{20}}{s\sqrt{c_{22}}} = \frac{-1.8616 + 2.5}{2.073\sqrt{0.0166}} = 2.390,$$
$$P = P(T > 2.390) = 0.04.$$

Decision: Reject H_0 and conclude that $\beta_2 > -2.5$.

Individual *t*-Tests for Variable Screening

The t-test most often used in multiple regression is the one that tests the importance of individual coefficients (i.e., $H_0\colon \beta_j = 0$ against the alternative $H_1\colon \beta_j \neq 0$). These tests often contribute to what is termed **variable screening**, where the analyst attempts to arrive at the most useful model (i.e., the choice of which regressors to use). It should be emphasized here that if a coefficient is found insignificant (i.e., the hypothesis $H_0\colon \beta_j = 0$ **is not rejected**), the conclusion drawn is that the **variable** is insignificant (i.e., explains an insignificant amount of variation in y), **in the presence of the other regressors in the model**. This point will be reaffirmed in a future discussion.

Inferences on Mean Response and Prediction

One of the most useful inferences that can be made regarding the quality of the predicted response y_0 corresponding to the values $x_{10}, x_{20}, \ldots, x_{k0}$ is the confidence interval on the mean response $\mu_{Y|x_{10},x_{20},\ldots,x_{k0}}$. We are interested in constructing a confidence interval on the mean response for the set of conditions given by

$$\mathbf{x}_0' = [1, x_{10}, x_{20}, \ldots, x_{k0}].$$

We augment the conditions on the x's by the number 1 in order to facilitate the matrix notation. Normality in the ϵ_i produces normality in the b_j and the mean and variance are still the same as indicated in Section 12.4. So is the covariance between b_i and b_j, for $i \neq j$. Hence,

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_{j0}$$

is likewise normally distributed and is, in fact, an unbiased estimator for the **mean response** on which we are attempting to attach a confidence interval. The variance of $\hat{y}_0$, written in matrix notation simply as a function of σ^2, $(\mathbf{X}'\mathbf{X})^{-1}$, and the condition vector $\mathbf{x}_0'$, is

$$\sigma^2_{\hat{y}_0} = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0.$$

If this expression is expanded for a given case, say $k = 2$, it is readily seen that it appropriately accounts for the variance of the b_j and the covariance of b_i and b_j, for $i \neq j$. After σ^2 is replaced by s^2 as given by Theorem 12.1, the $100(1-\alpha)\%$ confidence interval on $\mu_{Y|x_{10},x_{20},\ldots,x_{k0}}$ can be constructed from the statistic

$$T = \frac{\hat{y}_0 - \mu_{Y|x_{10},x_{20},\ldots,x_{k0}}}{s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}},$$

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

Confidence Interval for $\mu_{Y|x_{10},x_{20},\ldots,x_{k0}}$

A $100(1-\alpha)\%$ confidence interval for the **mean response** $\mu_{Y|x_{10},x_{20},\ldots,x_{k0}}$ is

$$\hat{y}_0 - t_{\alpha/2}s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} < \mu_{Y|x_{10},x_{20},\ldots,x_{k0}} < \hat{y}_0 + t_{\alpha/2}s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n-k-1$ degrees of freedom.

The quantity $s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$ is often called the **standard error of prediction** and appears on the printout of many regression computer packages.

Example 12.6: Using the data of Example 12.4, construct a 95% confidence interval for the mean response when $x_1 = 3\%$, $x_2 = 8\%$, and $x_3 = 9\%$.

***Solution*:** From the regression equation of Example 12.4, the estimated percent survival when $x_1 = 3\%$, $x_2 = 8\%$, and $x_3 = 9\%$ is

$$\hat{y} = 39.1574 + (1.0161)(3) - (1.8616)(8) - (0.3433)(9) = 24.2232.$$

Next, we find that

$$\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = [1, 3, 8, 9]\begin{bmatrix} 8.0648 & -0.0826 & -0.0942 & -0.7905 \\ -0.0826 & 0.0085 & 0.0017 & 0.0037 \\ -0.0942 & 0.0017 & 0.0166 & -0.0021 \\ -0.7905 & 0.0037 & -0.0021 & 0.0886 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 8 \\ 9 \end{bmatrix}$$
$$= 0.1267.$$

Using the mean square error, $s^2 = 4.298$ or $s = 2.073$, and Table A.4, we see that $t_{0.025} = 2.262$ for 9 degrees of freedom. Therefore, a 95% confidence interval for the mean percent survival for $x_1 = 3\%$, $x_2 = 8\%$, and $x_3 = 9\%$ is given by

$$24.2232 - (2.262)(2.073)\sqrt{0.1267} < \mu_{Y|3,8,9}$$
$$< 24.2232 + (2.262)(2.073)\sqrt{0.1267},$$

or simply $22.5541 < \mu_{Y|3,8,9} < 25.8923$. ▟

As in the case of simple linear regression, we need to make a clear distinction between the confidence interval on a mean response and the prediction interval on an *observed response*. The latter provides a bound within which we can say with a preselected degree of certainty that a new observed response will fall.

A prediction interval for a single predicted response y_0 is once again established by considering the difference $\hat{y}_0 - y_0$. The sampling distribution can be shown to be normal with mean

$$\mu_{\hat{y}_0 - y_0} = 0$$

and variance

$$\sigma^2_{\hat{y}_0 - y_0} = \sigma^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0].$$

Thus, a $100(1-\alpha)\%$ prediction interval for a single prediction value y_0 can be constructed from the statistic

$$T = \frac{\hat{y}_0 - y_0}{s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}},$$

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

Prediction Interval for y_0

A $100(1-\alpha)\%$ prediction interval for a **single response** y_0 is given by

$$\hat{y}_0 - t_{\alpha/2}s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} < y_0 < \hat{y}_0 + t_{\alpha/2}s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0},$$

where $t_{\alpha/2}$ is a value of the t-distribution with $n-k-1$ degrees of freedom.

Example 12.7: Using the data of Example 12.4, construct a 95% prediction interval for an individual percent survival response when $x_1 = 3\%$, $x_2 = 8\%$, and $x_3 = 9\%$.

Solution: Referring to the results of Example 12.6, we find that the 95% prediction interval for the response y_0, when $x_1 = 3\%$, $x_2 = 8\%$, and $x_3 = 9\%$, is

$$24.2232 - (2.262)(2.073)\sqrt{1.1267} < y_0 < 24.2232 + (2.262)(2.073)\sqrt{1.1267},$$

which reduces to $19.2459 < y_0 < 29.2005$. Notice, as expected, that the prediction interval is considerably wider than the confidence interval for mean percent survival found in Example 12.6.

Annotated Printout for Data of Example 12.4

Figure 12.1 shows an annotated computer printout for a multiple linear regression fit to the data of Example 12.4. The package used is *SAS*.

Note the model parameter estimates, the standard errors, and the t-statistics shown in the output. The standard errors are computed from square roots of diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}s^2$. In this illustration, the variable x_3 is insignificant in the presence of x_1 and x_2 based on the t-test and the corresponding P-value of 0.5916. The terms CLM and CLI are confidence intervals on mean response and prediction limits on an individual observation, respectively. The f-test in the analysis of variance indicates that a significant amount of variability is explained. As an example of the interpretation of CLM and CLI, consider observation 10. With an observation of 25.2000 and a predicted value of 26.0676, we are 95% confident that the mean response is between 24.5024 and 27.6329, and a new observation will fall between 21.1238 and 31.0114 with probability 0.95. The R^2 value of 0.9117 implies that the model explains 91.17% of the variability in the response. More discussion about R^2 appears in Section 12.6.

```
                         Sum of        Mean
Source           DF     Squares      Square  F Value  Pr > F
Model             3   399.45437   133.15146    30.98  <.0001
Error             9    38.67640     4.29738
Corrected Total  12   438.13077

Root MSE         2.07301    R-Square     0.9117
Dependent Mean  29.03846    Adj R-Sq     0.8823
Coeff Var        7.13885

                         Parameter       Standard
Variable     DF           Estimate          Error    t Value    Pr > |t|
Intercept     1           39.15735        5.88706       6.65      <.0001
x1            1            1.01610        0.19090       5.32      0.0005
x2            1           -1.86165        0.26733      -6.96      <.0001
x3            1           -0.34326        0.61705      -0.56      0.5916

    Dependent Predicted    Std Error
Obs  Variable       Value Mean Predict     95% CL Mean        95% CL Predict   Residual
  1   25.5000   27.3514        1.4152 24.1500   30.5528   21.6734   33.0294  -1.8514
  2   31.2000   32.2623        0.7846 30.4875   34.0371   27.2482   37.2764  -1.0623
  3   25.9000   27.3495        1.3588 24.2757   30.4234   21.7425   32.9566  -1.4495
  4   38.4000   38.3096        1.2818 35.4099   41.2093   32.7960   43.8232   0.0904
  5   18.4000   15.5447        1.5789 11.9730   19.1165    9.6499   21.4395   2.8553
  6   26.7000   26.1081        1.0358 23.7649   28.4512   20.8658   31.3503   0.5919
  7   26.4000   28.2532        0.8094 26.4222   30.0841   23.2189   33.2874  -1.8532
  8   25.9000   26.2219        0.9732 24.0204   28.4233   21.0414   31.4023  -0.3219
  9   32.0000   32.0882        0.7828 30.3175   33.8589   27.0755   37.1008  -0.0882
 10   25.2000   26.0676        0.6919 24.5024   27.6329   21.1238   31.0114  -0.8676
 11   39.7000   37.2524        1.3070 34.2957   40.2090   31.7086   42.7961   2.4476
 12   35.7000   32.4879        1.4648 29.1743   35.8015   26.7459   38.2300   3.2121
 13   26.5000   28.2032        0.9841 25.9771   30.4294   23.0122   33.3943  -1.7032
```

Figure 12.1: *SAS* printout for data in Example 12.4.

More on Analysis of Variance in Multiple Regression (Optional)

In Section 12.4, we discussed briefly the partition of the total sum of squares $\sum_{i=1}^{n}(y_i-\bar{y})^2$ into its two components, the regression model and error sums of squares (illustrated in Figure 12.1). The analysis of variance leads to a test of

$$H_0\colon\ \beta_1=\beta_2=\beta_3=\cdots=\beta_k=0.$$

Rejection of the null hypothesis has an important interpretation for the scientist or engineer. (For those who are interested in more extensive treatment of the subject using matrices, it is useful to discuss the development of these sums of squares used in ANOVA.)

First, recall in Section 12.3, **b**, the vector of least squares estimators, is given by

$$\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

A partition of the **uncorrected sum of squares**

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{n} \mathbf{y}_i^2$$

into two components is given by

$$\begin{aligned}\mathbf{y}'\mathbf{y} &= \mathbf{b}'\mathbf{X}'\mathbf{y} + (\mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y})\\ &= \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + [\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}].\end{aligned}$$

The second term (in brackets) on the right-hand side is simply the error sum of squares $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. The reader should see that an alternative expression for the error sum of squares is

$$SSE = \mathbf{y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}.$$

The term $\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is called the **regression sum of squares**. However, it is not the expression $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ used for testing the "importance" of the terms $b_1, b_2, \ldots, b_k$ but, rather,

$$\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \sum_{i=1}^{n} \hat{y}_i^2,$$

which is a regression sum of squares uncorrected for the mean. As such, it would only be used in testing if the regression equation differs significantly from zero, that is,

$$H_0\text{:}\ \ \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k = 0.$$

In general, this is not as important as testing

$$H_0\text{:}\ \ \beta_1 = \beta_2 = \cdots = \beta_k = 0,$$

since the latter states that the mean response is a constant, not necessarily zero.

Degrees of Freedom

Thus, the partition of sums of squares and degrees of freedom reduces to

Source	**Sum of Squares**	**d.f.**
Regression	$\sum_{i=1}^{n} \hat{y}_i^2 = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$	$k+1$
Error	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \mathbf{y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$	$n-(k+1)$
Total	$\sum_{i=1}^{n} y_i^2 = \mathbf{y}'\mathbf{y}$	n

Hypothesis of Interest

Now, of course, the hypotheses of interest for an ANOVA must eliminate the role of the intercept described previously. Strictly speaking, if $H_0\colon\ \beta_1 = \beta_2 = \cdots = \beta_k = 0$, then the estimated regression line is merely $\hat{y}_i = \bar{y}$. As a result, we are actually seeking evidence that the regression equation "varies from a constant." Thus, the total and regression sums of squares must be corrected for the mean. As a result, we have

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

In matrix notation this is simply

$$\mathbf{y}'[\mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']\mathbf{y} = \mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']\mathbf{y} + \mathbf{y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}.$$

In this expression, $\mathbf{1}$ is a vector of n ones. As a result, we are merely subtracting

$$\mathbf{y}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

from $\mathbf{y}'\mathbf{y}$ and from $\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (i.e., correcting the total and regression sums of squares for the mean).

Finally, the appropriate partitioning of sums of squares with degrees of freedom is as follows:

Source	**Sum of Squares**	**d.f.**
Regression	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}]\mathbf{y}$	k
Error	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \mathbf{y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$	$n - (k + 1)$
Total	$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathbf{y}'[\mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']\mathbf{y}$	$n - 1$

This is the ANOVA table that appears in the computer printout of Figure 12.1. The expression $\mathbf{y}'[\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']\mathbf{y}$ is often called the **regression sum of squares associated with the mean**, and 1 degree of freedom is allocated to it.

Exercises

12.17 For the data of Exercise 12.2 on page 450, estimate σ^2.

12.18 For the data of Exercise 12.1 on page 450, estimate σ^2.

12.19 For the data of Exercise 12.5 on page 450, estimate σ^2.

12.20 Obtain estimates of the variances and the covariance of the estimators b_1 and b_2 of Exercise 12.2 on page 450.

12.21 Referring to Exercise 12.5 on page 450, find the estimate of

(a) $\sigma^2_{b_2}$;

(b) $\mathrm{Cov}(b_1, b_4)$.

12.22 For the model of Exercise 12.7 on page 451, test the hypothesis that $\beta_2 = 0$ at the 0.05 level of significance against the alternative that $\beta_2 \neq 0$.

12.23 For the model of Exercise 12.2 on page 450, test the hypothesis that $\beta_1 = 0$ at the 0.05 level of significance against the alternative that $\beta_1 \neq 0$.

12.24 For the model of Exercise 12.1 on page 450, test the hypotheses that $\beta_1 = 2$ against the alternative that $\beta_1 \neq 2$. Use a P-value in your conclusion.

12.25 Using the data of Exercise 12.2 on page 450 and the estimate of σ^2 from Exercise 12.17, compute 95% confidence intervals for the predicted response and the mean response when $x_1 = 900$ and $x_2 = 1.00$.

12.26 For Exercise 12.8 on page 451, construct a 90% confidence interval for the mean compressive strength when the concentration is $x = 19.5$ and a quadratic model is used.

12.27 Using the data of Exercise 12.5 on page 450 and the estimate of σ^2 from Exercise 12.19, compute 95% confidence intervals for the predicted response and the mean response when $x_1 = 75$, $x_2 = 24$, $x_3 = 90$, and $x_4 = 98$.

12.28 Consider the following data from Exercise 12.13 on page 452.

y (wear)	**x_1 (oil viscosity)**	**x_2 (load)**
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

(a) Estimate σ^2 using multiple regression of y on x_1 and x_2.

(b) Compute predicted values, a 95% confidence interval for mean wear, and a 95% prediction interval for observed wear if $x_1 = 20$ and $x_2 = 1000$.

12.29 Using the data from Exercise 12.28, test the following at the 0.05 level.

(a) H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$;

(b) H_0: $\beta_2 = 0$ versus H_1: $\beta_2 \neq 0$.

(c) Do you have any reason to believe that the model in Exercise 12.28 should be changed? Why or why not?

12.30 Use the data from Exercise 12.16 on page 453.

(a) Estimate σ^2 using the multiple regression of y on x_1, x_2, and x_3,

(b) Compute a 95% prediction interval for the observed gain with the three regressors at $x_1 = 15.0$, $x_2 = 220.0$, and $x_3 = 6.0$.

12.6 Choice of a Fitted Model through Hypothesis Testing

In many regression situations, individual coefficients are of importance to the experimenter. For example, in an economics application, $\beta_1, \beta_2, \ldots$ might have some particular significance, and thus confidence intervals and tests of hypotheses on these parameters would be of interest to the economist. However, consider an industrial chemical situation in which the postulated model assumes that reaction yield is linearly dependent on reaction temperature and concentration of a certain catalyst. It is probably known that this is not the true model but an adequate approximation, so interest is likely to be not in the individual parameters but rather in the ability of the entire function to predict the true response in the range of the variables considered. Therefore, in this situation, one would put more emphasis on $\sigma^2_{\hat{Y}}$, confidence intervals on the mean response, and so forth, and likely deemphasize inferences on individual parameters.

The experimenter using regression analysis is also interested in deletion of variables when the situation dictates that, in addition to arriving at a workable prediction equation, he or she must find the "best regression" involving only variables that are useful predictors. There are a number of computer programs that sequentially arrive at the so-called best regression equation depending on certain criteria. We discuss this further in Section 12.9.

One criterion that is commonly used to illustrate the adequacy of a fitted regression model is the **coefficient of determination**, or $\boldsymbol{R^2}$.

Coefficient of Determination, or R^2

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}.$$

Note that this parallels the description of R^2 in Chapter 11. At this point the explanation might be clearer since we now focus on SSR as the **variability explained**. The quantity R^2 merely indicates what proportion of the total variation in the response Y is explained by the fitted model. Often an experimenter will report $R^2 \times 100\%$ and interpret the result as percent variation explained by the postulated model. The square root of R^2 is called the **multiple correlation coefficient** between Y and the set $x_1, x_2, \ldots, x_k$. The value of R^2 for the case in Example 12.4, indicating the proportion of variation explained by the three independent variables x_1, x_2, and x_3, is

$$R^2 = \frac{SSR}{SST} = \frac{399.45}{438.13} = 0.9117,$$

which means that 91.17% of the variation in percent survival has been explained by the linear regression model.

The regression sum of squares can be used to give some indication concerning whether or not the model is an adequate explanation of the true situation. We can test the hypothesis H_0 that the **regression is not significant** by merely forming the ratio

$$f = \frac{SSR/k}{SSE/(n-k-1)} = \frac{SSR/k}{s^2}$$

and rejecting H_0 at the α-level of significance when $f > f_\alpha(k, n-k-1)$. For the data of Example 12.4, we obtain

$$f = \frac{399.45/3}{4.298} = 30.98.$$

From the printout in Figure 12.1, the P-value is less than 0.0001. This should not be misinterpreted. Although it does indicate that the regression explained by the model is significant, this does not rule out the following possibilities:

1. The linear regression model for this set of x's is not the only model that can be used to explain the data; indeed, there may be other models with transformations on the x's that give a larger value of the F-statistic.
2. The model might have been more effective with the inclusion of other variables in addition to x_1, x_2, and x_3 or perhaps with the deletion of one or more of the variables in the model, say x_3, which has a $P = 0.5916$.

The reader should recall the discussion in Section 11.5 regarding the pitfalls in the use of R^2 as a criterion for comparing competing models. These pitfalls are certainly relevant in multiple linear regression. In fact, in its employment in multiple regression, the dangers are even more pronounced since the temptation to overfit is so great. One should always keep in mind that $R^2 \approx 1.0$ can always be achieved at the expense of error degrees of freedom when an excess of model terms is employed. However, $R^2 = 1$, describing a model with a near perfect fit, does not always result in a model that predicts well.

The Adjusted Coefficient of Determination (R^2_{adj})

In Chapter 11, several figures displaying computer printout from both *SAS* and *MINITAB* featured a statistic called *adjusted* R^2 or adjusted coefficient of determination. Adjusted R^2 is a variation on R^2 that provides an **adjustment for degrees of freedom**. The coefficient of determination as defined on page 407 cannot decrease as terms are added to the model. In other words, R^2 does not decrease as the error degrees of freedom $n - k - 1$ are reduced, the latter result being produced by an increase in k, the number of model terms. Adjusted R^2 is computed by dividing SSE and SST by their respective degrees of freedom as follows.

Adjusted R^2

$$R^2_{\text{adj}} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}.$$

To illustrate the use of R^2_{adj}, Example 12.4 will be revisited.

How Are R^2 and R^2_{adj} Affected by Removal of x_3?

The t-test (or corresponding F-test) for x_3 suggests that a simpler model involving only x_1 and x_2 may well be an improvement. In other words, the complete model with all the regressors may be an overfitted model. It is certainly of interest to investigate R^2 and R^2_{adj} for both the full (x_1, x_2, x_3) and the reduced (x_1, x_2) models. We already know that $R^2_{\text{full}} = 0.9117$ from Figure 12.1. The SSE for the reduced model is 40.01, and thus $R^2_{\text{reduced}} = 1 - \frac{40.01}{438.13} = 0.9087$. Thus, more variability is explained with x_3 in the model. However, as we have indicated, this will occur even if the model is an overfitted model. Now, of course, R^2_{adj} is designed to provide a statistic that punishes an overfitted model, so we might expect it to favor the reduced model. Indeed, for the full model

$$R^2_{\text{adj}} = 1 - \frac{38.6764/9}{438.1308/12} = 1 - \frac{4.2974}{36.5109} = 0.8823,$$

whereas for the reduced model (deletion of x_3)

$$R^2_{\text{adj}} = 1 - \frac{40.01/10}{438.1308/12} = 1 - \frac{4.001}{36.5109} = 0.8904.$$

Thus, R^2_{adj} does indeed favor the reduced model and confirms the evidence produced by the t- and F-tests, suggesting that the reduced model is preferable to the model containing all three regressors. The reader may expect that other statistics would suggest rejection of the overfitted model. See Exercise 12.40 on page 471.

Test on an Individual Coefficient

The addition of any single variable to a regression system *will increase the regression sum of squares* and thus *reduce the error sum of squares.* Consequently, we must decide whether the increase in regression is sufficient to warrant using the variable in the model. As we might expect, the use of unimportant variables can reduce the effectiveness of the prediction equation by increasing the variance of the estimated response. We shall pursue this point further by considering the importance of x_3 in Example 12.4. Initially, we can test

$$H_0\colon\ \beta_3 = 0,$$
$$H_1\colon\ \beta_3 \neq 0$$

by using the t-distribution with 9 degrees of freedom. We have

$$t = \frac{b_3 - 0}{s\sqrt{c_{33}}} = \frac{-0.3433}{2.073\sqrt{0.0886}} = -0.556,$$

which indicates that β_3 does not differ significantly from zero, and hence we may very well feel justified in removing x_3 from the model. Suppose that we consider the regression of Y on the set (x_1, x_2), the least squares normal equations now reducing to

$$\begin{bmatrix} 13.0 & 59.43 & 81.82 \\ 59.43 & 394.7255 & 360.6621 \\ 81.82 & 360.6621 & 576.7264 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 377.50 \\ 1877.5670 \\ 2246.6610 \end{bmatrix}.$$

The estimated regression coefficients for this reduced model are

$$b_0 = 36.094, \quad b_1 = 1.031, \quad b_2 = -1.870,$$

and the resulting regression sum of squares with 2 degrees of freedom is

$$R(\beta_1, \beta_2) = 398.12.$$

Here we use the notation $R(\beta_1, \beta_2)$ to indicate the regression sum of squares of the restricted model; it should not be confused with SSR, the regression sum of squares of the original model with 3 degrees of freedom. The new error sum of squares is then

$$SST - R(\beta_1, \beta_2) = 438.13 - 398.12 = 40.01,$$

and the resulting mean square error with 10 degrees of freedom becomes

$$s^2 = \frac{40.01}{10} = 4.001.$$

Does a Single Variable *t*-Test Have an *F* Counterpart?

From Example 12.4, the amount of variation in the percent survival that is attributed to x_3, in the presence of the variables x_1 and x_2, is

$$R(\beta_3 \mid \beta_1, \beta_2) = SSR - R(\beta_1, \beta_2) = 399.45 - 398.12 = 1.33,$$

which represents a small proportion of the entire regression variation. This amount of added regression is statistically insignificant, as indicated by our previous test on β_3. An equivalent test involves the formation of the ratio

$$f = \frac{R(\beta_3 \mid \beta_1, \beta_2)}{s^2} = \frac{1.33}{4.298} = 0.309,$$

which is a value of the F-distribution with 1 and 9 degrees of freedom. Recall that the basic relationship between the t-distribution with v degrees of freedom and the F-distribution with 1 and v degrees of freedom is

$$t^2 = f(1, v),$$

and note that the f-value of 0.309 is indeed the square of the t-value of -0.56.

To generalize the concepts above, we can assess the work of an independent variable x_i in the general multiple linear regression model

$$\mu_{Y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

by observing the amount of regression attributed to x_i **over and above that attributed to the other variables**, that is, the regression on x_i *adjusted for the other variables.* For example, we say that x_1 is assessed by calculating

$$R(\beta_1 \mid \beta_2, \beta_3, \ldots, \beta_k) = SSR - R(\beta_2, \beta_3, \ldots, \beta_k),$$

where $R(\beta_2, \beta_3, \ldots, \beta_k)$ is the regression sum of squares with $\beta_1 x_1$ removed from the model. To test the hypothesis

$$H_0\colon\ \beta_1 = 0,$$
$$H_1\colon\ \beta_1 \neq 0,$$

we compute

$$f = \frac{R(\beta_1 \mid \beta_2, \beta_3, \ldots, \beta_k)}{s^2},$$

and compare it with $f_\alpha(1, n-k-1)$.

Partial F-Tests on Subsets of Coefficients

In a similar manner, we can test for the significance of a *set* of the variables. For example, to investigate simultaneously the importance of including x_1 and x_2 in the model, we test the hypothesis

$$H_0\colon\ \beta_1 = \beta_2 = 0,$$
$$H_1\colon\ \beta_1 \text{ and } \beta_2 \text{ are not both zero,}$$

by computing

$$f = \frac{[R(\beta_1, \beta_2 \mid \beta_3, \beta_4, \ldots, \beta_k)]/2}{s^2} = \frac{[SSR - R(\beta_3, \beta_4, \ldots, \beta_k)]/2}{s^2}$$

and comparing it with $f_\alpha(2, n-k-1)$. The number of degrees of freedom associated with the numerator, in this case 2, equals the number of variables in the set being investigated.

Suppose we wish to test the hypothesis

$$H_0\colon \beta_2 = \beta_3 = 0,$$
$$H_1\colon \beta_2 \text{ and } \beta_3 \text{ are not both zero}$$

for Example 12.4. If we develop the regression model

$$y = \beta_0 + \beta_1 x_1 + \epsilon,$$

we can obtain $R(\beta_1) = SSR_{\text{reduced}} = 187.31179$. From Figure 12.1 on page 459, we have $s^2 = 4.29738$ for the full model. Hence, the f-value for testing the hypothesis is

$$\begin{aligned} f &= \frac{R(\beta_2, \beta_3 \mid \beta_1)/2}{s^2} = \frac{[R(\beta_1, \beta_2, \beta_3) - R(\beta_1)]/2}{s^2} = \frac{[SSR_{\text{full}} - SSR_{\text{reduced}}]/2}{s^2} \\ &= \frac{(399.45437 - 187.31179)/2}{4.29738} = 24.68278. \end{aligned}$$

This implies that β_2 and β_3 are not simultaneously zero. Using statistical software such as *SAS* one can directly obtain the above result with a P-value of 0.0002. Readers should note that in statistical software package output there are P-values associated with each individual model coefficient. The null hypothesis for each is that the coefficient is zero. However, it should be noted that the insignificance of any coefficient does not necessarily imply that it does not belong in the final model. It merely suggests that it is insignificant in the presence of all other variables in the problem. The case study at the end of this chapter illustrates this further.

12.7 Special Case of Orthogonality (Optional)

Prior to our original development of the general linear regression problem, the assumption was made that the independent variables are measured without error and are often controlled by the experimenter. Quite often they occur as a result of an *elaborately designed experiment*. In fact, we can increase the effectiveness of the resulting prediction equation with the use of a suitable experimental plan.

Suppose that we once again consider the $\mathbf{X}$ matrix as defined in Section 12.3. We can rewrite it as

$$\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k],$$

where $\mathbf{1}$ represents a column of ones and $\mathbf{x}_j$ is a column vector representing the levels of x_j. If

$$\mathbf{x}'_p \mathbf{x}_q = \mathbf{0}, \quad \text{for } p \neq q,$$

the variables x_p and x_q are said to be *orthogonal* to each other. There are certain obvious advantages to having a completely orthogonal situation where $\mathbf{x}'_p \mathbf{x}_q = \mathbf{0}$

for all possible p and q, $p \neq q$, and, in addition,

$$\sum_{i=1}^{n} x_{ji} = 0, \quad j = 1, 2, \ldots, k.$$

The resulting $\mathbf{X}'\mathbf{X}$ is a diagonal matrix, and the normal equations in Section 12.3 reduce to

$$nb_0 = \sum_{i=1}^{n} y_i, \quad b_1 \sum_{i=1}^{n} x_{1i}^2 = \sum_{i=1}^{n} x_{1i}y_i, \cdots, b_k \sum_{i=1}^{n} x_{ki}^2 = \sum_{i=1}^{n} x_{ki}y_i.$$

An important advantage is that one is easily able to partition SSR into **single-degree-of-freedom components**, each of which corresponds to the amount of variation in Y accounted for by a given controlled variable. In the orthogonal situation, we can write

$$\begin{aligned} SSR &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} (b_0 + b_1 x_{1i} + \cdots + b_k x_{ki} - b_0)^2 \\ &= b_1^2 \sum_{i=1}^{n} x_{1i}^2 + b_2^2 \sum_{i=1}^{n} x_{2i}^2 + \cdots + b_k^2 \sum_{i=1}^{n} x_{ki}^2 \\ &= R(\beta_1) + R(\beta_2) + \cdots + R(\beta_k). \end{aligned}$$

The quantity $R(\beta_i)$ is the amount of the regression sum of squares associated with a model involving a single independent variable x_i.

To test simultaneously for the significance of a set of m variables in an orthogonal situation, the regression sum of squares becomes

$$R(\beta_1, \beta_2, \ldots, \beta_m \mid \beta_{m+1}, \beta_{m+2}, \ldots, \beta_k) = R(\beta_1) + R(\beta_2) + \cdots + R(\beta_m),$$

and thus we have the further simplification

$$R(\beta_1 \mid \beta_2, \beta_3, \ldots, \beta_k) = R(\beta_1)$$

when evaluating a single independent variable. Therefore, the contribution of a given variable or set of variables is essentially found by *ignoring* the other variables in the model. Independent evaluations of the worth of the individual variables are accomplished using analysis-of-variance techniques, as given in Table 12.4. The total variation in the response is partitioned into single-degree-of-freedom components plus the error term with $n-k-1$ degrees of freedom. Each computed f-value is used to test one of the hypotheses

$$\left.\begin{aligned} H_0&: \beta_i = 0 \\ H_1&: \beta_i \neq 0 \end{aligned}\right\} \qquad i = 1, 2, \ldots, k,$$

by comparing with the critical point $f_\alpha(1, n-k-1)$ or merely interpreting the P-value computed from the f-distribution.

Table 12.4: Analysis of Variance for Orthogonal Variables

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
β_1	$R(\beta_1) = b_1^2 \sum_{i=1}^{n} x_{1i}^2$	1	$R(\beta_1)$	$\frac{R(\beta_1)}{s^2}$
β_2	$R(\beta_2) = b_2^2 \sum_{i=1}^{n} x_{2i}^2$	1	$R(\beta_2)$	$\frac{R(\beta_2)}{s^2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
β_k	$R(\beta_k) = b_k^2 \sum_{i=1}^{n} x_{ki}^2$	1	$R(\beta_k)$	$\frac{R(\beta_k)}{s^2}$
Error	SSE	$n-k-1$	$s^2 = \frac{SSE}{n-k-1}$	
Total	$SST = S_{yy}$	$n-1$		

Example 12.8: Suppose that a scientist takes experimental data on the radius of a propellant grain Y as a function of powder temperature x_1, extrusion rate x_2, and die temperature x_3. Fit a linear regression model for predicting grain radius, and determine the effectiveness of each variable in the model. The data are given in Table 12.5.

Table 12.5: Data for Example 12.8

Grain Radius	**Powder Temperature**	**Extrusion Rate**	**Die Temperature**
82	150 (−1)	12 (−1)	220 (−1)
93	190 (+1)	12 (−1)	220 (−1)
114	150 (−1)	24 (+1)	220 (−1)
124	150 (−1)	12 (−1)	250 (+1)
111	190 (+1)	24 (+1)	220 (−1)
129	190 (+1)	12 (−1)	250 (+1)
157	150 (−1)	24 (+1)	250 (+1)
164	190 (+1)	24 (+1)	250 (+1)

Solution: Note that each variable is controlled at two levels, and the experiment is composed of the eight possible combinations. The data on the independent variables are coded for convenience by means of the following formulas:

$$x_1 = \frac{\text{powder temperature} - 170}{20},$$

$$x_2 = \frac{\text{extrusion rate} - 18}{6},$$

$$x_3 = \frac{\text{die temperature} - 235}{15}.$$

The resulting levels of x_1, x_2, and x_3 take on the values -1 and $+1$ as indicated in the table of data. This particular experimental design affords the orthogonality that we want to illustrate here. (A more thorough treatment of this type of experimental layout appears in Chapter 15.) The **X** matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

and the orthogonality conditions are readily verified.

We can now compute coefficients

$$b_0 = \frac{1}{8}\sum_{i=1}^{8} y_i = 121.75, \qquad b_1 = \frac{1}{8}\sum_{i=1}^{8} x_{1i}y_i = \frac{20}{8} = 2.5,$$

$$b_2 = \frac{\sum\limits_{i=1}^{8} x_{2i}y_i}{8} = \frac{118}{8} = 14.75, \qquad b_3 = \frac{\sum\limits_{i=1}^{8} x_{3i}y_i}{8} = \frac{174}{8} = 21.75,$$

so in terms of the coded variables, the prediction equation is

$$\hat{y} = 121.75 + 2.5\,x_1 + 14.75\,x_2 + 21.75\,x_3.$$

The analysis of variance in Table 12.6 shows independent contributions to SSR for each variable. The results, when compared to the $f_{0.05}(1,4)$ critical point of 7.71, indicate that x_1 does not contribute significantly at the 0.05 level, whereas variables x_2 and x_3 are significant. In this example, the estimate for σ^2 is 23.1250. As for the single independent variable case, it should be pointed out that this estimate does not solely contain experimental error variation unless the postulated model is correct. Otherwise, the estimate is "contaminated" by lack of fit in addition to pure error, and the lack of fit can be separated out only if we obtain multiple experimental observations for the various (x_1, x_2, x_3) combinations.

Table 12.6: Analysis of Variance for Grain Radius Data

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Squares**	**Computed f**	***P*-Value**
β_1	$(2.5)^2(8) = 50.00$	1	50.00	2.16	0.2156
β_2	$(14.75)^2(8) = 1740.50$	1	1740.50	75.26	0.0010
β_3	$(21.75)^2(8) = 3784.50$	1	3784.50	163.65	0.0002
Error	92.50	4	23.13		
Total	5667.50	7			

Since x_1 is not significant, it can simply be eliminated from the model without altering the effects of the other variables. Note that x_2 and x_3 both impact the grain radius in a positive fashion, with x_3 being the more important factor based on the smallness of its P-value. ▟

Exercises

12.31 Compute and interpret the coefficient of multiple determination for the variables of Exercise 12.1 on page 450.

12.32 Test whether the regression explained by the model in Exercise 12.1 on page 450 is significant at the 0.01 level of significance.

12.33 Test whether the regression explained by the model in Exercise 12.5 on page 450 is significant at the 0.01 level of significance.

12.34 For the model of Exercise 12.5 on page 450, test the hypothesis

$$H_0\text{: } \beta_1 = \beta_2 = 0,$$
$$H_1\text{: } \beta_1 \text{ and } \beta_2 \text{ are not both zero.}$$

12.35 Repeat Exercise 12.17 on page 461 using an F-statistic.

12.36 A small experiment was conducted to fit a multiple regression equation relating the yield y to temperature x_1, reaction time x_2, and concentration of one of the reactants x_3. Two levels of each variable were chosen, and measurements corresponding to the coded independent variables were recorded as follows:

y	x_1	x_2	x_3
7.6	−1	−1	−1
8.4	1	−1	−1
9.2	−1	1	−1
10.3	−1	−1	1
9.8	1	1	−1
11.1	1	−1	1
10.2	−1	1	1
12.6	1	1	1

(a) Using the coded variables, estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2,x_3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

(b) Partition SSR, the regression sum of squares, into three single-degree-of-freedom components attributable to x_1, x_2, and x_3, respectively. Show an analysis-of-variance table, indicating significance tests on each variable.

12.37 Consider the electric power data of Exercise 12.5 on page 450. Test H_0: $\beta_1 = \beta_2 = 0$, making use of $R(\beta_1, \beta_2 \mid \beta_3, \beta_4)$. Give a P-value, and draw conclusions.

12.38 Consider the data for Exercise 12.36. Compute the following:

$$\begin{array}{ll} R(\beta_1 \mid \beta_0), & R(\beta_1 \mid \beta_0, \beta_2, \beta_3), \\ R(\beta_2 \mid \beta_0, \beta_1), & R(\beta_2 \mid \beta_0, \beta_1, \beta_3), \\ R(\beta_3 \mid \beta_0, \beta_1, \beta_2), & R(\beta_1, \beta_2 \mid \beta_3). \end{array}$$

Comment.

12.39 Consider the data of Exercise 11.55 on page 437. Fit a regression model using weight and drive ratio as explanatory variables. Compare this model with the SLR (simple linear regression) model using weight alone. Use R^2, R^2_{adj}, and any t-statistics (or F-statistics) you may need to compare the SLR with the multiple regression model.

12.40 Consider Example 12.4. Figure 12.1 on page 459 displays a *SAS* printout of an analysis of the model containing variables x_1, x_2, and x_3. Focus on the confidence interval of the mean response μ_Y at the (x_1, x_2, x_3) locations representing the 13 data points. Consider an item in the printout indicated by C.V. This is the **coefficient of variation**, which is defined by

$$\text{C.V.} = \frac{s}{\bar{y}} \cdot 100,$$

where $s = \sqrt{s^2}$ is the **root mean squared error**. The coefficient of variation is often used as yet another criterion for comparing competing models. It is a scale-free quantity which expresses the estimate of σ, namely s, as a percent of the average response $\bar{y}$. In competition for the "best" among a group of competing models, one strives for the model with a small value of C.V. Do a regression analysis of the data set shown in Example 12.4 but eliminate x_3. Compare the full (x_1, x_2, x_3) model with the restricted (x_1, x_2) model and focus on two criteria: (i) C.V.; (ii) the widths of the confidence intervals on μ_Y. For the second criterion you may want to use the average width. Comment.

12.41 Consider Example 12.3 on page 447. Compare the two competing models.

$$\begin{aligned} \text{First order: } \quad & y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \\ \text{Second order: } \quad & y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \\ & \quad + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \epsilon_i. \end{aligned}$$

Use R^2_{adj} in your comparison. Test H_0: $\beta_{11} = \beta_{22} = \beta_{12} = 0$. In addition, use the C.V. discussed in Exercise 12.40.

12.42 In Example 12.8, a case is made for eliminating x_1, powder temperature, from the model since the P-value based on the F-test is 0.2156 while P-values for x_2 and x_3 are near zero.

(a) Reduce the model by eliminating x_1, thereby producing a full and a restricted (or reduced) model, and compare them on the basis of R^2_{adj}.

(b) Compare the full and restricted models using the width of the 95% prediction intervals on a new observation. The better of the two models would be that with the tightened prediction intervals. Use the average of the width of the prediction intervals.

12.43 Consider the data of Exercise 12.13 on page 452. Can the response, wear, be explained adequately by a single variable (either viscosity or load) in an SLR rather than with the full two-variable regression? Justify your answer thoroughly through tests of hypotheses as well as comparison of the three competing models.

12.44 For the data set given in Exericise 12.16 on page 453, can the response be explained adequately by any two regressor variables? Discuss.

12.8 Categorical or Indicator Variables

An extremely important special-case application of multiple linear regression occurs when one or more of the regressor variables are **categorical**, **indicator**, or **dummy variables**. In a chemical process, the engineer may wish to model the process yield against regressors such as process temperature and reaction time. However, there is interest in using two different catalysts and somehow including "the catalyst" in the model. The catalyst effect cannot be measured on a continuum and is hence a categorical variable. An analyst may wish to model the price of homes against regressors that include square feet of living space x_1, the land acreage x_2, and age of the house x_3. These regressors are clearly continuous in nature. However, it is clear that cost of homes may vary substantially from one area of the country to another. If data are collected on homes in the east, midwest, south, and west, we have an indicator variable with **four categories**. In the chemical process example, if two catalysts are used, we have an indicator variable with two categories. In a biomedical example in which a drug is to be compared to a placebo, all subjects are evaluated on several continuous measurements such as age, blood pressure, and so on, as well as gender, which of course is categorical with two categories. So, included along with the continuous variables are two indicator variables: treatment with two categories (active drug and placebo) and gender with two categories (male and female).

Model with Categorical Variables

Let us use the chemical processing example to illustrate how indicator variables are involved in the model. Suppose y = yield and x_1 = temperature and x_2 = reaction time. Now let us denote the indicator variable by z. Let $z = 0$ for catalyst 1 and $z = 1$ for catalyst 2. The assignment of the $(0, 1)$ indicator to the catalyst is arbitrary. As a result, the model becomes

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 z_i + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

Three Categories

The estimation of coefficients by the method of least squares continues to apply. In the case of three levels or categories of a single indicator variable, the model will include **two** regressors, say z_1 and z_2, where the $(0, 1)$ assignment is as follows:

$$\begin{array}{c} \begin{matrix} z_1 & \qquad & z_2 \end{matrix} \\ \begin{bmatrix} \mathbf{1} & \qquad & \mathbf{0} \\ \mathbf{0} & \qquad & \mathbf{1} \\ \mathbf{0} & \qquad & \mathbf{0} \end{bmatrix} \end{array},$$

where **0** and **1** are vectors of 0's and 1's, respectively. In other words, if there are ℓ categories, the model includes $\ell - 1$ actual model terms.

It may be instructive to look at a graphical representation of the model with three categories. For the sake of simplicity, let us assume a single continuous variable x. As a result, the model is given by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_{1i} + \beta_3 z_{2i} + \epsilon_i.$$

Thus, Figure 12.2 reflects the nature of the model. The following are model expressions for the three categories.

$$\begin{aligned} E(Y) &= (\beta_0 + \beta_2) + \beta_1 x, && \text{category 1,} \\ E(Y) &= (\beta_0 + \beta_3) + \beta_1 x, && \text{category 2,} \\ E(Y) &= \beta_0 + \beta_1 x, && \text{category 3.} \end{aligned}$$

As a result, the model involving categorical variables essentially involves a **change in the intercept** as we change from one category to another. Here of course we are assuming that the **coefficients of continuous variables remain the same across the categories**.

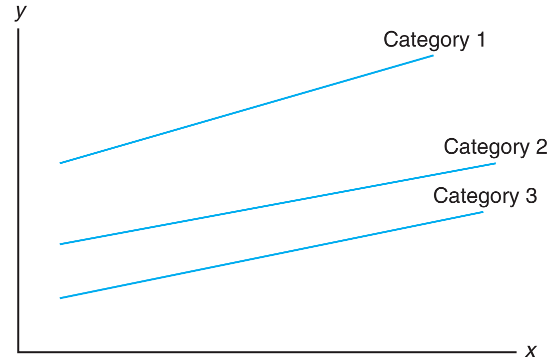


Figure 12.2: Case of three categories.

Example 12.9: Consider the data in Table 12.7. The response y is the amount of suspended solids in a coal cleansing system. The variable x is the pH of the system. Three different polymers are used in the system. Thus, "polymer" is categorical with three categories and hence produces two model terms. The model is given by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_{1i} + \beta_3 z_{2i} + \epsilon_i, \qquad i = 1, 2, \ldots, 18.$$

Here we have

$$z_1 = \begin{cases} 1, & \text{for polymer 1,} \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad z_2 = \begin{cases} 1, & \text{for polymer 2,} \\ 0, & \text{otherwise.} \end{cases}$$

From the analysis in Figure 12.3, the following conclusions are drawn. The coefficient b_1 for pH is the estimate of the **common slope** that is assumed in the regression analysis. All model terms are statistically significant. Thus, pH and the nature of the polymer have an impact on the amount of cleansing. The signs and magnitudes of the coefficients of z_1 and z_2 indicate that polymer 1 is most effective (producing higher suspended solids) for cleansing, followed by polymer 2. Polymer 3 is least effective. ▟

Table 12.7: Data for Example 12.9

x, (pH)	y, (amount of suspended solids)	Polymer
6.5	292	1
6.9	329	1
7.8	352	1
8.4	378	1
8.8	392	1
9.2	410	1
6.7	198	2
6.9	227	2
7.5	277	2
7.9	297	2
8.7	364	2
9.2	375	2
6.5	167	3
7.0	225	3
7.2	247	3
7.6	268	3
8.7	288	3
9.2	342	3

Slope May Vary with Indicator Categories

In the discussion given here, we have assumed that the indicator variable model terms enter the model in an additive fashion. This suggests that the slopes, as in Figure 12.2, are constant across categories. Obviously, this is not always going to be the case. We can account for the possibility of varying slopes and indeed test for this condition of **parallelism** by including product or **interaction** terms between indicator terms and continuous variables. For example, suppose a model with one continuous regressor and an indicator variable with two levels is chosen. The model is given by

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + \epsilon.$$

```
                                  Sum of
           Source   DF           Squares   Mean Square      F Value     Pr > F
            Model    3       80181.73127   26727.24376        73.68     <.0001
            Error   14        5078.71318     362.76523
 Corrected Total    17       85260.44444

  R-Square      Coeff Var       Root MSE        y Mean
  0.940433       6.316049       19.04640      301.5556
                                                  Standard
 Parameter      Estimate          Error  t Value  Pr > |t|
 Intercept -161.8973333  37.43315576    -4.32    0.0007
         x   54.2940260   4.75541126    11.42    <.0001
        z1   89.9980606  11.05228237     8.14    <.0001
        z2   27.1656970  11.01042883     2.47    0.0271
```

Figure 12.3: *SAS* printout for Example 12.9.

This model suggests that for category 1 ($z = 1$),

$$E(y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x,$$

while for category 2 ($z = 0$),

$$E(y) = \beta_0 + \beta_1 x.$$

Thus, we allow for varying intercepts and slopes for the two categories. Figure 12.4 displays the regression lines with varying slopes for the two categories.

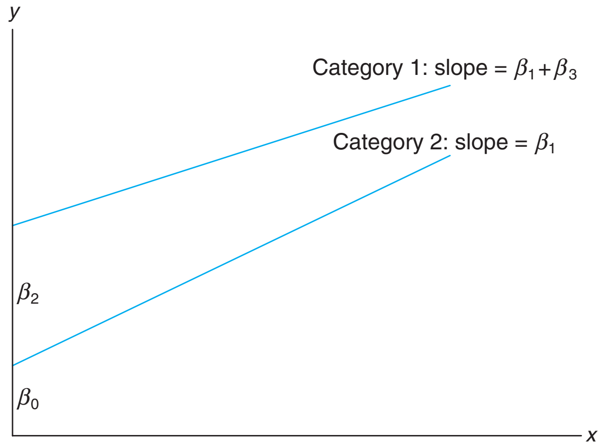


Figure 12.4: Nonparallelism in categorical variables.

In this case, β_0, β_1, and β_2 are positive while β_3 is negative with $|\beta_3| < \beta_1$. Obviously, if the interaction coefficient β_3 is insignificant, we are back to the common slope model.

Exercises

12.45 A study was done to assess the cost effectiveness of driving a four-door sedan instead of a van or an SUV (sports utility vehicle). The continuous variables are odometer reading and octane of the gasoline used. The response variable is miles per gallon. The data are presented here.

MPG	Car Type	Odometer	Octane
34.5	sedan	75,000	87.5
33.3	sedan	60,000	87.5
30.4	sedan	88,000	78.0
32.8	sedan	15,000	78.0
35.0	sedan	25,000	90.0
29.0	sedan	35,000	78.0
32.5	sedan	102,000	90.0
29.6	sedan	98,000	87.5
16.8	van	56,000	87.5
19.2	van	72,000	90.0
22.6	van	14,500	87.5
24.4	van	22,000	90.0
20.7	van	66,500	78.0
25.1	van	35,000	90.0
18.8	van	97,500	87.5
15.8	van	65,500	78.0
17.4	van	42,000	78.0
15.6	SUV	65,000	78.0
17.3	SUV	55,500	87.5
20.8	SUV	26,500	87.5
22.2	SUV	11,500	90.0
16.5	SUV	38,000	78.0
21.3	SUV	77,500	90.0
20.7	SUV	19,500	78.0
24.1	SUV	87,000	90.0

(a) Fit a linear regression model including two indicator variables. Use $(0, 0)$ to denote the four-door sedan.

(b) Which type of vehicle appears to get the best gas mileage?

(c) Discuss the difference between a van and an SUV in terms of gas mileage.

12.46 A study was done to determine whether the gender of the credit card holder was an important factor in generating profit for a certain credit card company. The variables considered were income, the number of family members, and the gender of the card holder. The data are as follows:

Profit	Income	Gender	Family Members
157	45,000	M	1
−181	55,000	M	2
−253	45,800	M	4
158	38,000	M	3
75	75,000	M	4
202	99,750	M	4
−451	28,000	M	1
146	39,000	M	2
89	54,350	M	1
−357	32,500	M	1
522	36,750	F	1
78	42,500	F	3
5	34,250	F	2
−177	36,750	F	3
123	24,500	F	2
251	27,500	F	1
−56	18,000	F	1
453	24,500	F	1
288	88,750	F	1
−104	19,750	F	2

(a) Fit a linear regression model using the variables available. Based on the fitted model, would the company prefer male or female customers?

(b) Would you say that income was an important factor in explaining the variability in profit?

12.9 Sequential Methods for Model Selection

At times, the significance tests outlined in Section 12.6 are quite adequate for determining which variables should be used in the final regression model. These tests are certainly effective if the experiment can be planned and the variables are orthogonal to each other. Even if the variables are not orthogonal, the individual t-tests can be of some use in many problems where the number of variables under investigation is small. However, there are many problems where it is necessary to use more elaborate techniques for screening variables, particularly when the experiment exhibits a substantial deviation from orthogonality. Useful measures of **multicollinearity** (linear dependency) among the independent variables are provided by the sample correlation coefficients $r_{x_i x_j}$. Since we are concerned only with linear dependency among independent variables, no confusion will result if we drop the x's from our notation and simply write $r_{x_i x_j} = r_{ij}$, where

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}.$$

Note that the r_{ij} do not give true estimates of population correlation coefficients in the strict sense, since the x's are actually not random variables in the context discussed here. Thus, the term *correlation*, although standard, is perhaps a misnomer.

When one or more of these sample correlation coefficients deviate substantially from zero, it can be quite difficult to find the most effective subset of variables for inclusion in our prediction equation. In fact, for some problems the multicollinearity will be so extreme that a suitable predictor cannot be found unless all possible subsets of the variables are investigated. Informative discussions of model selection in regression by Hocking (1976) are cited in the Bibliography. Procedures for detection of multicollinearity are discussed in the textbook by Myers (1990), also cited.

The user of multiple linear regression attempts to accomplish one of three objectives:

1. Obtain estimates of individual coefficients in a complete model.
2. Screen variables to determine which have a significant effect on the response.
3. Arrive at the most effective prediction equation.

In (1) it is known a priori that all variables are to be included in the model. In (2) prediction is secondary, while in (3) individual regression coefficients are not as important as the quality of the estimated response $\hat{y}$. For each of the situations above, multicollinearity in the experiment can have a profound effect on the success of the regression.

In this section, some standard sequential procedures for selecting variables are discussed. They are based on the notion that a single variable or a collection of variables should not appear in the estimating equation unless the variables result in a significant increase in the regression sum of squares or, equivalently, a significant increase in R^2, the coefficient of multiple determination.

Illustration of Variable Screening in the Presence of Collinearity

Example 12.10: Consider the data of Table 12.8, where measurements were taken for nine infants. The purpose of the experiment was to arrive at a suitable estimating equation relating the length of an infant to all or a subset of the independent variables. The sample correlation coefficients, indicating the linear dependency among the independent variables, are displayed in the symmetric matrix

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ \end{array}$$
$$\begin{bmatrix} 1.0000 & 0.9523 & 0.5340 & 0.3900 \\ 0.9523 & 1.0000 & 0.2626 & 0.1549 \\ 0.5340 & 0.2626 & 1.0000 & 0.7847 \\ 0.3900 & 0.1549 & 0.7847 & 1.0000 \end{bmatrix}$$

Table 12.8: Data Relating to Infant Length*

Infant Length, y (cm)	**Age, x_1 (days)**	**Length at Birth, x_2 (cm)**	**Weight at Birth, x_3 (kg)**	**Chest Size at Birth, x_4 (cm)**
57.5	78	48.2	2.75	29.5
52.8	69	45.5	2.15	26.3
61.3	77	46.3	4.41	32.2
67.0	88	49.0	5.52	36.5
53.5	67	43.0	3.21	27.2
62.7	80	48.0	4.32	27.7
56.2	74	48.0	2.31	28.3
68.5	94	53.0	4.30	30.3
69.2	102	58.0	3.71	28.7

*Data analyzed by the Statistical Consulting Center, Virginia Tech, Blacksburg, Virginia.

Note that there appears to be an appreciable amount of multicollinearity. Using the least squares technique outlined in Section 12.2, the estimated regression equation was fitted using the complete model and is

$$\hat{y} = 7.1475 + 0.1000x_1 + 0.7264x_2 + 3.0758x_3 - 0.0300x_4.$$

The value of s^2 with 4 degrees of freedom is 0.7414, and the value for the coefficient of determination for this model is found to be 0.9908. Regression sums of squares, measuring the variation attributed to each individual variable in the presence of the others, and the corresponding t-values are given in Table 12.9.

Table 12.9: t-Values for the Regression Data of Table 12.8

Variable x_1	**Variable x_2**	**Variable x_3**	**Variable x_4**
$R(\beta_1 \mid \beta_2, \beta_3, \beta_4)$ $= 0.0644$	$R(\beta_2 \mid \beta_1, \beta_3, \beta_4)$ $= 0.6334$	$R(\beta_3 \mid \beta_1, \beta_2, \beta_4)$ $= 6.2523$	$R(\beta_4 \mid \beta_1, \beta_2, \beta_3)$ $= 0.0241$
$t = 0.2947$	$t = 0.9243$	$t = 2.9040$	$t = -0.1805$

A two-tailed critical region with 4 degrees of freedom at the 0.05 level of significance is given by $|t| > 2.776$. Of the four computed t-values, **only variable x_3 appears to be significant**. However, recall that although the t-statistic described in Section 12.6 measures the worth of a variable adjusted for all other variables, it does not detect the potential importance of a variable in combination with a subset of the variables. For example, consider the model with only the variables x_2 and x_3 in the equation. The data analysis gives the regression function

$$\hat{y} = 2.1833 + 0.9576x_2 + 3.3253x_3,$$

with $R^2 = 0.9905$, certainly not a substantial reduction from $R^2 = 0.9907$ for the complete model. However, unless the performance characteristics of this particular combination had been observed, one would not be aware of its predictive potential. This, of course, lends support for a methodology that observes *all possible regressions* or a systematic sequential procedure designed to test subsets. ■

Stepwise Regression

One standard procedure for searching for the "optimum subset" of variables in the absence of orthogonality is a technique called **stepwise regression**. It is based on the procedure of sequentially introducing the variables into the model one at a time. Given a predetermined size α, the description of the stepwise routine will be better understood if the methods of **forward selection** and **backward elimination** are described first.

Forward selection is based on the notion that variables should be inserted one at a time until a satisfactory regression equation is found. The procedure is as follows:

STEP 1. Choose the variable that gives the largest regression sum of squares when performing a simple linear regression with y or, equivalently, that which gives the largest value of R^2. We shall call this initial variable x_1. If x_1 is insignificant, the procedure is terminated.

STEP 2. Choose the variable that, when inserted in the model, gives the largest increase in R^2, in the presence of x_1, over the R^2 found in step 1. This, of course, is the variable x_j for which

$$R(\beta_j|\beta_1) = R(\beta_1, \beta_j) - R(\beta_1)$$

is largest. Let us call this variable x_2. The regression model with x_1 and x_2 is then fitted and R^2 observed. If x_2 is insignificant, the procedure is terminated.

STEP 3. Choose the variable x_j that gives the largest value of

$$R(\beta_j \mid \beta_1, \beta_2) = R(\beta_1, \beta_2, \beta_j) - R(\beta_1, \beta_2),$$

again resulting in the largest increase of R^2 over that given in step 2. Calling this variable x_3, we now have a regression model involving x_1, x_2, and x_3. If x_3 is insignificant, the procedure is terminated.

This process is continued until the most recent variable inserted fails to induce a significant increase in the explained regression. Such an increase can be determined at each step by using the appropriate partial F-test or t-test. For example, in step 2 the value

$$f = \frac{R(\beta_2|\beta_1)}{s^2}$$

can be determined to test the appropriateness of x_2 in the model. Here the value of s^2 is the mean square error for the model containing the variables x_1 and x_2. Similarly, in step 3 the ratio

$$f = \frac{R(\beta_3 \mid \beta_1, \beta_2)}{s^2}$$

tests the appropriateness of x_3 in the model. Now, however, the value for s^2 is the mean square error for the model that contains the three variables x_1, x_2, and x_3. If $f < f_\alpha(1, n-3)$ at step 2, for a prechosen significance level, x_2 is not included and the process is terminated, resulting in a simple linear equation relating y and x_1. However, if $f > f_\alpha(1, n-3)$, we proceed to step 3. Again, if $f < f_\alpha(1, n-4)$ at step 3, x_3 is not included and the process is terminated with the appropriate regression equation containing the variables x_1 and x_2.

Backward elimination involves the same concepts as forward selection except that one begins with all the variables in the model. Suppose, for example, that there are five variables under consideration. The steps are as follows:

STEP 1. Fit a regression equation with all five variables included in the model. Choose the variable that gives the smallest value of the regression sum of squares **adjusted for the others**. Suppose that this variable is x_2. Remove x_2 from the model if

$$f = \frac{R(\beta_2 \mid \beta_1, \beta_3, \beta_4, \beta_5)}{s^2}$$

is insignificant.

STEP 2. Fit a regression equation using the remaining variables x_1, x_3, x_4, and x_5, and repeat step 1. Suppose that variable x_5 is chosen this time. Once again, if

$$f = \frac{R(\beta_5 \mid \beta_1, \beta_3, \beta_4)}{s^2}$$

is insignificant, the variable x_5 is removed from the model. At each step, the s^2 used in the F-test is the mean square error for the regression model at that stage.

This process is repeated until at some step the variable with the smallest adjusted regression sum of squares results in a significant f-value for some predetermined significance level.

Stepwise regression is accomplished with a slight but important modification of the forward selection procedure. The modification involves further testing at each stage to ensure the continued effectiveness of variables that had been inserted into the model at an earlier stage. This represents an improvement over forward selection, since it is quite possible that a variable entering the regression equation at an early stage might have been rendered unimportant or redundant because of relationships that exist between it and other variables entering at later stages. Therefore, at a stage in which a new variable has been entered into the regression equation through a significant increase in R^2 as determined by the F-test, all the variables already in the model are subjected to F-tests (or, equivalently, to t-tests) in light of this new variable and are deleted if they do not display a significant f-value. The procedure is continued until a stage is reached where no additional variables can be inserted or deleted. We illustrate the stepwise procedure in the following example.

Example 12.11: Using the techniques of stepwise regression, find an appropriate linear regression model for predicting the length of infants for the data of Table 12.8.

Solution: STEP 1. Considering each variable separately, four individual simple linear regression equations are fitted. The following pertinent regression sums of squares are computed:

$$R(\beta_1) = 288.1468, \qquad R(\beta_2) = 215.3013,$$
$$R(\beta_3) = 186.1065, \qquad R(\beta_4) = 100.8594.$$

Variable x_1 clearly gives the largest regression sum of squares. The mean square error for the equation involving only x_1 is $s^2 = 4.7276$, and since

$$f = \frac{R(\beta_1)}{s^2} = \frac{288.1468}{4.7276} = 60.9500,$$

which exceeds $f_{0.05}(1, 7) = 5.59$, the variable x_1 is significant and is entered into the model.

STEP 2. Three regression equations are fitted at this stage, all containing x_1. The important results for the combinations (x_1, x_2), (x_1, x_3), and (x_1, x_4) are

$$R(\beta_2|\beta_1) = 23.8703, \quad R(\beta_3|\beta_1) = 29.3086, \quad R(\beta_4|\beta_1) = 13.8178.$$

Variable x_3 displays the largest regression sum of squares in the presence of x_1. The regression involving x_1 and x_3 gives a new value of $s^2 = 0.6307$, and since

$$f = \frac{R(\beta_3|\beta_1)}{s^2} = \frac{29.3086}{0.6307} = 46.47,$$

which exceeds $f_{0.05}(1, 6) = 5.99$, the variable x_3 is significant and is included along with x_1 in the model. Now we must subject x_1 in the presence of x_3 to a significance test. We find that $R(\beta_1 \mid \beta_3) = 131.349$, and hence

$$f = \frac{R(\beta_1|\beta_3)}{s^2} = \frac{131.349}{0.6307} = 208.26,$$

which is highly significant. Therefore, x_1 is retained along with x_3.

STEP 3. With x_1 and x_3 already in the model, we now require $R(\beta_2 \mid \beta_1, \beta_3)$ and $R(\beta_4 \mid \beta_1, \beta_3)$ in order to determine which, if any, of the remaining two variables is entered at this stage. From the regression analysis using x_2 along with x_1 and x_3, we find $R(\beta_2 \mid \beta_1, \beta_3) = 0.7948$, and when x_4 is used along with x_1 and x_3, we obtain $R(\beta_4 \mid \beta_1, \beta_3) = 0.1855$. The value of s^2 is 0.5979 for the (x_1, x_2, x_3) combination and 0.7198 for the (x_1, x_2, x_4) combination. Since neither f-value is significant at the $\alpha = 0.05$ level, the final regression model includes only the variables x_1 and x_3. The estimating equation is found to be

$$\hat{y} = 20.1084 + 0.4136x_1 + 2.0253x_3,$$

and the coefficient of determination for this model is $R^2 = 0.9882$.

Although (x_1, x_3) is the combination chosen by stepwise regression, it is not necessarily the combination of two variables that gives the largest value of R^2. In fact, we have already observed that the combination (x_2, x_3) gives $R^2 = 0.9905$. Of course, the stepwise procedure never observed this combination. A rational argument could be made that there is actually a negligible difference in performance between these two estimating equations, at least in terms of percent variation explained. It is interesting to observe, however, that the backward elimination procedure gives the combination (x_2, x_3) in the final equation (see Exercise 12.49 on page 494). ▟

Summary

The main function of each of the procedures explained in this section is to expose the variables to a systematic methodology designed to ensure the eventual inclusion of the best combinations of the variables. Obviously, there is no assurance that this will happen in all problems, and, of course, it is possible that the multicollinearity is so extensive that one has no alternative but to resort to estimation procedures other than least squares. These estimation procedures are discussed in Myers (1990), listed in the Bibliography.

The sequential procedures discussed here represent three of many such methods that have been put forth in the literature and appear in various regression computer packages that are available. These methods are designed to be computationally efficient but, of course, do not give results for all possible subsets of the variables. As a result, the procedures are most effective for data sets that involve a **large number of variables**. For regression problems involving a relatively small number of variables, modern regression computer packages allow for the computation and summarization of quantitative information on all models for every possible subset of the variables. Illustrations are provided in Section 12.11.

Choice of *P*-Values

As one might expect, the choice of the final model with these procedures may depend dramatically on what P-value is chosen. In addition, a procedure is most successful when it is forced to test a large number of candidate variables. For this reason, any forward procedure will be most useful when a relatively large P-value is used. Thus, some software packages use a default P-value of 0.50.

12.10 Study of Residuals and Violation of Assumptions (Model Checking)

It was suggested earlier in this chapter that the residuals, or errors in the regression fit, often carry information that can be very informative to the data analyst. The $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$, which are the numerical counterpart to the ϵ_i, the model errors, often shed light on the possible violation of assumptions or the presence of "suspect" data points. Suppose that we let the vector $\mathbf{x}_i$ denote the values of the regressor variables corresponding to the ith data point, supplemented by a 1 in the initial position. That is,

$$\mathbf{x}_i' = [1, x_{1i}, x_{2i}, \ldots, x_{ki}].$$

Consider the quantity

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \qquad i = 1, 2, \ldots, n.$$

The reader should recognize that h_{ii} was used in the computation of the confidence intervals on the mean response in Section 12.5. Apart from σ^2, h_{ii} represents the variance of the fitted value $\hat{y}_i$. The h_{ii} values are the diagonal elements of the **HAT matrix**

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

which plays an important role in any study of residuals and in other modern aspects of regression analysis (see Myers, 1990, listed in the Bibliography). The term *HAT matrix* is derived from the fact that $\mathbf{H}$ generates the "y-hats," or the fitted values when multiplied by the vector $\mathbf{y}$ of observed responses. That is, $\hat{\mathbf{y}} = \mathbf{Xb}$, and thus

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{Hy},$$

where $\hat{\mathbf{y}}$ is the vector whose ith element is $\hat{y}_i$.

If we make the usual assumptions that the ϵ_i are independent and normally distributed with mean 0 and variance σ^2, the statistical properties of the residuals are readily characterized. Then

$$E(e_i) = E(y_i - \hat{y}_i) = 0 \quad \text{and} \quad \sigma^2_{\epsilon_i} = (1 - h_{ii})\sigma^2,$$

for $i = 1, 2, \ldots, n$. (See Myers, 1990, for details.) It can be shown that the HAT diagonal values are bounded according to the inequality

$$\frac{1}{n} \le h_{ii} \le 1.$$

In addition, $\sum\limits_{i=1}^{n} h_{ii} = k + 1$, the number of regression parameters. As a result, any data point whose HAT diagonal element is large, that is, well above the average value of $(k + 1)/n$, is in a position in the data set where the variance of $\hat{y}_i$ is relatively large and the variance of a residual is relatively small. As a result, the data analyst can gain some insight into how large a residual may become before its deviation from zero can be attributed to something other than mere chance. Many of the commercial regression computer packages produce the set of **studentized residuals**.

Studentized Residual

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}, \quad i = 1, 2, \ldots, n$$

Here each residual has been **divided by an estimate of its standard deviation**, creating a *t-like* statistic that is designed to give the analyst a scale-free quantity providing information regarding the *size* of the residual. In addition, standard computer packages often provide values of another set of studentized-type residuals called the ***R*-Student values**.

R-Student Residual

$$t_i = \frac{e_i}{s_{-i}\sqrt{1 - h_{ii}}}, \quad i = 1, 2, \ldots, n,$$

where s_{-i} is an estimate of the error standard deviation, calculated with the ***i*th data point deleted**.

There are three types of violations of assumptions that are readily detected through use of residuals or *residual plots*. While plots of the raw residuals, the e_i, can be helpful, it is often more informative to plot the studentized residuals. The three violations are as follows:

1. Presence of outliers
2. Heterogeneous error variance
3. Model misspecification

In case 1, we choose to define an **outlier** as a data point where there is a deviation from the usual assumption $E(\epsilon_i) = 0$ for a specific value of i. If there is a reason to believe that a specific data point is an outlier exerting a large influence on the fitted model, r_i or t_i may be informative. The R-Student values can be expected to be more sensitive to outliers than the r_i values.

In fact, under the condition that $E(\epsilon_i) = 0$, t_i is a value of a random variable following a t-distribution with $n-1-(k+1) = n-k-2$ degrees of freedom. Thus, a two-sided t-test can be used to provide information for detecting whether or not the ith point is an outlier.

Although the R-Student statistic t_i produces an exact t-test for detection of an outlier at a specific data location, the t-distribution would not apply for simultaneously testing for outliers at all locations. As a result, the studentized residuals or R-Student values should be used strictly as diagnostic tools *without* formal hypothesis testing as the mechanism. The implication is that these statistics highlight data points where the error of fit is larger than what is expected by chance. R-Student values large in magnitude suggest a need for "checking" the data with whatever resources are possible. The practice of eliminating observations from regression data sets should not be done indiscriminately. (For further information regarding the use of outlier diagnostics, see Myers, 1990, in the Bibliography.)

Illustration of Outlier Detection

Case Study 12.1: **Method for Capturing Grasshoppers**: In a biological experiment conducted at Virginia Tech by the Department of Entomology, n experimental runs were made with two different methods for capturing grasshoppers. The methods were drop net catch and sweep net catch. The average number of grasshoppers caught within a set of field quadrants on a given date was recorded for each of the two methods. An additional regressor variable, the average plant height in the quadrants, was also recorded. The experimental data are given in Table 12.10.

The goal is to be able to estimate grasshopper catch by using only the sweep net method, which is less costly. There was some concern about the validity of the fourth data point. The observed catch that was reported using the net drop method seemed unusually high given the other conditions and, indeed, it was felt that the figure might be erroneous. Fit a model of the type

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

to the 17 data points and study the residuals to determine if data point 4 is an outlier.

Table 12.10: Data Set for Case Study 12.1

Observation	Drop Net Catch, y	Sweep Net Catch, x_1	Plant Height, x_2 (cm)
1	18.0000	4.15476	52.705
2	8.8750	2.02381	42.069
3	2.0000	0.15909	34.766
4	20.0000	2.32812	27.622
5	2.3750	0.25521	45.879
6	2.7500	0.57292	97.472
7	3.3333	0.70139	102.062
8	1.0000	0.13542	97.790
9	1.3333	0.12121	88.265
10	1.7500	0.10937	58.737
11	4.1250	0.56250	42.386
12	12.8750	2.45312	31.274
13	5.3750	0.45312	31.750
14	28.0000	6.68750	35.401
15	4.7500	0.86979	64.516
16	1.7500	0.14583	25.241
17	0.1333	0.01562	36.354

***Solution*:** A computer package generated the fitted regression model

$$\hat{y} = 3.6870 + 4.1050x_1 - 0.0367x_2$$

along with the statistics $R^2 = 0.9244$ and $s^2 = 5.580$. The residuals and other diagnostic information were also generated and recorded in Table 12.11.

As expected, the residual at the fourth location appears to be unusually high, namely 7.769. The vital issue here is whether or not this residual is larger than one would expect by chance. The residual standard error for point 4 is 2.209. The R-Student value t_4 is found to be 9.9315. Viewing this as a value of a random variable having a t-distribution with 13 degrees of freedom, one would certainly conclude that the residual of the fourth observation is estimating something greater than 0 and that the suspected measurement error is supported by the study of residuals. Notice that no other residual results in an R-Student value that produces any cause for alarm. ■

Plotting Residuals for Case Study 12.1

In Chapter 11, we discussed, in some detail, the usefulness of plotting residuals in regression analysis. Violation of model assumptions can often be detected through these plots. In multiple regression, normal probability plotting of residuals or plotting of residuals against $\hat{y}$ may be useful. However, it is often preferable to plot studentized residuals.

Keep in mind that the preference for the studentized residuals over ordinary residuals for plotting purposes stems from the fact that since the variance of the

Table 12.11: Residual Information for the Data Set of Case Study 12.1

Obs.	$\boldsymbol{y_i}$	$\boldsymbol{\hat{y}_i}$	$\boldsymbol{y_i - \hat{y}_i}$	$\boldsymbol{h_{ii}}$	$\boldsymbol{s\sqrt{1-h_{ii}}}$	$\boldsymbol{r_i}$	$\boldsymbol{t_i}$
1	18.000	18.809	−0.809	0.2291	2.074	−0.390	−0.3780
2	8.875	10.452	−1.577	0.0766	2.270	−0.695	−0.6812
3	2.000	3.065	−1.065	0.1364	2.195	−0.485	−0.4715
4	20.000	12.231	7.769	0.1256	2.209	3.517	9.9315
5	2.375	3.052	−0.677	0.0931	2.250	−0.301	−0.2909
6	2.750	2.464	0.286	0.2276	2.076	0.138	0.1329
7	3.333	2.823	0.510	0.2669	2.023	0.252	0.2437
8	1.000	0.656	0.344	0.2318	2.071	0.166	0.1601
9	1.333	0.947	0.386	0.1691	2.153	0.179	0.1729
10	1.750	1.982	−0.232	0.0852	2.260	−0.103	−0.0989
11	4.125	4.442	−0.317	0.0884	2.255	−0.140	−0.1353
12	12.875	12.610	0.265	0.1152	2.222	0.119	0.1149
13	5.375	4.383	0.992	0.1339	2.199	0.451	0.4382
14	28.000	29.841	−1.841	0.6233	1.450	−1.270	−1.3005
15	4.750	4.891	−0.141	0.0699	2.278	−0.062	−0.0598
16	1.750	3.360	−1.610	0.1891	2.127	−0.757	−0.7447
17	0.133	2.418	−2.285	0.1386	2.193	−1.042	−1.0454

ith residual depends on the ith HAT diagonal, variances of residuals will differ if there is a dispersion in the HAT diagonals. Thus, the appearance of a plot of residuals may seem to suggest heterogeneity because the residuals themselves do not behave, in general, in an ideal way. The purpose of using studentized residuals is to provide a type of *standardization*. Clearly, if σ were known, then under ideal conditions (i.e., a correct model and homogeneous variance), we would have

$$E\left(\frac{e_i}{\sigma\sqrt{1-h_{ii}}}\right) = 0 \quad \text{and} \quad \text{Var}\left(\frac{e_i}{\sigma\sqrt{1-h_{ii}}}\right) = 1.$$

So the studentized residuals produce a set of statistics that behave in a standard way under ideal conditions. Figure 12.5 shows a plot of the ***R*-Student** values for the grasshopper data of Case Study 12.1. Note how the value for observation 4 stands out from the rest. The R-Student plot was generated by *SAS* software. The plot shows the residuals against the $\hat{y}$-values.

Normality Checking

The reader should recall the importance of normality checking through the use of normal probability plotting, as discussed in Chapter 11. The same recommendation holds for the case of multiple linear regression. Normal probability plots can be generated using standard regression software. Again, however, they can be more effective when one does not use ordinary residuals but, rather, studentized residuals or R-Student values.

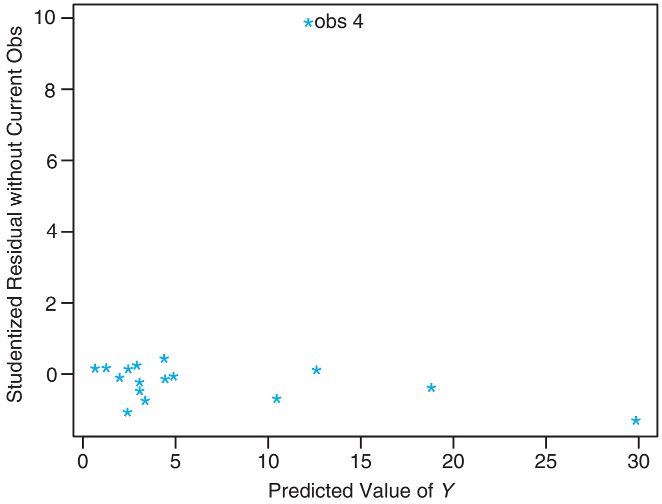


Figure 12.5: R-Student values plotted against predicted values for grasshopper data of Case Study 12.1.

12.11 Cross Validation, C_p, and Other Criteria for Model Selection

For many regression problems, the experimenter must choose among various alternative models or model forms that are developed from the same data set. Quite often, the model that best predicts or estimates mean response is required. The experimenter should take into account the relative sizes of the s^2-values for the candidate models and certainly the general nature of the confidence intervals on the mean response. One must also consider how well the model predicts response values that were **not used in building the candidate models**. The models should be subjected to **cross validation**. What are required, then, are cross-validation errors rather than fitting errors. Such errors in prediction are the **PRESS residuals**

$$\delta_i = y_i - \hat{y}_{i,-i}, \quad i = 1, 2, \ldots, n,$$

where $\hat{y}_{i,-i}$ is the prediction of the ith data point by a model that did not make use of the ith point in the calculation of the coefficients. These PRESS residuals are calculated from the formula

$$\delta_i = \frac{e_i}{1 - h_{ii}}, \quad i = 1, 2, \ldots, n.$$

(The derivation can be found in Myers, 1990.)

Use of the PRESS Statistic

The motivation for PRESS and the utility of PRESS residuals are very simple to understand. The purpose of extracting or *setting aside* data points one at a time is to allow the use of separate methodologies for fitting and assessment of a specific model. For assessment of a model, the "$-i$" indicates that the PRESS residual gives a prediction error where the observation being predicted is *independent of the model fit.*

Criteria that make use of the PRESS residuals are given by

$$\sum_{i=1}^{n} |\delta_i| \quad \text{and} \quad \text{PRESS} = \sum_{i=1}^{n} \delta_i^2.$$

(The term **PRESS** is an acronym for **prediction sum of squares**.) We suggest that both of these criteria be used. It is possible for PRESS to be dominated by one or only a few large PRESS residuals. Clearly, the criterion on $\sum\limits_{i=1}^{n} |\delta_i|$ is less sensitive to a small number of large values.

In addition to the PRESS statistic itself, the analyst can simply compute an R^2-like statistic reflecting prediction performance. The statistic is often called R^2_{pred} and is given as follows:

R^2 of Prediction Given a fitted model with a specific value for PRESS, R^2_{pred} is given by

$$R^2_{\text{pred}} = 1 - \frac{\text{PRESS}}{\sum\limits_{i=1}^{n} (y_i - \bar{y})^2}.$$

Note that R^2_{pred} is merely the ordinary R^2 statistic with SSE replaced by the PRESS statistic.

In the following case study, an illustration is provided in which many candidate models are fit to a set of data and the best model is chosen. The sequential procedures described in Section 12.9 are not used. Rather, the role of the PRESS residuals and other statistical values in selecting the best regression equation is illustrated.

Case Study 12.2: **Football Punting**: Leg strength is a necessary characteristic of a successful punter in American football. One measure of the quality of a good punt is the "hang time." This is the time that the ball hangs in the air before being caught by the punt returner. To determine what leg strength factors influence hang time and to develop an empirical model for predicting this response, a study on *The Relationship Between Selected Physical Performance Variables and Football Punting Ability* was conducted by the Department of Health, Physical Education, and Recreation at Virginia Tech. Thirteen punters were chosen for the experiment, and each punted a football 10 times. The average hang times, along with the strength measures used in the analysis, were recorded in Table 12.12.

Each regressor variable is defined as follows:

1. **RLS**, right leg strength (pounds)
2. **LLS**, left leg strength (pounds)
3. **RHF,** right hamstring muscle flexibility (degrees)
4. **LHF,** left hamstring muscle flexibility (degrees)

5. **Power**, overall leg strength (foot-pounds)

Determine the most appropriate model for predicting hang time.

Table 12.12: Data for Case Study 12.2

Punter	**Hang Time,** y **(sec)**	**RLS,** x_1	**LLS,** x_2	**RHF,** x_3	**LHF,** x_4	**Power,** x_5
1	4.75	170	170	106	106	240.57
2	4.07	140	130	92	93	195.49
3	4.04	180	170	93	78	152.99
4	4.18	160	160	103	93	197.09
5	4.35	170	150	104	93	266.56
6	4.16	150	150	101	87	260.56
7	4.43	170	180	108	106	219.25
8	3.20	110	110	86	92	132.68
9	3.02	120	110	90	86	130.24
10	3.64	130	120	85	80	205.88
11	3.68	120	140	89	83	153.92
12	3.60	140	130	92	94	154.64
13	3.85	160	150	95	95	240.57

***Solution*:** In the search for the best of the candidate models for predicting hang time, the information in Table 12.13 was obtained from a regression computer package. The models are ranked in ascending order of the values of the PRESS statistic. This display provides enough information on all possible models to enable the user to eliminate from consideration all but a few models. The model containing x_2 and x_5 (LLS and Power), denoted by x_2x_5, appears to be superior for predicting punter hang time. Also note that all models with low PRESS, low s^2, low $\sum\limits_{i=1}^{n} |\delta_i|$, and high R^2-values contain these two variables.

In order to gain some insight from the residuals of the fitted regression

$$\hat{y}_i = b_0 + b_2 x_{2i} + b_5 x_{5i},$$

the residuals and PRESS residuals were generated. The actual prediction model (see Exercise 12.47 on page 494) is given by

$$\hat{y} = 1.10765 + 0.01370 x_2 + 0.00429 x_5.$$

Residuals, HAT diagonal values, and PRESS values are listed in Table 12.14.

Note the relatively good fit of the two-variable regression model to the data. The PRESS residuals reflect the capability of the regression equation to predict hang time if independent predictions were to be made. For example, for punter number 4, the hang time of 4.180 would encounter a prediction error of 0.039 if the model constructed by using the remaining 12 punters were used. For this model, the average prediction error or cross-validation error is

$$\frac{1}{13} \sum_{i=1}^{n} |\delta_i| = 0.1489 \text{ second,}$$

Table 12.13: Comparing Different Regression Models

Model	s^2	$\sum \lvert \delta_i \rvert$	**PRESS**	R^2
x_2x_5	0.036907	1.93583	0.54683	0.871300
$x_1x_2x_5$	0.041001	2.06489	0.58998	0.871321
$x_2x_4x_5$	0.037708	2.18797	0.59915	0.881658
$x_2x_3x_5$	0.039636	2.09553	0.66182	0.875606
$x_1x_2x_4x_5$	0.042265	2.42194	0.67840	0.882093
$x_1x_2x_3x_5$	0.044578	2.26283	0.70958	0.875642
$x_2x_3x_4x_5$	0.042421	2.55789	0.86236	0.881658
$x_1x_3x_5$	0.053664	2.65276	0.87325	0.831580
$x_1x_4x_5$	0.056279	2.75390	0.89551	0.823375
x_1x_5	0.059621	2.99434	0.97483	0.792094
x_2x_3	0.056153	2.95310	0.98815	0.804187
x_1x_3	0.059400	3.01436	0.99697	0.792864
$x_1x_2x_3x_4x_5$	0.048302	2.87302	1.00920	0.882096
x_2	0.066894	3.22319	1.04564	0.743404
x_3x_5	0.065678	3.09474	1.05708	0.770971
x_1x_2	0.068402	3.09047	1.09726	0.761474
x_3	0.074518	3.06754	1.13555	0.714161
$x_1x_3x_4$	0.065414	3.36304	1.15043	0.794705
$x_2x_3x_4$	0.062082	3.32392	1.17491	0.805163
x_2x_4	0.063744	3.59101	1.18531	0.777716
$x_1x_2x_3$	0.059670	3.41287	1.26558	0.812730
x_3x_4	0.080605	3.28004	1.28314	0.718921
x_1x_4	0.069965	3.64415	1.30194	0.756023
x_1	0.080208	3.31562	1.30275	0.692334
$x_1x_3x_4x_5$	0.059169	3.37362	1.36867	0.834936
$x_1x_2x_4$	0.064143	3.89402	1.39834	0.798692
$x_3x_4x_5$	0.072505	3.49695	1.42036	0.772450
$x_1x_2x_3x_4$	0.066088	3.95854	1.52344	0.815633
x_5	0.111779	4.17839	1.72511	0.571234
x_4x_5	0.105648	4.12729	1.87734	0.631593
x_4	0.186708	4.88870	2.82207	0.283819

which is small compared to the average hang time for the 13 punters.

We indicated in Section 12.9 that the use of all possible subset regressions is often advisable when searching for the best model. Most commercial statistics software packages contain an *all possible regressions* routine. These algorithms compute various criteria for all subsets of model terms. Obviously, criteria such as R^2, s^2, and PRESS are reasonable for choosing among candidate subsets. Another very popular and useful statistic, particularly for areas in the physical sciences and engineering, is the C_p statistic, described below.

Table 12.14: PRESS Residuals

Punter	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	h_{ii}	δ_i
1	4.750	4.470	0.280	0.198	0.349
2	4.070	3.728	0.342	0.118	0.388
3	4.040	4.094	−0.054	0.444	−0.097
4	4.180	4.146	0.034	0.132	0.039
5	4.350	4.307	0.043	0.286	0.060
6	4.160	4.281	−0.121	0.250	−0.161
7	4.430	4.515	−0.085	0.298	−0.121
8	3.200	3.184	0.016	0.294	0.023
9	3.020	3.174	−0.154	0.301	−0.220
10	3.640	3.636	0.004	0.231	0.005
11	3.680	3.687	−0.007	0.152	−0.008
12	3.600	3.553	0.047	0.142	0.055
13	3.850	4.196	−0.346	0.154	−0.409

The C_p Statistic

Quite often, the choice of the most appropriate model involves many considerations. Obviously, the number of model terms is important; the matter of parsimony is a consideration that cannot be ignored. On the other hand, the analyst cannot be pleased with a model that is too simple, to the point where there is serious underspecification. A single statistic that represents a nice compromise in this regard is the C_p statistic. (See Mallows, 1973, in the Bibliography.)

The C_p statistic appeals nicely to common sense and is developed from considerations of the proper compromise between excessive bias incurred when one underfits (chooses too few model terms) and excessive prediction variance produced when one overfits (has redundancies in the model). The C_p statistic is a simple function of the total number of parameters in the candidate model and the mean square error s^2.

We will not present the entire development of the C_p statistic. (For details, the reader is referred to Myers, 1990, in the Bibliography.) The C_p for a particular subset model is *an estimate* of the following:

$$\Gamma_{(p)} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\text{Var}(\hat{y}_i) + \frac{1}{\sigma^2}\sum_{i=1}^{n}(\text{Bias } \hat{y}_i)^2.$$

It turns out that under the standard least squares assumptions indicated earlier in this chapter, and assuming that the "true" model is the model containing all candidate variables,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\text{Var}(\hat{y}_i) = p \qquad \text{(number of parameters in the candidate model)}$$

(see Review Exercise 12.63) and an unbiased estimate of

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(\text{Bias } \hat{y}_i)^2 \text{ is given by } \frac{1}{\sigma^2}\sum_{i=1}^{n}(\widehat{\text{Bias}}\ \hat{y}_i)^2 = \frac{(s^2-\sigma^2)(n-p)}{\sigma^2}.$$

In the above, s^2 is the mean square error for the candidate model and σ^2 is the population error variance. Thus, if we assume that some estimate $\hat{\sigma}^2$ is available for σ^2, C_p is given by the following equation:

C_p Statistic

$$C_p = p + \frac{(s^2-\hat{\sigma}^2)(n-p)}{\hat{\sigma}^2},$$

where p is the number of model parameters, s^2 is the mean square error for the candidate model, and $\hat{\sigma}^2$ is an estimate of σ^2.

Obviously, the scientist should adopt models with small values of C_p. The reader should note that, unlike the PRESS statistic, C_p is scale-free. In addition, one can gain some insight concerning the adequacy of a candidate model by observing its value of C_p. For example, $C_p > p$ indicates a model that is biased due to being an underfitted model, whereas $C_p \approx p$ indicates a reasonable model.

There is often confusion concerning where $\hat{\sigma}^2$ comes from in the formula for C_p. Obviously, the scientist or engineer does not have access to the population quantity σ^2. In applications where replicated runs are available, say in an experimental design situation, a model-independent estimate of σ^2 is available (see Chapters 11 and 15). However, most software packages use $\hat{\sigma}^2$ as the *mean square error from the most complete model.* Obviously, if this is not a good estimate, the bias portion of the C_p statistic can be negative. Thus, C_p can be less than p.

Example 12.12: Consider the data set in Table 12.15, in which a maker of asphalt shingles is interested in the relationship between sales for a particular year and factors that influence sales. (The data were taken from Kutner et al., 2004, in the Bibliography.)

Of the possible subset models, three are of particular interest. These three are x_2x_3, $x_1x_2x_3$, and $x_1x_2x_3x_4$. The following represents pertinent information for comparing the three models. We include the PRESS statistics for the three models to supplement the decision making.

Model	R^2	R^2_{pred}	s^2	**PRESS**	C_p
x_2x_3	0.9940	0.9913	44.5552	782.1896	11.4013
$x_1x_2x_3$	0.9970	0.9928	24.7956	643.3578	3.4075
$x_1x_2x_3x_4$	0.9971	0.9917	26.2073	741.7557	5.0

It seems clear from the information in the table that the model x_1, x_2, x_3 is preferable to the other two. Notice that, for the full model, $C_p = 5.0$. This occurs since the *bias portion* is zero, and $\hat{\sigma}^2 = 26.2073$ is the mean square error from the full model.

Figure 12.6 is a *SAS* PROC REG printout showing information for all possible regressions. Here we are able to show comparisons of other models with (x_1, x_2, x_3). Note that (x_1, x_2, x_3) appears to be quite good when compared to all models.

As a final check on the model (x_1, x_2, x_3), Figure 12.7 shows a normal probability plot of the residuals for this model.

Table 12.15: Data for Example 12.12

District	Promotional Accounts, x_1	Active Accounts, x_2	Competing Brands, x_3	Potential, x_4	Sales, y (thousands)
1	5.5	31	10	8	$ 79.3
2	2.5	55	8	6	200.1
3	8.0	67	12	9	163.2
4	3.0	50	7	16	200.1
5	3.0	38	8	15	146.0
6	2.9	71	12	17	177.7
7	8.0	30	12	8	30.9
8	9.0	56	5	10	291.9
9	4.0	42	8	4	160.0
10	6.5	73	5	16	339.4
11	5.5	60	11	7	159.6
12	5.0	44	12	12	86.3
13	6.0	50	6	6	237.5
14	5.0	39	10	4	107.2
15	3.5	55	10	4	155.0

```
                  Dependent Variable: sales
Number in                      Adjusted
    Model      C(p)  R-Square  R-Square           MSE  Variables in Model

        3    3.4075    0.9970    0.9961      24.79560  x1 x2 x3
        4    5.0000    0.9971    0.9959      26.20728  x1 x2 x3 x4
        2   11.4013    0.9940    0.9930      44.55518  x2 x3
        3   13.3770    0.9940    0.9924      48.54787  x2 x3 x4
        3  1053.643    0.6896    0.6049    2526.96144  x1 x3 x4
        2  1082.670    0.6805    0.6273    2384.14286  x3 x4
        2  1215.316    0.6417    0.5820    2673.83349  x1 x3
        1  1228.460    0.6373    0.6094    2498.68333  x3
        3  1653.770    0.5140    0.3814    3956.75275  x1 x2 x4
        2  1668.699    0.5090    0.4272    3663.99357  x1 x2
        2  1685.024    0.5042    0.4216    3699.64814  x2 x4
        1  1693.971    0.5010    0.4626    3437.12846  x2
        2  3014.641    0.1151    -.0324    6603.45109  x1 x4
        1  3088.650    0.0928    0.0231    6248.72283  x4
        1  3364.884    0.0120    -.0640    6805.59568  x1
```

Figure 12.6: *SAS* printout of all possible subsets on sales data for Example 12.12.

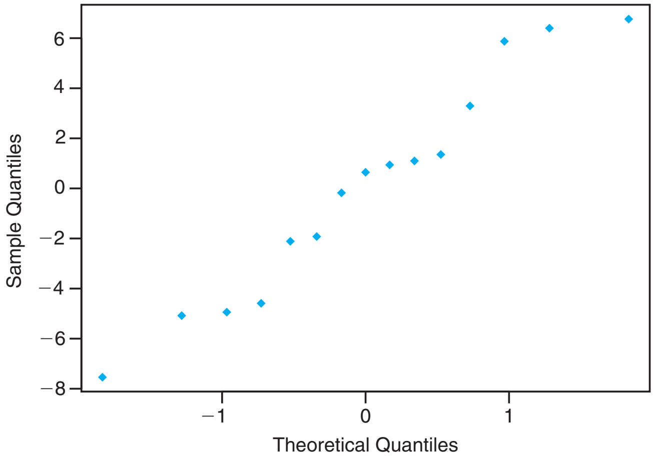


Figure 12.7: Normal probability plot of residuals using the model $x_1x_2x_3$ for Example 12.12.

Exercises

12.47 Consider the "hang time" punting data given in Case Study 12.2, using only the variables x_2 and x_3.

(a) Verify the regression equation shown on page 489.

(b) Predict punter hang time for a punter with LLS = 180 pounds and Power = 260 foot-pounds.

(c) Construct a 95% confidence interval for the mean hang time of a punter with LLS = 180 pounds and Power = 260 foot-pounds.

12.48 For the data of Exercise 12.15 on page 452, use the techniques of

(a) *forward selection* with a 0.05 level of significance to choose a linear regression model;

(b) *backward elimination* with a 0.05 level of significance to choose a linear regression model;

(c) *stepwise regression* with a 0.05 level of significance to choose a linear regression model.

12.49 Use the techniques of *backward elimination* with $\alpha = 0.05$ to choose a prediction equation for the data of Table 12.8.

12.50 For the punter data in Case Study 12.2, an additional response, "punting distance," was also recorded. The average distance values for each of the 13 punters are given.

(a) Using the distance data rather than the hang times, estimate a multiple linear regression model of the type

$$\mu_{Y|x_1,x_2,x_3,x_4,x_5} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

for predicting punting distance.

(b) Use stepwise regression with a significance level of 0.10 to select a combination of variables.

(c) Generate values for s^2, R^2, PRESS, and $\sum_{i=1}^{13} |\delta_i|$ for the entire set of 31 models. Use this information to determine the best combination of variables for predicting punting distance.

(d) For the final model you choose, plot the standardized residuals against Y and do a normal probability plot of the ordinary residuals. Comment.

Punter	**Distance, y (ft)**
1	162.50
2	144.00
3	147.50
4	163.50
5	192.00
6	171.75
7	162.00
8	104.93
9	105.67
10	117.59
11	140.25
12	150.17
13	165.16

12.51 The following is a set of data for y, the amount of money (in thousands of dollars) contributed to the alumni association at Virginia Tech by the Class of 1960, and x, the number of years following graduation:

y	x	y	x
812.52	1	2755.00	11
822.50	2	4390.50	12
1211.50	3	5581.50	13
1348.00	4	5548.00	14
1301.00	8	6086.00	15
2567.50	9	5764.00	16
2526.50	10	8903.00	17

(a) Fit a regression model of the type

$$\mu_{Y|x} = \beta_0 + \beta_1 x.$$

(b) Fit a quadratic model of the type

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_{11} x^2.$$

(c) Determine which of the models in (a) or (b) is preferable. Use s^2, R^2, and the PRESS residuals to support your decision.

12.52 For the model of Exercise 12.50(a), test the hypothesis

$$H_0\text{: } \beta_4 = 0,$$
$$H_1\text{: } \beta_4 \neq 0.$$

Use a P-value in your conclusion.

12.53 For the quadratic model of Exercise 12.51(b), give estimates of the variances and covariances of the estimates of β_1 and β_{11}.

12.54 A client from the Department of Mechanical Engineering approached the Consulting Center at Virginia Tech for help in analyzing an experiment dealing with gas turbine engines. The voltage output of engines was measured at various combinations of blade speed and sensor extension.

y (volts)	Speed, x_1 (in./sec)	Extension, x_2 (in.)
1.95	6336	0.000
2.50	7099	0.000
2.93	8026	0.000
1.69	6230	0.000
1.23	5369	0.000
3.13	8343	0.000
1.55	6522	0.006
1.94	7310	0.006
2.18	7974	0.006
2.70	8501	0.006
1.32	6646	0.012
1.60	7384	0.012
1.89	8000	0.012
2.15	8545	0.012
1.09	6755	0.018
1.26	7362	0.018
1.57	7934	0.018
1.92	8554	0.018

(a) Fit a multiple linear regression to the data.

(b) Compute t-tests on coefficients. Give P-values.

(c) Comment on the quality of the fitted model.

12.55 Rayon whiteness is an important factor for scientists dealing in fabric quality. Whiteness is affected by pulp quality and other processing variables. Some of the variables include acid bath temperature, °C (x_1); cascade acid concentration, % (x_2); water temperature, °C (x_3); sulfide concentration, % (x_4); amount of chlorine bleach, lb/min (x_5); and blanket finish temperature, °C (x_6). A set of data from rayon specimens is given here. The response, y, is the measure of whiteness.

y	x_1	x_2	x_3	x_4	x_5	x_6
88.7	43	0.211	85	0.243	0.606	48
89.3	42	0.604	89	0.237	0.600	55
75.5	47	0.450	87	0.198	0.527	61
92.1	46	0.641	90	0.194	0.500	65
83.4	52	0.370	93	0.198	0.485	54
44.8	50	0.526	85	0.221	0.533	60
50.9	43	0.486	83	0.203	0.510	57
78.0	49	0.504	93	0.279	0.489	49
86.8	51	0.609	90	0.220	0.462	64
47.3	51	0.702	86	0.198	0.478	63
53.7	48	0.397	92	0.231	0.411	61
92.0	46	0.488	88	0.211	0.387	88
87.9	43	0.525	85	0.199	0.437	63
90.3	45	0.486	84	0.189	0.499	58
94.2	53	0.527	87	0.245	0.530	65
89.5	47	0.601	95	0.208	0.500	67

(a) Use the criteria *MSE*, C_p, and PRESS to find the "best" model from among all subset models.

(b) Plot standardized residuals against Y and do a normal probability plot of residuals for the "best" model. Comment.

12.56 In an effort to model executive compensation for the year 1979, 33 firms were selected, and data were gathered on compensation, sales, profits, and employment. The following data were gathered for the year 1979.

Firm	Compensation, y (thousands)	Sales, x_1 (millions)	Profits, x_2 (millions)	Employment, x_3
1	\$450	\$4600.6	\$128.1	48,000
2	387	9255.4	783.9	55,900
3	368	1526.2	136.0	13,783
4	277	1683.2	179.0	27,765
5	676	2752.8	231.5	34,000
6	454	2205.8	329.5	26,500
7	507	2384.6	381.8	30,800
8	496	2746.0	237.9	41,000
9	487	1434.0	222.3	25,900

(cont.)

Firm	**Compensation, y (thousands)**	**Sales, x_1 (millions)**	**Profits, x_2 (millions)**	**Employment, x_3**
10	\$383	\$470.6	\$63.7	8600
11	311	1508.0	149.5	21,075
12	271	464.4	30.0	6874
13	524	9329.3	577.3	39,000
14	498	2377.5	250.7	34,300
15	343	1174.3	82.6	19,405
16	354	409.3	61.5	3586
17	324	724.7	90.8	3905
18	225	578.9	63.3	4139
19	254	966.8	42.8	6255
20	208	591.0	48.5	10,605
21	518	4933.1	310.6	65,392
22	406	7613.2	491.6	89,400
23	332	3457.4	228.0	55,200
24	340	545.3	54.6	7800
25	698	22,862.8	3011.3	337,119
26	306	2361.0	203.0	52,000
27	613	2614.1	201.0	50,500
28	302	1013.2	121.3	18,625
29	540	4560.3	194.6	97,937
30	293	855.7	63.4	12,300
31	528	4211.6	352.1	71,800
32	456	5440.4	655.2	87,700
33	417	1229.9	97.5	14,600

Consider the model

$$y_i = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \beta_3 \ln x_{3i} + \epsilon_i, \qquad i = 1, 2, \ldots, 33.$$

(a) Fit the regression with the model above.

(b) Is a model with a subset of the variables preferable to the full model?

12.57 The pull strength of a wire bond is an important characteristic. The following data give information on pull strength y, die height x_1, post height x_2, loop height x_3, wire length x_4, bond width on the die x_5, and bond width on the post x_6. (From Myers, Montgomery, and Anderson-Cook, 2009.)

(a) Fit a regression model using all independent variables.

(b) Use stepwise regression with input significance level 0.25 and removal significance level 0.05. Give your final model.

(c) Use all possible regression models and compute R^2, C_p, s^2, and adjusted R^2 for all models.

(d) Give the final model.

(e) For your model in part (d), plot studentized residuals (or R-Student) and comment.

y	x_1	x_2	x_3	x_4	x_5	x_6
8.0	5.2	19.6	29.6	94.9	2.1	2.3
8.3	5.2	19.8	32.4	89.7	2.1	1.8
8.5	5.8	19.6	31.0	96.2	2.0	2.0
8.8	6.4	19.4	32.4	95.6	2.2	2.1
9.0	5.8	18.6	28.6	86.5	2.0	1.8
9.3	5.2	18.8	30.6	84.5	2.1	2.1
9.3	5.6	20.4	32.4	88.8	2.2	1.9
9.5	6.0	19.0	32.6	85.7	2.1	1.9
9.8	5.2	20.8	32.2	93.6	2.3	2.1
10.0	5.8	19.9	31.8	86.0	2.1	1.8
10.3	6.4	18.0	32.6	87.1	2.0	1.6
10.5	6.0	20.6	33.4	93.1	2.1	2.1
10.8	6.2	20.2	31.8	83.4	2.2	2.1
11.0	6.2	20.2	32.4	94.5	2.1	1.9
11.3	6.2	19.2	31.4	83.4	1.9	1.8
11.5	5.6	17.0	33.2	85.2	2.1	2.1
11.8	6.0	19.8	35.4	84.1	2.0	1.8
12.3	5.8	18.8	34.0	86.9	2.1	1.8
12.5	5.6	18.6	34.2	83.0	1.9	2.0

12.58 For Exercise 12.57, test H_0: $\beta_1 = \beta_6 = 0$. Give P-values and comment.

12.59 In Exercise 12.28, page 462, we have the following data concerning wear of a bearing:

y (wear)	**x_1 (oil viscosity)**	**x_2 (load)**
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

(a) The following model may be considered to describe the data:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i,$$

for $i = 1, 2, \ldots, 6$. The $x_1 x_2$ is an "interaction" term. Fit this model and estimate the parameters.

(b) Use the models (x_1), (x_1, x_2), (x_2), (x_1, x_2, x_1x_2) and compute PRESS, C_p, and s^2 to determine the "best" model.

12.12 Special Nonlinear Models for Nonideal Conditions

In much of the preceding material in this chapter and in Chapter 11, we have benefited substantially from the assumption that the model errors, the ϵ_i, are normal with mean 0 and constant variance σ^2. However, there are many real-life situations in which the response is clearly nonnormal. For example, a wealth of applications exist where the **response is binary** (0 or 1) and hence Bernoulli in nature. In the social sciences, the problem may be to develop a model to predict whether or not an individual is a good credit risk (0 or 1) as a function of certain socioeconomic regressors such as income, age, gender, and level of education. In a biomedical drug trial, the response is often whether or not the patient responds positively to a drug while regressors may include drug dosage as well as biological factors such as age, weight, and blood pressure. Again the response is binary in nature. Applications are also abundant in manufacturing areas where certain controllable factors influence whether a manufactured item is **defective or not**.

A second type of nonnormal application on which we will touch briefly has to do with **count data**. Here the assumption of a Poisson response is often convenient. In biomedical applications, the number of cancer cell colonies may be the response which is modeled against drug dosages. In the textile industry, the number of imperfections per yard of cloth may be a reasonable response which is modeled against certain process variables.

Nonhomogeneous Variance

The reader should note the comparison of the ideal (i.e., the normal response) situation with that of the Bernoulli (or binomial) or the Poisson response. We have become accustomed to the fact that the normal case is very special in that the variance is **independent of the mean**. Clearly this is not the case for either Bernoulli or Poisson responses. For example, if the response is 0 or l, suggesting a Bernoulli response, then the model is of the form

$$p = f(\mathbf{x}, \beta),$$

where p is the **probability of a success** (say response = 1). The parameter p plays the role of $\mu_{Y|x}$ in the normal case. However, the Bernoulli variance is $p(1 - p)$, which, of course, is also a function of the regressor $\mathbf{x}$. As a result, the variance is not constant. This rules out the use of standard least squares, which we have utilized in our linear regression work up to this point. The same is true for the Poisson case since the model is of the form

$$\lambda = f(\mathbf{x}, \beta),$$

with $\text{Var}(y) = \mu_y = \lambda$, which varies with $\mathbf{x}$.

Binary Response (Logistic Regression)

The most popular approach to modeling binary responses is a technique entitled **logistic regression**. It is used extensively in the biological sciences, biomedical research, and engineering. Indeed, even in the social sciences binary responses are found to be plentiful. The basic distribution for the response is either Bernoulli or binomial. The former is found in observational studies where there are no repeated runs at each regressor level, while the latter will be the case when an experiment is designed. For example, in a clinical trial in which a new drug is being evaluated, the goal might be to determine the dose of the drug that provides efficacy. So certain doses will be employed in the experiment, and more than one subject will be used for each dose. This case is called the **grouped case**.

What Is the Model for Logistic Regression?

In the case of binary responses ,the mean response is a probability. In the preceding clinical trial illustration, we might say that we wish to estimate the probability that the patient responds properly to the drug, P(success). Thus, the model is written in terms of a probability. Given regressors $\mathbf{x}$, the logistic function is given by

$$p = \frac{1}{1 + e^{-\mathbf{x}'\beta}}.$$

The portion $\mathbf{x}'\beta$ is called the **linear predictor**, and in the case of a single regressor x it might be written $\mathbf{x}'\beta = \beta_0 + \beta_1 x$. Of course, we do not rule out involving multiple regressors and polynomial terms in the so-called linear predictor. In the grouped case, the model involves modeling the mean of a binomial rather than a Bernoulli, and thus we have the mean given by

$$np = \frac{n}{1 + e^{-\mathbf{x}'\beta}}.$$

Characteristics of Logistic Function

A plot of the logistic function reveals a great deal about its characteristics and why it is utilized for this type of problem. First, the function is nonlinear. In addition, the plot in Figure 12.8 reveals the S-shape with the function approaching $p = 1.0$ as an asymptote. In this case, $\beta_1 > 0$. Thus, we would never experience an estimated probability exceeding 1.0.

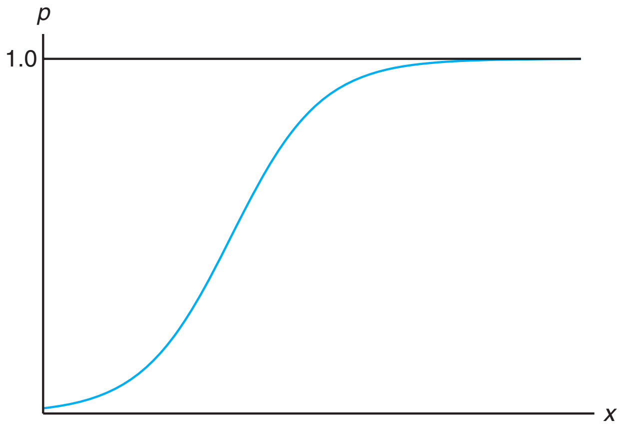


Figure 12.8: The logistic function.

The regression coefficients in the linear predictor can be estimated by the method of maximum likelihood, as described in Chapter 9. The solution to the likelihood equations involves an iterative methodology that will not be described here. However, we will present an example and discuss the computer printout and conclusions.

Example 12.13: The data set in Table 12.16 will be used to illustrate the use of logistic regression to analyze a single-agent quantal bioassay of a toxicity experiment. The results show the effect of different doses of nicotine on the common fruit fly.

Table 12.16: Data Set for Example 12.13

x Concentration (grams/100 cc)	n_i Number of Insects	y Number Killed	Percent Killed
0.10	47	8	17.0
0.15	53	14	26.4
0.20	55	24	43.6
0.30	52	32	61.5
0.50	46	38	82.6
0.70	54	50	92.6
0.95	52	50	96.2

The purpose of the experiment was to arrive at an appropriate model relating probability of "kill" to concentration. In addition, the analyst sought the so-called **effective dose** (ED), that is, the concentration of nicotine that results in a certain probability. Of particular interest was the ED_{50}, the concentration that produces a 0.5 probability of "insect kill."

This example is grouped, and thus the model is given by

$$E(Y_i) = n_i p_i = \frac{n_i}{1 + e^{-(\beta_0 + \beta_1 x_i)}}.$$

Estimates of β_0 and β_1 and their standard errors are found by the method of maximum likelihood. Tests on individual coefficients are found using χ^2-statistics rather than t-statistics since there is no common variance σ^2. The χ^2-statistic is derived from $\left(\frac{\text{coeff}}{\text{standard error}}\right)^2$.

Thus, we have the following from a *SAS* PROC LOGIST printout.

	Analysis of Parameter Estimates				
	df	**Estimate**	**Standard Error**	**Chi-Squared**	***P*-Value**
β_0	1	-1.7361	0.2420	51.4482	< 0.0001
β_1	1	6.2954	0.7422	71.9399	< 0.0001

Both coefficients are significantly different from zero. Thus, the fitted model used to predict the probability of "kill" is given by

$$\hat{p} = \frac{1}{1 + e^{-(-1.7361 + 6.2954x)}}.$$

Estimate of Effective Dose

The estimate of ED_{50} for Example 12.13 is found very simply from the estimates b_0 for β_0 and b_1 for β_1. From the logistic function, we see that

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x.$$

As a result, for $p = 0.5$, an estimate of x is found from

$$b_0 + b_1 x = 0.$$

Thus, ED_{50} is given by

$$x = -\left(\frac{b_0}{b_1}\right) = 0.276 \text{ gram}/100 \text{ cc}.$$

Concept of Odds Ratio

Another form of inference that is conveniently accomplished using logistic regression is derived from the use of the odds ratio. The odds ratio is designed to determine how the **odds of success**, $\frac{p}{1-p}$, increases as certain changes in regressor values occur. For example, in the case of Example 12.13 we may wish to know how the odds would increase if one were to increase dosage by, say, 0.2 gram/100 cc.

Definition 12.1: In logistic regression, an **odds ratio** is the ratio of odds of success at condition 2 to that of condition 1 in the regressors, that is,

$$\frac{[p/(1-p)]_2}{[p/(1-p)]_1}.$$

This allows the analyst to ascertain a sense of the utility of changing the regressor by a certain number of units. Now, since $\left(\frac{p}{1-p}\right) = e^{\beta_0+\beta_1 x}$, for Example 12.13, the ratio reflecting the increase in odds of success when the dosage of nicotine is increased by 0.2 gram/100 cc is given by

$$e^{0.2b_1} = e^{(0.2)(6.2954)} = 3.522.$$

The implication of an odds ratio of 3.522 is that the odds of success is enhanced by a factor of 3.522 when the nicotine dose is increased by 0.2 gram/100 cc.

Exercises

12.60 From a set of streptonignic dose-response data, an experimenter desires to develop a relationship between the proportion of lymphoblasts sampled that contain aberrations and the dosage of streptonignic. Five dosage levels were applied to the rabbits used for the experiment. The data are as follows (see Myers, 1990, in the Bibliography):

Dose (mg/kg)	Number of Lymphoblasts	Number with Aberrations
0	600	15
30	500	96
60	600	187
75	300	100
90	300	145

(a) Fit a logistic regression to the data set and thus estimate β_0 and β_1 in the model

$$p = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}},$$

where n is the number of lymphoblasts, x is the dose, and p is the probability of an aberration.

(b) Show results of χ^2-tests revealing the significance of the regression coefficients β_0 and β_1.

(c) Estimate ED_{50} and give an interpretation.

12.61 In an experiment to ascertain the effect of load, x, in lb/inches^2, on the probability of failure of specimens of a certain fabric type, an experiment was conducted in which numbers of specimens were exposed to loads ranging from 5 lb/in.^{2} to 90 lb/in.^{2}. The numbers of "failures" were observed. The data are as follows:

Load	**Number of Specimens**	**Number of Failures**
5	600	13
35	500	95
70	600	189
80	300	95
90	300	130

(a) Use logistic regression to fit the model

$$p = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}},$$

where p is the probability of failure and x is load.

(b) Use the odds ratio concept to determine the increase in odds of failure that results by increasing the load from 20 lb/in.^{2}.

Review Exercises

12.62 In the Department of Fisheries and Wildlife at Virginia Tech, an experiment was conducted to study the effect of stream characteristics on fish biomass. The regressor variables are as follows: average depth (of 50 cells), x_1; area of in-stream cover (i.e., undercut banks, logs, boulders, etc.), x_2; percent canopy cover (average of 12), x_3; and area $\geq$ 25 centimeters in depth, x_4. The response is y, the fish biomass. The data are as follows:

Obs.	y	x_1	x_2	x_3	x_4
1	100	14.3	15.0	12.2	48.0
2	388	19.1	29.4	26.0	152.2
3	755	54.6	58.0	24.2	469.7
4	1288	28.8	42.6	26.1	485.9
5	230	16.1	15.9	31.6	87.6
6	0	10.0	56.4	23.3	6.9
7	551	28.5	95.1	13.0	192.9
8	345	13.8	60.6	7.5	105.8
9	0	10.7	35.2	40.3	0.0
10	348	25.9	52.0	40.3	116.6

(a) Fit a multiple linear regression including all four regression variables.

(b) Use C_p, R^2, and s^2 to determine the best subset of variables. Compute these statistics for all possible subsets.

(c) Compare the appropriateness of the models in parts (a) and (b) for predicting fish biomass.

12.63 Show that, in a multiple linear regression data set,

$$\sum_{i=1}^{n} h_{ii} = p.$$

12.64 A small experiment was conducted to fit a multiple regression equation relating the yield y to temperature x_1, reaction time x_2, and concentration of one of the reactants x_3. Two levels of each variable were chosen, and measurements corresponding to the coded independent variables were recorded as follows:

y	x_1	x_2	x_3
7.6	−1	−1	−1
5.5	1	−1	−1
9.2	−1	1	−1
10.3	−1	−1	1
11.6	1	1	−1
11.1	1	−1	1
10.2	−1	1	1
14.0	1	1	1

(a) Using the coded variables, estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2,x_3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

(b) Partition *SSR*, the regression sum of squares, into three single-degree-of-freedom components attributable to x_1, x_2, and x_3, respectively. Show an analysis-of-variance table, indicating significance tests on each variable. Comment on the results.

12.65 In a chemical engineering experiment dealing with heat transfer in a shallow fluidized bed, data are collected on the following four regressor variables: fluidizing gas flow rate, lb/hr (x_1); supernatant gas flow rate, lb/hr (x_2); supernatant gas inlet nozzle opening, millimeters (x_3); and supernatant gas inlet temperature, °F (x_4). The responses measured are heat transfer efficiency (y_1) and thermal efficiency (y_2). The data are as follows:

Obs.	y_1	y_2	x_1	x_2	x_3	x_4
1	41.852	38.75	69.69	170.83	45	219.74
2	155.329	51.87	113.46	230.06	25	181.22
3	99.628	53.79	113.54	228.19	65	179.06
4	49.409	53.84	118.75	117.73	65	281.30
5	72.958	49.17	119.72	117.69	25	282.20
6	107.702	47.61	168.38	173.46	45	216.14
7	97.239	64.19	169.85	169.85	45	223.88
8	105.856	52.73	169.85	170.86	45	222.80
9	99.348	51.00	170.89	173.92	80	218.84
10	111.907	47.37	171.31	173.34	25	218.12
11	100.008	43.18	171.43	171.43	45	219.20
12	175.380	71.23	171.59	263.49	45	168.62
13	117.800	49.30	171.63	171.63	45	217.58
14	217.409	50.87	171.93	170.91	10	219.92
15	41.725	54.44	173.92	71.73	45	296.60
16	151.139	47.93	221.44	217.39	65	189.14
17	220.630	42.91	222.74	221.73	25	186.08
18	131.666	66.60	228.90	114.40	25	285.80
19	80.537	64.94	231.19	113.52	65	286.34
20	152.966	43.18	236.84	167.77	45	221.72

Consider the model for predicting the heat transfer coefficient response

$$y_{1i} = \beta_0 + \sum_{j=1}^{4}\beta_j x_{ji} + \sum_{i=1}^{4}\beta_{jj}x_{ji}^2 + \sum\sum_{j\neq l}\beta_{jl}x_{ji}x_{li} + \epsilon_i, \quad i = 1, 2, \ldots, 20.$$

(a) Compute PRESS and $\sum_{i=1}^{n}|y_i - \hat{y}_{i,-i}|$ for the least squares regression fit to the model above.

(b) Fit a second-order model with x_4 completely eliminated (i.e., deleting all terms involving x_4). Compute the prediction criteria for the reduced model. Comment on the appropriateness of x_4 for prediction of the heat transfer coefficient.

(c) Repeat parts (a) and (b) for thermal efficiency.

12.66 In exercise physiology, an objective measure of aerobic fitness is the oxygen consumption in volume per unit body weight per unit time. Thirty-one individuals were used in an experiment in order to be able to model oxygen consumption against age in years (x_1), weight in kilograms (x_2), time to run $1\frac{1}{2}$ miles (x_3), resting pulse rate (x_4), pulse rate at the end of run (x_5), and maximum pulse rate during run (x_6).

(a) Do a stepwise regression with input significance level 0.25. Quote the final model.

(b) Do all possible subsets using s^2, C_p, R^2, and R^2_{adj}. Make a decision and quote the final model.

ID	y	x_1	x_2	x_3	x_4	x_5	x_6
1	44.609	44	89.47	11.37	62	178	182
2	45.313	40	75.07	10.07	62	185	185
3	54.297	44	85.84	8.65	45	156	168
4	59.571	42	68.15	8.17	40	166	172
5	49.874	38	89.02	9.22	55	178	180
6	44.811	47	77.45	11.63	58	176	176
7	45.681	40	75.98	11.95	70	176	180
8	49.091	43	81.19	10.85	64	162	170
9	39.442	44	81.42	13.08	63	174	176
10	60.055	38	81.87	8.63	48	170	186
11	50.541	44	73.03	10.13	45	168	168
12	37.388	45	87.66	14.03	56	186	192
13	44.754	45	66.45	11.12	51	176	176
14	47.273	47	79.15	10.60	47	162	164
15	51.855	54	83.12	10.33	50	166	170
16	49.156	49	81.42	8.95	44	180	185
17	40.836	51	69.63	10.95	57	168	172
18	46.672	51	77.91	10.00	48	162	168
19	46.774	48	91.63	10.25	48	162	164
20	50.388	49	73.37	10.08	76	168	168
21	39.407	57	73.37	12.63	58	174	176
22	46.080	54	79.38	11.17	62	156	165
23	45.441	52	76.32	9.63	48	164	166
24	54.625	50	70.87	8.92	48	146	155
25	45.118	51	67.25	11.08	48	172	172
26	39.203	54	91.63	12.88	44	168	172
27	45.790	51	73.71	10.47	59	186	188
28	50.545	57	59.08	9.93	49	148	155
29	48.673	49	76.32	9.40	56	186	188
30	47.920	48	61.24	11.50	52	170	176
31	47.467	52	82.78	10.50	53	170	172

12.67 Consider the data of Review Exercise 12.64. Suppose it is of interest to add some "interaction" terms. Namely, consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + \beta_{123} x_{1i} x_{2i} x_{3i} + \epsilon_i.$$

(a) Do we still have orthogonality? Comment.

(b) With the fitted model in part (a), can you find prediction intervals and confidence intervals on the mean response? Why or why not?

(c) Consider a model with $\beta_{123}x_1x_2x_3$ removed. To determine if interactions (as a whole) are needed, test

$$H_0\colon\ \beta_{12} = \beta_{13} = \beta_{23} = 0.$$

Give the P-value and conclusions.

12.68 A carbon dioxide (CO_2) flooding technique is used to extract crude oil. The CO_2 floods oil pockets and displaces the crude oil. In an experiment, flow tubes are dipped into sample oil pockets containing a known amount of oil. Using three different values of flow pressure and three different values of dipping angles, the oil pockets are flooded with CO_2, and the percentage of oil displaced recorded. Consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \epsilon_i.$$

Fit the model above to the data, and suggest any model editing that may be needed.

Pressure (lb/in^2), x_1	**Dipping Angle, x_2**	**Oil Recovery (%), y**
1000	0	60.58
1000	15	72.72
1000	30	79.99
1500	0	66.83
1500	15	80.78
1500	30	89.78
2000	0	69.18
2000	15	80.31
2000	30	91.99

Source: Wang, G. C. "Microscopic Investigations of CO_2 Flooding Process," *Journal of Petroleum Technology*, Vol. 34, No. 8, Aug. 1982.

12.69 An article in the *Journal of Pharmaceutical Sciences* (Vol. 80, 1991) presents data on the mole fraction solubility of a solute at a constant temperature. Also measured are the dispersion x_1 and dipolar and hydrogen bonding solubility parameters x_2 and x_3. A portion of the data is shown in the table below. In the model, y is the negative logarithm of the mole fraction. Fit the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i,$$

for $i = 1, 2, \ldots, 20$.

Obs.	**y**	**x_1**	**x_2**	**x_3**
1	0.2220	7.3	0.0	0.0
2	0.3950	8.7	0.0	0.3
3	0.4220	8.8	0.7	1.0
4	0.4370	8.1	4.0	0.2
5	0.4280	9.0	0.5	1.0
6	0.4670	8.7	1.5	2.8
7	0.4440	9.3	2.1	1.0
8	0.3780	7.6	5.1	3.4
9	0.4940	10.0	0.0	0.3
10	0.4560	8.4	3.7	4.1
11	0.4520	9.3	3.6	2.0
12	0.1120	7.7	2.8	7.1
13	0.4320	9.8	4.2	2.0
14	0.1010	7.3	2.5	6.8
15	0.2320	8.5	2.0	6.6
16	0.3060	9.5	2.5	5.0
17	0.0923	7.4	2.8	7.8
18	0.1160	7.8	2.8	7.7
19	0.0764	7.7	3.0	8.0
20	0.4390	10.3	1.7	4.2

(a) Test H_0: $\beta_1 = \beta_2 = \beta_3 = 0$.

(b) Plot studentized residuals against x_1, x_2, and x_3 (three plots). Comment.

(c) Consider two additional models that are competitors to the models above:

Model 2: Add x_1^2, x_2^2, x_3^2.

Model 3: Add $x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$.

Use PRESS and C_p with these three models to arrive at the best among the three.

12.70 A study was conducted to determine whether lifestyle changes could replace medication in reducing blood pressure among hypertensives. The factors considered were a healthy diet with an exercise program, the typical dosage of medication for hypertension, and no intervention. The pretreatment body mass index (BMI) was also calculated because it is known to affect blood pressure. The response considered in this study was change in blood pressure. The variable "group" had the following levels.

1 = Healthy diet and an exercise program
2 = Medication
3 = No intervention

(a) Fit an appropriate model using the data below. Does it appear that exercise and diet could be effectively used to lower blood pressure? Explain your answer from the results.

(b) Would exercise and diet be an effective alternative to medication?

(*Hint*: You may wish to form the model in more than one way to answer both of these questions.)

Change in Blood Pressure	**Group**	**BMI**
−32	1	27.3
−21	1	22.1
−26	1	26.1
−16	1	27.8
−11	2	19.2
−19	2	26.1
−23	2	28.6
−5	2	23.0
−6	3	28.1
5	3	25.3
−11	3	26.7
14	3	22.3

12.71 Show that in choosing the so-called best subset model from a series of candidate models, choosing the model with the smallest s^2 is equivalent to choosing the model with the smallest R^2_{adj}.

12.72 Case Study: Consider the data set for Exercise 12.12, page 452 (hospital data), repeated here.

Site	x_1	x_2	x_3	x_4	x_5	y
1	15.57	2463	472.92	18.0	4.45	566.52
2	44.02	2048	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	6505	568.33	36.7	3.90	1003.62
5	49.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11,520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.l5	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20,106	3655.08	180.5	6.15	3503.93
11	96.00	13,313	2912.00	60.9	5.88	3571.59
12	131.42	10,771	3921.00	103.7	4.88	3741.40
13	127.21	15,543	3865.67	126.8	5.50	4026.52
14	252.90	36,194	7684.10	157.7	7.00	10,343.81
15	409.20	34,703	12,446.33	169.4	10.75	11,732.17
16	463.70	39,204	14,098.40	331.4	7.05	15,414.94
17	510.22	86,533	15,524.00	371.6	6.35	18,854.45

(a) The *SAS* PROC REG outputs provided in Figures 12.9 and 12.10 supply a considerable amount of information. Goals are to do outlier detection and eventually determine which model terms are to be used in the final model.

(b) Often the role of a single regressor variable is not apparent when it is studied in the presence of several other variables. This is due to multicollinearity. With this in mind, comment on the importance of x_2 and x_3 in the full model as opposed to their importance in a model in which they are the only variables.

(c) Comment on what other analyses should be run.

(d) Run appropriate analyses and write your conclusions concerning the final model.

```
Dependent Variable: y
                     Analysis of Variance
                                  Sum of           Mean
Source                DF         Squares         Square      F Value     Pr > F
Model                  5       490177488       98035498       237.79     <.0001
Error                 11         4535052         412277
Corrected Total       16       494712540
             Root MSE            642.08838    R-Square     0.9908
             Dependent Mean     4978.48000    Adj R-Sq     0.9867
             Coeff Var            12.89728
                                  Parameter Estimates
                                            Parameter  Standard
Variable  Label                         DF   Estimate       Error t Value  Pr > |t|
Intercept Intercept                      1  1962.94816 1071.36170    1.83    0.0941
x1        Average Daily Patient Load     1   -15.85167   97.65299   -0.16    0.8740
x2        Monthly X-Ray Exposure         1     0.05593    0.02126    2.63    0.0234
x3        Monthly Occupied Bed Days      1     1.58962    3.09208    0.51    0.6174
x4        Eligible Population in the     1    -4.21867    7.17656   -0.59    0.5685
          Area/100
x5        Average Length of Patients     1  -394.31412 209.63954   -1.88    0.0867
          Stay in Days
```

Figure 12.9: *SAS* output for Review Exercise 12.72; part I.

```
     Dependent  Predicted     Std Error
Obs   Variable      Value  Mean Predict    95% CL Mean      95% CL Predict
  1   566.5200   775.0251      241.2323  244.0765    1306  -734.6494   2285
  2   696.8200   740.6702      331.1402   11.8355    1470  -849.4275   2331
  3       1033       1104      278.5116  490.9234    1717  -436.5244   2644
  4       1604       1240      268.1298  650.3459    1831  -291.0028   2772
  5       1611       1564      211.2372      1099    2029    76.6816   3052
  6       1613       2151      279.9293      1535    2767   609.5796   3693
  7       1854       1690      218.9976      1208    2172   196.5345   3183
  8       2161       1736      468.9903  703.9948    2768   -13.8306   3486
  9       2306       2737      290.4749      2098    3376       1186   4288
 10       3504       3682      585.2517      2394    4970       1770   5594
 11       3572       3239      189.0989      2823    3655       1766   4713
 12       3741       4353      328.8507      3630    5077       2766   5941
 13       4027       4257      314.0481      3566    4948       2684   5830
 14      10344       8768      252.2617      8213    9323       7249  10286
 15      11732      12237      573.9168     10974   13500      10342  14133
 16      15415      15038      585.7046     13749   16328      13126  16951
 17      18854      19321      599.9780     18000   20641      17387  21255

                        Std Error     Student
      Obs   Residual     Residual    Residual      -2-1 0 1 2
        1  -208.5051        595.0      -0.350    |      |      |
        2   -43.8502        550.1     -0.0797    |      |      |
        3   -70.7734        578.5      -0.122    |      |      |
        4   363.1244        583.4       0.622    |      |*     |
        5    46.9483        606.3      0.0774    |      |      |
        6  -538.0017        577.9      -0.931    |     *|      |
        7   164.4696        603.6       0.272    |      |      |
        8   424.3145        438.5       0.968    |      |*     |
        9  -431.4090        572.6      -0.753    |     *|      |
       10  -177.9234        264.1      -0.674    |     *|      |
       11   332.6011        613.6       0.542    |      |*     |
       12  -611.9330        551.5      -1.110    |    **|      |
       13  -230.5684        560.0      -0.412    |      |      |
       14       1576        590.5       2.669    |      |***** |
       15  -504.8574        287.9      -1.753    |   ***|      |
       16   376.5491        263.1       1.431    |      |**    |
       17  -466.2470        228.7      -2.039    |  ****|      |
```

Figure 12.10: *SAS* output for Review Exercise 12.72; part II.

12.13 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

There are several procedures discussed in this chapter for use in the "attempt" to find the best model. However, one of the most important misconceptions under which naïve scientists or engineers labor is that there is a **true linear model** and that it can be found. In most scientific phenomena, relationships between scientific variables are nonlinear in nature and the true model is unknown. Linear statistical models are **empirical approximations**.

At times, the choice of the model to be adopted may depend on what information needs to be derived from the model. Is it to be used for prediction? Is it to be used for the purpose of explaining the role of each regressor? This "choice" can be made difficult in the presence of collinearity. It is true that for many regression problems there are multiple models that are very similar in performance. See the Myers reference (1990) for details.

One of the most damaging misuses of the material in this chapter is to assign too much importance to R^2 in the choice of the so-called best model. It is important to remember that for any data set, one can obtain an R^2 as large as one desires, within the constraint $0 \leq R^2 \leq 1$. **Too much attention to $\boldsymbol{R^2}$ often leads to overfitting**.

Much attention was given in this chapter to outlier detection. A classical serious misuse of statistics centers around the decision made concerning the detection of outliers. We hope it is clear that the analyst should absolutely not carry out the exercise of detecting outliers, eliminating them from the data set, fitting a new model, reporting outlier detection, and so on. This is a tempting and disastrous procedure for arriving at a model that fits the data well, with the result being an example of **how to lie with statistics**. If an outlier is detected, the history of the data should be checked for possible clerical or procedural error before it is eliminated from the data set. One must remember that an outlier by definition is a data point that the model did not fit well. The problem may not be in the data but rather in the model selection. A changed model may result in the point not being detected as an outlier.

There are many types of responses that occur naturally in practice but can't be used in an analysis of standard least squares because classic least squares assumptions do not hold. The assumptions that often fail are those of normal errors and homogeneous variance. For example, if the response is a proportion, say proportion defective, the response distribution is related to the binomial distribution. A second response that occurs often in practice is that of Poisson counts. Clearly the distribution is not normal, and the response variance, which is equal to the Poisson mean, will vary from observation to observation. For more details on these nonideal conditions, see Myers et al. (2008) in the Bibliography.

Chapter 13

One-Factor Experiments: General

13.1 Analysis-of-Variance Technique

In the estimation and hypothesis testing material covered in Chapters 9 and 10, we were restricted in each case to considering no more than two population parameters. Such was the case, for example, in testing for the equality of two population means using independent samples from normal populations with common but unknown variance, where it was necessary to obtain a pooled estimate of σ^2.

This material dealing in two-sample inference represents a special case of what we call the *one-factor problem.* For example, in Exercise 10.35 on page 357, the survival time was measured for two samples of mice, where one sample received a new serum for leukemia treatment and the other sample received no treatment. In this case, we say that there is *one factor*, namely *treatment*, and the factor is at *two levels.* If several competing treatments were being used in the sampling process, more samples of mice would be necessary. In this case, the problem would involve one factor with more than two levels and thus more than two samples.

In the $k > 2$ sample problem, it will be assumed that there are k samples from k populations. One very common procedure used to deal with testing population means is called the **analysis of variance**, or **ANOVA**.

The analysis of variance is certainly not a new technique to the reader who has followed the material on regression theory. We used the analysis-of-variance approach to partition the total sum of squares into a portion due to regression and a portion due to error.

Suppose in an industrial experiment that an engineer is interested in how the mean absorption of moisture in concrete varies among 5 different concrete aggregates. The samples are exposed to moisture for 48 hours. It is decided that 6 samples are to be tested for each aggregate, requiring a total of 30 samples to be tested. The data are recorded in Table 13.1.

The model for this situation may be set up as follows. There are 6 observations taken from each of 5 populations with means $\mu_1, \mu_2, \ldots, \mu_5$, respectively. We may wish to test

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$

$$H_1\colon\ \text{At least two of the means are not equal.}$$

Table 13.1: Absorption of Moisture in Concrete Aggregates

Aggregate:	**1**	**2**	**3**	**4**	**5**	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	16,854
Mean	553.33	569.33	610.50	465.17	610.67	561.80

In addition, we may be interested in making individual comparisons among these 5 population means.

Two Sources of Variability in the Data

In the analysis-of-variance procedure, it is assumed that whatever variation exists among the aggregate averages is attributed to (1) variation in absorption among observations *within* aggregate types and (2) variation *among* aggregate types, that is, due to differences in the chemical composition of the aggregates. The **within-aggregate variation** is, of course, brought about by various causes. Perhaps humidity and temperature conditions were not kept entirely constant throughout the experiment. It is possible that there was a certain amount of heterogeneity in the batches of raw materials that were used. At any rate, we shall consider the within-sample variation to be **chance or random variation**. Part of the goal of the analysis of variance is to determine if the differences among the 5 sample means are what we would expect due to random variation alone or, rather, due to variation beyond merely random effects, i.e., differences in the chemical composition of the aggregates.

Many pointed questions appear at this stage concerning the preceding problem. For example, how many samples must be tested for each aggregate? This is a question that continually haunts the practitioner. In addition, what if the within-sample variation is so large that it is difficult for a statistical procedure to detect the systematic differences? Can we systematically control extraneous sources of variation and thus remove them from the portion we call random variation? We shall attempt to answer these and other questions in the following sections.

13.2 The Strategy of Experimental Design

In Chapters 9 and 10, the notions of estimation and testing for the two-sample case were covered under the important backdrop of the way the experiment is conducted. This falls into the broad category of design of experiments. For example, for the **pooled *t*-test** discussed in Chapter 10, it is assumed that the factor levels (treatments in the mice example) are assigned randomly to the experimental units (mice). The notion of experimental units was discussed in Chapters 9 and 10 and illustrated through examples. Simply put, experimental units are the units (mice, patients, concrete specimens, time) that **provide the heterogeneity that leads to experimental error** in a scientific investigation. The random assignment eliminates bias that could result with systematic assignment. The goal is to distribute uniformly among the factor levels the risks brought about by the heterogeneity of the experimental units. Random assignment best simulates the conditions that are assumed by the model. In Section 13.7, we discuss **blocking** in experiments. The notion of blocking was presented in Chapters 9 and 10, when comparisons between means were accomplished with **pairing**, that is, the division of the experimental units into homogeneous pairs called **blocks**. The factor levels or treatments are then assigned randomly within blocks. The purpose of blocking is to reduce the effective experimental error. In this chapter, we naturally extend the pairing to larger block sizes, with analysis of variance being the primary analytical tool.

13.3 One-Way Analysis of Variance: Completely Randomized Design (One-Way ANOVA)

Random samples of size n are selected from each of k populations. The k different populations are classified on the basis of a single criterion such as different treatments or groups. Today the term **treatment** is used generally to refer to the various classifications, whether they be different aggregates, different analysts, different fertilizers, or different regions of the country.

Assumptions and Hypotheses in One-Way ANOVA

It is assumed that the k populations are independent and normally distributed with means $\mu_1, \mu_2, \ldots, \mu_k$ and common variance σ^2. As indicated in Section 13.2, these assumptions are made more palatable by randomization. We wish to derive appropriate methods for testing the hypothesis

$$\begin{aligned} &H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k,\\ &H_1\colon\ \text{At least two of the means are not equal.}\end{aligned}$$

Let y_{ij} denote the jth observation from the ith treatment and arrange the data as in Table 13.2. Here, $Y_{i.}$ is the total of all observations in the sample from the ith treatment, $\bar{y}_{i.}$ is the mean of all observations in the sample from the ith treatment, $Y_{..}$ is the total of all nk observations, and $\bar{y}_{..}$ is the mean of all nk observations.

Model for One-Way ANOVA

Each observation may be written in the form

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where ϵ_{ij} measures the deviation of the jth observation of the ith sample from the corresponding treatment mean. The ϵ_{ij}-term represents random error and plays the same role as the error terms in the regression models. An alternative and

Table 13.2: k Random Samples

Treatment:	**1**	**2**	$\cdots$	i	$\cdots$	k	
	y_{11}	y_{21}	$\cdots$	y_{i1}	$\cdots$	y_{k1}	
	y_{12}	y_{22}	$\cdots$	y_{i2}	$\cdots$	y_{k2}	
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
	y_{1n}	y_{2n}	$\cdots$	y_{in}	$\cdots$	y_{kn}	
Total	$Y_{1.}$	$Y_{2.}$	$\cdots$	$Y_{i.}$	$\cdots$	$Y_{k.}$	$Y_{..}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{i.}$	$\cdots$	$\bar{y}_{k.}$	$\bar{y}_{..}$

preferred form of this equation is obtained by substituting $\mu_i = \mu + \alpha_i$, subject to the constraint $\sum_{i=1}^{k} \alpha_i = 0$. Hence, we may write

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where μ is just the **grand mean** of all the μ_i, that is,

$$\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i,$$

and α_i is called the **effect** of the ith treatment.

The null hypothesis that the k population means are equal against the alternative that at least two of the means are unequal may now be replaced by the equivalent hypothesis

$$H_0\colon\ \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0,$$
$$H_1\colon\ \text{At least one of the } \alpha_i \text{ is not equal to zero.}$$

Resolution of Total Variability into Components

Our test will be based on a comparison of two independent estimates of the common population variance σ^2. These estimates will be obtained by partitioning the total variability of our data, designated by the double summation

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

into two components.

Theorem 13.1: **Sum-of-Squares Identity**

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

It will be convenient in what follows to identify the terms of the sum-of-squares identity by the following notation:

Three Important Measures of Variability

$$\begin{aligned} SST &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \text{total sum of squares}, \\ SSA &= n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = \text{treatment sum of squares}, \\ SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = \text{error sum of squares}. \end{aligned}$$

The sum-of-squares identity can then be represented symbolically by the equation

$$SST = SSA + SSE.$$

The identity above expresses how between-treatment and within-treatment variation add to the total sum of squares. However, much insight can be gained by investigating the **expected value of both *SSA* and *SSE***. Eventually, we shall develop variance estimates that formulate the ratio to be used to test the equality of population means.

Theorem 13.2:

$$\boxed{E(SSA) = (k-1)\sigma^2 + n\sum_{i=1}^{k}\alpha_i^2}$$

The proof of the theorem is left as an exercise (see Review Exercise 13.53 on page 556).

If H_0 is true, an estimate of σ^2, based on $k-1$ degrees of freedom, is provided by this expression:

Treatment Mean Square

$$s_1^2 = \frac{SSA}{k-1}$$

If H_0 is true and thus each α_i in Theorem 13.2 is equal to zero, we see that

$$E\left(\frac{SSA}{k-1}\right) = \sigma^2,$$

and s_1^2 is an unbiased estimate of σ^2. However, if H_1 is true, we have

$$E\left(\frac{SSA}{k-1}\right) = \sigma^2 + \frac{n}{k-1}\sum_{i=1}^{k}\alpha_i^2,$$

and s_1^2 estimates σ^2 plus an additional term, which measures variation due to the systematic effects.

A second and independent estimate of σ^2, based on $k(n-1)$ degrees of freedom, is this familiar formula:

Error Mean Square

$$s^2 = \frac{SSE}{k(n-1)}$$

It is instructive to point out the importance of the expected values of the mean squares indicated above. In the next section, we discuss the use of an **F-ratio** with the treatment mean square residing in the numerator. It turns out that when H_1 is true, the presence of the condition $E(s_1^2) > E(s^2)$ suggests that the F-ratio be used in the context of a **one-sided upper-tailed test**. That is, when H_1 is true, we would expect the numerator s_1^2 to exceed the denominator.

Use of *F*-Test in ANOVA

The estimate s^2 is unbiased regardless of the truth or falsity of the null hypothesis (see Review Exercise 13.52 on page 556). It is important to note that the sum-of-squares identity has partitioned not only the total variability of the data, but also the total number of degrees of freedom. That is,

$$nk - 1 = k - 1 + k(n - 1).$$

F-Ratio for Testing Equality of Means

When H_0 is true, the ratio $f = s_1^2/s^2$ is a value of the random variable F having the F-distribution with $k-1$ and $k(n-1)$ degrees of freedom (see Theorem 8.8). Since s_1^2 overestimates σ^2 when H_0 is false, we have a one-tailed test with the critical region entirely in the right tail of the distribution.

The null hypothesis H_0 is rejected at the α-level of significance when

$$f > f_\alpha[k-1, k(n-1)].$$

Another approach, the P-value approach, suggests that the evidence in favor of or against H_0 is

$$P = P\{f[k-1, k(n-1)] > f\}.$$

The computations for an analysis-of-variance problem are usually summarized in tabular form, as shown in Table 13.3.

Table 13.3: Analysis of Variance for the One-Way ANOVA

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed *f***
Treatments	SSA	$k-1$	$s_1^2 = \frac{SSA}{k-1}$	$\frac{s_1^2}{s^2}$
Error	SSE	$k(n-1)$	$s^2 = \frac{SSE}{k(n-1)}$	
Total	SST	$kn-1$		

Example 13.1: Test the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_5$ at the 0.05 level of significance for the data of Table 13.1 on absorption of moisture by various types of cement aggregates.

***Solution*:** The hypotheses are

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$
$$\alpha = 0.05.$$

Critical region: $f > 2.76$ with $v_1 = 4$ and $v_2 = 25$ degrees of freedom. The sum-of-squares computations give

$$SST = 209{,}377, \quad SSA = 85{,}356,$$
$$SSE = 209{,}377 - 85{,}356 = 124{,}021.$$

These results and the remaining computations are exhibited in Figure 13.1 in the *SAS* ANOVA procedure.

```
                         The GLM Procedure
Dependent Variable: moisture

                                                Sum of
Source             DF          Squares     Mean Square    F Value     Pr > F
Model               4       85356.4667      21339.1167       4.30     0.0088
Error              25      124020.3333       4960.8133
Corrected Total    29      209376.8000

  R-Square      Coeff Var       Root MSE     moisture Mean
  0.407669       12.53703       70.43304          561.8000

Source             DF        Type I SS     Mean Square    F Value     Pr > F
aggregate           4      85356.46667     21339.11667       4.30     0.0088
```

Figure 13.1: *SAS* output for the analysis-of-variance procedure.

Decision: Reject H_0 and conclude that the aggregates do not have the same mean absorption. The P-value for $f = 4.30$ is 0.0088, which is smaller than 0.05. ◾

In addition to the ANOVA, a box plot was constructed for each aggregate. The plots are shown in Figure 13.2. From these plots it is evident that the absorption is not the same for all aggregates. In fact, it appears as if aggregate 4 stands out from the rest. A more formal analysis showing this result will appear in Exercise 13.21 on page 531.

During experimental work, one often loses some of the desired observations. Experimental animals may die, experimental material may be damaged, or human subjects may drop out of a study. The previous analysis for equal sample size will still be valid if we slightly modify the sum of squares formulas. We now assume the k random samples to be of sizes $n_1, n_2, \ldots, n_k$, respectively.

Sum of Squares, Unequal Sample Sizes

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2, \ SSA = \sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2, \ SSE = SST - SSA$$

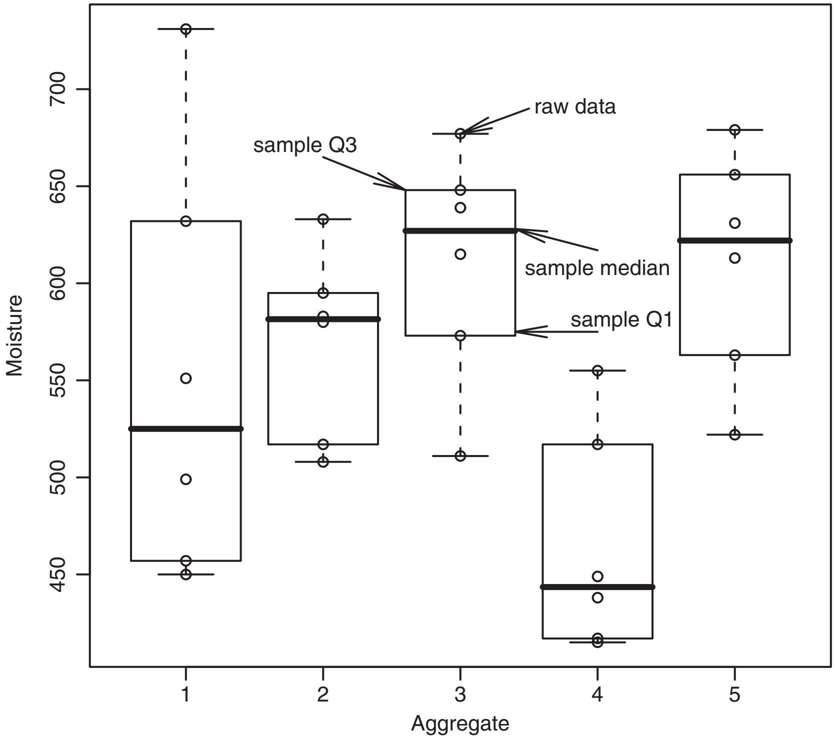


Figure 13.2: Box plots for the absorption of moisture in concrete aggregates.

The degrees of freedom are then partitioned as before: $N - 1$ for SST, $k - 1$ for SSA, and $N - 1 - (k - 1) = N - k$ for SSE, where $N = \sum_{i=1}^{k} n_i$.

Example 13.2: Part of a study conducted at Virginia Tech was designed to measure serum alkaline phosphatase activity levels (in Bessey-Lowry units) in children with seizure disorders who were receiving anticonvulsant therapy under the care of a private physician. Forty-five subjects were found for the study and categorized into four drug groups:

G-1: Control (not receiving anticonvulsants and having no history of seizure disorders)

G-2: Phenobarbital

G-3: Carbamazepine

G-4: Other anticonvulsants

From blood samples collected from each subject, the serum alkaline phosphatase activity level was determined and recorded as shown in Table 13.4. Test the hypothesis at the 0.05 level of significance that the average serum alkaline phosphatase activity level is the same for the four drug groups.

Table 13.4: Serum Alkaline Phosphatase Activity Level

G-1		**G-2**	**G-3**	**G-4**
49.20	97.50	97.07	62.10	110.60
44.54	105.00	73.40	94.95	57.10
45.80	58.05	68.50	142.50	117.60
95.84	86.60	91.85	53.00	77.71
30.10	58.35	106.60	175.00	150.00
36.50	72.80	0.57	79.50	82.90
82.30	116.70	0.79	29.50	111.50
87.85	45.15	0.77	78.40	
105.00	70.35	0.81	127.50	
95.22	77.40			

***Solution*:** With the level of significance at 0.05, the hypotheses are

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4,$$
$$H_1\text{: At least two of the means are not equal.}$$

Critical region: $f > 2.836$, from interpolating in Table A.6.
Computations: $Y_{1.} = 1460.25$, $Y_{2.} = 440.36$, $Y_{3.} = 842.45$, $Y_{4.} = 707.41$, and $Y_{..} = 3450.47$. The analysis of variance is shown in the *MINITAB* output of Figure 13.3.

```
 One-way ANOVA: G-1, G-2, G-3, G-4

 Source  DF     SS    MS     F      P
 Factor   3  13939  4646  3.57  0.022
 Error   41  53376  1302
 Total   44  67315

S = 36.08   R-Sq = 20.71%   R-Sq(adj) = 14.90%

                          Individual 95% CIs For Mean Based on
                          Pooled StDev
 Level   N    Mean  StDev  --+---------+---------+---------+-------
 G-1    20   73.01  25.75             (----*-----)
 G-2     9   48.93  47.11  (-------*-------)
 G-3     9   93.61  46.57                 (-------*-------)
 G-4     7  101.06  30.76                   (--------*--------)
                           --+---------+---------+---------+-------
                           30        60        90       120

Pooled StDev = 36.08
```

Figure 13.3: *MINITAB* analysis of data in Table 13.4.

Decision: Reject H_0 and conclude that the average serum alkaline phosphatase activity levels for the four drug groups are not all the same. The calculated P-value is 0.022. ▟

In concluding our discussion on the analysis of variance for the one-way classification, we state the advantages of choosing equal sample sizes over the choice of unequal sample sizes. The first advantage is that the f-ratio is insensitive to slight departures from the assumption of equal variances for the k populations when the samples are of equal size. Second, the choice of equal sample sizes minimizes the probability of committing a type II error.

13.4 Tests for the Equality of Several Variances

Although the f-ratio obtained from the analysis-of-variance procedure is insensitive to departures from the assumption of equal variances for the k normal populations when the samples are of equal size, we may still prefer to exercise caution and run a preliminary test for homogeneity of variances. Such a test would certainly be advisable in the case of unequal sample sizes if there was a reasonable doubt concerning the homogeneity of the population variances. Suppose, therefore, that we wish to test the null hypothesis

$$H_0\colon\ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

against the alternative

$$H_1\colon\ \text{The variances are not all equal.}$$

The test that we shall use, called **Bartlett's test**, is based on a statistic whose sampling distribution provides exact critical values when the sample sizes are equal. These critical values for equal sample sizes can also be used to yield highly accurate approximations to the critical values for unequal sample sizes.

First, we compute the k sample variances $s_1^2, s_2^2, \ldots, s_k^2$ from samples of size $n_1, n_2, \ldots, n_k$, with $\sum\limits_{i=1}^{k} n_i = N$. Second, we combine the sample variances to give the pooled estimate

$$s_p^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)s_i^2.$$

Now

$$b = \frac{[(s_1^2)^{n_1-1}(s_2^2)^{n_2-1}\cdots(s_k^2)^{n_k-1}]^{1/(N-k)}}{s_p^2}$$

is a value of a random variable B having the **Bartlett distribution**. For the special case where $n_1 = n_2 = \cdots = n_k = n$, we reject H_0 at the α-level of significance if

$$b < b_k(\alpha; n),$$

where $b_k(\alpha; n)$ is the critical value leaving an area of size α in the left tail of the Bartlett distribution. Table A.10 gives the critical values, $b_k(\alpha; n)$, for $\alpha = 0.01$ and 0.05; $k = 2, 3, \dots, 10$; and selected values of n from 3 to 100.

When the sample sizes are unequal, the null hypothesis is rejected at the α-level of significance if

$$b < b_k(\alpha; n_1, n_2, \dots, n_k),$$

where

$$b_k(\alpha; n_1, n_2, \dots, n_k) \approx \frac{n_1 b_k(\alpha; n_1) + n_2 b_k(\alpha; n_2) + \cdots + n_k b_k(\alpha; n_k)}{N}.$$

As before, all the $b_k(\alpha; n_i)$ for sample sizes $n_1, n_2, \dots, n_k$ are obtained from Table A.10.

Example 13.3: Use Bartlett's test to test the hypothesis at the 0.01 level of significance that the population variances of the four drug groups of Example 13.2 are equal.

***Solution*:** We have the hypotheses

$$H_0\colon\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2,$$
$$H_1\colon\ \text{The variances are not equal,}$$
$$\text{with } \alpha = 0.01.$$

Critical region: Referring to Example 13.2, we have $n_1 = 20$, $n_2 = 9$, $n_3 = 9$, $n_4 = 7$, $N = 45$, and $k = 4$. Therefore, we reject when

$$\begin{aligned} b &< b_4(0.01; 20, 9, 9, 7) \\ &\approx \frac{(20)(0.8586) + (9)(0.6892) + (9)(0.6892) + (7)(0.6045)}{45} \\ &= 0.7513. \end{aligned}$$

Computations: First compute

$$s_1^2 = 662.862,\ s_2^2 = 2219.781,\ s_3^2 = 2168.434,\ s_4^2 = 946.032,$$

and then

$$\begin{aligned} s_p^2 &= \frac{(19)(662.862) + (8)(2219.781) + (8)(2168.434) + (6)(946.032)}{41} \\ &= 1301.861. \end{aligned}$$

Now

$$b = \frac{[(662.862)^{19}(2219.781)^8(2168.434)^8(946.032)^6]^{1/41}}{1301.861} = 0.8557.$$

Decision: Do not reject the hypothesis, and conclude that the population variances of the four drug groups are not significantly different. ┘

Although Bartlett's test is most often used for testing of homogeneity of variances, other methods are available. A method due to Cochran provides a computationally simple procedure, but it is restricted to situations in which the sample sizes are equal. **Cochran's test** is particularly useful for detecting if one variance is much larger than the others. The statistic that is used is

$$G = \frac{\text{largest } S_i^2}{\sum\limits_{i=1}^{k} S_i^2},$$

and the hypothesis of equality of variances is rejected if $g > g_\alpha$, where the value of g_α is obtained from Table A.11.

To illustrate Cochran's test, let us refer again to the data of Table 13.1 on moisture absorption in concrete aggregates. Were we justified in assuming equal variances when we performed the analysis of variance in Example 13.1? We find that

$$s_1^2 = 12{,}134,\ s_2^2 = 2303,\ s_3^2 = 3594,\ s_4^2 = 3319,\ s_5^2 = 3455.$$

Therefore,

$$g = \frac{12{,}134}{24{,}805} = 0.4892,$$

which does not exceed the table value $g_{0.05} = 0.5065$. Hence, we conclude that the assumption of equal variances is reasonable.

Exercises

13.1 Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. The following are the tensile-strength measurements in kilograms per square centimeter × 10^{-1}:

Machine					
1	**2**	**3**	**4**	**5**	**6**
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1

Perform the analysis of variance at the 0.05 level of significance and indicate whether or not the mean tensile strengths differ significantly for the six machines.

13.2 The data in the following table represent the number of hours of relief provided by five different brands of headache tablets administered to 25 subjects experiencing fevers of 38°C or more. Perform the analysis of variance and test the hypothesis at the 0.05 level of significance that the mean number of hours of relief provided by the tablets is the same for all five brands. Discuss the results.

Tablet				
A	***B***	***C***	***D***	***E***
5.2	9.1	3.2	2.4	7.1
4.7	7.1	5.8	3.4	6.6
8.1	8.2	2.2	4.1	9.3
6.2	6.0	3.1	1.0	4.2
3.0	9.1	7.2	4.0	7.6

13.3 In an article "Shelf-Space Strategy in Retailing," published in *Proceedings: Southern Marketing Association*, the effect of shelf height on the supermarket sales of canned dog food is investigated. An experiment was conducted at a small supermarket for a period of 8 days on the sales of a single brand of dog food, referred to as Arf dog food, involving three levels of shelf height: knee level, waist level, and eye level. During each day, the shelf height of the canned dog food was randomly changed on three different occasions. The remaining sections of the gondola that housed the given brand were filled with a mixture of dog food brands that were both familiar and unfamiliar to customers in this particular geographic area. Sales, in hundreds of dollars, of Arf dog food per day for the three shelf heights are given. Based on the data, is there a significant difference in the average daily sales of this dog food based on shelf height? Use a 0.01 level of significance.

Shelf Height		
Knee Level	**Waist Level**	**Eye Level**
77	88	85
82	94	85
86	93	87
78	90	81
81	91	80
86	94	79
77	90	87
81	87	93

13.4 Immobilization of free-ranging white-tailed deer by drugs allows researchers the opportunity to closely examine the deer and gather valuable physiological information. In the study *Influence of Physical Restraint and Restraint Facilitating Drugs on Blood Measurements of White-Tailed Deer and Other Selected Mammals*, conducted at Virginia Tech, wildlife biologists tested the "knockdown" time (time from injection to immobilization) of three different immobilizing drugs. Immobilization, in this case, is defined as the point where the animal no longer has enough muscle control to remain standing. Thirty male white-tailed deer were randomly assigned to each of three treatments. Group A received 5 milligrams of liquid succinylcholine chloride (SCC); group B received 8 milligrams of powdered SCC; and group C received 200 milligrams of phencyclidine hydrochloride. Knockdown times, in minutes, were recorded. Perform an analysis of variance at the 0.01 level of significance and determine whether or not the average knockdown time for the three drugs is the same.

Group		
A	***B***	***C***
11	10	4
5	7	4
14	16	6
7	7	3
10	7	5
7	5	6
23	10	8
4	10	3
11	6	7
11	12	3

13.5 The mitochondrial enzyme NADPH:NAD transhydrogenase of the common rat tapeworm (*Hymenolepiasis diminuta*) catalyzes hydrogen in the transfer from NADPH to NAD, producing NADH. This enzyme is known to serve a vital role in the tapeworm's anaerobic metabolism, and it has recently been hypothesized that it may serve as a proton exchange pump, transferring protons across the mitochondrial membrane. A study on *Effect of Various Substrate Concentrations on the Conformational Variation of the NADPH:NAD Transhydrogenase of Hymenolepiasis diminuta*, conducted at Bowling Green State University, was designed to assess the ability of this enzyme to undergo conformation or shape changes. Changes in the specific activity of the enzyme caused by variations in the concentration of NADP could be interpreted as supporting the theory of conformational change. The enzyme in question is located in the inner membrane of the tapeworm's mitochondria. Tapeworms were homogenized, and through a series of centrifugations, the enzyme was isolated. Various concentrations of NADP were then added to the isolated enzyme solution, and the mixture was then incubated in a water bath at 56°C for 3 minutes. The enzyme was then analyzed on a dual-beam spectrophotometer, and the results shown were calculated, with the specific activity of the enzyme given in nanomoles per minute per milligram of protein. Test the hypothesis at the 0.01 level that the average specific activity is the same for the four concentrations.

NADP Concentration (nm)				
0	**80**	**160**	**360**	
11.01	11.38	11.02	6.04	10.31
12.09	10.67	10.67	8.65	8.30
10.55	12.33	11.50	7.76	9.48
11.26	10.08	10.31	10.13	8.89
			9.36	

13.6 A study measured the sorption (either absorption or adsorption) rates of three different types of organic chemical solvents. These solvents are used to clean industrial fabricated-metal parts and are potential hazardous waste. Independent samples from each type of solvent were tested, and their sorption rates were recorded as a mole percentage. (See McClave, Dietrich, and Sincich, 1997.)

Aromatics		**Chloroalkanes**		**Esters**		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45	0.91	0.06	0.51	0.09
0.82	1.15	0.57	0.83	0.44	0.10	0.17
0.89	1.12	1.16	0.43	0.55	0.53	0.17
1.05				0.61	0.34	0.60

Is there a significant difference in the mean sorption rates for the three solvents? Use a P-value for your conclusions. Which solvent would you use?

13.7 It has been shown that the fertilizer magnesium ammonium phosphate, $MgNH_4PO_4$, is an effective supplier of the nutrients necessary for plant growth. The compounds supplied by this fertilizer are highly soluble in water, allowing the fertilizer to be applied directly on the soil surface or mixed with the growth substrate during the potting process. A study on the *Effect of Magnesium Ammonium Phosphate on Height of Chrysanthemums* was conducted at George Mason University to determine a possible optimum level of fertilization, based on the enhanced vertical growth response of the chrysanthemums. Forty chrysanthemum seedlings were divided into four groups, each containing 10 plants. Each was planted in a similar pot containing a uniform growth medium. To each group of plants an increasing concentration of $MgNH_4PO_4$, measured in grams per bushel, was added. The four groups of plants were grown under uniform conditions in a greenhouse for a period of four weeks. The treatments and the respective changes in heights, measured in centimeters, are shown next.

Treatment							
50 g/bu		**100 g/bu**		**200 g/bu**		**400 g/bu**	
13.2	12.4	16.0	12.6	7.8	14.4	21.0	14.8
12.8	17.2	14.8	13.0	20.0	15.8	19.1	15.8
13.0	14.0	14.0	23.6	17.0	27.0	18.0	26.0
14.2	21.6	14.0	17.0	19.6	18.0	21.1	22.0
15.0	20.0	22.2	24.4	20.2	23.2	25.0	18.2

Can we conclude at the 0.05 level of significance that different concentrations of $MgNH_4PO_4$ affect the average attained height of chrysanthemums? How much $MgNH_4PO_4$ appears to be best?

13.8 For the data set in Exercise 13.7, use Bartlett's test to check whether the variances are equal. Use $\alpha = 0.05$.

13.9 Use Bartlett's test at the 0.01 level of significance to test for homogeneity of variances in Exercise 13.5 on page 519.

13.10 Use Cochran's test at the 0.01 level of significance to test for homogeneity of variances in Exercise 13.4 on page 519.

13.11 Use Bartlett's test at the 0.05 level of significance to test for homogeneity of variances in Exercise 13.6 on page 519.

13.5 Single-Degree-of-Freedom Comparisons

The analysis of variance in a one-way classification, or a one-factor experiment, as it is often called, merely indicates whether or not the hypothesis of equal treatment means can be rejected. Usually, an experimenter would prefer his or her analysis to probe deeper. For instance, in Example 13.1, by rejecting the null hypothesis we concluded that the means are not all equal, but we still do not know where the differences exist among the aggregates. The engineer might have the feeling *a priori* that aggregates 1 and 2 should have similar absorption properties and that the same is true for aggregates 3 and 5. However, it is of interest to study the difference between the two groups. It would seem, then, appropriate to test the hypothesis

$$H_0\colon\ \mu_1 + \mu_2 - \mu_3 - \mu_5 = 0,$$
$$H_1\colon\ \mu_1 + \mu_2 - \mu_3 - \mu_5 \neq 0.$$

We notice that the hypothesis is a linear function of the population means where the coefficients sum to zero.

Definition 13.1: Any linear function of the form

$$\omega = \sum_{i=1}^{k} c_i \mu_i,$$

where $\sum_{i=1}^{k} c_i = 0$, is called a **comparison** or **contrast** in the treatment means.

The experimenter can often make multiple comparisons by testing the significance of contrasts in the treatment means, that is, by testing a hypothesis of the following type:

Hypothesis for a Contrast

$$H_0\colon \sum_{i=1}^{k} c_i\mu_i = 0,$$

$$H_1\colon \sum_{i=1}^{k} c_i\mu_i \neq 0,$$

where $\sum\limits_{i=1}^{k} c_i = 0$.

The test is conducted by first computing a similar contrast in the sample means,

$$w = \sum_{i=1}^{k} c_i \bar{y}_{i.}.$$

Since $\bar{Y}_{1.}, \bar{Y}_{2.}, \ldots, \bar{Y}_{k.}$ are independent random variables having normal distributions with means $\mu_1, \mu_2, \ldots, \mu_k$ and variances $\sigma_1^2/n_1, \sigma_2^2/n_2, \ldots, \sigma_k^2/n_k$, respectively, Theorem 7.11 assures us that w is a value of the normal random variable W with

$$\text{mean } \mu_W = \sum_{i=1}^{k} c_i\mu_i \text{ and variance } \sigma_W^2 = \sigma^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}.$$

Therefore, when H_0 is true, $\mu_W = 0$ and, by Example 7.5, the statistic

$$\frac{W^2}{\sigma_W^2} = \frac{\left(\sum\limits_{i=1}^{k} c_i \bar{Y}_{i.}\right)^2}{\sigma^2 \sum\limits_{i=1}^{k} (c_i^2/n_i)}$$

is distributed as a chi-squared random variable with 1 degree of freedom.

Test Statistic for Testing a Contrast

Our hypothesis is tested at the α-level of significance by computing

$$f = \frac{\left(\sum\limits_{i=1}^{k} c_i \bar{y}_{i.}\right)^2}{s^2 \sum\limits_{i=1}^{k} (c_i^2/n_i)} = \frac{\left[\sum\limits_{i=1}^{k} (c_i Y_{i.}/n_i)\right]^2}{s^2 \sum\limits_{i=1}^{k} (c_i^2/n_i)} = \frac{SSw}{s^2}.$$

Here f is a value of the random variable F having the F-distribution with 1 and $N-k$ degrees of freedom.

When the sample sizes are all equal to n,

$$SSw = \frac{\left(\sum\limits_{i=1}^{k} c_i Y_{i.}\right)^2}{n \sum\limits_{i=1}^{k} c_i^2}.$$

The quantity SSw, called the **contrast sum of squares**, indicates the portion of SSA that is explained by the contrast in question.

This sum of squares will be used to test the hypothesis that

$$\sum_{i=1}^{k} c_i \mu_i = 0.$$

It is often of interest to test multiple contrasts, particularly contrasts that are linearly independent or orthogonal. As a result, we need the following definition:

Definition 13.2: The two contrasts

$$\omega_1 = \sum_{i=1}^{k} b_i \mu_i \qquad \text{and} \qquad \omega_2 = \sum_{i=1}^{k} c_i \mu_i$$

are said to be **orthogonal** if $\sum_{i=1}^{k} b_i c_i / n_i = 0$ or, when the n_i are all equal to n, if

$$\sum_{i=1}^{k} b_i c_i = 0.$$

If ω_1 and ω_2 are orthogonal, then the quantities SSw_1 and SSw_2 are components of SSA, each with a single degree of freedom. The treatment sum of squares with $k-1$ degrees of freedom can be partitioned into at most $k-1$ independent single-degree-of-freedom contrast sums of squares satisfying the identity

$$SSA = SSw_1 + SSw_2 + \cdots + SSw_{k-1},$$

if the contrasts are orthogonal to each other.

Example 13.4: Referring to Example 13.1, find the contrast sum of squares corresponding to the orthogonal contrasts

$$\omega_1 = \mu_1 + \mu_2 - \mu_3 - \mu_5, \qquad \omega_2 = \mu_1 + \mu_2 + \mu_3 - 4\mu_4 + \mu_5,$$

and carry out appropriate tests of significance. In this case, it is of interest *a priori* to compare the two groups $(1, 2)$ and $(3, 5)$. An important and independent contrast is the comparison between the set of aggregates $(1, 2, 3, 5)$ and aggregate 4.

Solution: It is obvious that the two contrasts are orthogonal, since

$$(1)(1) + (1)(1) + (-1)(1) + (0)(-4) + (-1)(1) = 0.$$

The second contrast indicates a comparison between aggregates $(1, 2, 3,$ and $5)$ and aggregate 4. We can write two additional contrasts orthogonal to the first two, namely

$$\begin{aligned} \omega_3 &= \mu_1 - \mu_2 && \text{(aggregate 1 versus aggregate 2)}, \\ \omega_4 &= \mu_3 - \mu_5 && \text{(aggregate 3 versus aggregate 5)}. \end{aligned}$$

From the data of Table 13.1, we have

$$SSw_1 = \frac{(3320 + 3416 - 3663 - 3664)^2}{6[(1)^2 + (1)^2 + (-1)^2 + (-1)^2]} = 14,553,$$

$$SSw_2 = \frac{[3320 + 3416 + 3663 + 3664 - 4(2791)]^2}{6[(1)^2 + (1)^2 + (1)^2 + (1)^2 + (-4)^2]} = 70,035.$$

A more extensive analysis-of-variance table is shown in Table 13.5. We note that the two contrast sums of squares account for nearly all the aggregate sum of squares. There is a significant difference between aggregates in their absorption properties, and the contrast ω_1 is marginally significant. However, the f-value of 14.12 for ω_2 is highly significant, and the hypothesis

$$H_0\colon\ \mu_1 + \mu_2 + \mu_3 + \mu_5 = 4\mu_4$$

is rejected.

Table 13.5: Analysis of Variance Using Orthogonal Contrasts

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Aggregates	85,356	4	21,339	4.30
$(1,2)$ vs. $(3,5)$	14,553	1	14,553	2.93
$(1,2,3,5)$ vs. 4	70,035	1	70,035	14.12
Error	124,021	25	4961	
Total	209,377	29		

Orthogonal contrasts allow the practitioner to partition the treatment variation into independent components. Normally, the experimenter would have certain contrasts that were of interest to him or her. Such was the case in our example, where *a priori* considerations suggested that aggregates $(1, 2)$ and $(3, 5)$ constituted distinct groups with different absorption properties, a postulation that was not strongly supported by the significance test. However, the second comparison supported the conclusion that aggregate 4 seemed to "stand out" from the rest. In this case, the complete partitioning of SSA was not necessary, since two of the four possible independent comparisons accounted for a majority of the variation in treatments.

Figure 13.4 shows a *SAS* GLM procedure that displays a complete set of orthogonal contrasts. Note that the sums of squares for the four contrasts add to the aggregate sum of squares. Also, note that the latter two contrasts (1 versus 2, 3 versus 5) reveal insignificant comparisons. ▟

13.6 Multiple Comparisons

The analysis of variance is a powerful procedure for testing the homogeneity of a set of means. However, if we reject the null hypothesis and accept the stated alternative—that the means are not all equal—we still do not know which of the population means are equal and which are different.

```
                          The GLM Procedure
Dependent Variable: moisture
                              Sum of
 Source              DF      Squares   Mean Square  F Value  Pr > F
 Model                4   85356.4667    21339.1167     4.30  0.0088
 Error               25  124020.3333     4960.8133
 Corrected Total     29  209376.8000

        R-Square     Coeff Var      Root MSE    moisture Mean
        0.407669      12.53703      70.43304         561.8000

 Source             DF     Type I SS     Mean Square    F Value    Pr > F
 aggregate           4   85356.46667     21339.11667       4.30    0.0088

 Source             DF   Type III SS     Mean Square    F Value    Pr > F
 aggregate           4   85356.46667     21339.11667       4.30    0.0088

 Contrast           DF   Contrast SS     Mean Square    F Value    Pr > F
 (1,2,3,5) vs. 4     1   70035.00833     70035.00833      14.12    0.0009
 (1,2) vs. (3,5)     1   14553.37500     14553.37500       2.93    0.0991
 1 vs. 2             1     768.00000       768.00000       0.15    0.6973
 3 vs. 5             1       0.08333         0.08333       0.00    0.9968
```

Figure 13.4: A set of orthogonal contrasts

Often it is of interest to make several (perhaps all possible) **paired comparisons** among the treatments. Actually, a paired comparison may be viewed as a simple contrast, namely, a test of

$$\begin{aligned} H_0\colon\ \mu_i - \mu_j &= 0, \\ H_1\colon\ \mu_i - \mu_j &\neq 0, \end{aligned}$$

for all $i \neq j$. Making all possible paired comparisons among the means can be very beneficial when particular complex contrasts are not known *a priori*. For example, in the aggregate data of Table 13.1, suppose that we wish to test

$$\begin{aligned} H_0\colon\ \mu_1 - \mu_5 &= 0, \\ H_1\colon\ \mu_1 - \mu_5 &\neq 0. \end{aligned}$$

The test is developed through use of an F, t, or confidence interval approach. Using t, we have

$$t = \frac{\bar{y}_{1.} - \bar{y}_{5.}}{s\sqrt{2/n}},$$

where s is the square root of the mean square error and $n = 6$ is the sample size per treatment. In this case,

$$t = \frac{553.33 - 610.67}{\sqrt{4961}\sqrt{1/3}} = -1.41.$$

The P-value for the t-test with 25 degrees of freedom is 0.17. Thus, there is not sufficient evidence to reject H_0.

Relationship between T and F

In the foregoing, we displayed the use of a pooled t-test along the lines of that discussed in Chapter 10. The pooled estimate was taken from the mean squared error in order to enjoy the degrees of freedom that are pooled across all five samples. In addition, we have tested a contrast. The reader should note that if the t-value is squared, the result is exactly of the same form as the value of f for a test on a contrast, discussed in the preceding section. In fact,

$$f = \frac{(\bar{y}_{1.} - \bar{y}_{5.})^2}{s^2(1/6 + 1/6)} = \frac{(553.33 - 610.67)^2}{4961(1/3)} = 1.988,$$

which, of course, is t^2.

Confidence Interval Approach to a Paired Comparison

It is straightforward to solve the same problem of a paired comparison (or a contrast) using a confidence interval approach. Clearly, if we compute a $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_5$, we have

$$\bar{y}_{1.} - \bar{y}_{5.} \pm t_{\alpha/2} s\sqrt{\frac{2}{6}},$$

where $t_{\alpha/2}$ is the upper $100(1-\alpha/2)\%$ point of a t-distribution with 25 degrees of freedom (degrees of freedom coming from s^2). This straightforward connection between hypothesis testing and confidence intervals should be obvious from discussions in Chapters 9 and 10. The test of the simple contrast $\mu_1 - \mu_5$ involves no more than observing whether or not the confidence interval above covers zero. Substituting the numbers, we have as the 95% confidence interval

$$(553.33 - 610.67) \pm 2.060\sqrt{4961}\sqrt{\frac{1}{3}} = -57.34 \pm 83.77.$$

Thus, since the interval covers zero, the contrast is not significant. In other words, we do not find a significant difference between the means of aggregates 1 and 5.

Experiment-wise Error Rate

Serious difficulties occur when the analyst attempts to make many or all possible paired comparisons. For the case of k means, there will be, of course, $r = k(k-1)/2$ possible paired comparisons. Assuming independent comparisons, the **experiment-wise error rate** or **family error rate** (i.e., the probability of false rejection of at least one of the hypotheses) is given by $1 - (1-\alpha)^r$, where α is the selected probability of a type I error for a specific comparison. Clearly, this measure of experiment-wise type I error can be quite large. For example, even if there are only 6 comparisons, say, in the case of 4 means, and $\alpha = 0.05$, the experiment-wise rate is

$$1 - (0.95)^6 \approx 0.26.$$

When many paired comparisons are being tested, there is usually a need to make the effective contrast on a single comparison more conservative. That is, with the confidence interval approach, the confidence intervals would be much wider than the $\pm t_{\alpha/2} s\sqrt{2/n}$ used for the case where only a single comparison is being made.

Tukey's Test

There are several standard methods for making paired comparisons that sustain the credibility of the type I error rate. We shall discuss and illustrate two of them here. The first one, called **Tukey's procedure**, allows formation of simultaneous $100(1-\alpha)\%$ confidence intervals for all paired comparisons. The method is based on the *studentized* range distribution. The appropriate percentile point is a function of α, k, and $v =$ degrees of freedom for s^2. A list of upper percentage points for $\alpha = 0.05$ is shown in Table A.12. The method of paired comparisons by Tukey involves finding a significant difference between means i and j $(i \neq j)$ if $|\bar{y}_{i.} - \bar{y}_{j.}|$ exceeds $q(\alpha, k, v)\sqrt{\frac{s^2}{n}}$.

Tukey's procedure is easily illustrated. Consider a hypothetical example where we have 6 treatments in a one-factor completely randomized design, with 5 observations taken per treatment. Suppose that the mean square error taken from the analysis-of-variance table is $s^2 = 2.45$ (24 degrees of freedom). The sample means are in ascending order:

$\bar{y}_{2.}$	$\bar{y}_{5.}$	$\bar{y}_{1.}$	$\bar{y}_{3.}$	$\bar{y}_{6.}$	$\bar{y}_{4.}$
14.50	16.75	19.84	21.12	22.90	23.20.

With $\alpha = 0.05$, the value of $q(0.05, 6, 24)$ is 4.37. Thus, all absolute differences are to be compared to

$$4.37\sqrt{\frac{2.45}{5}} = 3.059.$$

As a result, the following represent means found to be significantly different using Tukey's procedure:

4 and 1, 4 and 5, 4 and 2, 6 and 1, 6 and 5,
6 and 2, 3 and 5, 3 and 2, 1 and 5, 1 and 2.

Where Does the α-Level Come From in Tukey's Test?

We briefly alluded to the concept of **simultaneous confidence intervals** being employed for Tukey's procedure. The reader will gain a useful insight into the notion of multiple comparisons if he or she gains an understanding of what is meant by simultaneous confidence intervals.

In Chapter 9, we saw that if we compute a 95% confidence interval on, say, a mean μ, then the probability that the interval covers the true mean μ is 0.95.

However, as we have discussed, for the case of multiple comparisons, the effective probability of interest is tied to the experiment-wise error rate, and it should be emphasized that the confidence intervals of the type $\bar{y}_{i.} - \bar{y}_{j.} \pm q(\alpha, k, v)s\sqrt{1/n}$ are not independent since they all involve s and many involve the use of the same averages, the $\bar{y}_{i.}$. Despite the difficulties, if we use $q(0.05, k, v)$, the simultaneous confidence level is controlled at 95%. The same holds for $q(0.01, k, v)$; namely, the confidence level is controlled at 99%. In the case of $\alpha = 0.05$, there is a probability of 0.05 that at least one pair of measures will be falsely found to be different (false rejection of at least one null hypothesis). In the $\alpha = 0.01$ case, the corresponding probability will be 0.01.

Duncan's Test

The second procedure we shall discuss is called **Duncan's procedure** or **Duncan's multiple-range test**. This procedure is also based on the general notion of studentized range. The range of any subset of p sample means must exceed a certain value before any of the p means are found to be different. This value is called the **least significant range** for the p means and is denoted by R_p, where

$$R_p = r_p\sqrt{\frac{s^2}{n}}.$$

The values of the quantity r_p, called the **least significant studentized range**, depend on the desired level of significance and the number of degrees of freedom of the mean square error. These values may be obtained from Table A.13 for $p = 2, 3, \ldots, 10$ means.

To illustrate the multiple-range test procedure, let us consider the hypothetical example where 6 treatments are compared, with 5 observations per treatment. This is the same example used to illustrate Tukey's test. We obtain R_p by multiplying each r_p by 0.70. The results of these computations are summarized as follows:

p	2	3	4	5	6
r_p	2.919	3.066	3.160	3.226	3.276
R_p	2.043	2.146	2.212	2.258	2.293

Comparing these least significant ranges with the differences in ordered means, we arrive at the following conclusions:

1. Since $\bar{y}_{4.} - \bar{y}_{2.} = 8.70 > R_6 = 2.293$, we conclude that μ_4 and μ_2 are significantly different.
2. Comparing $\bar{y}_{4.} - \bar{y}_{5.}$ and $\bar{y}_{6.} - \bar{y}_{2.}$ with R_5, we conclude that μ_4 is significantly greater than μ_5 and μ_6 is significantly greater than μ_2.
3. Comparing $\bar{y}_{4.} - \bar{y}_{1.}$, $\bar{y}_{6.} - \bar{y}_{5.}$, and $\bar{y}_{3.} - \bar{y}_{2.}$ with R_4, we conclude that each difference is significant.
4. Comparing $\bar{y}_{4.} - \bar{y}_{3.}$, $\bar{y}_{6.} - \bar{y}_{1.}$, $\bar{y}_{3.} - \bar{y}_{5.}$, and $\bar{y}_{1.} - \bar{y}_{2.}$ with R_3, we find all differences significant except for $\mu_4 - \mu_3$. Therefore, μ_3, μ_4, and μ_6 constitute a subset of homogeneous means.
5. Comparing $\bar{y}_{3.} - \bar{y}_{1.}$, $\bar{y}_{1.} - \bar{y}_{5.}$, and $\bar{y}_{5.} - \bar{y}_{2.}$ with R_2, we conclude that only μ_3 and μ_1 are not significantly different.

It is customary to summarize the conclusions above by drawing a line under any subsets of adjacent means that are not significantly different. Thus, we have

$$\begin{array}{cccccc} \bar{y}_{2.} & \bar{y}_{5.} & \bar{y}_{1.} & \bar{y}_{3.} & \bar{y}_{6.} & \bar{y}_{4.} \\ \underline{14.50} & \underline{16.75} & 19.84 & 21.12 & 22.90 & 23.20 \end{array}$$

It is clear that in this case the results from Tukey's and Duncan's procedures are very similar. Tukey's procedure did not detect a difference between 2 and 5, whereas Duncan's did.

Dunnett's Test: Comparing Treatment with a Control

In many scientific and engineering problems, one is not interested in drawing inferences regarding all possible comparisons among the treatment means of the type $\mu_i - \mu_j$. Rather, the experiment often dictates the need to simultaneously compare each *treatment* with a *control*. A test procedure developed by C. W. Dunnett determines significant differences between each treatment mean and the control, at a single joint significance level α. To illustrate Dunnett's procedure, let us consider the experimental data of Table 13.6 for a one-way classification where the effect of three catalysts on the yield of a reaction is being studied. A fourth treatment, no catalyst, is used as a control.

Table 13.6: Yield of Reaction

Control	**Catalyst 1**	**Catalyst 2**	**Catalyst 3**
50.7	54.1	52.7	51.2
51.5	53.8	53.9	50.8
49.2	53.1	57.0	49.7
53.1	52.5	54.1	48.0
52.7	54.0	52.5	47.2
$\bar{y}_{0.} = 51.44$	$\bar{y}_{1.} = 53.50$	$\bar{y}_{2.} = 54.04$	$\bar{y}_{3.} = 49.38$

In general, we wish to test the k hypotheses

$$\left.\begin{array}{l} H_0\colon\ \mu_0 = \mu_i \\ H_1\colon\ \mu_0 \neq \mu_i \end{array}\right\} \qquad i = 1, 2, \ldots, k,$$

where μ_0 represents the mean yield for the population of measurements in which the control is used. The usual analysis-of-variance assumptions, as outlined in Section 13.3, are expected to remain valid. To test the null hypotheses specified by H_0 against two-sided alternatives for an experimental situation in which there are k treatments, excluding the control, and n observations per treatment, we first calculate the values

$$d_i = \frac{\bar{y}_{i.} - \bar{y}_{0.}}{\sqrt{2s^2/n}}, \qquad i = 1, 2, \ldots, k.$$

The sample variance s^2 is obtained, as before, from the mean square error in the analysis of variance. Now, the critical region for rejecting H_0, at the α-level of significance, is established by the inequality

$$|d_i| > d_{\alpha/2}(k, v),$$

where v is the number of degrees of freedom for the mean square error. The values of the quantity $d_{\alpha/2}(k, v)$ for a two-tailed test are given in Table A.14 for $\alpha = 0.05$ and $\alpha = 0.01$ for various values of k and v.

Example 13.5: For the data of Table 13.6, test hypotheses comparing each catalyst with the control, using two-sided alternatives. Choose $\alpha = 0.05$ as the joint significance level.

Solution: The mean square error with 16 degrees of freedom is obtained from the analysis-of-variance table, using all $k + 1$ treatments. The mean square error is given by

$$s^2 = \frac{36.812}{16} = 2.30075 \text{ and } \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{(2)(2.30075)}{5}} = 0.9593.$$

Hence,

$$d_1 = \frac{53.50 - 51.44}{0.9593} = 2.147, \quad d_2 = \frac{54.04 - 51.44}{0.9593} = 2.710,$$

$$d_3 = \frac{49.38 - 51.44}{0.9593} = -2.147.$$

From Table A.14 the critical value for $\alpha = 0.05$ is found to be $d_{0.025}(3, 16) = 2.59$. Since $|d_1| < 2.59$ and $|d_3| < 2.59$, we conclude that only the mean yield for catalyst 2 is significantly different from the mean yield of the reaction using the control. ▟

Many practical applications dictate the need for a one-tailed test for comparing treatments with a control. Certainly, when a pharmacologist is concerned with the effect of various dosages of a drug on cholesterol level and his control is zero dosage, it is of interest to determine if each dosage produces a significantly larger reduction than the control. Table A.15 shows the critical values of $d_\alpha(k, v)$ for one-sided alternatives.

Exercises

13.12 Consider the data of Review Exercise 13.45 on page 555. Make significance tests on the following contrasts:

(a) B versus A, C, and D;

(b) C versus A and D;

(c) A versus D.

13.13 The purpose of the study *The Incorporation of a Chelating Agent into a Flame Retardant Finish of a Cotton Flannelette and the Evaluation of Selected Fabric Properties* conducted at Virginia Tech was to evaluate the use of a chelating agent as part of the flame-retardant finish of cotton flannelette by determining its effects upon flammability after the fabric is laundered under specific conditions. Two baths were prepared, one with carboxymethyl cellulose and one without. Twelve pieces of fabric were laundered 5 times in bath I, and 12 other pieces of fabric were laundered 10 times in bath I. This was repeated using 24 additional pieces of cloth in bath II. After the washings the lengths of fabric that burned and the burn times were measured. For convenience, let us define the following treatments:

Treatment 1: 5 launderings in bath I,

Treatment 2: 5 launderings in bath II,

Treatment 3: 10 launderings in bath I,

Treatment 4: 10 launderings in bath II.

Burn times, in seconds, were recorded as follows:

Treatment			
1	**2**	**3**	**4**
13.7	6.2	27.2	18.2
23.0	5.4	16.8	8.8
15.7	5.0	12.9	14.5
25.5	4.4	14.9	14.7
15.8	5.0	17.1	17.1
14.8	3.3	13.0	13.9
14.0	16.0	10.8	10.6
29.4	2.5	13.5	5.8
9.7	1.6	25.5	7.3
14.0	3.9	14.2	17.7
12.3	2.5	27.4	18.3
12.3	7.1	11.5	9.9

(a) Perform an analysis of variance, using a 0.01 level of significance, and determine whether there are any significant differences among the treatment means.

(b) Use single-degree-of-freedom contrasts with $\alpha = 0.01$ to compare the mean burn time of treatment 1 versus treatment 2 and also treatment 3 versus treatment 4.

13.14 The study *Loss of Nitrogen Through Sweat by Preadolescent Boys Consuming Three Levels of Dietary Protein* was conducted by the Department of Human Nutrition and Foods at Virginia Tech to determine perspiration nitrogen loss at various dietary protein levels. Twelve preadolescent boys ranging in age from 7 years, 8 months to 9 years, 8 months, all judged to be clinically healthy, were used in the experiment. Each boy was subjected to one of three controlled diets in which 29, 54, or 84 grams of protein were consumed per day. The following data represent the body perspiration nitrogen loss, in milligrams, during the last two days of the experimental period:

Protein Level		
29 Grams	**54 Grams**	**84 Grams**
190	318	390
266	295	321
270	271	396
	438	399
	402	

(a) Perform an analysis of variance at the 0.05 level of significance to show that the mean perspiration nitrogen losses at the three protein levels are different.

(b) Use Tukey's test to determine which protein levels are significantly different from each other in mean nitrogen loss.

13.15 Use Tukey's test, with a 0.05 level of significance, to analyze the means of the five different brands of headache tablets in Exercise 13.2 on page 518.

13.16 An investigation was conducted to determine the source of reduction in yield of a certain chemical product. It was known that the loss in yield occurred in the mother liquor, that is, the material removed at the filtration stage. It was thought that different blends of the original material might result in different yield reductions at the mother liquor stage. The following are the percent reductions for 3 batches at each of 4 preselected blends:

Blend			
1	**2**	**3**	**4**
25.6	25.2	20.8	31.6
24.3	28.6	26.7	29.8
27.9	24.7	22.2	34.3

(a) Perform the analysis of variance at the $\alpha = 0.05$ level of significance.

(b) Use Duncan's multiple-range test to determine which blends differ.

(c) Do part (b) using Tukey's test.

13.17 In the study *An Evaluation of the Removal Method for Estimating Benthic Populations and Diversity* conducted by Virginia Tech on the Jackson River, 5 different sampling procedures were used to determine the species counts. Twenty samples were selected at random, and each of the 5 sampling procedures was repeated 4 times. The species counts were recorded as follows:

Sampling Procedure				
Depletion	**Modified Hess**	**Surber**	**Substrate Removal Kicknet**	**Kicknet**
85	75	31	43	17
55	45	20	21	10
40	35	9	15	8
77	67	37	27	15

(a) Is there a significant difference in the average species counts for the different sampling procedures? Use a *P*-value in your conclusion.

(b) Use Tukey's test with $\alpha = 0.05$ to find which sampling procedures differ.

13.18 The following data are values of pressure (psi) in a torsion spring for several settings of the angle between the legs of the spring in a free position:

Angle (°)					
67	**71**	**75**		**79**	**83**
83	84	86	87	89	90
85	85	87	87	90	92
	85	88	88	90	
	86	88	88	91	
	86	88	89		
	87	90			

Compute a one-way analysis of variance for this experiment and state your conclusion concerning the effect of angle on the pressure in the spring. (From C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart and Winston, New York, 1973.)

13.19 It is suspected that the environmental temperature at which batteries are activated affects their life. Thirty homogeneous batteries were tested, six at each of five temperatures, and the data are shown below (activated life in seconds). Analyze and interpret the data. (From C. R. Hicks, *Fundamental Concepts in Design of Experiments*, Holt, Rinehart and Winston, New York, 1973.)

Temperature (°C)				
0	**25**	**50**	**75**	**100**
55	60	70	72	65
55	61	72	72	66
57	60	72	72	60
54	60	68	70	64
54	60	77	68	65
56	60	77	69	65

13.20 The following table (from A. Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, New York, 1952) gives tensile strengths (in deviations from 340) for wires taken from nine cables to be used for a high-voltage network. Each cable is made from 12 wires. We want to know whether the mean strengths of the wires in the nine cables are the same. If the cables are different, which ones differ? Use a P-value in your analysis of variance.

Cable	**Tensile Strength**											
1	5	−13	−5	−2	−10	−6	−5	0	−3	2	−7	−5
2	−11	−13	−8	8	−3	−12	−12	−10	5	−6	−12	−10
3	0	−10	−15	−12	−2	−8	−5	0	−4	−1	−5	−11
4	−12	4	2	10	−5	−8	−12	0	−5	−3	−3	0
5	7	1	5	0	10	6	5	2	0	−1	−10	−2
6	1	0	−5	−4	−1	0	2	5	1	−2	6	7
7	−1	0	2	1	−4	2	7	5	1	0	−4	2
8	−1	0	7	5	10	8	1	2	−3	6	0	5
9	2	6	7	8	15	11	−7	7	10	7	8	1

13.21 The printout in Figure 13.5 on page 532 gives information on Duncan's test, using PROC GLM in *SAS*, for the aggregate data in Example 13.1. Give conclusions regarding paired comparisons using Duncan's test results.

13.22 Do Duncan's test for paired comparisons for the data of Exercise 13.6 on page 519. Discuss the results.

13.23 In a biological experiment, four concentrations of a certain chemical are used to enhance the growth of a certain type of plant over time. Five plants are used at each concentration, and the growth in each plant is measured in centimeters. The following growth data are taken. A control (no chemical) is also applied.

	Concentration			
Control	**1**	**2**	**3**	**4**
6.8	8.2	7.7	6.9	5.9
7.3	8.7	8.4	5.8	6.1
6.3	9.4	8.6	7.2	6.9
6.9	9.2	8.1	6.8	5.7
7.1	8.6	8.0	7.4	6.1

Use Dunnett's two-sided test at the 0.05 level of significance to simultaneously compare the concentrations with the control.

13.24 The financial structure of a firm refers to the way the firm's assets are divided into equity and debt, and the financial leverage refers to the percentage of assets financed by debt. In the paper *The Effect of Financial Leverage on Return*, Tai Ma of Virginia Tech claims that financial leverage can be used to increase the rate of return on equity. To say it another way, stockholders can receive higher returns on equity with the same amount of investment through the use of financial leverage. The following data show the rates of return on equity using 3 different levels of financial leverage and a control level (zero debt) for 24 randomly selected firms:

Financial Leverage			
Control	**Low**	**Medium**	**High**
2.1	6.2	9.6	10.3
5.6	4.0	8.0	6.9
3.0	8.4	5.5	7.8
7.8	2.8	12.6	5.8
5.2	4.2	7.0	7.2
2.6	5.0	7.8	12.0

Source: Standard & Poor's *Machinery Industry Survey*, 1975.

(a) Perform the analysis of variance at the 0.05 level of significance.

(b) Use Dunnett's test at the 0.01 level of significance to determine whether the mean rates of return on equity are higher at the low, medium, and high levels of financial leverage than at the control level.

```
                    The GLM Procedure
           Duncan's Multiple Range Test for moisture
 NOTE: This test controls the Type I comparisonwise error rate,
       not the experimentwise error rate.
              Alpha                          0.05
              Error Degrees of Freedom         25
              Error Mean Square          4960.813
   Number of Means          2          3          4          5
   Critical Range       83.75      87.97      90.69      92.61

 Means with the same letter are not significantly different.
         Duncan Grouping          Mean       N    aggregate
                      A         610.67       6    5
                      A
                      A         610.50       6    3
                      A
                      A         569.33       6    2
                      A
                      A         553.33       6    1

                      B         465.17       6    4
```

Figure 13.5: *SAS* printout for Exercise 13.21.

13.7 Comparing a Set of Treatments in Blocks

In Section 13.2, we discussed the idea of blocking, that is, isolating sets of experimental units that are reasonably homogeneous and randomly assigning treatments to these units. This is an extension of the "pairing" concept discussed in Chapters 9 and 10, and it is done to reduce experimental error, since the units in a block have more common characteristics than units in different blocks.

The reader should not view blocks as a second factor, although this is a tempting way of visualizing the design. In fact, the main factor (treatments) still carries the major thrust of the experiment. Experimental units are still the source of error, just as in the completely randomized design. We merely treat sets of these units more systematically when blocking is accomplished. In this way, we say there are restrictions in randomization. Before we turn to a discussion of blocking, let us look at two examples of a **completely randomized design**. The first example is a chemical experiment designed to determine if there is a difference in mean reaction yield among four catalysts. Samples of materials to be tested are drawn from the same batches of raw materials, while other conditions, such as temperature and concentration of reactants, are held constant. In this case, the time of day for the experimental runs might represent the experimental units, and if the experimenter believed that there could possibly be a slight time effect, he or she would randomize the assignment of the catalysts to the runs to counteract the possible trend. As a second example of such a design, consider an experiment to compare four methods of measuring a particular physical property of a fluid substance. Suppose the sampling process is destructive; that is, once a sample of the substance has been measured by one method, it cannot be measured again by any of the other methods. If it is decided that five measurements are to be taken for each method, then 20 samples of the material are selected from a large batch at random and are used in the experiment to compare the four measuring methods. The experimental units are the randomly selected samples. Any variation from sample to sample will appear in the error variation, as measured by s^2 in the analysis.

What Is the Purpose of Blocking?

If the variation due to heterogeneity in experimental units is so large that the sensitivity with which treatment differences are detected is reduced due to an inflated value of s^2, a better plan might be to "block off" variation due to these units and thus reduce the extraneous variation to that accounted for by smaller or more homogeneous blocks. For example, suppose that in the previous catalyst illustration it is known *a priori* that there definitely is a significant day-to-day effect on the yield and that we can measure the yield for four catalysts on a given day. Rather than assign the four catalysts to the 20 test runs completely at random, we choose, say, five days and run each of the four catalysts on each day, randomly assigning the catalysts to the runs within days. In this way, the day-to-day variation is removed from the analysis, and consequently the experimental error, which still includes any time trend *within days*, more accurately represents chance variation. Each day is referred to as a **block**.

The most straightforward of the randomized block designs is one in which we randomly assign each treatment once to every block. Such an experimental layout is called a **randomized complete block (RCB) design**, each block constituting a single replication of the treatments.

13.8 Randomized Complete Block Designs

A typical layout for the randomized complete block design using 3 measurements in 4 blocks is as follows:

Block 1	Block 2	Block 3	Block 4
t_2	t_1	t_3	t_2
t_1	t_3	t_2	t_1
t_3	t_2	t_1	t_3

The t's denote the assignment to blocks of each of the 3 treatments. Of course, the true allocation of treatments to units within blocks is done at random. Once the experiment has been completed, the data can be recorded in the following 3 × 4 array:

Treatment	**Block:**	**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}

where y_{11} represents the response obtained by using treatment 1 in block l, y_{12} represents the response obtained by using treatment 1 in block 2, ..., and y_{34} represents the response obtained by using treatment 3 in block 4.

Let us now generalize and consider the case of k treatments assigned to b blocks. The data may be summarized as shown in the $k \times b$ rectangular array of Table 13.7. It will be assumed that the y_{ij}, $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, b$, are values of independent random variables having normal distributions with mean μ_{ij} and common variance σ^2.

Table 13.7: $k \times b$ Array for the RCB Design

	Block							
Treatment	**1**	**2**	$\cdots$	$\boldsymbol{j}$	$\cdots$	$\boldsymbol{b}$	**Total**	**Mean**
1	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1b}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2b}	$T_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{ib}	$T_{i.}$	$\bar{y}_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
k	y_{k1}	y_{k2}	$\cdots$	y_{kj}	$\cdots$	y_{kb}	$T_{k.}$	$\bar{y}_{k.}$
Total	$T_{.1}$	$T_{.2}$	$\cdots$	$T_{.j}$	$\cdots$	$T_{.b}$	$T_{..}$	
Mean	$\bar{y}_{.1}$	$\bar{y}_{.2}$	$\cdots$	$\bar{y}_{.j}$	$\cdots$	$\bar{y}_{.b}$		$\bar{y}_{..}$

Let $\mu_{i.}$ represent the average (rather than the total) of the b population means for the ith treatment. That is,

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b} \mu_{ij}, \text{ for } i = 1, \ldots, k.$$

Similarly, the average of the population means for the jth block, $\mu_{.j}$, is defined by

$$\mu_{.j} = \frac{1}{k}\sum_{i=1}^{k} \mu_{ij}, \text{ for } j = 1, \ldots, b$$

and the average of the bk population means, μ, is defined by

$$\mu = \frac{1}{bk}\sum_{i=1}^{k}\sum_{j=1}^{b} \mu_{ij}.$$

To determine if part of the variation in our observations is due to differences among the treatments, we consider the following test:

Hypothesis of Equal Treatment Means

$$H_0\colon\ \mu_{1.} = \mu_{2.} = \cdots \mu_{k.} = \mu,$$
$$H_1\colon\ \text{The } \mu_{i.} \text{ are not all equal.}$$

Model for the RCB Design

Each observation may be written in the form

$$y_{ij} = \mu_{ij} + \epsilon_{ij},$$

where ϵ_{ij} measures the deviation of the observed value y_{ij} from the population mean μ_{ij}. The preferred form of this equation is obtained by substituting

$$\mu_{ij} = \mu + \alpha_i + \beta_j,$$

where α_i is, as before, the effect of the ith treatment and β_j is the effect of the jth block. It is assumed that the treatment and block effects are additive. Hence, we may write

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}.$$

Notice that the model resembles that of the one-way classification, the essential difference being the introduction of the block effect β_j. The basic concept is much like that of the one-way classification except that we must account in the analysis for the additional effect due to blocks, since we are now systematically controlling variation *in two directions.* If we now impose the restrictions that

$$\sum_{i=1}^{k} \alpha_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{b} \beta_j = 0,$$

then

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b}(\mu + \alpha_i + \beta_j) = \mu + \alpha_i, \text{ for } i = 1, \ldots, k,$$

and

$$\mu_{.j} = \frac{1}{k}\sum_{i=1}^{k}(\mu + \alpha_i + \beta_j) = \mu + \beta_j, \text{ for } j = 1, \ldots, b.$$

The null hypothesis that the k treatment means $\mu_{i.}$ are equal, and therefore equal to μ, is now **equivalent to testing the hypothesis**

$$H_0\colon\ \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0,$$
$$H_1\colon\ \text{At least one of the } \alpha_i \text{ is not equal to zero.}$$

Each of the tests on treatments will be based on a comparison of independent estimates of the common population variance σ^2. These estimates will be obtained by splitting the total sum of squares of our data into three components by means of the following identity.

Theorem 13.3: **Sum-of-Squares Identity**

$$\sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 = b\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 + k\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

The proof is left to the reader.

The sum-of-squares identity may be presented symbolically by the equation

$$SST = SSA + SSB + SSE,$$

where

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 \quad = \text{total sum of squares,}$$

$$SSA = b\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 \quad = \text{treatment sum of squares,}$$

$$SSB = k\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 \quad = \text{block sum of squares,}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2 = \text{error sum of squares.}$$

Following the procedure outlined in Theorem 13.2, where we interpreted the sums of squares as functions of the independent random variables $Y_{11}, Y_{12}, \ldots, Y_{kb}$, we can show that the expected values of the treatment, block, and error sums of squares are given by

$$E(SSA) = (k-1)\sigma^2 + b\sum_{i=1}^{k}\alpha_i^2, \quad E(SSB) = (b-1)\sigma^2 + k\sum_{j=1}^{b}\beta_j^2,$$
$$E(SSE) = (b-1)(k-1)\sigma^2.$$

As in the case of the one-factor problem, we have the treatment mean square

$$s_1^2 = \frac{SSA}{k-1}.$$

If the treatment effects $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, s_1^2 is an unbiased estimate of σ^2. However, if the treatment effects are not all zero, we have the following:

Expected Treatment Mean Square

$$E\left(\frac{SSA}{k-1}\right) = \sigma^2 + \frac{b}{k-1}\sum_{i=1}^{k}\alpha_i^2$$

In this case, s_1^2 overestimates σ^2. A second estimate of σ^2, based on $b-1$ degrees of freedom, is

$$s_2^2 = \frac{SSB}{b-1}.$$

The estimate s_2^2 is an unbiased estimate of σ^2 if the block effects $\beta_1 = \beta_2 = \cdots = \beta_b = 0$. If the block effects are not all zero, then

$$E\left(\frac{SSB}{b-1}\right) = \sigma^2 + \frac{k}{b-1}\sum_{j=1}^{b}\beta_j^2,$$

and s_2^2 will overestimate σ^2. A third estimate of σ^2, based on $(k-1)(b-1)$ degrees of freedom and independent of s_1^2 and s_2^2, is

$$s^2 = \frac{SSE}{(k-1)(b-1)},$$

which is unbiased regardless of the truth or falsity of either null hypothesis.

To test the null hypothesis that the treatment effects are all equal to zero, we compute the ratio $f_1 = s_1^2/s^2$, which is a value of the random variable F_1 having an F-distribution with $k-1$ and $(k-1)(b-1)$ degrees of freedom when the null hypothesis is true. The null hypothesis is rejected at the α-level of significance when

$$f_1 > f_\alpha[k-1, (k-1)(b-1)].$$

In practice, we first compute SST, SSA, and SSB and then, using the sum-of-squares identity, obtain SSE by subtraction. The degrees of freedom associated with SSE are also usually obtained by subtraction; that is,

$$(k-1)(b-1) = kb - 1 - (k-1) - (b-1).$$

The computations in an analysis-of-variance problem for a randomized complete block design may be summarized as shown in Table 13.8.

Example 13.6: Four different machines, M_1, M_2, M_3, and M_4, are being considered for the assembling of a particular product. It was decided that six different operators would be used in a randomized block experiment to compare the machines. The machines were assigned in a random order to each operator. The operation of the machines requires physical dexterity, and it was anticipated that there would be a difference among the operators in the speed with which they operated the machines. The amounts of time (in seconds) required to assemble the product are shown in Table 13.9.

Test the hypothesis H_0, at the 0.05 level of significance, that the machines perform at the same mean rate of speed.

Table 13.8: Analysis of Variance for the Randomized Complete Block Design

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Treatments	SSA	$k-1$	$s_1^2 = \frac{SSA}{k-1}$	$f_1 = \frac{s_1^2}{s^2}$
Blocks	SSB	$b-1$	$s_2^2 = \frac{SSB}{b-1}$	
Error	SSE	$(k-1)(b-1)$	$s^2 = \frac{SSE}{(k-1)(b-1)}$	
Total	SST	$kb-1$		

Table 13.9: Time, in Seconds, to Assemble Product

	Operator						
Machine	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	42.5	39.3	39.6	39.9	42.9	43.6	247.8
2	39.8	40.1	40.5	42.3	42.5	43.1	248.3
3	40.2	40.5	41.3	43.4	44.9	45.1	255.4
4	41.3	42.2	43.5	44.2	45.9	42.3	259.4
Total	163.8	162.1	164.9	169.8	176.2	174.1	1010.9

***Solution*:** The hypotheses are

$$H_0\text{: } \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \qquad \text{(machine effects are zero),}$$
$$H_1\text{: At least one of the } \alpha_i \text{ is not equal to zero.}$$

The sum-of-squares formulas shown on page 536 and the degrees of freedom are used to produce the analysis of variance in Table 13.10. The value $f = 3.34$ is significant at $P = 0.048$. If we use $\alpha = 0.05$ as at least an approximate yardstick, we conclude that the machines do not perform at the same mean rate of speed. ▟

Table 13.10: Analysis of Variance for the Data of Table 13.9

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Machines	15.93	3	5.31	3.34
Operators	42.09	5	8.42	
Error	23.84	15	1.59	
Total	81.86	23		

Further Comments Concerning Blocking

In Chapter 10, we presented a procedure for comparing means when the observations were *paired*. The procedure involved "subtracting out" the effect due to the homogeneous pair and thus working with differences. This is a special case of a randomized complete block design with $k = 2$ treatments. The n homogeneous units to which the treatments were assigned take on the role of blocks.

If there is heterogeneity in the experimental units, the experimenter should not be misled into believing that it is always advantageous to reduce the experimental error through the use of small homogeneous blocks. Indeed, there may be instances where it would not be desirable to block. The purpose in reducing the error variance is to increase the *sensitivity* of the test for detecting differences in the treatment means. This is reflected in the power of the test procedure. (The power of the analysis-of-variance test procedure is discussed more extensively in Section 13.11.) The power to detect certain differences among the treatment means increases with a decrease in the error variance. However, the power is also affected by the degrees of freedom with which this variance is estimated, and blocking reduces the degrees of freedom that are available from $k(b-1)$ for the one-way classification to $(k-1)(b-1)$. So one could lose power by blocking if there is not a significant reduction in the error variance.

Interaction between Blocks and Treatments

Another important assumption that is implicit in writing the model for a randomized complete block design is that the treatment and block effects are additive. This is equivalent to stating that

$$\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'} \qquad \text{or} \qquad \mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'},$$

for every value of i, i', j, and j'. That is, the difference between the population means for blocks j and j' is the same for every treatment and the difference between the population means for treatments i and i' is the same for every block. The parallel lines of Figure 13.6(a) illustrate a set of mean responses for which the treatment and block effects are additive, whereas the intersecting lines of Figure 13.6(b) show a situation in which treatment and block effects are said to **interact**. Referring to Example 13.6, if operator 3 is 0.5 second faster on the average than operator 2 when machine 1 is used, then operator 3 will still be 0.5 second faster on the average than operator 2 when machine 2, 3, or 4 is used. In many experiments, the assumption of additivity does not hold and the analysis described in this section leads to erroneous conclusions. Suppose, for instance, that operator 3 is 0.5 second faster on the average than operator 2 when machine 1 is used but is 0.2 second slower on the average than operator 2 when machine 2 is used. The operators and machines are now interacting.

An inspection of Table 13.9 suggests the possible presence of interaction. This apparent interaction may be real or it may be due to experimental error. The analysis of Example 13.6 was based on the assumption that the apparent interaction was due entirely to experimental error. If the total variability of our data was in part due to an interaction effect, this source of variation remained a part of the error sum of squares, **causing the mean square error to overestimate $\boldsymbol{\sigma^2}$** and thereby increasing the probability of committing a type II error. We have, in fact, assumed an incorrect model. If we let $(\alpha\beta)_{ij}$ denote the interaction effect of the ith treatment and the jth block, we can write a more appropriate model in the

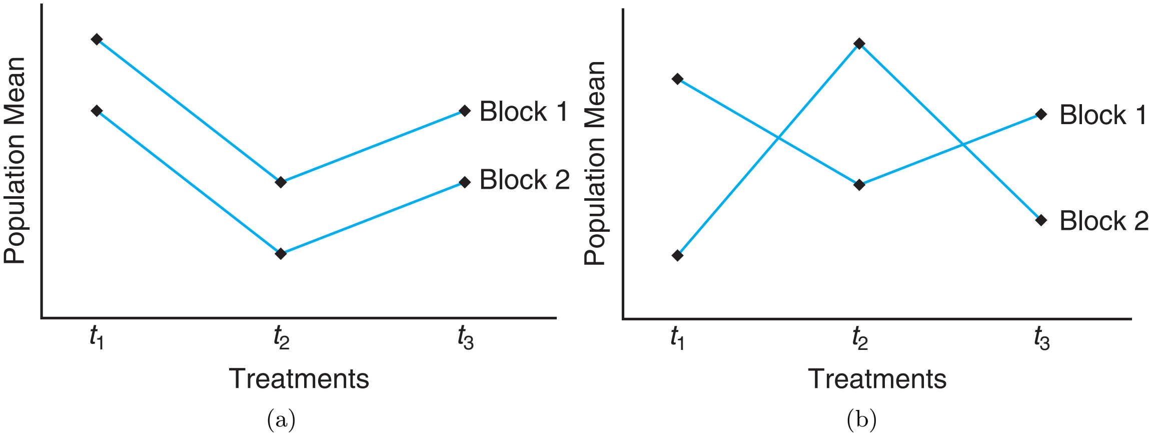


Figure 13.6: Population means for (a) additive results and (b) interacting effects.

form

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij},$$

on which we impose the additional restrictions

$$\sum_{i=1}^{k}(\alpha\beta)_{ij} = \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0, \text{ for } i = 1, \ldots, k \text{ and } j = 1, \ldots, b.$$

We can now readily verify that

$$E\left[\frac{SSE}{(b-1)(k-1)}\right] = \sigma^2 + \frac{1}{(b-1)(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2.$$

Thus, the mean square error is seen to be a **biased estimate of σ^2 when existing interaction has been ignored**. It would seem necessary at this point to arrive at a procedure for the detection of interaction for cases where there is suspicion that it exists. Such a procedure requires the availability of an unbiased and independent estimate of σ^2. Unfortunately, the randomized block design does not lend itself to such a test unless the experimental setup is altered. This subject is discussed extensively in Chapter 14.

13.9 Graphical Methods and Model Checking

In several chapters, we make reference to graphical procedures displaying data and analytical results. In early chapters, we used stem-and-leaf and box-and-whisker plots as visuals to aid in summarizing samples. We used similar diagnostics to better understand the data in two sample problems in Chapter 10. In Chapter 11 we introduced the notion of residual plots to detect violations of standard assumptions. In recent years, much attention in data analysis has centered on **graphical**

methods. Like regression, analysis of variance lends itself to graphics that aid in summarizing data as well as detecting violations. For example, a simple plotting of the raw observations around each treatment mean can give the analyst a feel for variability between sample means and within samples. Figure 13.7 depicts such a plot for the aggregate data of Table 13.1. From the appearance of the plot one may even gain a graphical insight into which aggregates (if any) stand out from the others. It is clear that aggregate 4 stands out from the others. Aggregates 3 and 5 certainly form a homogeneous group, as do aggregates 1 and 2.

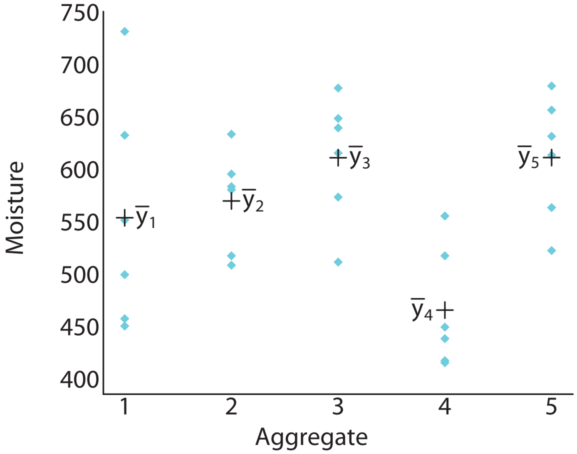


Figure 13.7: Plot of data around the mean for the aggregate data of Table 13.1.

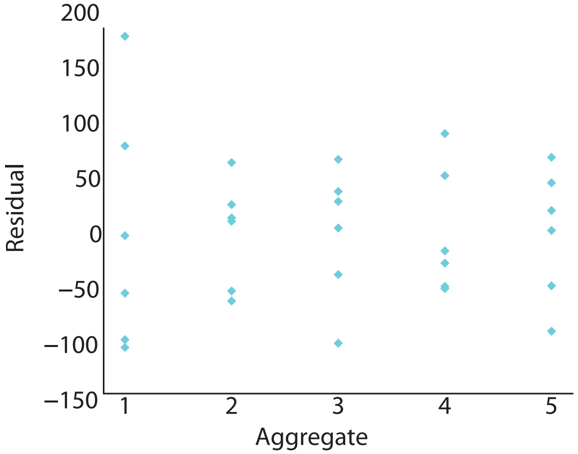


Figure 13.8: Plot of residuals for five aggregates, using data in Table 13.1.

As in the case of regression, residuals can be helpful in analysis of variance in providing a diagnostic that may detect violations of assumptions. To form the residuals, we merely need to consider the model of the one-factor problem, namely

$$y_{ij} = \mu_i + \epsilon_{ij}.$$

It is straightforward to determine that the estimate of μ_i is $\bar{y}_{i.}$. Hence, the ijth residual is $y_{ij} - \bar{y}_{i.}$. This is easily extendable to the randomized complete block model. It may be instructive to have the residuals plotted for each aggregate in order to gain some insight regarding the homogeneous variance assumption. This plot is shown in Figure 13.8.

Trends in plots such as these may reveal difficulties in some situations, particularly when the violation of a particular assumption is graphic. In the case of Figure 13.8, the residuals seem to indicate that the *within-treatment* variances are reasonably homogeneous apart from aggregate 1. There is some graphical evidence that the variance for aggregate 1 is larger than the rest.

What Is a Residual for an RCB Design?

The randomized complete block design is another experimental situation in which graphical displays can make the analyst feel comfortable with an "ideal picture" or perhaps highlight difficulties. Recall that the model for the randomized complete block design is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \qquad i = 1, \dots, k, \qquad j = 1, \dots, b,$$

with the imposed constraints

$$\sum_{i=1}^{k} \alpha_i = 0, \qquad \sum_{j=1}^{b} \beta_j = 0.$$

To determine what indeed constitutes a residual, consider that

$$\alpha_i = \mu_{i.} - \mu, \qquad \beta_j = \mu_{.j} - \mu$$

and that μ is estimated by $\bar{y}_{..}$, $\mu_{i.}$ is estimated by $\bar{y}_{i.}$, and $\mu_{.j}$ is estimated by $\bar{y}_{.j}$. As a result, the predicted or *fitted value* $\hat{y}_{ij}$ is given by

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..},$$

and thus the residual at the (i, j) observation is given by

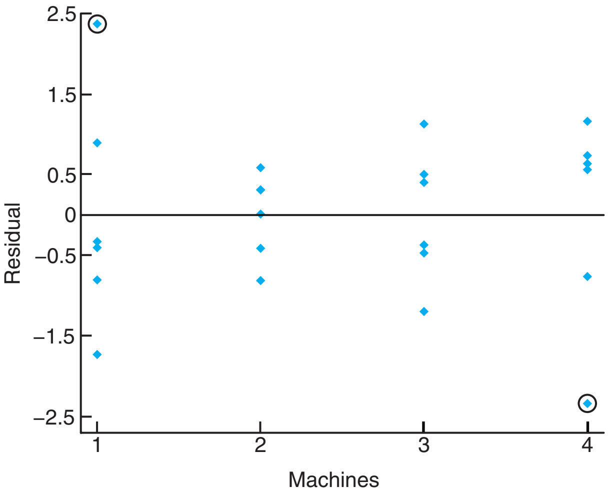
$$y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}.$$

Note that $\hat{y}_{ij}$, the fitted value, is an estimate of the mean μ_{ij}. This is consistent with the partitioning of variability given in Theorem 13.3, where the error sum of squares is

$$SSE = \sum_{i}^{k} \sum_{j}^{b} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2.$$

The visual displays in the randomized complete block design involve plotting the residuals separately for each treatment and for each block. The analyst should expect roughly equal variability if the homogeneous variance assumption holds. The reader should recall that in Chapter 12 we discussed plotting residuals for the purpose of detecting model misspecification. In the case of the randomized complete block design, the serious model misspecification may be related to our assumption of additivity (i.e., no interaction). If no interaction is present, a random pattern should appear.

Consider the data of Example 13.6, in which treatments are four machines and blocks are six operators. Figures 13.9 and 13.10 give the residual plots for separate treatments and separate blocks. Figure 13.11 shows a plot of the residuals against the fitted values. Figure 13.9 reveals that the error variance may not be the same for all machines. The same may be true for error variance for each of the six operators. However, two unusually large residuals appear to produce the apparent difficulty. Figure 13.11 is a plot of residuals that shows reasonable evidence of random behavior. However, the two large residuals displayed earlier still stand out.



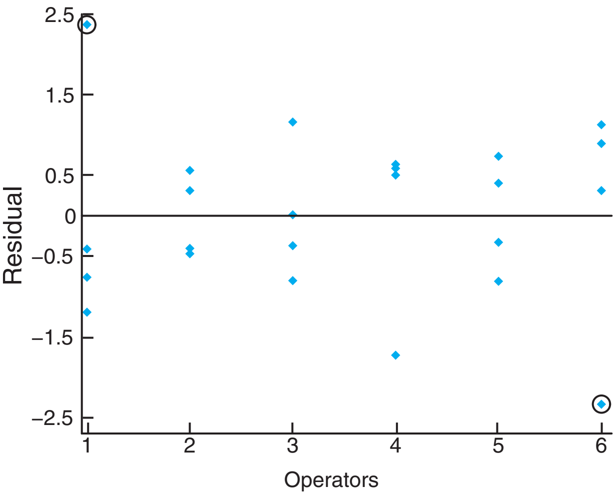


Figure 13.9: Residual plot for the four machines for the data of Example 13.6.

Figure 13.10: Residual plot for the six operators for the data of Example 13.6.

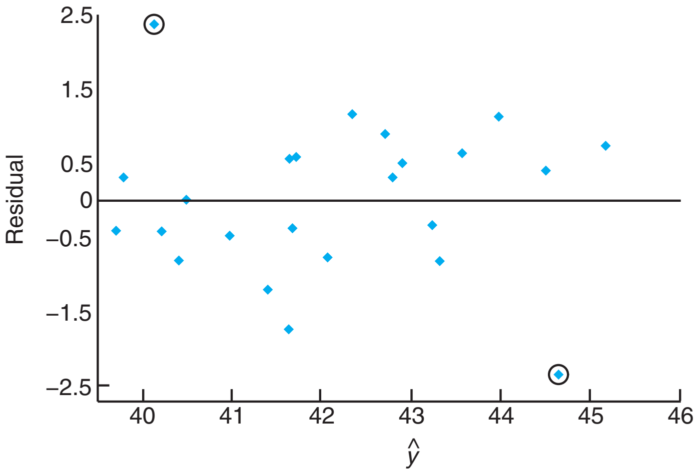


Figure 13.11: Residuals plotted against fitted values for the data of Example 13.6.

13.10 Data Transformations in Analysis of Variance

In Chapter 11, considerable attention was given to transformation of the response y in situations where a linear regression model was being fit to a set of data. Obviously, the same concept applies to multiple linear regression, though it was not discussed in Chapter 12. In the regression modeling discussion, emphasis was placed on the transformations of y that would produce a model that fit the data better than the model in which y enters linearly. For example, if the "time" structure is exponential in nature, then a log transformation on y linearizes the structure and thus more success is anticipated when one uses the transformed response.

While the primary purpose for data transformation discussed thus far has been to improve the fit of the model, there are certainly other reasons to transform or reexpress the response y, and many of them are related to assumptions that are being made (i.e., assumptions on which the validity of the analysis depends). One very important assumption in analysis of variance is the homogeneous variance assumption discussed early in Section 13.4. We assume a **common variance $\boldsymbol{\sigma^2}$**. If the variance differs a great deal from treatment to treatment and we perform the standard ANOVA discussed in this chapter (and future chapters), the results can be substantially flawed. In other words, the analysis of variance is not **robust** to the assumption of homogeneous variance. As we have discussed thus far, this is the centerpiece of motivation for the residual plots discussed in the previous section and illustrated in Figures 13.9, 13.10, and 13.11. These plots allow us to detect nonhomogeneous variance problems. However, what do we do about them? How can we accommodate them?

Where Does Nonhomogeneous Variance Come From?

Often, but not always, nonhomogeneous variance in ANOVA is present because of the distribution of the responses. Now, of course we assume normality in the response. But there certainly are situations in which tests on means are needed even though the distribution of the response is one of the nonnormal distributions discussed in Chapters 5 and 6, such as Poisson, lognormal, exponential, or gamma. ANOVA-type problems certainly exist with count data, time to failure data, and so on.

We demonstrated in Chapters 5 and 6 that, apart from the normal case, the variance of a distribution will often be a function of the mean, say $\sigma_i^2 = g(\mu_i)$. For example, in the Poisson case $\text{Var}(Y_i) = \mu_i = \sigma_i^2$ (i.e., the *variance is equal to the mean*). In the case of the exponential distribution, $\text{Var}(Y_i) = \sigma_i^2 = \mu_i^2$ (i.e., the *variance is equal to the square of the mean*). For the case of the lognormal, a log transformation produces a normal distribution with constant variance σ^2.

The same concepts that we used in Chapter 4 to determine the variance of a nonlinear function can be used as an aid to determine the nature of the *variance stabilizing transformation* $g(y_i)$. Recall that the first order Taylor series expansion of $g(y_i)$ around $y_i = \mu_i$ where $g'(\mu_i) = \left[\frac{\partial g(y_i)}{\partial y_i}\right]_{y_i=\mu_i}$. The transformation function $g(y)$ must be independent of μ in order to suffice as the variance stabilizing transformation. From the above,

$$\text{Var}[g(y_i)] \approx [g'(\mu_i)]^2\sigma_i^2.$$

As a result, $g(y_i)$ must be such that $g'(\mu_i) \propto \frac{1}{\sigma}$. Thus, if we suspect that the response is Poisson distributed, $\sigma_i = \mu_i^{1/2}$, so $g'(\mu_i) \propto \frac{1}{\mu_i^{1/2}}$. Thus, the variance stabilizing transformation is $g(y_i) = y_i^{1/2}$. From this illustration and similar manipulation for the exponential and gamma distributions, we have the following.

Distribution	**Variance Stabilizing Transformations**
Poisson	$g(y) = y^{1/2}$
Exponential	$g(y) = \ln y$
Gamma	$g(y) = \ln y$

Exercises

13.25 Four kinds of fertilizer f_1, f_2, f_3, and f_4 are used to study the yield of beans. The soil is divided into 3 blocks, each containing 4 homogeneous plots. The yields in kilograms per plot and the corresponding treatments are as follows:

Block 1	Block 2	Block 3
$f_1 = 42.7$	$f_3 = 50.9$	$f_4 = 51.1$
$f_3 = 48.5$	$f_1 = 50.0$	$f_2 = 46.3$
$f_4 = 32.8$	$f_2 = 38.0$	$f_1 = 51.9$
$f_2 = 39.3$	$f_4 = 40.2$	$f_3 = 53.5$

Conduct an analysis of variance at the 0.05 level of significance using the randomized complete block model.

13.26 Three varieties of potatoes are being compared for yield. The experiment is conducted by assigning each variety at random to one of 3 equal-size plots at each of 4 different locations. The following yields for varieties A, B, and C, in 100 kilograms per plot, were recorded:

Location 1	Location 2	Location 3	Location 4
B: 13	C: 21	C: 9	A: 11
A: 18	A: 20	B: 12	C: 10
C: 12	B: 23	A: 14	B: 17

Perform a randomized complete block analysis of variance to test the hypothesis that there is no difference in the yielding capabilities of the 3 varieties of potatoes. Use a 0.05 level of significance. Draw conclusions.

13.27 The following data are the percents of foreign additives measured by 5 analysts for 3 similar brands of strawberry jam, A, B, and C:

Analyst 1	Analyst 2	Analyst 3	Analyst 4	Analyst 5
B: 2.7	C: 7.5	B: 2.8	A: 1.7	C: 8.1
C: 3.6	A: 1.6	A: 2.7	B: 1.9	A: 2.0
A: 3.8	B: 5.2	C: 6.4	C: 2.6	B: 4.8

Perform a randomized complete block analysis of variance to test the hypothesis, at the 0.05 level of significance, that the percent of foreign additives is the same for all 3 brands of jam. Which brand of jam appears to have fewer additives?

13.28 The following data represent the final grades obtained by 5 students in mathematics, English, French, and biology:

	Subject			
Student	**Math**	**English**	**French**	**Biology**
1	68	57	73	61
2	83	94	91	86
3	72	81	63	59
4	55	73	77	66
5	92	68	75	87

Test the hypothesis that the courses are of equal difficulty. Use a P-value in your conclusions and discuss your findings.

13.29 In a study on *The Periphyton of the South River, Virginia: Mercury Concentration, Productivity, and Autotropic Index Studies*, conducted by the Department of Environmental Sciences and Engineering at Virginia Tech, the total mercury concentration in periphyton total solids was measured at 6 different stations on 6 different days. Determine whether the mean mercury content is significantly different between the stations by using the following recorded data. Use a P-value and discuss your findings.

	Station					
Date	***CA***	***CB***	***E1***	***E2***	***E3***	***E4***
April 8	0.45	3.24	1.33	2.04	3.93	5.93
June 23	0.10	0.10	0.99	4.31	9.92	6.49
July 1	0.25	0.25	1.65	3.13	7.39	4.43
July 8	0.09	0.06	0.92	3.66	7.88	6.24
July 15	0.15	0.16	2.17	3.50	8.82	5.39
July 23	0.17	0.39	4.30	2.91	5.50	4.29

13.30 A nuclear power facility produces a vast amount of heat, which is usually discharged into aquatic systems. This heat raises the temperature of the aquatic system, resulting in a greater concentration of chlorophyll a, which in turn extends the growing season. To study this effect, water samples were collected monthly at 3 stations for a period of 12 months. Station A is located closest to a potential heated water discharge, station C is located farthest away from the discharge, and station B is located halfway between stations A and C. The following concentrations of chlorophyll a were recorded.

	Station		
Month	***A***	***B***	***C***
January	9.867	3.723	4.410
February	14.035	8.416	11.100
March	10.700	20.723	4.470
April	13.853	9.168	8.010
May	7.067	4.778	34.080
June	11.670	9.145	8.990
July	7.357	8.463	3.350
August	3.358	4.086	4.500
September	4.210	4.233	6.830
October	3.630	2.320	5.800
November	2.953	3.843	3.480
December	2.640	3.610	3.020

Perform an analysis of variance and test the hypothesis, at the 0.05 level of significance, that there is no difference in the mean concentrations of chlorophyll a at the 3 stations.

13.31 In a study conducted by the Department of Health and Physical Education at Virginia Tech, 3 diets were assigned for a period of 3 days to each of 6 subjects in a randomized complete block design. The subjects, playing the role of blocks, were assigned the following 3 diets in a random order:

Diet 1:	mixed fat and carbohydrates,
Diet 2:	high fat,
Diet 3:	high carbohydrates.

At the end of the 3-day period, each subject was put on a treadmill and the time to exhaustion, in seconds, was measured. Perform the analysis of variance, separating out the diet, subject, and error sum of squares. Use a P-value to determine if there are significant differences among the diets, using the following recorded data.

	Subject					
Diet	**1**	**2**	**3**	**4**	**5**	**6**
1	84	35	91	57	56	45
2	91	48	71	45	61	61
3	122	53	110	71	91	122

13.32 Organic arsenicals are used by forestry personnel as silvicides. The amount of arsenic that the body takes in when exposed to these silvicides is a major health problem. It is important that the amount of exposure be determined quickly so that a field worker with a high level of arsenic can be removed from the job. In an experiment reported in the paper "A Rapid Method for the Determination of Arsenic Concentrations in Urine at Field Locations," published in the *American Industrial Hygiene Association Journal* (Vol. 37, 1976), urine specimens from 4 forest service personnel were divided equally into 3 samples each so that each individual's urine could be analyzed for arsenic by a university laboratory, by a chemist using a portable system, and by a forest-service employee after a brief orientation. The following arsenic levels, in parts per million, were recorded:

	Analyst		
Individual	**Employee**	**Chemist**	**Laboratory**
1	0.05	0.05	0.04
2	0.05	0.05	0.04
3	0.04	0.04	0.03
4	0.15	0.17	0.10

Perform an analysis of variance and test the hypothesis, at the 0.05 level of significance, that there is no difference in the arsenic levels for the 3 methods of analysis.

13.33 Scientists in the Department of Plant Pathology at Virginia Tech devised an experiment in which 5 different treatments were applied to 6 different locations in an apple orchard to determine if there were significant differences in growth among the treatments. Treatments 1 through 4 represent different herbicides and treatment 5 represents a control. The growth period was from May to November in 1982, and the amounts of new growth, measured in centimeters, for samples selected from the 6 locations in the orchard were recorded as follows:

	Locations					
Treatment	**1**	**2**	**3**	**4**	**5**	**6**
1	455	72	61	215	695	501
2	622	82	444	170	437	134
3	695	56	50	443	701	373
4	607	650	493	257	490	262
5	388	263	185	103	518	622

Perform an analysis of variance, separating out the treatment, location, and error sum of squares. Determine if there are significant differences among the treatment means. Quote a P-value.

13.34 In the paper "Self-Control and Therapist Control in the Behavioral Treatment of Overweight Women," published in *Behavioral Research and Therapy* (Vol. 10, 1972), two reduction treatments and a control treatment were studied for their effects on the weight change of obese women. The two reduction treatments were a self-induced weight reduction program and a therapist-controlled reduction program. Each of 10 subjects was assigned to one of the 3 treatment programs in a random order and measured for weight loss. The following weight changes were recorded:

	Treatment		
Subject	**Control**	**Self-induced**	**Therapist**
1	1.00	−2.25	−10.50
2	3.75	−6.00	−13.50
3	0.00	−2.00	0.75
4	−0.25	−1.50	−4.50
5	−2.25	−3.25	−6.00
6	−1.00	−1.50	4.00
7	−1.00	−10.75	−12.25
8	3.75	−0.75	−2.75
9	1.50	0.00	−6.75
10	0.50	−3.75	−7.00

Perform an analysis of variance and test the hypothesis, at the 0.01 level of significance, that there is no difference in the mean weight losses for the 3 treatments. Which treatment was best?

13.35 In the book *Design of Experiments for the Quality Improvement*, published by the Japanese Standards Association (1989), a study on the amount of dye needed to get the best color for a certain type of fabric was reported. The three amounts of dye, $\frac{1}{3}$% wof ($\frac{1}{3}$% of the weight of a fabric), 1% wof, and 3% wof, were each administered at two different plants. The color density of the fabric was then observed four times for each level of dye at each plant.

	Amount of Dye					
	1/3%		**1%**		**3%**	
Plant 1	5.2	6.0	12.3	10.5	22.4	17.8
	5.9	5.9	12.4	10.9	22.5	18.4
Plant 2	6.5	5.5	14.5	11.8	29.0	23.2
	6.4	5.9	16.0	13.6	29.7	24.0

Perform an analysis of variance to test the hypothesis, at the 0.05 level of significance, that there is no difference in the color density of the fabric for the three levels of dye. Consider plants to be blocks.

13.36 An experiment was conducted to compare three types of coating materials for copper wire. The purpose of the coating is to eliminate "flaws" in the wire. Ten different specimens of length 5 millimeters were randomly assigned to receive each coating, and the thirty specimens were subjected to an abrasive wear type process. The number of flaws was measured for each, and the results are as follows:

Material											
1				**2**				**3**			
6	8	4	5	3	3	5	4	12	8	7	14
7	7	9	6	2	4	4	5	18	6	7	18
7	8			4	3			8	5		

Suppose it is assumed that the Poisson process applies and thus the model is $Y_{ij} = \mu_i + \epsilon_{ij}$, where μ_i is the mean of a Poisson distribution and $\sigma^2_{Y_{ij}} = \mu_i$.

(a) Do an appropriate transformation on the data and perform an analysis of variance.

(b) Determine whether or not there is sufficient evidence to choose one coating material over the other. Show whatever findings suggest a conclusion.

(c) Do a plot of the residuals and comment.

(d) Give the purpose of your data transformation.

(e) What additional assumption is made here that may not have been completely satisfied by your transformation?

(f) Comment on (e) after doing a normal probability plot on the residuals.

13.11 Random Effects Models

Throughout this chapter, we deal with analysis-of-variance procedures in which the primary goal is to study the effect on some response of certain fixed or predetermined treatments. Experiments in which the treatments or treatment levels are preselected by the experimenter as opposed to being chosen randomly are called **fixed effects experiments**. For the fixed effects model, inferences are made only on those particular treatments used in the experiment.

It is often important that the experimenter be able to draw inferences about a population of treatments by means of an experiment in which the treatments used are chosen randomly from the population. For example, a biologist may be interested in whether or not there is significant variance in some physiological characteristic due to animal type. The animal types actually used in the experiment are then chosen randomly and represent the treatment effects. A chemist may be interested in studying the effect of analytical laboratories on the chemical analysis of a substance. She is not concerned with particular laboratories but rather with a large population of laboratories. She might then select a group of laboratories at random and allocate samples to each for analysis. The statistical inference would then involve (1) testing whether or not the laboratories contribute a nonzero variance to the analytical results and (2) estimating the variance due to laboratories and the variance within laboratories.

Model and Assumptions for Random Effects Model

The one-way **random effects model** is written like the fixed effects model but with the terms taking on different meanings. The response $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ is now a value of the random variable

$$Y_{ij} = \mu + A_i + \epsilon_{ij}, \text{ with } i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n,$$

where the A_i are independently and normally distributed with mean 0 and variance σ_α^2 and are independent of the ϵ_{ij}. As for the fixed effects model, the ϵ_{ij} are also independently and normally distributed with mean 0 and variance σ^2. Note that for a random effects experiment, the constraint that $\sum\limits_{i=1}^{k} \alpha_i = 0$ no longer applies.

Theorem 13.4: For the one-way random effects analysis-of-variance model,

$$E(SSA) = (k-1)\sigma^2 + n(k-1)\sigma_\alpha^2 \qquad \text{and} \qquad E(SSE) = k(n-1)\sigma^2.$$

Table 13.11 shows the expected mean squares for both a fixed effects and a random effects experiment. The computations for a random effects experiment are carried out in exactly the same way as for a fixed effects experiment. That is, the sum-of-squares, degrees-of-freedom, and mean-square columns in an analysis-of-variance table are the same for both models.

Table 13.11: Expected Mean Squares for the One-Factor Experiment

Source of Variation	**Degrees of Freedom**	**Mean Squares**	**Expected Mean Squares**	
			Fixed Effects	**Random Effects**
Treatments	$k-1$	s_1^2	$\sigma^2 + \frac{n}{k-1}\sum\limits_i \alpha_i^2$	$\sigma^2 + n\sigma_\alpha^2$
Error	$k(n-1)$	s^2	σ^2	σ^2
Total	$nk-1$			

For the random effects model, the hypothesis that the treatment effects are all zero is written as follows:

Hypothesis for a Random Effects Experiment

$$H_0\colon\ \sigma_\alpha^2 = 0,$$
$$H_1\colon\ \sigma_\alpha^2 \neq 0.$$

This hypothesis says that the different treatments contribute nothing to the variability of the response. It is obvious from Table 13.11 that s_1^2 and s^2 are both estimates of σ^2 when H_0 is true and that the ratio

$$f = \frac{s_1^2}{s^2}$$

is a value of the random variable F having the F-distribution with $k-1$ and $k(n-1)$ degrees of freedom. The null hypothesis is rejected at the α-level of significance when

$$f > f_\alpha[k-1, k(n-1)].$$

In many scientific and engineering studies, interest is not centered on the F-test. The scientist knows that the random effect does, indeed, have a significant effect. What is more important is estimation of the various variance components. This produces a *ranking* in terms of what factors produce the most variability and by how much. In the present context, it may be of interest to quantify how much larger the *single-factor variance component* is than that produced by chance (random variation).

Estimation of Variance Components

Table 13.11 can also be used to estimate the **variance components** σ^2 and σ_α^2. Since s_1^2 estimates $\sigma^2 + n\sigma_\alpha^2$ and s^2 estimates σ^2,

$$\hat{\sigma}^2 = s^2, \qquad \hat{\sigma}_\alpha^2 = \frac{s_1^2 - s^2}{n}.$$

Example 13.7: The data in Table 13.12 are coded observations on the yield of a chemical process, using five batches of raw material selected randomly. Show that the batch variance component is significantly greater than zero and obtain its estimate.

Table 13.12: Data for Example 13.7

Batch:	**1**	**2**	**3**	**4**	**5**	
	9.7	10.4	15.9	8.6	9.7	
	5.6	9.6	14.4	11.1	12.8	
	8.4	7.3	8.3	10.7	8.7	
	7.9	6.8	12.8	7.6	13.4	
	8.2	8.8	7.9	6.4	8.3	
	7.7	9.2	11.6	5.9	11.7	
	8.1	7.6	9.8	8.1	10.7	
Total	55.6	59.7	80.7	58.4	75.3	329.7

***Solution*:** The total, batch, and error sums of squares are, respectively,

$$SST = 194.64,\ SSA = 72.60,\ \text{and}\ SSE = 194.64 - 72.60 = 122.04.$$

These results, with the remaining computations, are shown in Table 13.13.

Table 13.13: Analysis of Variance for Example 13.7

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed *f***
Batches	72.60	4	18.15	4.46
Error	122.04	30	4.07	
Total	194.64	34		

The f-ratio is significant at the $\alpha = 0.05$ level, indicating that the hypothesis of a zero batch component is rejected. An estimate of the batch variance component is

$$\hat{\sigma}_\alpha^2 = \frac{18.15 - 4.07}{7} = 2.01.$$

Note that while the **batch variance component** is significantly different from zero, when gauged against the estimate of σ^2, namely $\hat{\sigma}^2 = MSE = 4.07$, it appears as if the batch variance component is not appreciably large. ▟

If the result using the formula for σ_α^2 appears negative, (i.e., when s_1^2 is smaller than s^2), $\hat{\sigma}_\alpha^2$ is then set to zero. This is a biased estimator. In order to have a better estimator of σ_α^2, a method called **restricted** (or **residual**) **maximum likelihood (REML)** is commonly used (see Harville, 1977, in the Bibliography). Such an estimator can be found in many statistical software packages. The details for this estimation procedure are beyond the scope of this text.

Randomized Block Design with Random Blocks

In a randomized complete block experiment where the blocks represent days, it is conceivable that the experimenter would like the results to apply not only to the actual days used in the analysis but to every day in the year. He or she would then select at random the days on which to run the experiment as well as the treatments and use the random effects model

$$Y_{ij} = \mu + A_i + B_j + \epsilon_{ij}, \text{ for } i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, b,$$

with the A_i, B_j, and ϵ_{ij} being independent random variables with means 0 and variances σ_α^2, σ_β^2, and σ^2, respectively. The expected mean squares for a random effects randomized complete block design are obtained, using the same procedure as for the one-factor problem, and are presented along with those for a fixed effects experiment in Table 13.14.

Again the computations for the individual sums of squares and degrees of freedom are identical to those of the fixed effects model. The hypothesis

$$H_0\colon\ \sigma_\alpha^2 = 0,$$
$$H_1\colon\ \sigma_\alpha^2 \neq 0$$

is carried out by computing

$$f = \frac{s_1^2}{s^2}$$

Table 13.14: Expected Mean Squares for the Randomized Complete Block Design

Source of Variation	Degrees of Freedom	Mean Squares	Expected Mean Squares	
			Fixed Effects	**Random Effects**
Treatments	$k-1$	s_1^2	$\sigma^2 + \frac{b}{k-1}\sum_i \alpha_i^2$	$\sigma^2 + b\sigma_\alpha^2$
Blocks	$b-1$	s_2^2	$\sigma^2 + \frac{k}{b-1}\sum_j \beta_j^2$	$\sigma^2 + k\sigma_\beta^2$
Error	$(k-1)(b-1)$	s^2	σ^2	σ^2
Total	$kb-1$			

and rejecting H_0 when $f > f_\alpha[k-1,(b-1)(k-1)]$.

The unbiased estimates of the variance components are

$$\hat{\sigma}^2 = s^2, \qquad \hat{\sigma}_\alpha^2 = \frac{s_1^2 - s^2}{b}, \qquad \hat{\sigma}_\beta^2 = \frac{s_2^2 - s^2}{k}.$$

Tests of hypotheses concerning the various variance components are made by computing the ratios of appropriate mean squares, as indicated in Table 13.14, and comparing them with corresponding f-values from Table A.6.

13.12 Case Study

Case Study 13.1: **Chemical Analysis**: Personnel in the Chemistry Department of Virginia Tech were called upon to analyze a data set that was produced to compare 4 different methods of analysis of aluminum in a certain solid igniter mixture. To get a broad range of analytical laboratories involved, 5 laboratories were used in the experiment. These laboratories were selected because they are generally adept in doing these types of analyses. Twenty samples of igniter material containing 2.70% aluminum were assigned randomly, 4 to each laboratory, and directions were given on how to carry out the chemical analysis using all 4 methods. The data retrieved are as follows:

	Laboratory					
Method	**1**	**2**	**3**	**4**	**5**	**Mean**
A	2.67	2.69	2.62	2.66	2.70	2.668
B	2.71	2.74	2.69	2.70	2.77	2.722
C	2.76	2.76	2.70	2.76	2.81	2.758
D	2.65	2.69	2.60	2.64	2.73	2.662

The laboratories are not considered as random effects since they were not selected randomly from a larger population of laboratories. The data were analyzed as a randomized complete block design. Plots of the data were sought to determine if an additive model of the type

$$y_{ij} = \mu + m_i + l_j + \epsilon_{ij}$$

is appropriate: in other words, a model with additive effects. The randomized block is not appropriate when interaction between laboratories and methods exists. Consider the plot shown in Figure 13.12. Although this plot is a bit difficult to interpret because each point is a single observation, there appears to be no appreciable interaction between methods and laboratories.

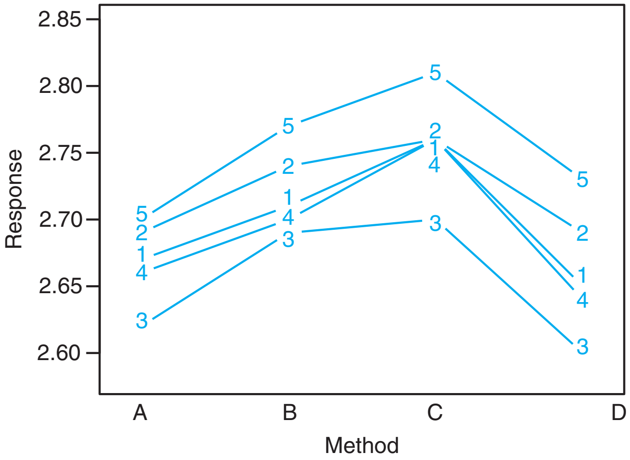
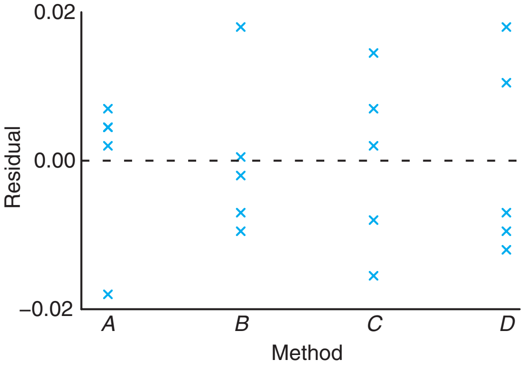


Figure 13.12: Interaction plot for data of Case Study 13.1.

Residual Plots

Residual plots were used as diagnostic indicators regarding the homogeneous variance assumption. Figure 13.13 shows a plot of residuals against analytical methods. The variability depicted in the residuals seems to be remarkably homogeneous. For completeness, a normal probability plot of the residuals is shown in Figure 13.14.



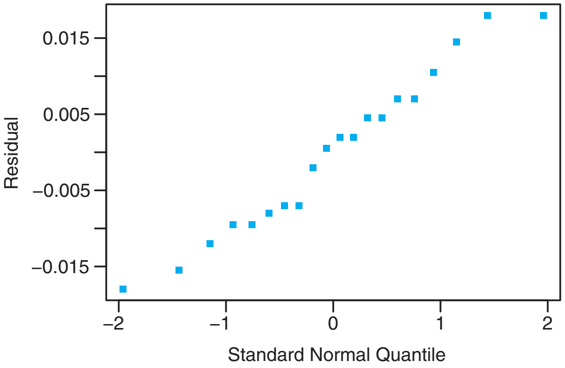


Figure 13.13: Plot of residuals against method for the data of Case Study 13.1.

Figure 13.14: Normal probability plot of residuals for the data of Case Study 13.1.

The residual plots show no difficulty with either the assumption of normal errors or the assumption of homogeneous variance. *SAS* PROC GLM was used to conduct the analysis of variance. Figure 13.15 shows the annotated computer printout.

The computed f- and P-values do indicate a significant difference between analytical methods. This analysis can be followed by a multiple comparison analysis to determine where the differences are among the methods.

Exercises

13.37 Testing patient blood samples for HIV antibodies, a spectrophotometer determines the optical density of each sample. Optical density is measured as the absorbance of light at a particular wavelength. The blood sample is positive if it exceeds a certain cutoff value that is determined by the control samples for that run. Researchers are interested in comparing the laboratory variability for the positive control values. The data represent positive control values for 10 different runs at 4 randomly selected laboratories.

	Laboratory			
Run	**1**	**2**	**3**	**4**
1	0.888	1.065	1.325	1.232
2	0.983	1.226	1.069	1.127
3	1.047	1.332	1.219	1.051
4	1.087	0.958	0.958	0.897
5	1.125	0.816	0.819	1.222
6	0.997	1.015	1.140	1.125
7	1.025	1.071	1.222	0.990
8	0.969	0.905	0.995	0.875
9	0.898	1.140	0.928	0.930
10	1.018	1.051	1.322	0.775

(a) Write an appropriate model for this experiment.

(b) Estimate the laboratory variance component and the variance within laboratories.

13.38 An experiment is conducted in which 4 treatments are to be compared in 5 blocks. The data are given below.

	Block				
Treatment	**1**	**2**	**3**	**4**	**5**
1	12.8	10.6	11.7	10.7	11.0
2	11.7	14.2	11.8	9.9	13.8
3	11.5	14.7	13.6	10.7	15.9
4	12.6	16.5	15.4	9.6	17.1

(a) Assuming a random effects model, test the hypothesis, at the 0.05 level of significance, that there is no difference between treatment means.

(b) Compute estimates of the treatment and block variance components.

13.39 The following data show the effect of 4 operators, chosen randomly, on the output of a particular machine.

Operator			
1	**2**	**3**	**4**
175.4	168.5	170.1	175.2
171.7	162.7	173.4	175.7
173.0	165.0	175.7	180.1
170.5	164.1	170.7	183.7

(a) Perform a random effects analysis of variance at the 0.05 level of significance.

(b) Compute an estimate of the operator variance component and the experimental error variance component.

13.40 Five "pours" of metals have had 5 core samples each analyzed for the amount of a trace element. The data for the 5 randomly selected pours are as follows:

	Pour				
Core	**1**	**2**	**3**	**4**	**5**
1	0.98	0.85	1.12	1.21	1.00
2	1.02	0.92	1.68	1.19	1.21
3	1.57	1.16	0.99	1.32	0.93
4	1.25	1.43	1.26	1.08	0.86
5	1.16	0.99	1.05	0.94	1.41

(a) The intent is that the pours be identical. Thus, test that the "pour" variance component is zero. Draw conclusions.

(b) Show a complete ANOVA along with an estimate of the within-pour variance.

13.41 A textile company weaves a certain fabric on a large number of looms. The managers would like the looms to be homogeneous so that their fabric is of uniform strength. It is suspected that there may be significant variation in strength among looms. Consider the following data for 4 randomly selected looms. Each observation is a determination of strength of the fabric in pounds per square inch.

Loom			
1	**2**	**3**	**4**
99	97	94	93
97	96	95	94
97	92	90	90
96	98	92	92

(a) Write a model for the experiment.

(b) Does the loom variance component differ significantly from zero?

(c) Comment on the managers' suspicion.

```
                        The GLM Procedure
                     Class Level Information
                 Class         Levels    Values
                 Method             4    A B C D
                 Lab                5    1 2 3 4 5
               Number of Observations Read          20
               Number of Observations Used          20
Dependent Variable: Response
                         Sum of
Source          DF      Squares   Mean Square  F Value    Pr > F
Model            7   0.05340500    0.00762929    42.19    <.0001
Error           12   0.00217000    0.00018083
Corrected Total 19   0.05557500

R-Square     Coeff Var      Root MSE    Response Mean
0.960954      0.497592      0.013447         2.702500

Source          DF   Type III SS   Mean Square  F Value     Pr > F
Method           3    0.03145500    0.01048500    57.98     <.0001
Lab              4    0.02195000    0.00548750    30.35     <.0001

Observation       Observed          Predicted          Residual
          1     2.67000000         2.66300000        0.00700000
          2     2.71000000         2.71700000       -0.00700000
          3     2.76000000         2.75300000        0.00700000
          4     2.65000000         2.65700000       -0.00700000
          5     2.69000000         2.68550000        0.00450000
          6     2.74000000         2.73950000        0.00050000
          7     2.76000000         2.77550000       -0.01550000
          8     2.69000000         2.67950000        0.01050000
          9     2.62000000         2.61800000        0.00200000
         10     2.69000000         2.67200000        0.01800000
         11     2.70000000         2.70800000       -0.00800000
         12     2.60000000         2.61200000       -0.01200000
         13     2.66000000         2.65550000        0.00450000
         14     2.70000000         2.70950000       -0.00950000
         15     2.76000000         2.74550000        0.01450000
         16     2.64000000         2.64950000       -0.00950000
         17     2.70000000         2.71800000       -0.01800000
         18     2.77000000         2.77200000       -0.00200000
         19     2.81000000         2.80800000        0.00200000
         20     2.73000000         2.71200000        0.01800000
```

Figure 13.15: *SAS* printout for data of Case Study 13.1.

Review Exercises

13.42 An analysis was conducted by the Statistics Consulting Center at Virginia Tech in conjunction with the Department of Forestry. A certain treatment was applied to a set of tree stumps in which the chemical Garlon was used with the purpose of regenerating the roots of the stumps. A spray was used with four levels of Garlon concentration. After a period of time, the height of the shoots was observed. Perform a one-factor analysis of variance on the following data. Test to see if the concentration of Garlon has a significant impact on the height of the shoots. Use $\alpha = 0.05$.

Garlon Level							
1		**2**		**3**		**4**	
2.87	2.31	3.27	2.66	2.39	1.91	3.05	0.91
3.91	2.04	3.15	2.00	2.89	1.89	2.43	0.01

13.43 Consider the aggregate data of Example 13.1. Perform Bartlett's test, at level $\alpha = 0.1$, to determine if there is heterogeneity of variance among the aggregates.

13.44 Three catalysts are used in a chemical process; a control (no catalyst) is also included. The following are yield data from the process:

	Catalyst		
Control	**1**	**2**	**3**
74.5	77.5	81.5	78.1
76.1	82.0	82.3	80.2
75.9	80.6	81.4	81.5
78.1	84.9	79.5	83.0
76.2	81.0	83.0	82.1

Use Dunnett's test at the $\alpha = 0.01$ level of significance to determine if a significantly higher yield is obtained with the catalysts than with no catalyst.

13.45 Four laboratories are being used to perform chemical analysis. Samples of the same material are sent to the laboratories for analysis as part of a study to determine whether or not they give, on the average, the same results. The analytical results for the four laboratories are as follows:

Laboratory			
A	***B***	***C***	***D***
58.7	62.7	55.9	60.7
61.4	64.5	56.1	60.3
60.9	63.1	57.3	60.9
59.1	59.2	55.2	61.4
58.2	60.3	58.1	62.3

(a) Use Bartlett's test to show that the within-laboratory variances are not significantly different at the $\alpha = 0.05$ level of significance.

(b) Perform the analysis of variance and give conclusions concerning the laboratories.

(c) Do a normal probability plot of residuals.

13.46 An experiment was designed for personnel in the Department of Animal Science at Virginia Tech to study urea and aqueous ammonia treatment of wheat straw. The purpose was to improve nutritional value for male sheep. The diet treatments were control, urea at feeding, ammonia-treated straw, and urea-treated straw. Twenty-four sheep were used in the experiment, and they were separated according to relative weight. There were four sheep in each homogeneous group (by weight) and each of them was given one of the four diets in random order. For each of the 24 sheep, the percent dry matter digested was measured. The data follow.

	Group by Weight (block)					
Diet	**1**	**2**	**3**	**4**	**5**	**6**
Control	32.68	36.22	36.36	40.95	34.99	33.89
Urea at feeding	35.90	38.73	37.55	34.64	37.36	34.35
Ammonia treated	49.43	53.50	52.86	45.00	47.20	49.76
Urea treated	46.58	42.82	45.41	45.08	43.81	47.40

(a) Use a randomized complete block type of analysis to test for differences between the diets. Use $\alpha = 0.05$.

(b) Use Dunnett's test to compare the three diets with the control. Use $\alpha = 0.05$.

(c) Do a normal probability plot of residuals.

13.47 In a study that was analyzed for personnel in the Department of Biochemistry at Virginia Tech, three diets were given to groups of rats in order to study the effect of each on dietary residual zinc in the bloodstream. Five pregnant rats were randomly assigned to each diet group, and each was given the diet on day 22 of pregnancy. The amount of zinc in parts per million was measured. The data are as follows:

	1	0.50	0.42	0.65	0.47	0.44
Diet	2	0.42	0.40	0.73	0.47	0.69
	3	1.06	0.82	0.72	0.72	0.82

Determine if there is a significant difference in residual dietary zinc among the three diets. Use $\alpha = 0.05$. Perform a one-way ANOVA.

13.48 An experiment was conducted to compare three types of paint for evidence of differences in their wearing qualities. They were exposed to abrasive action and the time in hours until abrasion was noticed was observed. Six specimens were used for each type of paint. The data are as follows.

Paint Type								
1			**2**			**3**		
158	97	282	515	264	544	317	662	213
315	220	115	525	330	525	536	175	614

(a) Do an analysis of variance to determine if the evidence suggests that wearing quality differs for the three paints. Use a P-value in your conclusion.

(b) If significant differences are found, characterize what they are. Is there one paint that stands out? Discuss your findings.

(c) Do whatever graphical analysis you need to determine if assumptions used in (a) are valid. Discuss your findings.

(d) Suppose it is determined that the data for each treatment follow an exponential distribution. Does this suggest an alternative analysis? If so, do the alternative analysis and give findings.

13.49 A company that stamps gaskets out of sheets of rubber, plastic, and cork wants to compare the mean number of gaskets produced per hour for the three types of material. Two randomly selected stamping machines are chosen as blocks. The data represent the number of gaskets (in thousands) produced per hour. The data is given below. In addition, the printout analysis is given in Figure 13.16 on page 557.

	Material								
Machine	**Cork**			**Rubber**			**Plastic**		
A	4.31	4.27	4.40	3.36	3.42	3.48	4.01	3.94	3.89
B	3.94	3.81	3.99	3.91	3.80	3.85	3.48	3.53	3.42

(a) Why would the stamping machines be chosen as blocks?

(b) Plot the six means for machine and material combinations.

(c) Is there a single material that is best?

(d) Is there an interaction between treatments and blocks? If so, is the interaction causing any serious difficulty in arriving at a proper conclusion? Explain.

13.50 A study is conducted to compare gas mileage for 3 competing brands of gasoline. Four different automobile models of varying size are randomly selected. The data, in miles per gallon, follow. The order of testing is random for each model.

	Gasoline Brand		
Model	***A***	***B***	***C***
A	32.4	35.6	38.7
B	28.8	28.6	29.9
C	36.5	37.6	39.1
D	34.4	36.2	37.9

(a) Discuss the need for the use of more than a single model of car.

(b) Consider the ANOVA from the *SAS* printout in Figure 13.17 on page 558. Does brand of gasoline matter?

(c) Which brand of gasoline would you select? Consult the result of Duncan's test.

13.51 Four different locations in the northeast were used for collecting ozone measurements in parts per million. Amounts of ozone were collected in 5 samples at each location.

Location			
1	**2**	**3**	**4**
0.09	0.15	0.10	0.10
0.10	0.12	0.13	0.07
0.08	0.17	0.08	0.05
0.08	0.18	0.08	0.08
0.11	0.14	0.09	0.09

(a) Is there sufficient information here to suggest that there are differences in the mean ozone levels across locations? Be guided by a P-value.

(b) If significant differences are found in (a), characterize the nature of the differences. Use whatever methods you have learned.

13.52 Show that the mean square error

$$s^2 = \frac{SSE}{k(n-1)}$$

for the analysis of variance in a one-way classification is an unbiased estimate of σ^2.

13.53 Prove Theorem 13.2.

13.54 Show that the computing formula for SSB, in the analysis of variance of the randomized complete block design, is equivalent to the corresponding term in the identity of Theorem 13.3.

13.55 For the randomized block design with k treatments and b blocks, show that

$$E(SSB) = (b-1)\sigma^2 + k\sum_{j=1}^{b} \beta_j^2.$$

```
                          The GLM Procedure

Dependent Variable: gasket
                                Sum of
 Source            DF         Squares   Mean Square  F Value   Pr > F
 Model              5      1.68122778    0.33624556    76.52   <.0001
 Error             12      0.05273333    0.00439444
 Corrected Total   17      1.73396111
 R-Square     Coeff Var      Root MSE    gasket Mean
 0.969588      1.734095      0.066291       3.822778

 Source            DF   Type III SS    Mean Square  F Value    Pr > F
 material           2    0.81194444     0.40597222    92.38    <.0001
 machine            1    0.10125000     0.10125000    23.04    0.0004
 material*machine   2    0.76803333     0.38401667    87.39    <.0001
  Level of     Level of            ------------gasket-----------
  material     machine      N              Mean          Std Dev
  cork         A            3        4.32666667       0.06658328
  cork         B            3        3.91333333       0.09291573
  plastic      A            3        3.94666667       0.06027714
  plastic      B            3        3.47666667       0.05507571
  rubber       A            3        3.42000000       0.06000000
  rubber       B            3        3.85333333       0.05507571

  Level of            ------------gasket-----------
  material     N              Mean          Std Dev
  cork         6        4.12000000       0.23765521
  plastic      6        3.71166667       0.26255793
  rubber       6        3.63666667       0.24287171

  Level of            ------------gasket-----------
  machine      N              Mean          Std Dev
  A            9        3.89777778       0.39798800
  B            9        3.74777778       0.21376259
```

Figure 13.16: *SAS* printout for Review Exercise 13.49.

```
                            The GLM Procedure
Dependent Variable: MPG
                            Sum of
 Source             DF      Squares   Mean Square   F Value    Pr > F
 Model               5  153.2508333    30.6501667     24.66    0.0006
 Error               6    7.4583333     1.2430556
 Corrected Total    11  160.7091667

 R-Square     Coeff Var      Root MSE      MPG Mean
 0.953591      3.218448      1.114924      34.64167

 Source             DF   Type III SS   Mean Square  F Value    Pr > F
 Model               3   130.3491667    43.4497222    34.95    0.0003
 Brand               2    22.9016667    11.4508333     9.21    0.0148
                 Duncan's Multiple Range Test for MPG
 NOTE: This test controls the Type I comparisonwise error rate, not
 the experimentwise error rate.

                   Alpha                         0.05
                   Error Degrees of Freedom        6
                   Error Mean Square        1.243056

                   Number of Means          2          3
                   Critical Range       1.929      1.999
        Means with the same letter are not significantly different.
               Duncan Grouping          Mean      N    Brand
                              A       36.4000      4    C
                              A
                         B    A       34.5000      4    B
                         B
                         B            33.0250      4    A
```

Figure 13.17: *SAS* printout for Review Exercise 13.50.

13.56 Group Project: It is of interest to determine which type of sports ball can be thrown the longest distance. The competition involves a tennis ball, a baseball, and a softball. Divide the class into teams of five individuals. Each team should design and conduct a separate experiment. Each team should also analyze the data from its own experiment. For a given team, each of the five individuals will throw each ball (after sufficient arm warmup). The experimental response will be the distance (in feet) that the ball is thrown. The data for each team will involve 15 observations. Important points:

(a) This is not a competition among teams. The competition is among the three types of sports balls. One would expect that the conclusion drawn by each team would be similar.

(b) Each team should be gender mixed.

(c) The experimental design for each team should be a randomized complete block design. The five individuals throwing are the blocks.

(d) Be sure to incorporate the appropriate randomization in conducting the experiment.

(e) The results should contain a description of the experiment with an ANOVA table complete with a P-value and appropriate conclusions. Use graphical techniques where appropriate. Use multiple comparisons where appropriate. Draw practical conclusions concerning differences between the ball types. Be thorough.

13.13 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

As in other procedures covered in previous chapters, the analysis of variance is reasonably robust to the normality assumption but less robust to the homogeneous variance assumption. Also we note here that Bartlett's test for equal variance is extremely nonrobust to normality.

This chapter is an extremely pivotal chapter in that it is essentially an "entry level" point for important topics such as design of experiments and analysis of variance. Chapter 14 will concern itself with the same topics, but the expansion will be to more than one factor, with the total analysis further complicated by the interpretation of interaction among factors. There are times when the role of interaction in a scientific experiment is more important than the role of the main factors (main effects). The presence of interaction results in even more emphasis placed on graphical displays. In Chapters 14 and 15, it will be necessary to give more details regarding the randomization process since the number of factor combinations can be large.

This page intentionally left blank

Chapter 14

Factorial Experiments (Two or More Factors)

14.1 Introduction

Consider a situation where it is of interest to study the effects of **two factors**, A and B, on some response. For example, in a chemical experiment, we would like to vary simultaneously the reaction pressure and reaction time and study the effect of each on the yield. In a biological experiment, it is of interest to study the effects of drying time and temperature on the amount of solids (percent by weight) left in samples of yeast. As in Chapter 13, the term **factor** is used in a general sense to denote any feature of the experiment such as temperature, time, or pressure that may be varied from trial to trial. We define the **levels** of a factor to be the actual values used in the experiment.

For each of these cases, it is important to determine not only if each of the two factors has an influence on the response, but also if there is a significant interaction between the two factors. As far as terminology is concerned, the experiment described here is a two-factor experiment and the experimental design may be either a completely randomized design, in which the various treatment combinations are assigned randomly to all the experimental units, or a randomized complete block design, in which factor combinations are assigned randomly within blocks. In the case of the yeast example, the various treatment combinations of temperature and drying time would be assigned randomly to the samples of yeast if we were using a completely randomized design.

Many of the concepts studied in Chapter 13 are extended in this chapter to two and three factors. The main thrust of this material is the use of the completely randomized design with a *factorial experiment.* A factorial experiment in two factors involves experimental trials (or a single trial) with all factor combinations. For example, in the temperature-drying-time example with, say, 3 levels of each and $n = 2$ runs at each of the 9 combinations, we have a *two-factor factorial experiment in a completely randomized design.* Neither factor is a blocking factor; we are interested in how each influences percent solids in the samples and whether or not they interact. The biologist would have available 18 physical samples of material which are experimental units. These would then be assigned randomly to the 18 combinations (9 treatment combinations, each duplicated).

Before we launch into analytical details, sums of squares, and so on, it may be of interest for the reader to observe the obvious connection between what we have described and the situation with the one-factor problem. Consider the yeast experiment. Explanation of degrees of freedom aids the reader or the analyst in visualizing the extension. We should initially view the 9 treatment combinations as if they represented one factor with 9 levels (8 degrees of freedom). Thus, an initial look at degrees of freedom gives

Treatment combinations	8
Error	9
Total	17

Main Effects and Interaction

The experiment could be analyzed as described in the above table. However, the F-test for combinations would probably not give the analyst the information he or she desires, namely, that which considers the role of temperature and drying time. Three drying times have 2 associated degrees of freedom; three temperatures have 2 degrees of freedom. The main factors, temperature and drying time, are called **main effects**. The main effects represent 4 of the 8 degrees of freedom for *factor combinations*. The additional 4 degrees of freedom are associated with *interaction* between the two factors. As a result, the analysis involves

Combinations		8
Temperature	2	
Drying time	2	
Interaction	4	
Error		9
Total		17

Recall from Chapter 13 that factors in an analysis of variance may be viewed as fixed or random, depending on the type of inference desired and how the levels were chosen. Here we must consider fixed effects, random effects, and even cases where effects are mixed. Most attention will be directed toward expected mean squares when we advance to these topics. In the following section, we focus on the concept of interaction.

14.2 Interaction in the Two-Factor Experiment

In the randomized block model discussed previously, it was assumed that one observation on each treatment is taken in each block. If the model assumption is correct, that is, if blocks and treatments are the only real effects and interaction does not exist, the expected value of the mean square error is the experimental error variance σ^2. Suppose, however, that there is interaction occurring between treatments and blocks as indicated by the model

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

of Section 13.8. The expected value of the mean square error is then given as

$$E\left[\frac{SSE}{(b-1)(k-1)}\right] = \sigma^2 + \frac{1}{(b-1)(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2.$$

The treatment and block effects do not appear in the expected mean square error, but the interaction effects do. Thus, if there is interaction in the model, the mean square error reflects variation due to experimental error plus an interaction contribution, and for this experimental plan, there is no way of separating them.

Interaction and the Interpretation of Main Effects

From an experimenter's point of view it should seem necessary to arrive at a significance test on the existence of interaction by separating true error variation from that due to interaction. The main effects, A and B, take on a different meaning in the presence of interaction. In the previous biological example, the effect that drying time has on the amount of solids left in the yeast might very well depend on the temperature to which the samples are exposed. In general, there could be experimental situations in which factor A has a positive effect on the response at one level of factor B, while at a different level of factor B the effect of A is negative. We use the term **positive effect** here to indicate that the yield or response increases as the levels of a given factor increase according to some defined order. In the same sense, a **negative effect** corresponds to a decrease in response for increasing levels of the factor.

Consider, for example, the following data on temperature (factor A at levels t_1, t_2, and t_3 in increasing order) and drying time d_1, d_2, and d_3 (also in increasing order). The response is percent solids. These data are completely hypothetical and given to illustrate a point.

	B			
A	d_1	d_2	d_3	**Total**
t_1	4.4	8.8	5.2	18.4
t_2	7.5	8.5	2.4	18.4
t_3	9.7	7.9	0.8	18.4
Total	21.6	25.2	8.4	55.2

Clearly the effect of temperature on percent solids is positive at the low drying time d_1 but negative for high drying time d_3. This **clear interaction** between temperature and drying time is obviously of interest to the biologist, but, based on the totals of the responses for temperatures t_1, t_2, and t_3, the temperature sum of squares, SSA, will yield a value of zero. We say then that the presence of interaction is **masking** the effect of temperature. Thus, if we consider the average effect of temperature, averaged over drying time, **there is no effect**. This then defines the main effect. But, of course, this is likely not what is pertinent to the biologist.

Before drawing any final conclusions resulting from tests of significance on the main effects and interaction effects, the **experimenter should first observe whether or not the test for interaction is significant**. If interaction is not significant, then the results of the tests on the main effects are meaningful. However, if interaction should be significant, then only those tests on the main effects that turn out to be significant are meaningful. Nonsignificant main effects in the presence of interaction might well be a result of masking and dictate the need to observe the influence of each factor at fixed levels of the other.

A Graphical Look at Interaction

The presence of interaction as well as its scientific impact can be interpreted nicely through the use of **interaction plots**. The plots clearly give a pictorial view of the tendency in the data to show the effect of changing one factor as one moves from one level to another of a second factor. Figure 14.1 illustrates the strong temperature by drying time interaction. The interaction is revealed in nonparallel lines.

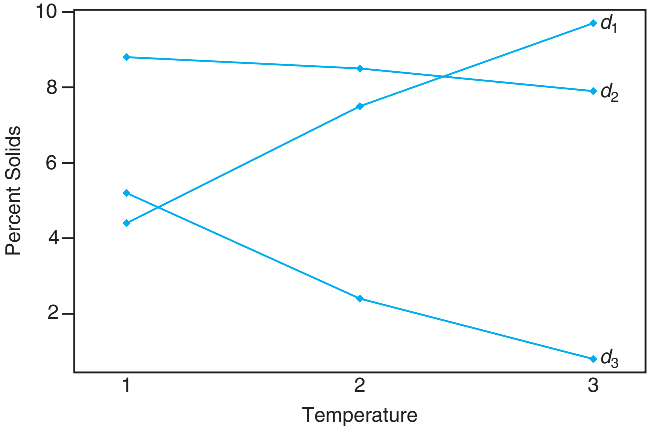


Figure 14.1: Interaction plot for temperature–drying time data.

The relatively strong *temperature effect* on percent solids at the lower drying time is reflected in the steep slope at d_1. At the middle drying time d_2 the temperature has very little effect, while at the high drying time d_3 the negative slope illustrates a negative effect of temperature. Interaction plots such as this set give the scientist a quick and meaningful interpretation of the interaction that is present. It should be apparent that **parallelism** in the plots signals an **absence of interaction**.

Need for Multiple Observations

Interaction and experimental error are separated in the two-factor experiment only if multiple observations are taken at the various treatment combinations. For maximum efficiency, there should be the same number n of observations at each combination. These should be true replications, not just repeated measurements. For example, in the yeast illustration, if we take $n = 2$ observations at each combination of temperature and drying time, there should be two separate samples and not merely repeated measurements on the same sample. This allows variability due to experimental units to appear in “error,” so the variation is not merely measurement error.

14.3 Two-Factor Analysis of Variance

To present general formulas for the analysis of variance of a two-factor experiment using repeated observations in a completely randomized design, we shall consider the case of n replications of the treatment combinations determined by a levels of factor A and b levels of factor B. The observations may be classified by means of a rectangular array where the rows represent the levels of factor A and the columns represent the levels of factor B. Each treatment combination defines a cell in our array. Thus, we have ab cells, each cell containing n observations. Denoting the kth observation taken at the ith level of factor A and the jth level of factor B by y_{ijk}, Table 14.1 shows the abn observations.

Table 14.1: Two-Factor Experiment with n Replications

	B					
A	**1**	**2**	$\cdots$	***b***	**Total**	**Mean**
1	y_{111}	y_{121}	$\cdots$	y_{1b1}	$Y_{1..}$	$\bar{y}_{1..}$
	y_{112}	y_{122}	$\cdots$	y_{1b2}		
	$\vdots$	$\vdots$		$\vdots$		
	y_{11n}	y_{12n}	$\cdots$	y_{1bn}		
2	y_{211}	y_{221}	$\cdots$	y_{2b1}	$Y_{2..}$	$\bar{y}_{2..}$
	y_{212}	y_{222}	$\cdots$	y_{2b2}		
	$\vdots$	$\vdots$		$\vdots$		
	y_{21n}	y_{22n}	$\cdots$	y_{2bn}		
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
a	y_{a11}	y_{a21}	$\cdots$	y_{ab1}	$Y_{a..}$	$\bar{y}_{a..}$
	y_{a12}	y_{a22}	$\cdots$	y_{ab2}		
	$\vdots$	$\vdots$		$\vdots$		
	y_{a1n}	y_{a2n}	$\cdots$	y_{abn}		
Total	$Y_{.1.}$	$Y_{.2.}$	$\cdots$	$Y_{.b.}$	$Y_{...}$	
Mean	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	$\cdots$	$\bar{y}_{.b.}$		$\bar{y}_{...}$

The observations in the (ij)th cell constitute a random sample of size n from a population that is assumed to be normally distributed with mean μ_{ij} and variance σ^2. All ab populations are assumed to have the same variance σ^2. Let us define the following useful symbols, some of which are used in Table 14.1:

$$
\begin{aligned}
Y_{ij.} &= \text{sum of the observations in the } (ij)\text{th cell},\\
Y_{i..} &= \text{sum of the observations for the } i\text{th level of factor } A,\\
Y_{.j.} &= \text{sum of the observations for the } j\text{th level of factor } B,\\
Y_{...} &= \text{sum of all } abn \text{ observations},\\
\bar{y}_{ij.} &= \text{mean of the observations in the } (ij)\text{th cell},\\
\bar{y}_{i..} &= \text{mean of the observations for the } i\text{th level of factor } A,\\
\bar{y}_{.j.} &= \text{mean of the observations for the } j\text{th level of factor } B,\\
\bar{y}_{...} &= \text{mean of all } abn \text{ observations}.
\end{aligned}
$$

Unlike in the one-factor situation covered at length in Chapter 13, here we are assuming that the **populations**, where n independent identically distributed observations are taken, are **combinations** of factors. Also we will assume throughout that an equal number (n) of observations are taken at each factor combination. In cases in which the sample sizes per combination are unequal, the computations are more complicated but the concepts are transferable.

Model and Hypotheses for the Two-Factor Problem

Each observation in Table 14.1 may be written in the form

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

where ϵ_{ijk} measures the deviations of the observed y_{ijk} values in the (ij)th cell from the population mean μ_{ij}. If we let $(\alpha\beta)_{ij}$ denote the interaction effect of the ith level of factor A and the jth level of factor B, α_i the effect of the ith level of factor A, β_j the effect of the jth level of factor B, and μ the overall mean, we can write

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij},$$

and then

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

on which we impose the restrictions

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \sum_{j=1}^{b} \beta_j = 0, \qquad \sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \qquad \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0.$$

The three hypotheses to be tested are as follows:

1. H_0': $\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$,
H_1': At least one of the α_i is not equal to zero.

2. H_0'': $\beta_1 = \beta_2 = \cdots = \beta_b = 0$,
H_1'': At least one of the β_j is not equal to zero.

3. $H_0''': \ (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{ab} = 0,$
$H_1''':$ At least one of the $(\alpha\beta)_{ij}$ is not equal to zero.

We warned the reader about the problem of masking of main effects when interaction is a heavy contributor in the model. It is recommended that the interaction test result be considered first. The interpretation of the main effect test follows, and the nature of the scientific conclusion depends on whether interaction is found. If interaction is ruled out, then hypotheses 1 and 2 above can be tested and the interpretation is quite simple. However, if interaction is found to be present the interpretation can be more complicated, as we have seen from the discussion of the drying time and temperature in the previous section. In what follows, the structure of the tests of hypotheses 1, 2, and 3 will be discussed. Interpretation of results will be incorporated in the discussion of the analysis in Example 14.1.

The tests of the hypotheses above will be based on a comparison of independent estimates of σ^2 provided by splitting the total sum of squares of our data into four components by means of the following identity.

Partitioning of Variability in the Two-Factor Case

Theorem 14.1: **Sum-of-Squares Identity**

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2$$
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

Symbolically, we write the sum-of-squares identity as

$$SST = SSA + SSB + SS(AB) + SSE,$$

where SSA and SSB are called the sums of squares for the main effects A and B, respectively, $SS(AB)$ is called the interaction sum of squares for A and B, and SSE is the error sum of squares. The degrees of freedom are partitioned according to the identity

$$abn - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1).$$

Formation of Mean Squares

If we divide each of the sums of squares on the right side of the sum-of-squares identity by its corresponding number of degrees of freedom, we obtain the four statistics

$$S_1^2 = \frac{SSA}{a-1}, \qquad S_2^2 = \frac{SSB}{b-1}, \qquad S_3^2 = \frac{SS(AB)}{(a-1)(b-1)}, \qquad S^2 = \frac{SSE}{ab(n-1)}.$$

All of these variance estimates are independent estimates of σ^2 under the condition that there are no effects α_i, β_j, and, of course, $(\alpha\beta)_{ij}$. If we interpret the sums of squares as functions of the independent random variables $y_{111}, y_{112}, \ldots, y_{abn}$, it is not difficult to verify that

$$\begin{aligned}
E(S_1^2) &= E\left[\frac{SSA}{a-1}\right] = \sigma^2 + \frac{nb}{a-1}\sum_{i=1}^{a}\alpha_i^2,\\
E(S_2^2) &= E\left[\frac{SSB}{b-1}\right] = \sigma^2 + \frac{na}{b-1}\sum_{j=1}^{b}\beta_j^2,\\
E(S_3^2) &= E\left[\frac{SS(AB)}{(a-1)(b-1)}\right] = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2,\\
E(S^2) &= E\left[\frac{SSE}{ab(n-1)}\right] = \sigma^2,
\end{aligned}$$

from which we immediately observe that all four estimates of σ^2 are unbiased when H_0', H_0'', and H_0''' are true.

To test the hypothesis H_0', that the effects of factors A are all equal to zero, we compute the following ratio:

F-Test for Factor A

$$f_1 = \frac{s_1^2}{s^2},$$

which is a value of the random variable F_1 having the F-distribution with $a-1$ and $ab(n-1)$ degrees of freedom when H_0' is true. The null hypothesis is rejected at the α-level of significance when $f_1 > f_\alpha[a-1, ab(n-1)]$.

Similarly, to test the hypothesis H_0'' that the effects of factor B are all equal to zero, we compute the following ratio:

F-Test for Factor B

$$f_2 = \frac{s_2^2}{s^2},$$

which is a value of the random variable F_2 having the F-distribution with $b-1$ and $ab(n-1)$ degrees of freedom when H_0'' is true. This hypothesis is rejected at the α-level of significance when $f_2 > f_\alpha[b-1, ab(n-1)]$.

Finally, to test the hypothesis H_0''', that the interaction effects are all equal to zero, we compute the following ratio:

F-Test for Interaction

$$f_3 = \frac{s_3^2}{s^2},$$

which is a value of the random variable F_3 having the F-distribution with $(a-1)(b-1)$ and $ab(n-1)$ degrees of freedom when H_0''' is true. We conclude that, at the α-level of significance, interaction is present when $f_3 > f_\alpha[(a-1)(b-1), ab(n-1)]$.

As indicated in Section 14.2, it is advisable to interpret the test for interaction before attempting to draw inferences on the main effects. If interaction is not significant, there is certainly evidence that the tests on main effects are interpretable. Rejection of hypothesis 1 on page 566 implies that the response means at the levels of factor A are significantly different, while rejection of hypothesis 2 implies a similar condition for the means at levels of factor B. However, a significant interaction could very well imply that the data should be analyzed in a somewhat different manner—**perhaps observing the effect of factor A at fixed levels of factor B**, and so forth.

The computations in an analysis-of-variance problem, for a two-factor experiment with n replications, are usually summarized as in Table 14.2.

Table 14.2: Analysis of Variance for the Two-Factor Experiment with n Replications

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Main effect:				
A	SSA	$a-1$	$s_1^2 = \frac{SSA}{a-1}$	$f_1 = \frac{s_1^2}{s^2}$
B	SSB	$b-1$	$s_2^2 = \frac{SSB}{b-1}$	$f_2 = \frac{s_2^2}{s^2}$
Two-factor interactions:				
AB	$SS(AB)$	$(a-1)(b-1)$	$s_3^2 = \frac{SS(AB)}{(a-1)(b-1)}$	$f_3 = \frac{s_3^2}{s^2}$
Error	SSE	$ab(n-1)$	$s^2 = \frac{SSE}{ab(n-1)}$	
Total	SST	$abn-1$		

Example 14.1: In an experiment conducted to determine which of 3 different missile systems is preferable, the propellant burning rate for 24 static firings was measured. Four different propellant types were used. The experiment yielded duplicate observations of burning rates at each combination of the treatments.

The data, after coding, are given in Table 14.3. Test the following hypotheses: (a) $H_0^{'}$: there is no difference in the mean propellant burning rates when different missile systems are used, (b) $H_0^{''}$: there is no difference in the mean propellant burning rates of the 4 propellant types, (c) $H_0^{'''}$: there is no interaction between the different missile systems and the different propellant types.

Table 14.3: Propellant Burning Rates

Missile	**Propellant Type**			
System	b_1	b_2	b_3	b_4
a_1	34.0	30.1	29.8	29.0
	32.7	32.8	26.7	28.9
a_2	32.0	30.2	28.7	27.6
	33.2	29.8	28.1	27.8
a_3	28.4	27.3	29.7	28.8
	29.3	28.9	27.3	29.1

Solution: 1. (a) $H_0^{'}$: $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

(b) $H_0^{''}$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.

(c) $H_0^{'''}$: $(\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{34} = 0$.

2. (a) $H_1^{'}$: At least one of the α_i is not equal to zero.
   (b) $H_1^{''}$: At least one of the β_j is not equal to zero.
   (c) $H_1^{'''}$: At least one of the $(\alpha\beta)_{ij}$ is not equal to zero.

The sum-of-squares formula is used as described in Theorem 14.1. The analysis of variance is shown in Table 14.4.

Table 14.4: Analysis of Variance for the Data of Table 14.3

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed *f***
Missile system	14.52	2	7.26	5.84
Propellant type	40.08	3	13.36	10.75
Interaction	22.16	6	3.69	2.97
Error	14.91	12	1.24	
Total	91.68	23		

The reader is directed to a *SAS* GLM Procedure (General Linear Models) for analysis of the burning rate data in Figure 14.2. Note how the "model" (11 degrees of freedom) is initially tested and the system, type, and system by type interaction are tested separately. The F-test on the model ($P = 0.0030$) is testing the accumulation of the two main effects and the interaction.

(a) Reject $H_0^{'}$ and conclude that different missile systems result in different mean propellant burning rates. The P-value is approximately 0.0169.

(b) Reject $H_0^{''}$ and conclude that the mean propellant burning rates are not the same for the four propellant types. The P-value is approximately 0.0010.

(c) Interaction is barely insignificant at the 0.05 level, but the P-value of approximately 0.0513 would indicate that interaction must be taken seriously.

At this point we should draw some type of interpretation of the interaction. It should be emphasized that statistical significance of a main effect merely implies that *marginal means are significantly different.* However, consider the two-way table of averages in Table 14.5.

Table 14.5: Interpretation of Interaction

	b_1	b_2	b_3	b_4	**Average**
a_1	33.35	31.45	28.25	28.95	30.50
a_2	32.60	30.00	28.40	27.70	29.68
a_3	28.85	28.10	28.50	28.95	28.60
Average	31.60	29.85	28.38	28.53	

It is apparent that more important information exists in the body of the table—trends that are inconsistent with the trend depicted by marginal averages. Table 14.5 certainly suggests that the effect of propellant type depends on the system

```
                           The GLM Procedure
Dependent Variable: rate
                                  Sum of
 Source              DF          Squares    Mean Square    F Value    Pr > F
 Model               11      76.76833333     6.97893939       5.62    0.0030
 Error               12      14.91000000     1.24250000
 Corrected Total     23      91.67833333

 R-Square     Coeff Var      Root MSE     rate Mean
 0.837366      3.766854      1.114675      29.59167

 Source              DF      Type III SS    Mean Square   F Value     Pr > F
 system               2      14.52333333     7.26166667      5.84     0.0169
 type                 3      40.08166667    13.36055556     10.75     0.0010
 system*type          6      22.16333333     3.69388889      2.97     0.0512
```

Figure 14.2: *SAS* printout of the analysis of the propellant rate data of Table 14.3.

being used. For example, for system 3 the propellant-type effect does not appear to be important, although it does have a large effect if either system 1 or system 2 is used. This explains the "significant" interaction between these two factors. More will be revealed subsequently concerning this interaction.

Example 14.2: Referring to Example 14.1, choose two orthogonal contrasts to partition the sum of squares for the missile systems into single-degree-of-freedom components to be used in comparing systems 1 and 2 versus 3, and system 1 versus system 2.

***Solution*:** The contrast for comparing systems 1 and 2 with 3 is

$$w_1 = \mu_{1.} + \mu_{2.} - 2\mu_{3.}.$$

A second contrast, orthogonal to w_1, for comparing system 1 with system 2, is given by $w_2 = \mu_{1.} - \mu_{2.}$. The single-degree-of-freedom sums of squares are

$$SSw_1 = \frac{[244.0 + 237.4 - (2)(228.8)]^2}{(8)[(1)^2 + (1)^2 + (-2)^2]} = 11.80$$

and

$$SSw_2 = \frac{(244.0 - 237.4)^2}{(8)[(1)^2 + (-1)^2]} = 2.72.$$

Notice that $SSw_1 + SSw_2 = SSA$, as expected. The computed f-values corresponding to w_1 and w_2 are, respectively,

$$f_1 = \frac{11.80}{1.24} = 9.5 \qquad \text{and} \qquad f_2 = \frac{2.72}{1.24} = 2.2.$$

Compared to the critical value $f_{0.05}(1, 12) = 4.75$, we find f_1 to be significant. In fact, the P-value is less than 0.01. Thus, the first contrast indicates that the hypothesis

$$H_0\colon\ \frac{1}{2}(\mu_{1.} + \mu_{2.}) = \mu_{3.}$$

is rejected. Since $f_2 < 4.75$, the mean burning rates of the first and second systems are not significantly different.

Impact of Significant Interaction in Example 14.1

If the hypothesis of no interaction in Example 14.1 is true, we could make the *general* comparisons of Example 14.2 regarding our missile systems rather than separate comparisons for each propellant. Similarly, we might make general comparisons among the propellants rather than separate comparisons for each missile system. For example, we could compare propellants 1 and 2 with 3 and 4 and also propellant 1 versus propellant 2. The resulting f-ratios, each with 1 and 12 degrees of freedom, turn out to be 24.81 and 7.39, respectively, and both are quite significant at the 0.05 level.

From propellant averages there appears to be evidence that propellant 1 gives the highest mean burning rate. A prudent experimenter might be somewhat cautious in drawing overall conclusions in a problem such as this one, where the f-ratio for interaction is barely below the 0.05 critical value. For example, the overall evidence, 31.60 versus 29.85 on the average for the two propellants, certainly indicates that propellant 1 is superior, in terms of a higher burning rate, to propellant 2. However, if we restrict ourselves to system 3, where we have an average of 28.85 for propellant 1 as opposed to 28.10 for propellant 2, there appears to be little or no difference between these two propellants. In fact, there appears to be a stabilization of burning rates for the different propellants if we operate with system 3. There is certainly overall evidence which indicates that system 1 gives a higher burning rate than system 3, but if we restrict ourselves to propellant 4, this conclusion does not appear to hold.

The analyst can conduct a simple t-test using average burning rates for system 3 in order to display conclusive evidence that interaction is *producing considerable difficulty in allowing broad conclusions on main effects.* Consider a comparison of propellant 1 against propellant 2 only using system 3. Borrowing an estimate of σ^2 from the overall analysis, that is, using $s^2 = 1.24$ with 12 degrees of freedom, we have

$$|t| = \frac{0.75}{\sqrt{2s^2/n}} = \frac{0.75}{\sqrt{1.24}} = 0.67,$$

which is not even close to being significant. This illustration suggests that one must be cautious about strict interpretation of main effects in the presence of interaction.

Graphical Analysis for the Two-Factor Problem of Example 14.1

Many of the same types of graphical displays that were suggested in the one-factor problems certainly apply in the two-factor case. Two-dimensional plots of cell means or treatment combination means can provide insight into the presence of interactions between the two factors. In addition, a plot of residuals against fitted values may well provide an indication of whether or not the homogeneous variance assumption holds. Often, of course, a violation of the homogeneous variance assumption involves an increase in the error variance as *the response numbers get larger*. As a result, this plot may point out the violation.

Figure 14.3 shows the plot of cell means in the case of the missile system propellant illustration in Example 14.1. Notice how graphically (in this case) the lack of parallelism shows through. Note the flatness of the part of the figure showing the propellant effect for system 3. This illustrates interaction among the factors. Figure 14.4 shows the plot of residuals against fitted values for the same data. There is no apparent sign of difficulty with the homogeneous variance assumption.

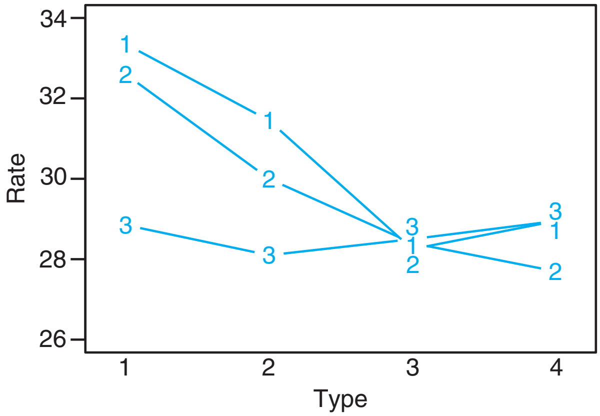


Figure 14.3: Plot of cell means for data of Example 14.1. Numbers represent missile systems.

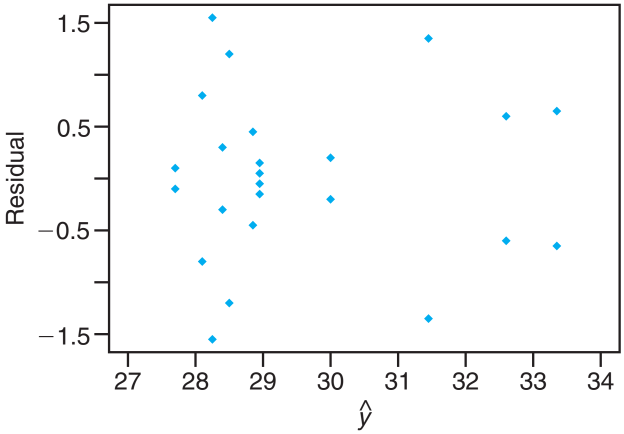


Figure 14.4: Residual plot of data of Example 14.1.

Example 14.3: An electrical engineer is investigating a plasma etching process used in semiconductor manufacturing. It is of interest to study the effects of two factors, the C_2F_6 gas flow rate (A) and the power applied to the cathode (B). The response is the etch rate. Each factor is run at 3 levels, and 2 experimental runs on etch rate are made for each of the 9 combinations. The setup is that of a completely randomized design. The data are given in Table 14.6. The etch rate is in A°/min.

Table 14.6: Data for Example 14.3

	Power Supplied		
C_2F_6 Flow Rate	**1**	**2**	**3**
1	288	488	670
	360	465	720
2	385	482	692
	411	521	724
3	488	595	761
	462	612	801

The levels of the factors are in ascending order, with level 1 being low level and level 3 being the highest.

(a) Show an analysis of variance table and draw conclusions, beginning with the test on interaction.

(b) Do tests on main effects and draw conclusions.

***Solution*:** A *SAS* output is given in Figure 14.5. From the output we learn the following.

```
                          The GLM Procedure
Dependent Variable: etchrate
                                 Sum of
 Source              DF         Squares    Mean Square    F Value    Pr > F
 Model                8     379508.7778     47438.5972      61.00    <.0001
 Error                9       6999.5000       777.7222
 Corrected Total     17     386508.2778

 R-Square     Coeff Var      Root MSE    etchrate Mean
 0.981890      5.057714      27.88767         551.3889

 Source              DF     Type III SS    Mean Square    F Value    Pr > F
 c2f6                 2      46343.1111     23171.5556      29.79    0.0001
 power                2     330003.4444    165001.7222     212.16    <.0001
 c2f6*power           4       3162.2222       790.5556       1.02    0.4485
```

Figure 14.5: *SAS* printout for Example 14.3.

(a) The P-value for the test of interaction is 0.4485. We can conclude that there is no significant interaction.

(b) There is a significant difference in mean etch rate for the 3 levels of C_2F_6 flow rate. Duncan's test shows that the mean etch rate for level 3 is significantly higher than that for level 2 and the rate for level 2 is significantly higher than that for level 1. See Figure 14.6(a).

There is a significant difference in mean etch rate based on the level of power to the cathode. Duncan's test revealed that the etch rate for level 3 is significantly higher than that for level 2 and the rate for level 2 is significantly higher than that for level 1. See Figure 14.6(b).

```
Duncan Grouping      Mean   N  c2f6          Duncan Grouping      Mean   N  power
              A    619.83   6  3                           A    728.00   6  3
              B    535.83   6  2                           B    527.17   6  2
              C    498.50   6  1                           C    399.00   6  1
                (a)                                          (b)
```

Figure 14.6: *SAS* output, for Example 14.3. (a) Duncan's test on gas flow rate; (b) Duncan's test on power.

Exercises

14.1 An experiment was conducted to study the effects of temperature and type of oven on the life of a particular component. Four types of ovens and 3 temperature levels were used in the experiment. Twenty-four pieces were assigned randomly, two to each combination of treatments, and the following results recorded.

	Oven			
Temperature (°F)	O_1	O_2	O_3	O_4
500	227	214	225	260
	221	259	236	229
550	187	181	232	246
	208	179	198	273
600	174	198	178	206
	202	194	213	219

Using a 0.05 level of significance, test the hypothesis that

(a) different temperatures have no effect on the life of the component;

(b) different ovens have no effect on the life of the component;

(c) the type of oven and temperature do not interact.

14.2 To ascertain the stability of vitamin C in reconstituted frozen orange juice concentrate stored in a refrigerator for a period of up to one week, the study *Vitamin C Retention in Reconstituted Frozen Orange Juice* was conducted by the Department of Human Nutrition and Foods at Virginia Tech. Three types of frozen orange juice concentrate were tested using 3 different time periods. The time periods refer to the number of days from when the orange juice was blended until it was tested. The results, in milligrams of ascorbic acid per liter, were recorded. Use a 0.05 level of significance to test the hypothesis that

(a) there is no difference in ascorbic acid contents among the different brands of orange juice concentrate;

(b) there is no difference in ascorbic acid contents for the different time periods;

(c) the brands of orange juice concentrate and the number of days from the time the juice was blended until it was tested do not interact.

	Time (days)					
Brand	**0**		**3**		**7**	
Richfood	52.6	54.2	49.4	49.2	42.7	48.8
	49.8	46.5	42.8	53.2	40.4	47.6
Sealed-Sweet	56.0	48.0	48.8	44.0	49.2	44.0
	49.6	48.4	44.0	42.4	42.0	43.2
Minute Maid	52.5	52.0	48.0	47.0	48.5	43.3
	51.8	53.6	48.2	49.6	45.2	47.6

14.3 Three strains of rats were studied under 2 environmental conditions for their performance in a maze test. The error scores for the 48 rats were recorded.

	Strain					
Environment	**Bright**		**Mixed**		**Dull**	
Free	28	12	33	83	101	94
	22	23	36	14	33	56
	25	10	41	76	122	83
	36	86	22	58	35	23
Restricted	72	32	60	89	136	120
	48	93	35	126	38	153
	25	31	83	110	64	128
	91	19	99	118	87	140

Use a 0.01 level of significance to test the hypothesis that

(a) there is no difference in error scores for different environments;

(b) there is no difference in error scores for different strains;

(c) the environments and strains of rats do not interact.

14.4 Corrosion fatigue in metals has been defined as the simultaneous action of cyclic stress and chemical attack on a metal structure. A widely used technique for minimizing corrosion fatigue damage in aluminum involves the application of a protective coating. A study conducted by the Department of Mechanical Engineering at Virginia Tech used 3 different levels of humidity

Low: 20–25% relative humidity

Medium: 55–60% relative humidity

High: 86–91% relative humidity

and 3 types of surface coatings

Uncoated: no coating

Anodized: sulfuric acid anodic oxide coating

Conversion: chromate chemical conversion coating

The corrosion fatigue data, expressed in thousands of cycles to failure, were recorded as follows:

	Relative Humidity					
Coating	**Low**		**Medium**		**High**	
	361	469	314	522	1344	1216
Uncoated	466	937	244	739	1027	1097
	1069	1357	261	134	1011	1011
	114	1032	322	471	78	466
Anodized	1236	92	306	130	387	107
	533	211	68	398	130	327
	130	1482	252	874	586	524
Conversion	841	529	105	755	402	751
	1595	754	847	573	846	529

(a) Perform an analysis of variance with $\alpha = 0.05$ to test for significant main and interaction effects.

(b) Use Duncan's multiple-range test at the 0.05 level of significance to determine which humidity levels result in different corrosion fatigue damage.

14.5 To determine which muscles need to be subjected to a conditioning program in order to improve one's performance on the flat serve used in tennis, a study was conducted by the Department of Health, Physical Education and Recreation at Virginia Tech. Five different muscles

1: anterior deltoid
2: pectorial major
3: posterior deltoid
4: middle deltoid
5: triceps

were tested on each of 3 subjects, and the experiment was carried out 3 times for each treatment combination. The electromyographic data, recorded during the serve, are presented here.

	Muscle				
Subject	**1**	**2**	**3**	**4**	**5**
1	32	5	58	10	19
	59	1.5	61	10	20
	38	2	66	14	23
2	63	10	64	45	43
	60	9	78	61	61
	50	7	78	71	42
3	43	41	26	63	61
	54	43	29	46	85
	47	42	23	55	95

Use a 0.01 level of significance to test the hypothesis that

(a) different subjects have equal electromyographic measurements;

(b) different muscles have no effect on electromyographic measurements;

(c) subjects and types of muscle do not interact.

14.6 An experiment was conducted to determine whether additives increase the adhesiveness of rubber products. Sixteen products were made with the new additive and another 16 without the new additive. The observed adhesiveness was as recorded below.

	Temperature (°C)			
	50	**60**	**70**	**80**
	2.3	3.4	3.8	3.9
Without Additive	2.9	3.7	3.9	3.2
	3.1	3.6	4.1	3.0
	3.2	3.2	3.8	2.7
	4.3	3.8	3.9	3.5
With Additive	3.9	3.8	4.0	3.6
	3.9	3.9	3.7	3.8
	4.2	3.5	3.6	3.9

Perform an analysis of variance to test for significant main and interaction effects.

14.7 The extraction rate of a certain polymer is known to depend on the reaction temperature and the amount of catalyst used. An experiment was conducted at four levels of temperature and five levels of the catalyst, and the extraction rate was recorded in the following table.

	Amount of Catalyst				
	0.5%	**0.6%**	**0.7%**	**0.8%**	**0.9%**
50°C	38	45	57	59	57
	41	47	59	61	58
60°C	44	56	70	73	61
	43	57	69	72	58
70°C	44	56	70	73	61
	47	60	67	61	59
80°C	49	62	70	62	53
	47	65	55	69	58

Perform an analysis of variance. Test for significant main and interaction effects.

14.8 In Myers, Montgomery, and Anderson-Cook (2009), a scenario is discussed involving an auto bumper plating process. The response is the thickness of the material. Factors that may impact the thickness include amount of nickel (A) and pH (B). A two-factor experiment is designed. The plan is a completely randomized design in which the individual bumpers are assigned randomly to the factor combinations. Three levels of pH and two levels of nickel content are involved in the experiment. The thickness data, in cm $\times$ 10^{-3}, are as follows:

Nickel Content (grams)	**pH** 5	5.5	6
18	250	211	221
	195	172	150
	188	165	170
10	115	88	69
	165	112	101
	142	108	72

(a) Display the analysis-of-variance table with tests for both main effects and interaction. Show P-values.

(b) Give engineering conclusions. What have you learned from the analysis of the data?

(c) Show a plot that depicts either a presence or an absence of interaction.

14.9 An engineer is interested in the effects of cutting speed and tool geometry on the life in hours of a machine tool. Two cutting speeds and two different geometries are used. Three experimental tests are accomplished at each of the four combinations. The data are as follows.

Tool	**Cutting Speed**					
Geometry	**Low**			**High**		
1	22	28	20	34	37	29
2	18	15	16	11	10	10

(a) Show an analysis-of-variance table with tests on interaction and main effects.

(b) Comment on the effect that interaction has on the test on cutting speed.

(c) Do secondary tests that will allow the engineer to learn the true impact of cutting speed.

(d) Show a plot that graphically displays the interaction effect.

14.10 Two factors in a manufacturing process for an integrated circuit are studied in a two-factor experiment. The purpose of the experiment is to learn their effect on the resistivity of the wafer. The factors are implant dose (2 levels) and furnace position (3 levels). Experimentation is costly so only one experimental run is made at each combination. The data are as follows.

Dose	**Position**		
1	15.5	14.8	21.3
2	27.2	24.9	26.1

It is to be assumed that no interaction exists between these two factors.

(a) Write the model and explain terms.

(b) Show the analysis-of-variance table.

(c) Explain the 2 "error" degrees of freedom.

(d) Use Tukey's test to do multiple-comparison tests on furnace position. Explain what the results show.

14.11 A study was done to determine the impact of two factors, method of analysis and the laboratory doing the analysis, on the level of sulfur content in coal. Twenty-eight coal specimens were randomly assigned to 14 factor combinations, the structure of the experimental units represented by combinations of seven laboratories and two methods of analysis with two specimens per factor combination. The data, expressed in percent of sulfur, are as follows:

	Method			
Laboratory	**1**		**2**	
1	0.109	0.105	0.105	0.108
2	0.129	0.122	0.127	0.124
3	0.115	0.112	0.109	0.111
4	0.108	0.108	0.117	0.118
5	0.097	0.096	0.110	0.097
6	0.114	0.119	0.116	0.122
7	0.155	0.145	0.164	0.160

(The data are taken from G. Taguchi, "Signal to Noise Ratio and Its Applications to Testing Material," *Reports of Statistical Application Research*, Union of Japanese Scientists and Engineers, Vol. 18, No. 4, 1971.)

(a) Do an analysis of variance and show results in an analysis-of-variance table.

(b) Is interaction significant? If so, discuss what it means to the scientist. Use a P-value in your conclusion.

(c) Are the individual main effects, laboratory, and method of analysis statistically significant? Discuss what is learned and let your answer be couched in the context of any significant interaction.

(d) Do an interaction plot that illustrates the effect of interaction.

(e) Do a test comparing methods 1 and 2 at laboratory 1 and do the same test at laboratory 7. Comment on what these results illustrate.

14.12 In an experiment conducted in the Civil Engineering Department at Virginia Tech, growth of a certain type of algae in water was observed as a function of time and the dosage of copper added to the water. The data are as follows. Response is in units of algae.

	Time in Days		
Copper	**5**	**12**	**18**
1	0.30	0.37	0.25
	0.34	0.36	0.23
	0.32	0.35	0.24
2	0.24	0.30	0.27
	0.23	0.32	0.25
	0.22	0.31	0.25
3	0.20	0.30	0.27
	0.28	0.31	0.29
	0.24	0.30	0.25

(a) Do an analysis of variance and show the analysis-of-variance table.

(b) Comment concerning whether the data are sufficient to show a time effect on algae concentration.

(c) Do the same for copper content. Does the level of copper impact algae concentration?

(d) Comment on the results of the test for interaction. How is the effect of copper content influenced by time?

14.13 In Myers, *Classical and Modern Regression with Applications* (Duxbury Classic Series, 2nd edition, 1990), an experiment is described in which the Environmental Protection Agency seeks to determine the effect of two water treatment methods on magnesium uptake. Magnesium levels in grams per cubic centimeter (cc) are measured, and two different time levels are incorporated into the experiment. The data are as follows:

	Treatment					
Time (hr)	**1**			**2**		
1	2.19	2.15	2.16	2.03	2.01	2.04
2	2.01	2.03	2.04	1.88	1.86	1.91

(a) Do an interaction plot. What is your impression?

(b) Do an analysis of variance and show tests for the main effects and interaction.

(c) Give scientific findings regarding how time and treatment influence magnesium uptake.

(d) Fit the appropriate regression model with treatment as a categorical variable. Include interaction in the model.

(e) Is interaction significant in the regression model?

14.14 Consider the data set in Exercise 14.12 and answer the following questions.

(a) Both factors, copper and time, are quantitative in nature. As a result, a regression model may be of interest. Describe what might be an appropriate model using x_1 = copper content and x_2 = time. Fit the model to the data, showing regression coefficients and a t-test on each.

(b) Fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon,$$

and compare it to the one you chose in (a). Which is more appropriate? Use R^2_{adj} as a criterion.

14.15 The purpose of the study *The Incorporation of a Chelating Agent into a Flame Retardant Finish of a Cotton Flannelette and the Evaluation of Selected Fabric Properties*, conducted at Virginia Tech, was to evaluate the use of a chelating agent as part of the flame retardant finish of cotton flannelette by determining its effect upon flammability after the fabric is laundered under specific conditions. There were two treatments at two levels. Two baths were prepared, one with carboxymethyl cellulose (bath I) and one without (bath II). Half of the fabric was laundered 5 times and half was laundered 10 times. There were 12 pieces of fabric in each bath/number of launderings combination. After the washings, the lengths of fabric that burned and the burn times were measured. Burn times (in seconds) were recorded as follows:

Launderings	**Bath I**			**Bath II**		
5	13.7	23.0	15.7	6.2	5.4	5.0
	25.5	15.8	14.8	4.4	5.0	3.3
	14.0	29.4	9.7	16.0	2.5	1.6
	14.0	12.3	12.3	3.9	2.5	7.1
10	27.2	16.8	12.9	18.2	8.8	14.5
	14.9	17.1	13.0	14.7	17.1	13.9
	10.8	13.5	25.5	10.6	5.8	7.3
	14.2	27.4	11.5	17.7	18.3	9.9

(a) Perform an analysis of variance. Is there a significant interaction term?

(b) Are there main effect differences? Discuss.

14.4 Three-Factor Experiments

In this section, we consider an experiment with three factors, A, B, and C, at a, b, and c levels, respectively, in a completely randomized experimental design. Assume again that we have n observations for each of the abc treatment combinations. We shall proceed to outline significance tests for the three main effects and interactions involved. It is hoped that the reader can then use the description given here to generalize the analysis to $k > 3$ factors.

Model for the Three-Factor Experiment

The model for the three-factor experiment is

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl},$$

$i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, c$; and $l = 1, 2, \dots, n$, where α_i, β_j, and γ_k are the main effects and $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ are the two-factor interaction effects that have the same interpretation as in the two-factor experiment.

The term $(\alpha\beta\gamma)_{ijk}$ is called the **three-factor interaction effect**, a term that represents a nonadditivity of the $(\alpha\beta)_{ij}$ over the different levels of the factor C. As before, the sum of all main effects is zero and the sum over any subscript of the two- and three-factor interaction effects is zero. In many experimental situations, these higher-order interactions are insignificant and their mean squares reflect only random variation, but we shall outline the analysis in its most general form.

Again, in order that valid significance tests can be made, we must assume that the errors are values of independent and normally distributed random variables, each with mean 0 and common variance σ^2.

The general philosophy concerning the analysis is the same as that discussed for the one- and two-factor experiments. The sum of squares is partitioned into eight terms, each representing a source of variation from which we obtain independent estimates of σ^2 when all the main effects and interaction effects are zero. If the effects of any given factor or interaction are not all zero, then the mean square will estimate the error variance plus a component due to the systematic effect in question.

Sum of Squares for a Three-Factor Experiment

$$SSA = bcn\sum_{i=1}^{a}(\bar{y}_{i...} - \bar{y}_{....})^2 \quad SS(AB) = cn\sum_{i}\sum_{j}(\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....})^2$$

$$SSB = acn\sum_{j=1}^{b}(\bar{y}_{.j..} - \bar{y}_{....})^2 \quad SS(AC) = bn\sum_{i}\sum_{k}(\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k.} + \bar{y}_{....})^2$$

$$SSC = abn\sum_{k=1}^{c}(\bar{y}_{..k.} - \bar{y}_{....})^2 \quad SS(BC) = an\sum_{j}\sum_{k}(\bar{y}_{.jk.} - \bar{y}_{.j..} - \bar{y}_{..k.} + \bar{y}_{....})^2$$

$$SS(ABC) = n\sum_{i}\sum_{j}\sum_{k}(\bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{..k.} - \bar{y}_{....})^2$$

$$SST = \sum_{i}\sum_{j}\sum_{k}\sum_{l}(y_{ijkl} - \bar{y}_{....})^2 \quad SSE = \sum_{i}\sum_{j}\sum_{k}\sum_{l}(y_{ijkl} - \bar{y}_{ijk.})^2$$

Although we emphasize interpretation of annotated computer printout in this section rather than being concerned with laborious computation of sums of squares, we do offer the following as the sums of squares for the three main effects and interactions. Notice the obvious extension from the two- to three-factor problem.

The averages in the formulas are defined as follows:

$\bar{y}_{....}$ = average of all $abcn$ observations,
$\bar{y}_{i...}$ = average of the observations for the ith level of factor A,
$\bar{y}_{.j..}$ = average of the observations for the jth level of factor B,
$\bar{y}_{..k.}$ = average of the observations for the kth level of factor C,
$\bar{y}_{ij..}$ = average of the observations for the ith level of A and the jth level of B,
$\bar{y}_{i.k.}$ = average of the observations for the ith level of A and the kth level of C,
$\bar{y}_{.jk.}$ = average of the observations for the jth level of B and the kth level of C,
$\bar{y}_{ijk.}$ = average of the observations for the (ijk)th treatment combination.

The computations in an analysis-of-variance table for a three-factor problem with n replicated runs at each factor combination are summarized in Table 14.7.

Table 14.7: ANOVA for the Three-Factor Experiment with n Replications

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed *f***
Main effect:				
A	SSA	$a-1$	s_1^2	$f_1 = \frac{s_1^2}{s^2}$
B	SSB	$b-1$	s_2^2	$f_2 = \frac{s_2^2}{s^2}$
C	SSC	$c-1$	s_3^2	$f_3 = \frac{s_3^2}{s^2}$
Two-factor interaction:				
AB	$SS(AB)$	$(a-1)(b-1)$	s_4^2	$f_4 = \frac{s_4^2}{s^2}$
AC	$SS(AC)$	$(a-1)(c-1)$	s_5^2	$f_5 = \frac{s_5^2}{s^2}$
BC	$SS(BC)$	$(b-1)(c-1)$	s_6^2	$f_6 = \frac{s_6^2}{s^2}$
Three-factor interaction:				
ABC	$SS(ABC)$	$(a-1)(b-1)(c-1)$	s_7^2	$f_7 = \frac{s_7^2}{s^2}$
Error	SSE	$abc(n-1)$	s^2	
Total	SST	$abcn-1$		

For the three-factor experiment with a single experimental run per combination, we may use the analysis of Table 14.7 by setting $n = 1$ and using the ABC interaction sum of squares for SSE. In this case, we are assuming that the $(\alpha\beta\gamma)_{ijk}$ interaction effects are all equal to zero so that

$$E\left[\frac{SS(ABC)}{(a-1)(b-1)(c-1)}\right] = \sigma^2 + \frac{n}{(a-1)(b-1)(c-1)} \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\alpha\beta\gamma)_{ijk}^2 = \sigma^2.$$

That is, $SS(ABC)$ represents variation due only to experimental error. Its mean square thereby provides an unbiased estimate of the error variance. With $n = 1$ and $SSE = SS(ABC)$, the error sum of squares is found by subtracting the sums of squares of the main effects and two-factor interactions from the total sum of squares.

Example 14.4: In the production of a particular material, three variables are of interest: A, the operator effect (three operators): B, the catalyst used in the experiment (three catalysts); and C, the washing time of the product following the cooling process (15 minutes and 20 minutes). Three runs were made at each combination of factors. It was felt that all interactions among the factors should be studied. The coded yields are in Table 14.8. Perform an analysis of variance to test for significant effects.

Table 14.8: Data for Example 14.4

	Washing Time, *C*					
	15 Minutes			**20 Minutes**		
	Catalyst, *B*			**Catalyst, *B***		
Operator, *A*	**1**	**2**	**3**	**1**	**2**	**3**
1	10.7	10.3	11.2	10.9	10.5	12.2
	10.8	10.2	11.6	12.1	11.1	11.7
	11.3	10.5	12.0	11.5	10.3	11.0
2	11.4	10.2	10.7	9.8	12.6	10.8
	11.8	10.9	10.5	11.3	7.5	10.2
	11.5	10.5	10.2	10.9	9.9	11.5
3	13.6	12.0	11.1	10.7	10.2	11.9
	14.1	11.6	11.0	11.7	11.5	11.6
	14.5	11.5	11.5	12.7	10.9	12.2

***Solution*:** Table 14.9 shows an analysis of variance of the data given above. None of the interactions show a significant effect at the $\alpha = 0.05$ level. However, the P-value for BC is 0.0610; thus, it should not be ignored. The operator and catalyst effects are significant, while the effect of washing time is not significant. ◾

Impact of Interaction *BC*

More should be discussed regarding Example 14.4, particularly about dealing with the effect that the interaction between catalyst and washing time is having on the test on the washing time main effect (factor C). Recall our discussion in Section 14.2. Illustrations were given of how the presence of interaction could change the interpretation that we make regarding main effects. In Example 14.4, the BC interaction is significant at approximately the 0.06 level. Suppose, however, that we observe a two-way table of means as in Table 14.10.

It is clear why washing time was found not to be significant. A non-thorough analyst may get the impression that washing time can be eliminated from any future study in which yield is being measured. However, it is obvious how the

Table 14.9: ANOVA for a Three-Factor Experiment in a Completely Randomized Design

Source	**df**	**Sum of Squares**	**Mean Square**	**F-Value**	**P-Value**
A	2	13.98	6.99	11.64	0.0001
B	2	10.18	5.09	8.48	0.0010
AB	4	4.77	1.19	1.99	0.1172
C	1	1.19	1.19	1.97	0.1686
AC	2	2.91	1.46	2.43	0.1027
BC	2	3.63	1.82	3.03	0.0610
ABC	4	4.91	1.23	2.04	0.1089
Error	36	21.61	0.60		
Total	53	63.19			

Table 14.10: Two-Way Table of Means for Example 14.4

	Washing Time, C	
Catalyst, B	**15 min**	**20 min**
1	12.19	11.29
2	10.86	10.50
3	11.09	11.46
Means	11.38	11.08

effect of washing time changes from a negative effect for the first catalyst to what appears to be a positive effect for the third catalyst. If we merely focus on the data for catalyst 1, a simple comparison between the means at the two washing times will produce a simple t-statistic:

$$t = \frac{12.19 - 11.29}{\sqrt{0.6(2/9)}} = 2.5,$$

which is significant at a level less than 0.02. Thus, an important negative effect of washing time for catalyst 1 might very well be ignored if the analyst makes the incorrect broad interpretation of the insignificant F-ratio for washing time.

Pooling in Multifactor Models

We have described the three-factor model and its analysis in the most general form by including all possible interactions in the model. Of course, there are many situations where it is known *a priori* that the model should not contain certain interactions. We can then take advantage of this knowledge by combining or pooling the sums of squares corresponding to negligible interactions with the error sum of squares to form a new estimator for σ^2 with a larger number of degrees of freedom. For example, in a metallurgy experiment designed to study the effect on film thickness of three important processing variables, suppose it is known that factor A, acid concentration, does not interact with factors B and C. The

Table 14.11: ANOVA with Factor A Noninteracting

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Main effect:				
A	SSA	$a-1$	s_1^2	$f_1 = \frac{s_1^2}{s^2}$
B	SSB	$b-1$	s_2^2	$f_2 = \frac{s_2^2}{s^2}$
C	SSC	$c-1$	s_3^2	$f_3 = \frac{s_3^2}{s^2}$
Two-factor interaction:				
BC	$SS(BC)$	$(b-1)(c-1)$	s_4^2	$f_4 = \frac{s_4^2}{s^2}$
Error	SSE	Subtraction	s^2	
Total	SST	$abcn-1$		

sums of squares SSA, SSB, SSC, and $SS(BC)$ are computed using the methods described earlier in this section. The mean squares for the remaining effects will now all independently estimate the error variance σ^2. Therefore, we form our new **mean square error by pooling** $SS(AB)$, $SS(AC)$, $SS(ABC)$, and SSE, along with the corresponding degrees of freedom. The resulting denominator for the significance tests is then the mean square error given by

$$s^2 = \frac{SS(AB) + SS(AC) + SS(ABC) + SSE}{(a-1)(b-1) + (a-1)(c-1) + (a-1)(b-1)(c-1) + abc(n-1)}.$$

Computationally, of course, one obtains the pooled sum of squares and the pooled degrees of freedom by subtraction once SST and the sums of squares for the existing effects are computed. The analysis-of-variance table would then take the form of Table 14.11.

Factorial Experiments in Blocks

In this chapter, we have assumed that the experimental design used is a completely randomized design. By interpreting the levels of factor A in Table 14.11 **as different blocks**, we then have the analysis-of-variance procedure for a two-factor experiment in a randomized block design. For example, if we interpret the operators in Example 14.4 as blocks and assume no interaction between blocks and the other two factors, the analysis of variance takes the form of Table 14.12 rather than that of Table 14.9. The reader can verify that the mean square error is also

$$s^2 = \frac{4.77 + 2.91 + 4.91 + 21.61}{4 + 2 + 4 + 36} = 0.74,$$

which demonstrates the pooling of the sums of squares for the nonexisting interaction effects. Note that factor B, catalyst, has a significant effect on yield.

Table 14.12: ANOVA for a Two-Factor Experiment in a Randomized Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-Value
Blocks	13.98	2	6.99		
Main effect:					
B	10.18	2	5.09	6.88	0.0024
C	1.18	1	1.18	1.59	0.2130
Two-factor interaction:					
BC	3.64	2	1.82	2.46	0.0966
Error	34.21	46	0.74		
Total	63.19	53			

Example 14.5: An experiment was conducted to determine the effects of temperature, pressure, and stirring rate on product filtration rate. This was done in a pilot plant. The experiment was run at two levels of each factor. In addition, it was decided that two batches of raw materials should be used, where batches were treated as blocks. Eight experimental runs were made in random order for each batch of raw materials. It is thought that all two-factor interactions may be of interest. No interactions with batches are assumed to exist. The data appear in Table 14.13. "L" and "H" imply low and high levels, respectively. The filtration rate is in gallons per hour.

(a) Show the complete ANOVA table. Pool all "interactions" with blocks into error.

(b) What interactions appear to be significant?

(c) Create plots to reveal and interpret the significant interactions. Explain what the plot means to the engineer.

Table 14.13: Data for Example 14.5

	Low Stirring Rate			High Stirring Rate	
Batch 1					
Temp.	**Pressure L**	**Pressure H**	**Temp.**	**Pressure L**	**Pressure H**
L	43	49	**L**	44	47
H	64	68	**H**	97	102
Batch 2					
Temp.	**Pressure L**	**Pressure H**	**Temp.**	**Pressure L**	**Pressure H**
L	49	57	**L**	51	55
H	70	76	**H**	103	106

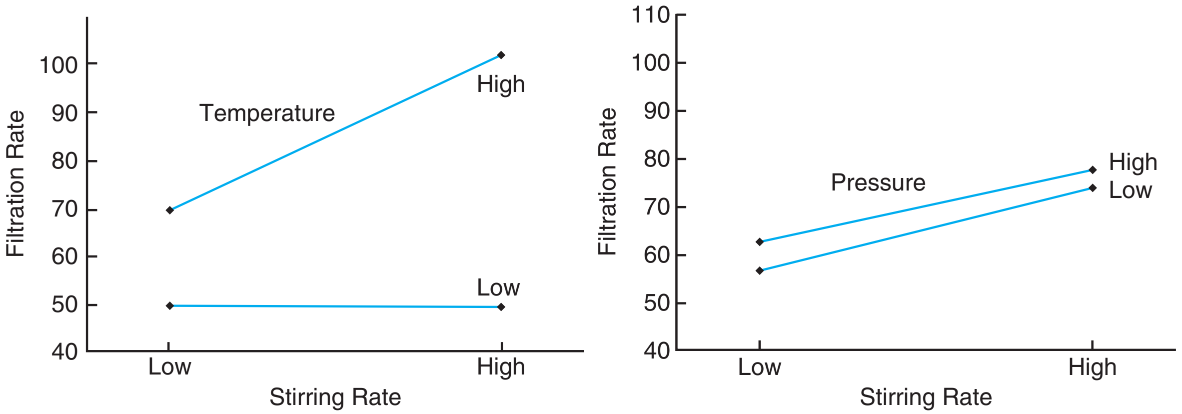
Solution: (a) The *SAS* printout is given in Figure 14.7.

(b) As seen in Figure 14.7, the temperature by stirring rate (strate) interaction appears to be highly significant. The pressure by stirring rate interaction also appears to be significant. Incidentally, if one were to do further pooling by combining the insignificant interactions with error, the conclusions would remain the same and the P-value for the pressure by stirring rate interaction would become stronger, namely 0.0517.

(c) The main effects for both stirring rate and temperature are highly significant, as shown in Figure 14.7. A look at the interaction plot of Figure 14.8(a) shows that the effect of stirring rate is dependent upon the level of temperature. At the low level of temperature the stirring rate effect is negligible, whereas at the high level of temperature stirring rate has a strong positive effect on mean filtration rate. In Figure 14.8(b), the interaction between pressure and stirring rate, though not as pronounced as that of Figure 14.8(a), still shows a slight inconsistency of the stirring rate effect across pressure.

```
Source                DF  Type III SS  Mean Square  F Value  Pr > F
batch                  1   175.562500   175.562500   177.14  <.0001
pressure               1    95.062500    95.062500    95.92  <.0001
temp                   1  5292.562500  5292.562500  5340.24  <.0001
pressure*temp          1     0.562500     0.562500     0.57  0.4758
strate                 1  1040.062500  1040.062500  1049.43  <.0001
pressure*strate        1     5.062500     5.062500     5.11  0.0583
temp*strate            1  1072.562500  1072.562500  1082.23  <.0001
pressure*temp*strate   1     1.562500     1.562500     1.58  0.2495
Error                  7     6.937500     0.991071
Corrected Total       15  7689.937500
```

Figure 14.7: ANOVA for Example 14.5, batch interaction pooled with error.



(a) Temperature versus stirring rate.

(b) Pressure versus stirring rate.

Figure 14.8: Interaction plots for Example 14.5.

Exercises

14.16 Consider an experimental situation involving factors A, B, and C, where we assume a three-way fixed effects model of the form $y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \epsilon_{ijkl}$. All other interactions are considered to be nonexistent or negligible. The data are presented here.

	B_1			B_2		
	C_1	C_2	C_3	C_1	C_2	C_3
A_1	4.0	3.4	3.9	4.4	3.1	3.1
	4.9	4.1	4.3	3.4	3.5	3.7
A_2	3.6	2.8	3.1	2.7	2.9	3.7
	3.9	3.2	3.5	3.0	3.2	4.2
A_3	4.8	3.3	3.6	3.6	2.9	2.9
	3.7	3.8	4.2	3.8	3.3	3.5
A_4	3.6	3.2	3.2	2.2	2.9	3.6
	3.9	2.8	3.4	3.5	3.2	4.3

(a) Perform a test of significance on the BC interaction at the $\alpha = 0.05$ level.

(b) Perform tests of significance on the main effects A, B, and C using a pooled mean square error at the $\alpha = 0.05$ level.

14.17 The following data are measurements from an experiment conducted using three factors A, B, and C, all fixed effects:

	C_1			C_2			C_3		
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
A_1	15.0	14.8	15.9	16.8	14.2	13.2	15.8	15.5	19.2
	18.5	13.6	14.8	15.4	12.9	11.6	14.3	13.7	13.5
	22.1	12.2	13.6	14.3	13.0	10.1	13.0	12.6	11.1
A_2	11.3	17.2	16.1	18.9	15.4	12.4	12.7	17.3	7.8
	14.6	15.5	14.7	17.3	17.0	13.6	14.2	15.8	11.5
	18.2	14.2	13.4	16.1	18.6	15.2	15.9	14.6	12.2

(a) Perform tests of significance on all interactions at the $\alpha = 0.05$ level.

(b) Perform tests of significance on the main effects at the $\alpha = 0.05$ level.

(c) Give an explanation of how a significant interaction has masked the effect of factor C.

14.18 The method of X-ray fluorescence is an important analytical tool for determining the concentration of material in solid missile propellants. In the paper *An X-ray Fluorescence Method for Analyzing Polybutadiene Acrylic Acid (PBAA) Propellants* (Quarterly Report, RK-TR-62-1, Army Ordinance Missile Command, 1962), it is postulated that the propellant mixing process and analysis time have an influence on the homogeneity of the material and hence on the accuracy of X-ray intensity measurements. An experiment was conducted using 3 factors: A, the mixing conditions (4 levels); B, the analysis time (2 levels); and C, the method of loading propellant into sample holders (hot and room temperature). The following data, which represent the weight percent of ammonium perchlorate in a particular propellant, were recorded.

	Method of Loading, C			
	Hot		**Room Temp.**	
A	B_1	B_2	B_1	B_2
1	38.62	38.45	39.82	39.82
	37.20	38.64	39.15	40.26
	38.02	38.75	39.78	39.72
2	37.67	37.81	39.53	39.56
	37.57	37.75	39.76	39.25
	37.85	37.91	39.90	39.04
3	37.51	37.21	39.34	39.74
	37.74	37.42	39.60	39.49
	37.58	37.79	39.62	39.45
4	37.52	37.60	40.09	39.36
	37.15	37.55	39.63	39.38
	37.51	37.91	39.67	39.00

(a) Perform an analysis of variance with $\alpha = 0.01$ to test for significant main and interaction effects.

(b) Discuss the influence of the three factors on the weight percent of ammonium perchlorate. Let your discussion involve the role of any significant interaction.

14.19 Corrosion fatigue in metals has been defined as the simultaneous action of cyclic stress and chemical attack on a metal structure. In the study *Effect of Humidity and Several Surface Coatings on the Fatigue Life of 2024-T351 Aluminum Alloy*, conducted by the Department of Mechanical Engineering at Virginia Tech, a technique involving the application of a protective chromate coating was used to minimize corrosion fatigue damage in aluminum. Three factors were used in the investigation, with 5 replicates for each treatment combination: coating, at 2 levels, and humidity and shear stress, both with 3 levels. The fatigue data, recorded in thousands of cycles to failure, are presented here.

(a) Perform an analysis of variance with $\alpha = 0.01$ to test for significant main and interaction effects.

(b) Make a recommendation for combinations of the three factors that would result in low fatigue damage.

Coating	Humidity	Shear Stress (psi) 13,000	17,000	20,000
Uncoated	**Low (20–25% RH)**	4580	5252	361
		10,126	897	466
		1341	1465	1069
		6414	2694	469
		3549	1017	937
	Medium (50–60% RH)	2858	799	314
		8829	3471	244
		10,914	685	261
		4067	810	522
		2595	3409	739
	High (86–91% RH)	6489	1862	1344
		5248	2710	1027
		6816	2632	663
		5860	2131	1216
		5901	2470	1097
Chromated	**Low (20–25% RH)**	5395	4035	130
		2768	2022	841
		1821	914	1595
		3604	2036	1482
		4106	3524	529
	Medium (50–60% RH)	4833	1847	252
		7414	1684	105
		10,022	3042	847
		7463	4482	874
		21,906	996	755
	High (86–91% RH)	3287	1319	586
		5200	929	402
		5493	1263	846
		4145	2236	524
		3336	1392	751

14.20 For a study of the hardness of gold dental fillings, five randomly chosen dentists were assigned combinations of three methods of condensation and two types of gold. The hardness was measured. (See Hoaglin, Mosteller, and Tukey, 1991.) Let the dentists play the role of blocks. The data are presented here.

(a) State the appropriate model with the assumptions.

(b) Is there a significant interaction between method of condensation and type of gold filling material?

(c) Is there one method of condensation that seems to be best? Explain.

Dentist	Method	Type Gold Foil	Goldent
1	**1**	792	824
	2	772	772
	3	782	803
2	**1**	803	803
	2	752	772
	3	715	707
3	**1**	715	724
	2	792	715
	3	762	606
4	**1**	673	946
	2	657	743
	3	690	245
5	**1**	634	715
	2	649	724
	3	724	627

(*cont.*)

14.21 Electronic copiers make copies by gluing black ink on paper, using static electricity. Heating and gluing the ink on the paper comprise the final stage of the copying process. The gluing power during this final process determines the quality of the copy. It is postulated that temperature, surface state of the gluing roller, and hardness of the press roller influence the gluing power of the copier. An experiment is run with treatments consisting of a combination of these three factors at each of three levels. The following data show the gluing power for each treatment combination. Perform an analysis of variance with $\alpha = 0.05$ to test for significant main and interaction effects.

	Surface State of Gluing Roller	Hardness of the Press Roller 20		40		60	
Low Temp.	**Soft**	0.52	0.44	0.54	0.52	0.60	0.55
		0.57	0.53	0.65	0.56	0.78	0.68
	Medium	0.64	0.59	0.79	0.73	0.49	0.48
		0.58	0.64	0.79	0.78	0.74	0.50
	Hard	0.67	0.77	0.58	0.68	0.55	0.65
		0.74	0.65	0.57	0.59	0.57	0.58
Medium Temp.	**Soft**	0.46	0.40	0.31	0.49	0.56	0.42
		0.58	0.37	0.48	0.66	0.49	0.49
	Medium	0.60	0.43	0.66	0.57	0.64	0.54
		0.62	0.61	0.72	0.56	0.74	0.56
	Hard	0.53	0.65	0.53	0.45	0.56	0.66
		0.66	0.56	0.59	0.47	0.71	0.67
High Temp.	**Soft**	0.52	0.44	0.54	0.52	0.65	0.49
		0.57	0.53	0.65	0.56	0.65	0.52
	Medium	0.53	0.65	0.53	0.45	0.49	0.48
		0.66	0.56	0.59	0.47	0.74	0.50
	Hard	0.43	0.43	0.48	0.31	0.55	0.65
		0.47	0.44	0.43	0.27	0.57	0.58

14.22 Consider the data set in Exercise 14.21.

(a) Construct an interaction plot for any two-factor interaction that is significant.

(b) Do a normal probability plot of residuals and comment.

14.23 Consider combinations of three factors in the removal of dirt from standard loads of laundry. The first factor is the brand of the detergent, X, Y, or Z. The second factor is the type of detergent, liquid or powder. The third factor is the temperature of the water, hot or warm. The experiment was replicated three times. Response is percent dirt removal. The data are as follows:

Brand	Type	Temperature			
X	**Powder**	**Hot**	85	88	80
		Warm	82	83	85
	Liquid	**Hot**	78	75	72
		Warm	75	75	73
Y	**Powder**	**Hot**	90	92	92
		Warm	88	86	88
	Liquid	**Hot**	78	76	70
		Warm	76	77	76
Z	**Powder**	**Hot**	85	87	88
		Warm	76	74	78
	Liquid	**Hot**	60	70	68
		Warm	55	57	54

(a) Are there significant interaction effects at the $\alpha = 0.05$ level?

(b) Are there significant differences between the three brands of detergent?

(c) Which combination of factors would you prefer to use?

14.24 A scientist collects experimental data on the radius of a propellant grain, y, as a function of powder temperature, extrusion rate, and die temperature. Results of the three-factor experiment are as follows:

	Powder Temp			
	150		**190**	
	Die Temp		**Die Temp**	
Rate	**220**	**250**	**220**	**250**
12	82	124	88	129
24	114	157	121	164

Resources are not available to make repeated experimental trials at the eight combinations of factors. It is believed that extrusion rate does not interact with die temperature and that the three-factor interaction should be negligible. Thus, these two interactions may be pooled to produce a 2 d.f. "error" term.

(a) Do an analysis of variance that includes the three main effects and two two-factor interactions. Determine what effects influence the radius of the propellant grain.

(b) Construct interaction plots for the powder temperature by die temperature and powder temperature by extrusion rate interactions.

(c) Comment on the consistency in the appearance of the interaction plots and the tests on the two interactions in the ANOVA.

14.25 In the book *Design of Experiments for Quality Improvement*, published by the Japanese Standards Association (1989), a study is reported on the extraction of polyethylene by using a solvent and how the amount of gel (proportion) is influenced by three factors: the type of solvent, extraction temperature, and extraction time. A factorial experiment was designed, and the following data were collected on proportion of gel.

		Time					
Solvent	**Temp.**	**4**		**8**		**16**	
Ethanol	**120**	94.0	94.0	93.8	94.2	91.1	90.5
	80	95.3	95.1	94.9	95.3	92.5	92.4
Toluene	**120**	94.6	94.5	93.6	94.1	91.1	91.0
	80	95.4	95.4	95.6	96.0	92.1	92.1

(a) Do an analysis of variance and determine what factors and interactions influence the proportion of gel.

(b) Construct an interaction plot for any two-factor interaction that is significant. In addition, explain what conclusion can be drawn from the presence of the interaction.

(c) Do a normal probability plot of residuals and comment.

14.5 Factorial Experiments for Random Effects and Mixed Models

In a two-factor experiment with random effects, we have the model

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk},$$

for $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; and $k = 1, 2, \ldots, n$, where the A_i, B_j, $(AB)_{ij}$, and ϵ_{ijk} are independent random variables with means 0 and variances σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$, and σ^2, respectively. The sums of squares for random effects experiments are computed in exactly the same way as for fixed effects experiments. We are now interested in testing hypotheses of the form

$$H_0^{'}\colon\ \sigma_\alpha^2 = 0, \qquad H_0^{''}\colon\ \sigma_\beta^2 = 0, \qquad H_0^{'''}\colon\ \sigma_{\alpha\beta}^2 = 0,$$
$$H_1^{'}\colon\ \sigma_\alpha^2 \neq 0, \qquad H_1^{''}\colon\ \sigma_\beta^2 \neq 0, \qquad H_1^{'''}\colon\ \sigma_{\alpha\beta}^2 \neq 0,$$

where the denominator in the *f*-ratio is not necessarily the mean square error. The appropriate denominator can be determined by examining the expected values of the various mean squares. These are shown in Table 14.14.

Table 14.14: Expected Mean Squares for a Two-Factor Random Effects Experiment

Source of Variation	**Degrees of Freedom**	**Mean Square**	**Expected Mean Square**
A	$a-1$	s_1^2	$\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_\alpha^2$
B	$b-1$	s_2^2	$\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2$
AB	$(a-1)(b-1)$	s_3^2	$\sigma^2 + n\sigma_{\alpha\beta}^2$
Error	$ab(n-1)$	s^2	σ^2
Total	$abn-1$		

From Table 14.14 we see that $H_0^{'}$ and $H_0^{''}$ are tested by using s_3^2 in the denominator of the *f*-ratio, whereas $H_0^{'''}$ is tested using s^2 in the denominator. The unbiased estimates of the variance components are

$$\hat{\sigma}^2 = s^2, \qquad \hat{\sigma}_{\alpha\beta}^2 = \frac{s_3^2 - s^2}{n}, \qquad \hat{\sigma}_\alpha^2 = \frac{s_1^2 - s_3^2}{bn}, \qquad \hat{\sigma}_\beta^2 = \frac{s_2^2 - s_3^2}{an}.$$

Table 14.15: Expected Mean Squares for a Three-Factor Random Effects Experiment

Source of Variation	**Degrees of Freedom**	**Mean Square**	**Expected Mean Square**
A	$a-1$	s_1^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + cn\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha\gamma}^2 + bcn\sigma_\alpha^2$
B	$b-1$	s_2^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + cn\sigma_{\alpha\beta}^2 + an\sigma_{\beta\gamma}^2 + acn\sigma_\beta^2$
C	$c-1$	s_3^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + bn\sigma_{\alpha\gamma}^2 + an\sigma_{\beta\gamma}^2 + abn\sigma_\gamma^2$
AB	$(a-1)(b-1)$	s_4^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + cn\sigma_{\alpha\beta}^2$
AC	$(a-1)(c-1)$	s_5^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + bn\sigma_{\alpha\gamma}^2$
BC	$(b-1)(c-1)$	s_6^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + an\sigma_{\beta\gamma}^2$
ABC	$(a-1)(b-1)(c-1)$	s_7^2	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$
Error	$abc(n-1)$	s^2	σ^2
Total	$abcn-1$		

The expected mean squares for the three-factor experiment with random effects in a completely randomized design are shown in Table 14.15. It is evident from the expected mean squares of Table 14.15 that one can form appropriate *f*-ratios for testing all two-factor and three-factor interaction variance components. However, to test a hypothesis of the form

$$H_0\colon\ \sigma_\alpha^2 = 0,$$
$$H_1\colon\ \sigma_\alpha^2 \neq 0,$$

there appears to be no appropriate f-ratio unless we have found one or more of the two-factor interaction variance components not significant. Suppose, for example, that we have compared s_5^2 (mean square AC) with s_7^2 (mean square ABC) and found $\sigma_{\alpha\gamma}^2$ to be negligible. We could then argue that the term $\sigma_{\alpha\gamma}^2$ should be dropped from all the expected mean squares of Table 14.15; then the ratio s_1^2/s_4^2 provides a test for the significance of the variance component σ_α^2. Therefore, if we are to test hypotheses concerning the variance components of the main effects, it is necessary first to investigate the significance of the two-factor interaction components. An approximate test derived by Satterthwaite (1946; see the Bibliography) may be used when certain two-factor interaction variance components are found to be significant and hence must remain a part of the expected mean square.

Example 14.6: In a study to determine which are the important sources of variation in an industrial process, 3 measurements are taken on yield for 3 operators chosen randomly and 4 batches of raw materials chosen randomly. It is decided that a statistical test should be made at the 0.05 level of significance to determine if the variance components due to batches, operators, and interaction are significant. In addition, estimates of variance components are to be computed. The data are given in Table 14.16, with the response being percent by weight.

Table 14.16: Data for Example 14.6

	Batch			
Operator	**1**	**2**	**3**	**4**
1	66.9	68.3	69.0	69.3
	68.1	67.4	69.8	70.9
	67.2	67.7	67.5	71.4
2	66.3	68.1	69.7	69.4
	65.4	66.9	68.8	69.6
	65.8	67.6	69.2	70.0
3	65.6	66.0	67.1	67.9
	66.3	66.9	66.2	68.4
	65.2	67.3	67.4	68.7

Solution: The sums of squares are found in the usual way, with the following results:

$$SST \text{ (total)} = 84.5564, \qquad SSE \text{ (error)} = 10.6733,$$
$$SSA \text{ (operators)} = 18.2106, \qquad SSB \text{ (batches)} = 50.1564,$$
$$SS(AB) \text{ (interaction)} = 5.5161.$$

All other computations are carried out and exhibited in Table 14.17. Since

$$f_{0.05}(2,6) = 5.14, \qquad f_{0.05}(3,6) = 4.76, \qquad \text{and} \qquad f_{0.05}(6,24) = 2.51,$$

we find the operator and batch variance components to be significant. Although the interaction variance is not significant at the $\alpha = 0.05$ level, the P-value is 0.095. Estimates of the main effect variance components are

$$\hat{\sigma}_\alpha^2 = \frac{9.1053 - 0.9194}{12} = 0.68, \qquad \hat{\sigma}_\beta^2 = \frac{16.7188 - 0.9194}{9} = 1.76.$$

Table 14.17: Analysis of Variance for Example 14.6

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed** f
Operators	18.2106	2	9.1053	9.90
Batches	50.1564	3	16.7188	18.18
Interaction	5.5161	6	0.9194	2.07
Error	10.6733	24	0.4447	
Total	84.5564	35		

Mixed Model Experiment

There are situations where the experiment dictates the assumption of a **mixed model** (i.e., a mixture of random and fixed effects). For example, for the case of two factors, we may have

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk},$$

for $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$. The A_i may be independent random variables, independent of ϵ_{ijk}, and the B_j may be fixed effects. The mixed nature of the model requires that the interaction terms be random variables. As a result, the relevant hypotheses are of the form

$$\begin{aligned} &H_0^{'}\colon\ \sigma_\alpha^2 = 0, \quad && H_0^{''}\colon\ B_1 = B_2 = \cdots = B_b = 0, \quad && H_0^{'''}\colon\ \sigma_{\alpha\beta}^2 = 0,\\ &H_1^{'}\colon\ \sigma_\alpha^2 \neq 0, \quad && H_1^{''}\colon\ \text{At least one the } B_j \text{ is not zero}, \quad && H_1^{'''}\colon\ \sigma_{\alpha\beta}^2 \neq 0. \end{aligned}$$

Again, the computations of sums of squares are identical to those of fixed and random effects situations, and the F-test is dictated by the expected mean squares. Table 14.18 provides the expected mean squares for the two-factor mixed model problem.

Table 14.18: Expected Mean Squares for Two-Factor Mixed Model Experiment

Factor	**Expected Mean Square**
A (random)	$\sigma^2 + bn\sigma_\alpha^2$
B (fixed)	$\sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{an}{b-1}\sum_j B_j^2$
AB (random)	$\sigma^2 + n\sigma_{\alpha\beta}^2$
Error	σ^2

From the nature of the expected mean squares it becomes clear that the **test on the random effect employs the mean square error s^2** as the denominator, whereas the **test on the fixed effect** uses the interaction mean square. Suppose we now consider three factors. Here, of course, we must take into account the situation where one factor is fixed and the situation in which two factors are fixed. Table 14.19 covers both situations.

Table 14.19: Expected Mean Squares for Mixed Model Factorial Experiments in Three Factors

	***A* Random**	***A* Random, *B* Random**
A	$\sigma^2 + bcn\sigma_\alpha^2$	$\sigma^2 + cn\sigma_{\alpha\beta}^2 + bcn\sigma_\alpha^2$
B	$\sigma^2 + cn\sigma_{\alpha\beta}^2 + acn\sum\limits_{j=1}^{b}\frac{B_j^2}{b-1}$	$\sigma^2 + cn\sigma_{\alpha\beta}^2 + acn\sigma_\beta^2$
C	$\sigma^2 + bn\sigma_{\alpha\gamma}^2 + abn\sum\limits_{k=1}^{c}\frac{C_k^2}{c-1}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + an\sigma_{\beta\gamma}^2 + bn\sigma_{\alpha\gamma}^2 + abn\sum\limits_{k=1}^{c}\frac{C_k^2}{c-1}$
AB	$\sigma^2 + cn\sigma_{\alpha\beta}^2$	$\sigma^2 + cn\sigma_{\alpha\beta}^2$
AC	$\sigma^2 + bn\sigma_{\alpha\gamma}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + bn\sigma_{\alpha\gamma}^2$
BC	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + an\sum\limits_{j}\sum\limits_{k}\frac{(BC)_{jk}^2}{(b-1)(c-1)}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + an\sigma_{\beta\gamma}^2$
ABC	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$
Error	σ^2	σ^2

Note that in the case of A random, all effects have proper f-tests. But in the case of A and B random, the main effect C must be tested using a Satterthwaite-type procedure similar to that used in the random effects experiment.

Exercises

14.26 Assuming a random effects experiment for Exercise 14.2 on page 575, estimate the variance components for brand of orange juice concentrate, for number of days from when orange juice was blended until it was tested, and for experimental error.

14.27 To estimate the various components of variability in a filtration process, the percent of material lost in the mother liquor is measured for 12 experimental conditions, with 3 runs on each condition. Three filters and 4 operators are selected at random for use in the experiment.

(a) Test the hypothesis of no interaction variance component between filters and operators at the $\alpha = 0.05$ level of significance.

(b) Test the hypotheses that the operators and the filters have no effect on the variability of the filtration process at the $\alpha = 0.05$ level of significance.

(c) Estimate the components of variance due to filters, operators, and experimental error.

	Operator			
Filter	**1**	**2**	**3**	**4**
1	16.2	15.9	15.6	14.9
	16.8	15.1	15.9	15.2
	17.1	14.5	16.1	14.9
2	16.6	16.0	16.1	15.4
	16.9	16.3	16.0	14.6
	16.8	16.5	17.2	15.9
3	16.7	16.5	16.4	16.1
	16.9	16.9	17.4	15.4
	17.1	16.8	16.9	15.6

14.28 A defense contractor is interested in studying an inspection process to detect failure or fatigue of transformer parts. Three levels of inspections are used by three randomly chosen inspectors. Five lots are used for each combination in the study. The factor levels are given in the data. The response is in failures per 1000 pieces.

(a) Write an appropriate model, with assumptions.

(b) Use analysis of variance to test the appropriate hypothesis for inspector, inspection level, and interaction.

	Inspection Level		
Inspector	**Full Military Inspection**	**Reduced Military Inspection**	**Commercial**
A	7.50 7.42	7.08 6.17	6.15 5.52
	5.85 5.89	5.65 5.30	5.48 5.48
	5.35	5.02	5.98
B	7.58 6.52	7.68 5.86	6.17 6.20
	6.54 5.64	5.28 5.38	5.44 5.75
	5.12	4.87	5.68
C	7.70 6.82	7.19 6.19	6.21 5.66
	6.42 5.39	5.85 5.35	5.36 5.90
	5.35	5.01	6.12

14.29 Consider the following analysis of variance for a random effects experiment:

Source of Variation	**Degrees of Freedom**	**Mean Square**
A	3	140
B	1	480
C	2	325
AB	3	15
AC	6	24
BC	2	18
ABC	6	2
Error	24	5
Total	47	

Test for significant variance components among all main effects and interaction effects at the 0.01 level of significance

(a) by using a pooled estimate of error when appropriate;

(b) by not pooling sums of squares of insignificant effects.

14.30 A plant manager would like to show that the yield of a woven fabric in the plant does not depend on machine operator or time of day and is consistently high. Four randomly selected operators and 3 randomly selected hours of the day are chosen for the study. The yield is measured in yards produced per minute. Samples are taken on 3 randomly chosen days.

(a) Write the appropriate model.

(b) Evaluate the variance components for operator and time.

(c) Draw conclusions.

	Operator			
Time	**1**	**2**	**3**	**4**
1	9.5	9.8	9.8	10.0
	9.8	10.1	10.3	9.7
	10.0	9.6	9.7	10.2
2	10.2	10.1	10.2	10.3
	9.9	9.8	9.8	10.1
	9.5	9.7	9.7	9.9
3	10.5	10.4	9.9	10.0
	10.2	10.2	10.3	10.1
	9.3	9.8	10.2	9.7

14.31 A manufacturer of latex house paint (brand A) would like to show that its paint is more robust to the material being painted than that of its two closest competitors. The response is the time, in years, until chipping occurs. The study involves the three brands of paint and three randomly chosen materials. Two pieces of material are used for each combination.

	Brand of Paint		
Material	***A***	***B***	***C***
A	5.50 5.15	4.75 4.60	5.10 5.20
B	5.60 5.55	5.50 5.60	5.40 5.50
C	5.40 5.48	5.05 4.95	4.50 4.55

(a) What is this type of model called?

(b) Analyze the data, using the appropriate model.

(c) Did the manufacturer of brand A support its claim with the data?

14.32 A process engineer wants to determine if the power setting on the machines used to fill certain types of cereal boxes results in a significant effect on the actual weight of the product. The study consists of 3 randomly chosen types of cereal manufactured by the company and 3 fixed power settings. Weight is measured for 4 different randomly selected boxes of cereal at each combination. The desired weight is 400 grams. The data are presented here.

Power	**Cereal Type**		
Setting	**1**	**2**	**3**
Low	395 390	392 392	402 405
	401 400	394 401	399 399
Current	396 399	390 392	404 403
	400 402	395 502	400 399
High	410 408	404 406	415 412
	408 407	401 400	413 415

(a) Give the appropriate model, and list the assumption being made.

(b) Is there a significant effect due to the power setting?

(c) Is there a significant variance component due to cereal type?

Review Exercises

14.33 The Statistics Consulting Center at Virginia Tech was involved in analyzing a set of data taken by personnel in the Human Nutrition and Foods Department in which it was of interest to study the effects of flour type and percent sweetener on certain physical attributes of a type of cake. All-purpose flour and cake flour were used, and the percent sweetener was varied at four levels. The following data show information on specific gravity of cake samples. Three cakes were prepared at each of the eight factor combinations.

Sweetener	**Flour**					
Concentration	**All-Purpose**			**Cake**		
0	0.90	0.87	0.90	0.91	0.90	0.80
50	0.86	0.89	0.91	0.88	0.82	0.83
75	0.93	0.88	0.87	0.86	0.85	0.80
100	0.79	0.82	0.80	0.86	0.85	0.85

(a) Treat the analysis as a two-factor analysis of variance. Test for differences between flour type. Test for differences between sweetener concentration.

(b) Discuss the effect of interaction, if any. Give *P*-values on all tests.

14.34 An experiment was conducted in the Department of Food Science at Virginia Tech. It was of interest to characterize the texture of certain types of fish in the herring family. The effect of sauce types used in preparing the fish was also studied. The response in the experiment was "texture value," measured with a machine that sliced the fish product. The following are data on texture values:

	Fish Type					
Sauce Type	**Unbleached Menhaden**		**Bleached Menhaden**		**Herring**	
Sour Cream	27.6	57.4	64.0	66.9	107.0	83.9
	47.8	71.1	66.5	66.8	110.4	93.4
	53.8		53.8		83.1	
Wine Sauce	49.8	31.0	48.3	62.2	88.0	95.2
	11.8	35.1	54.6	43.6	108.2	86.7
	16.1		41.8		105.2	

(a) Do an analysis of variance. Determine whether or not there is an interaction between sauce type and fish type.

(b) Based on your results from part (a) and on *F*-tests on main effects, determine if there is a significant difference in texture due to sauce types, and determine whether there is a significant difference due to fish types.

14.35 A study was made to determine if humidity conditions have an effect on the force required to pull apart pieces of glued plastic. Three types of plastic were tested using 4 different levels of humidity. The results, in kilograms, are as follows:

	Humidity			
Plastic Type	**30%**	**50%**	**70%**	**90%**
A	39.0	33.1	33.8	33.0
	42.8	37.8	30.7	32.9
B	36.9	27.2	29.7	28.5
	41.0	26.8	29.1	27.9
C	27.4	29.2	26.7	30.9
	30.3	29.9	32.0	31.5

(a) Assuming a fixed effects experiment, perform an analysis of variance and test the hypothesis of no interaction between humidity and plastic type at the 0.05 level of significance.

(b) Using only plastics A and B and the value of s^2 from part (a), once again test for the presence of interaction at the 0.05 level of significance.

14.36 Personnel in the Materials Engineering Department at Virginia Tech conducted an experiment to study the effects of environmental factors on the stability of a certain type of copper-nickel alloy. The basic response was the fatigue life of the material. The factors are level of stress and environment. The data are as follows:

	Stress Level		
Environment	**Low**	**Medium**	**High**
Dry Hydrogen	11.08	13.12	14.18
	10.98	13.04	14.90
	11.24	13.37	15.10
High Humidity (95%)	10.75	12.73	14.15
	10.52	12.87	14.42
	10.43	12.95	14.25

(a) Do an analysis of variance to test for interaction between the factors. Use $\alpha = 0.05$.

(b) Based on part (a), do an analysis on the two main effects and draw conclusions. Use a *P*-value approach in drawing conclusions.

14.37 In the experiment of Review Exercise 14.33, cake volume was also used as a response. The units are cubic inches. Test for interaction between factors and discuss main effects. Assume that both factors are fixed effects.

Sweetener	**Flour**					
Concentration	**All-Purpose**			**Cake**		
0	4.48	3.98	4.42	4.12	4.92	5.10
50	3.68	5.04	3.72	5.00	4.26	4.34
75	3.92	3.82	4.06	4.82	4.34	4.40
100	3.26	3.80	3.40	4.32	4.18	4.30

14.38 A control valve needs to be very sensitive to the input voltage, thus generating a good output voltage. An engineer turns the control bolts to change the input voltage. The book *SN-Ratio for the Quality Evaluation*, published by the Japanese Standards Association (1988), described a study on how these three factors (relative position of control bolts, control range of bolts, and input voltage) affect the sensitivity of a control valve. The factors and their levels are shown below. The data show the sensitivity of a control valve.

Factor A, relative position of control bolts:
center -0.5, center, and center $+0.5$
Factor B, control range of bolts:
2, 4.5, and 7 (mm)
Factor C, input voltage:
100, 120, and 150 (V)

		C		
$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C_1}$	$\boldsymbol{C_2}$	$\boldsymbol{C_3}$
$\boldsymbol{A_1}$	$\boldsymbol{B_1}$	151 135	151 135	151 138
$\boldsymbol{A_1}$	$\boldsymbol{B_2}$	178 171	180 173	181 174
$\boldsymbol{A_1}$	$\boldsymbol{B_3}$	204 190	205 190	206 192
$\boldsymbol{A_2}$	$\boldsymbol{B_1}$	156 148	158 149	158 150
$\boldsymbol{A_2}$	$\boldsymbol{B_2}$	183 168	183 170	183 172
$\boldsymbol{A_2}$	$\boldsymbol{B_3}$	210 204	211 203	213 204
$\boldsymbol{A_3}$	$\boldsymbol{B_1}$	161 145	162 148	163 148
$\boldsymbol{A_3}$	$\boldsymbol{B_2}$	189 182	191 184	192 183
$\boldsymbol{A_3}$	$\boldsymbol{B_3}$	215 202	216 203	217 205

Perform an analysis of variance with $\alpha = 0.05$ to test for significant main and interaction effects. Draw conclusions.

14.39 Exercise 14.25 on page 588 describes an experiment involving the extraction of polyethylene through use of a solvent.

		Time		
Solvent	**Temp.**	**4**	**8**	**16**
Ethanol	**120**	94.0 94.0	93.8 94.2	91.1 90.5
	80	95.3 95.1	94.9 95.3	92.5 92.4
Toluene	**120**	94.6 94.5	93.6 94.1	91.1 91.0
	80	95.4 95.4	95.6 96.0	92.1 92.1

(a) Do a different sort of analysis on the data. Fit an appropriate regression model with a solvent categorical variable, a temperature term, a time term, a temperature by time interaction, a solvent by temperature interaction, and a solvent by time interaction. Do t-tests on all coefficients and report your findings.

(b) Do your findings suggest that different models are appropriate for ethanol and toluene, or are they equivalent apart from the intercepts? Explain.

(c) Do you find any conclusions here that contradict conclusions drawn in your solution of Exercise 14.25? Explain.

14.40 In the book *SN-Ratio for the Quality Evaluation*, published by the Japanese Standards Association (1988), a study on how tire air pressure affects the maneuverability of an automobile was described. Three different tire air pressures were compared on three different driving surfaces. The three air pressures were both left- and right-side tires inflated to 6 kgf/cm^2, left-side tires inflated to 6 kgf/cm^2 and right-side tires inflated to 3 kgf/cm^2, and both left- and right-side tires inflated to 3 kgf/cm^2. The three driving surfaces were asphalt, dry asphalt, and dry cement. The turning radius of a test vehicle was observed twice for each level of tire pressure on each of the three different driving surfaces.

	Tire Air Pressure		
Driving Surface	**1**	**2**	**3**
Asphalt	44.0 25.5	34.2 37.2	27.4 42.8
Dry Asphalt	31.9 33.7	31.8 27.6	43.7 38.2
Dry Cement	27.3 39.5	46.6 28.1	35.5 34.6

Perform an analysis of variance of the above data. Comment on the interpretation of the main and interaction effects.

14.41 The manufacturer of a certain brand of freeze-dried coffee hopes to shorten the process time without jeopardizing the integrity of the product. The process engineer wants to use 3 temperatures for the drying chamber and 4 drying times. The current drying time is 3 hours at a temperature of -15°C. The flavor response is an average of scores of 4 professional judges. The score is on a scale from 1 to 10, with 10 being the best. The data are as shown in the following table.

	Temperature		
Time	**-20°C**	**-15°C**	**-10°C**
1 hr	9.60 9.63	9.55 9.50	9.40 9.43
1.5 hr	9.75 9.73	9.60 9.61	9.55 9.48
2 hr	9.82 9.93	9.81 9.78	9.50 9.52
3 hr	9.78 9.81	9.80 9.75	9.55 9.58

(a) What type of model should be used? State assumptions.

(b) Analyze the data appropriately.

(c) Write a brief report to the vice-president in charge and make a recommendation for future manufacturing of this product.

14.42 To ascertain the number of tellers needed during peak hours of operation, data were collected by an urban bank. Four tellers were studied during three "busy" times: (1) weekdays between 10:00 and 11:00 A.M., (2) weekday afternoons between 2:00 and 3:00 P.M., and (3) Saturday mornings between 11:00 A.M. and 12:00 noon. An analyst chose four randomly selected times within each of the three time periods for each of the four teller positions over a period of months, and the numbers of customers serviced were observed. The data are as follows:

	Time Period											
Teller	**1**				**2**				**3**			
1	18	24	17	22	25	29	23	32	29	30	21	34
2	16	11	19	14	23	32	25	17	27	29	18	16
3	12	19	11	22	27	33	27	24	25	20	29	15
4	11	9	13	8	10	7	19	8	11	9	17	9

It is assumed that the number of customers served is a Poisson random variable.

(a) Discuss the danger in doing a standard analysis of variance on the data above. What assumptions, if any, would be violated?

(b) Construct a standard ANOVA table that includes F-tests on main effects and interactions. If interactions and main effects are found to be significant, give scientific conclusions. What have we learned? Be sure to interpret any significant interaction. Use your own judgment regarding P-values.

(c) Do the entire analysis again using an appropriate transformation on the response. Do you see any differences in your findings? Comment.

14.6 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

One of the most confusing issues in the analysis of factorial experiments resides in the interpretation of main effects in the presence of interaction. The presence of a relatively large P-value for a main effect when interactions are clearly present may tempt the analyst to conclude "no significant main effect." However, one must understand that if a main effect is involved in a significant interaction, then the main effect is **influencing the response**. The nature of the effect is inconsistent across levels of other effects. The nature of the role of the main effect can be deduced from **interaction plots**.

In light of what is communicated in the preceding paragraph, there is danger of a substantial misuse of statistics when one employs a multiple comparison test on main effects in the clear presence of interaction among the factors.

One must be cautious in the analysis of a factorial experiment when the assumption of a complete randomized design is made when in fact complete randomization is not carried out. For example, it is common to encounter factors that are very **difficult to change**. As a result, factor levels may need to be held without change for long periods of time throughout the experiment. For instance, a temperature factor is a common example. Moving temperature up and down in a randomization scheme is a costly plan, and most experimenters will refuse to do it. Experimental designs with *restrictions in randomization* are quite common and are called **split plot designs**. They are beyond the scope of the book, but presentations are found in Montgomery (2008a).

Many of the concepts discussed in this chapter carry over into Chapter 15 (e.g., the importance of randomization and the role of interaction in the interpretation of results). However, there are two areas covered in Chapter 15 that represent an expansion of principles dealt with both in Chapter 13 and in this chapter. In Chapter 15, problem solving through the use of factorial experiments is done with regression analysis since most of the factors are assumed to be quantitative and measured on a continuum (e.g., temperature and time). Prediction equations are developed from the data of the designed experiment, and they are used for process improvement or even process optimization. In addition, development is given on the topic of fractional factorials, in which only a portion or fraction of the entire factorial experiment is implemented due to the prohibitive cost of doing the entire experiment.

Chapter 15

2^k Factorial Experiments and Fractions

15.1 Introduction

We have already been exposed to certain experimental design concepts. The sampling plan for the simple *t*-test on the mean of a normal population and the analysis of variance involve randomly allocating pre-chosen treatments to experimental units. The randomized block design, where treatments are assigned to units within relatively homogeneous blocks, involves restricted randomization.

In this chapter, we give special attention to experimental designs in which the experimental plan calls for the study of the effect on a response of k factors, each at two levels. These are commonly known as **2^k factorial experiments**. We often denote the levels as "high" and "low" even though this notation may be arbitrary in the case of qualitative variables. The complete factorial design requires that each level of every factor occur with each level of every other factor, giving a total of **2^k treatment combinations**.

Factor Screening and Sequential Experimentation

Often, when experimentation is conducted either on a research or on a development level, a well-planned experimental design is a **stage** of what is truly a **sequential plan** of experimentation. More often than not, the scientists and engineers at the outset of a study may not be aware of which factors are important or what are appropriate ranges for the potential factors on which experimentation should be conducted. For example, in the text *Response Surface Methodology* by Myers, Montgomery, and Anderson-Cook (2009), one example is given of an investigation of a pilot plant experiment in which four factors—temperature, pressure, concentration of formaldehyde, and steering rate—are varied in order to establish their influence on the response, filtration rate of a certain chemical product. Even at the pilot plant level, the scientists are not certain if all four factors should be involved in the model. In addition, the eventual goal is to determine the proper settings of contributing factors that maximize the filtration rate. Thus, there is a need to determine the **proper region of experimentation**. These questions can be answered only if the total experimental plan is done sequentially. Many experimental endeavors are plans that feature *iterative learning*, the type of learning that is consistent with the scientific method, with the word *iterative* implying stage-wise experimentation.

Generally, the initial stage of the ideal sequential plan is variable or **factor screening**, a procedure that involves an inexpensive experimental design using the **candidate factors**. This is particularly important when the plan involves a complex system like a manufacturing process. The information received from the results of a *screening design* is used to design one or more subsequent experiments in which adjustments in the important factors are made, the adjustments that provide improvements in the system or process.

The 2^k factorial experiments and fractions of the 2^k are powerful tools that are ideal screening designs. They are simple, practical, and intuitively appealing. Many of the general concepts discussed in Chapter 14 continue to apply. However, there are graphical methods that provide useful intuition in the analysis of the two-level designs.

Screening Designs for Large Numbers of Factors

When k is small, say $k = 2$ or even $k = 3$, the utility of the 2^k factorial for factor screening is clear. Analysis of variance and/or regression analysis as discussed and illustrated in Chapters 12, 13, and 14 remain useful as tools. In addition, graphical approaches are helpful.

If k is large, say as large as 6, 7, or 8, the number of factor combinations and thus experimental runs required for the 2^k factorial often becomes prohibitive. For example, suppose one is interested in carrying out a screening design involving $k = 8$ factors. There may be interest in gaining information on all $k = 8$ main effects as well as the $\frac{k(k-1)}{2} = 28$ two-factor interactions. However, including $2^8 = 256$ runs would appear to make the study much too large and be wasteful for studying $28 + 8 = 36$ effects. But, as we will illustrate in future sections, when k is large we can gain considerable information in an efficient manner by using only a fraction of the complete 2^k factorial experiment. This class of designs is the class of *fractional factorial designs*. The goal is to retain high-quality information on main effects and interesting interactions even though the size of the design is reduced considerably.

15.2 The 2^k Factorial: Calculation of Effects and Analysis of Variance

Consider initially a 2^2 factorial with factors A and B and n *experimental observations per factor combination*. It is useful to use the symbols (1), a, b, and ab to signify the design points, where the presence of a lowercase letter implies that the factor (A or B) is at the *high level*. Thus, absence of the lower case implies that the factor is at the *low level*. So ab is the design point $(+, +)$, a is $(+, -)$, b is $(-, +)$ and (1) is $(-, -)$. There are situations in which the notation also stands for the response data at the design point in question. As an introduction to the calculation of important **effects** that aid in the determination of the influence of the factors and **sums of squares** that are incorporated into analysis of variance computations, we have Table 15.1.

Table 15.1: A 2^2 Factorial Experiment

	A		**Mean**
B	b	ab	$\frac{b+ab}{2n}$
	(1)	a	$\frac{(1)+a}{2n}$
Mean	$\frac{(1)+b}{2n}$	$\frac{a+ab}{2n}$	

In this table, (1), a, b, and ab signify totals of the n response values at the individual design points. The simplicity of the 2^2 factorial lies in the fact that apart from experimental error, important information comes to the analyst in single-degree-of-freedom components, one each for the two main effects A and B and one degree of freedom for interaction AB. The information retrieved on all these takes the form of three **contrasts**. Let us define the following contrasts among the treatment totals:

$$\begin{aligned} A \text{ contrast} &= ab + a - b - (1), \\ B \text{ contrast} &= ab - a + b - (1), \\ AB \text{ contrast} &= ab - a - b + (1). \end{aligned}$$

The three **effects** from the experiment involve these contrasts and appeal to common sense and intuition. The two computed main effects are of the form

$$\text{effect} = \bar{y}_H - \bar{y}_L,$$

where $\bar{y}_H$ and $\bar{y}_L$ are average response at the high, or "+" level and average response at the low, or "−" level, respectively. As a result,

Calculation of Main Effects

$$A = \frac{ab + a - b - (1)}{2n} = \frac{A \text{ contrast}}{2n}$$

and

$$B = \frac{ab - a + b - (1)}{2n} = \frac{B \text{ contrast}}{2n}.$$

The quantity A is seen to be the *difference between the mean responses at the low and high levels of factor* A. In fact, we call A the **main effect** of factor A. Similarly, B is the main effect of factor B. Apparent interaction in the data is observed by inspecting the difference between $ab - b$ and $a - (1)$ or between $ab - a$ and $b - (1)$ in Table 15.1. If, for example,

$$ab - a \approx b - (1) \qquad \text{or} \qquad ab - a - b + (1) \approx 0,$$

a line connecting the responses for each level of factor A at the high level of factor B will be approximately parallel to a line connecting the responses for each level of factor A at the low level of factor B. The nonparallel lines of Figure 15.1 suggest the presence of interaction. To test whether this apparent interaction is significant, a third contrast in the treatment totals orthogonal to the main effect contrasts, called the **interaction effect**, is constructed by evaluating

Interaction Effect

$$AB = \frac{ab - a - b + (1)}{2n} = \frac{AB \text{ contrast}}{2n}.$$

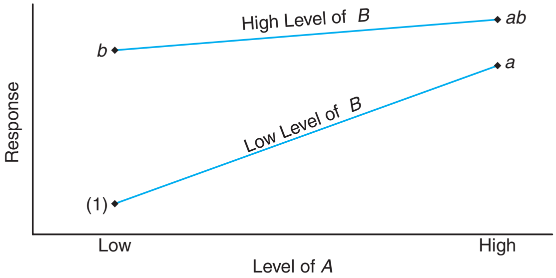


Figure 15.1: Response suggesting apparent interaction.

Example 15.1: Consider the data in Tables 15.2 and 15.3 with $n = 1$ for a 2^2 factorial experiment.

Table 15.2: 2^2 Factorial with No Interaction

	B	
A	−	+
+	50	70
−	80	100

Table 15.3: 2^2 Factorial with Interaction

	B	
A	−	+
+	50	70
−	80	40

The numbers in the cells in Tables 15.2 and 15.3 clearly illustrate how contrasts and the resulting calculation of the two main effects and resulting conclusions can be highly influenced by the presence of interaction. In Table 15.2, the effect of A is −30 at both the low and high levels of factor B and the effect of B is 20 at both the low and high levels of factor A. This "consistency of effect" (no interaction) can be very important information to the analyst. The main effects are

$$A = \frac{70+50}{2} - \frac{100+80}{2} = 60 - 90 = -30,$$
$$B = \frac{100+70}{2} - \frac{80+50}{2} = 85 - 65 = 20,$$

while the interaction effect is

$$AB = \frac{100+50}{2} - \frac{80+70}{2} = 75 - 75 = 0.$$

On the other hand, in Table 15.3 the effect of A is once again -30 at the low level of B but $+30$ at the high level of B. This "inconsistency of effect" (interaction) also is present for B across levels of A. In these cases, the main effects can be meaningless and, in fact, highly misleading. For example, the effect of A is

$$A = \frac{50+70}{2} - \frac{80+40}{2} = 0,$$

since there is a complete "masking" of the effect as one averages over levels of B. The strong interaction is illustrated by the calculated effect

$$AB = \frac{70+80}{2} - \frac{50+40}{2} = 30.$$

Here it is convenient to illustrate the scenarios of Tables 15.2 and 15.3 with interaction plots. Note the parallelism in the plot of Figure 15.2 and the interaction that is apparent in Figure 15.3.

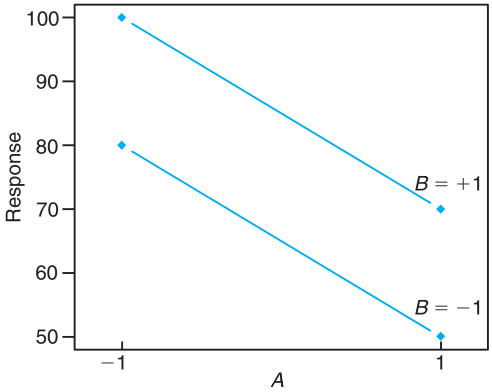


Figure 15.2: Interaction plot for data of Table 15.2.

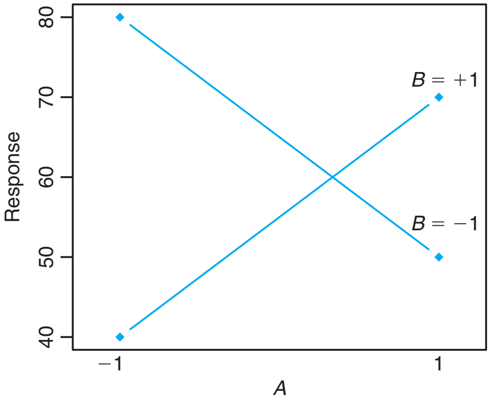


Figure 15.3: Interaction plot for data of Table 15.3.

Computation of Sums of Squares

We take advantage of the fact that in the 2^2 factorial, or for that matter in the general 2^k factorial experiment, each main effect and interaction effect has an associated **single degree of freedom**. Therefore, we can write $2^k - 1$ orthogonal single-degree-of-freedom contrasts in the treatment combinations, each accounting for variation due to some main or interaction effect. Thus, under the usual independence and normality assumptions in the experimental model, we can make tests to determine if the contrast reflects systematic variation or merely chance or random variation. The sums of squares for each contrast are found by following the procedures given in Section 13.5. Writing

$$Y_{1..} = b + (1), \qquad Y_{2..} = ab + a, \qquad c_1 = -1, \qquad \text{and} \qquad c_2 = 1,$$

where $Y_{1..}$ and $Y_{2..}$ are the total of $2n$ observations, we have

$$SSA = SSw_A = \frac{\left(\sum_{i=1}^{2} c_i Y_{i..}\right)^2}{2n \sum_{i=1}^{2} c_i^2} = \frac{[ab + a - b - (1)]^2}{2^2 n} = \frac{(A \text{ contrast})^2}{2^2 n},$$

with 1 degree of freedom. Similarly, we find that

$$SSB = \frac{[ab + b - a - (1)]^2}{2^2 n} = \frac{(B \text{ contrast})^2}{2^2 n}$$

and

$$SS(AB) = \frac{[ab + (1) - a - b]^2}{2^2 n} = \frac{(AB \text{ contrast})^2}{2^2 n}.$$

Each contrast has l degree of freedom, whereas the error sum of squares, with $2^2(n-1)$ degrees of freedom, is obtained by subtraction from the formula

$$SSE = SST - SSA - SSB - SS(AB).$$

In computing the sums of squares for the main effects A and B and the interaction effect AB, it is convenient to present the total responses of the treatment combinations along with the appropriate algebraic signs for each contrast, as in Table 15.4. The main effects are obtained as simple comparisons between the low and high levels. Therefore, we assign a positive sign to the treatment combination that is at the high level of a given factor and a negative sign to the treatment combination at the low level. The positive and negative signs for the interaction effect are obtained by multiplying the corresponding signs of the contrasts of the interacting factors.

Table 15.4: Signs for Contrasts in a 2^2 Factorial Experiment

Treatment	**Factorial Effect**		
Combination	***A***	***B***	***AB***
(1)	−	−	+
a	+	−	−
b	−	+	−
ab	+	+	+

The 2^3 Factorial

Let us now consider an experiment using three factors, A, B, and C, each with levels −1 and +1. This is a 2^3 factorial experiment giving the eight treatment combinations (1), a, b, c, ab, ac, bc, and abc. The treatment combinations and the appropriate algebraic signs for each contrast used in computing the sums of squares for the main effects and interaction effects are presented in Table 15.5.

Table 15.5: Signs for Contrasts in a 2^3 Factorial Experiment

Treatment Combination	**Factorial Effect (symbolic)**						
	$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C}$	$\boldsymbol{AB}$	$\boldsymbol{AC}$	$\boldsymbol{BC}$	$\boldsymbol{ABC}$
(1)	$-$	$-$	$-$	$+$	$+$	$+$	$-$
a	$+$	$-$	$-$	$-$	$-$	$+$	$+$
b	$-$	$+$	$-$	$-$	$+$	$-$	$+$
c	$-$	$-$	$+$	$+$	$-$	$-$	$+$
ab	$+$	$+$	$-$	$+$	$-$	$-$	$-$
ac	$+$	$-$	$+$	$-$	$+$	$-$	$-$
bc	$-$	$+$	$+$	$-$	$-$	$+$	$-$
abc	$+$	$+$	$+$	$+$	$+$	$+$	$+$

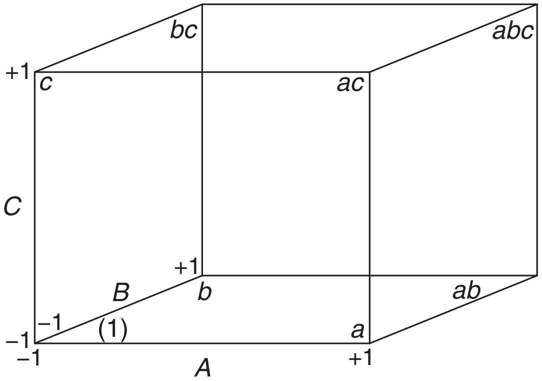


Figure 15.4: Geometric view of 2^3.

It is helpful to discuss and illustrate the geometry of the 2^3 factorial much as we illustrated that of the 2^2 factorial in Figure 15.1. For the 2^3, the **eight design points** represent the vertices of a cube, as shown in Figure 15.4.

The columns of Table 15.5 represent the signs that are used for the contrasts and thus computation of seven effects and corresponding sums of squares. These columns are analogous to those given in Table 15.4 for the case of the 2^2. Seven effects are available since there are eight design points. For example,

$$A = \frac{a + ab + ac + abc - (1) - b - c - bc}{4n},$$

$$AB = \frac{(1) + c + ab + abc - a - b - ac - bc}{4n},$$

and so on. The sums of squares are merely given by

$$SS(\text{effect}) = \frac{(\text{contrast})^2}{2^3 n}.$$

An inspection of Table 15.5 reveals that for the 2^3 experiment all contrasts among the seven are mutually orthogonal, and therefore the seven effects are assessed independently.

Effects and Sums of Squares for the 2^k

For a 2^k factorial experiment the single-degree-of-freedom sums of squares for the main effects and interaction effects are obtained by squaring the appropriate contrasts in the treatment totals and dividing by $2^k n$, where n is the number of replications of the treatment combinations.

As before, an effect is always calculated by subtracting the average response at the "low" level from the average response at the "high" level. The high and low for main effects are quite clear. The symbolic high and low for interactions are evident from information as in Table 15.5.

The orthogonality property has the same importance here as it does for the material on comparisons discussed in Chapter 13. Orthogonality of contrasts implies that the estimated effects and thus the sums of squares are independent. This independence is readily illustrated in the 2^3 factorial experiment if the responses, with factor A at its high level, are increased by an amount x in Table 15.5. Only the A contrast leads to a larger sum of squares, since the x effect cancels out in the formation of the six remaining contrasts as a result of the two positive and two negative signs associated with treatment combinations in which A is at the high level.

There are additional advantages produced by orthogonality. These are pointed out when we discuss the 2^k factorial experiment in regression situations.

15.3 Nonreplicated 2^k Factorial Experiment

The full 2^k factorial may often involve considerable experimentation, particularly when k is large. As a result, replication of each factor combination is often not feasible. If all effects, including all interactions, are included in the model of the experiment, no degrees of freedom are allowed for error. Often, when k is large, the data analyst will *pool* sums of squares and corresponding degrees of freedom for high-order interactions that are known or assumed to be negligible. This will produce F-tests for main effects and lower-order interactions.

Diagnostic Plotting with Nonreplicated 2^k Factorial Experiments

Normal probability plotting can be a very useful methodology for determining the relative importance of effects in a reasonably large two-level factored experiment when there is no replication. This type of diagnostic plot can be particularly useful when the data analyst is hesitant to pool high-order interactions for fear that some of the effects pooled in the "error" may truly be real effects and not merely random. The reader should bear in mind that all effects that are not real (i.e., they are independent *estimates of zero*) follow a normal distribution with mean near zero and constant variance. For example, in a 2^4 factorial experiment, we are reminded that all effects (keep in mind that $n = 1$) are of the form

$$AB = \frac{\text{contrast}}{8} = \bar{y}_H - \bar{y}_L,$$

where $\bar{y}_H$ is the average of eight independent experimental runs at the high, or "+," level and $\bar{y}_L$ is the average of eight independent runs at the low, or "−," level. Thus, the variance of each contrast is $\text{Var}(\bar{y}_H - \bar{y}_L) = \sigma^2/4$. For any real effects, $E(\bar{y}_H - \bar{y}_L) \neq 0$. Thus, normal probability plotting should reveal "significant" effects as those that fall off the straight line that depicts realizations of independent, identically distributed normal random variables.

The probability plotting can take one of many forms. The reader is referred to Chapter 8, where these plots were first presented. The empirical normal quantile-quantile plot may be used. The plotting procedure that makes use of normal probability paper may also be used. In addition, there are several other types of diagnostic normal probability plots. In summary, the procedure for diagnostic effect plots is as follows.

Probability Effect Plots for Nonreplicated 2^4 Factorial Experiments

1. Calculate effects as

$$\text{effect} = \frac{\text{contrast}}{2^{k-1}}.$$

2. Construct a normal probability plot of all effects.

3. Effects that fall off the straight line should be considered real effects.

Further comments regarding normal probability plotting of effects are in order. First, the data analyst may feel frustrated if he or she uses these plots with a small experiment. On the other hand, the plotting is likely to give satisfying results when there is *effect sparsity*—many effects that are truly not real. This sparsity will be evident in large experiments where high-order interactions are not likely to be real.

Case Study 15.1: **Injection Molding:** Many manufacturing companies in the United States and abroad use molded parts as components. Shrinkage is often a major problem. Often, a molded die for a part is built larger than nominal to allow for part shrinkage. In the following experimental situation, a new die is being produced, and ultimately it is important to find the proper process settings to minimize shrinkage. In the following experiment, the response values are deviations from nominal (i.e., shrinkage). The factors and levels are as follows:

	Coded Levels	
	−1	**+1**
***A.* Injection velocity (ft/sec)**	1.0	2.0
***B.* Mold temperature (°C)**	100	150
***C.* Mold pressure (psi)**	500	1000
***D.* Back pressure (psi)**	75	120

The purpose of the experiment was to determine what effects (main effects and interaction effects) influence shrinkage. The experiment was considered a preliminary screening experiment from which the factors for a more complete analysis might be determined. Also, it was hoped that some insight might be gained into how the important factors impact shrinkage. The data from a nonreplicated 2^4 factorial experiment are given in Table 15.6.

Table 15.6: Data for Case Study 15.1

Factor Combination	**Response (cm × 10^4)**	**Factor Combination**	**Response (cm × 10^4)**
(1)	72.68	d	73.52
a	71.74	ad	75.97
b	76.09	bd	74.28
ab	93.19	abd	92.87
c	71.25	cd	79.34
ac	70.59	acd	75.12
bc	70.92	bcd	79.67
abc	104.96	$abcd$	97.80

Initially, effects were calculated and placed on a normal probability plot. The calculated effects are as follows:

$$\begin{aligned}
A &= 10.5613, & BD &= -2.2787, & B &= 12.4463,\\
C &= 2.4138, & D &= 2.1438, & AB &= 11.4038,\\
AC &= 1.2613, & AD &= -1.8238, & BC &= 1.8163,\\
CD &= 1.4088, & ABC &= 2.8588, & ABD &= -1.7813,\\
ACD &= -3.0438, & BCD &= -0.4788, & ABCD &= -1.3063.
\end{aligned}$$

The normal quantile-quantile plot is shown in Figure 15.5. The plot seems to imply that effects A, B, and AB stand out as being important. The signs of the important effects indicate the preliminary conclusions.

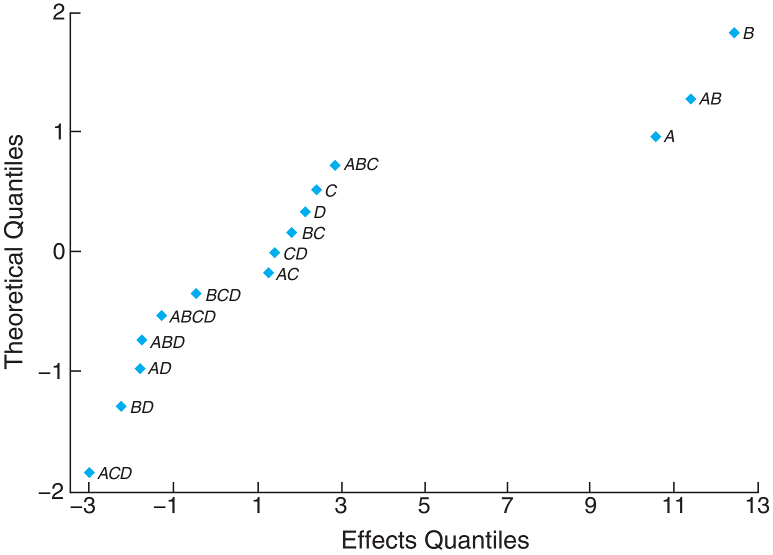


Figure 15.5: Normal quantile-quantile plot of effects for Case Study 15.1.

1. An increase in injection velocity from 1.0 to 2.0 increases shrinkage.
2. An increase in mold temperature from 100°C to 150°C increases shrinkage.
3. There is an interaction between injection velocity and mold temperature; although both main effects are important, it is crucial that we understand the impact of the two-factor interaction.

Interpretation of Two-Factor Interaction

As one would expect, a two-way table of means provides ease in interpretation of the AB interaction. Consider the two-factor situation in Table 15.7.

Table 15.7: Illustration of Two-Factor Interaction

	B (temperature)	
A (velocity)	**100**	**150**
2	73.355	97.205
1	74.1975	75.240

Notice that the large sample mean at high velocity and high temperature created the significant interaction. The **shrinkage increases in a nonadditive manner**. Mold temperature appears to have a positive effect despite the velocity level. But the effect is greatest at high velocity. The velocity effect is very slight at low temperature but clearly is positive at high mold temperature. To control shrinkage at a low level, *one should avoid using high injection velocity and high mold temperature simultaneously.* All of these results are illustrated graphically in Figure 15.6.

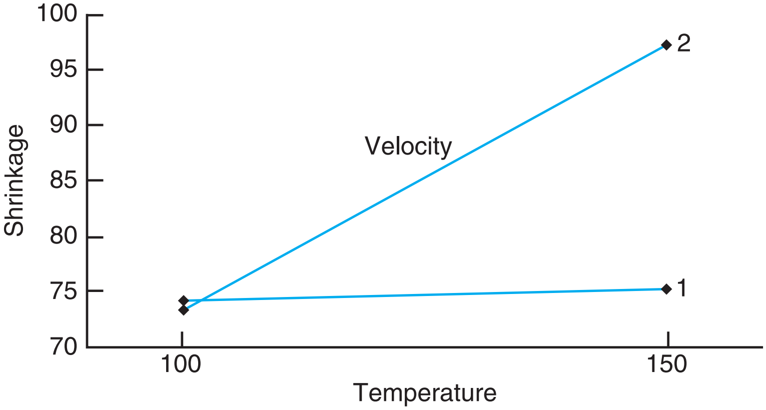


Figure 15.6: Interaction plot for Case Study 15.1.

Analysis with Pooled Mean Square Error: Annotated Computer Printout

It may be of interest to observe an analysis of variance of the injection molding data with high-order interactions pooled to form a mean square error. Interactions of order three and four are pooled. Figure 15.7 shows a *SAS* PROC GLM printout. The analysis of variance reveals essentially the same conclusion as that of the normal probability plot.

The tests and P-values shown in Figure 15.7 require interpretation. A significant P-value suggests that the effect differs significantly from zero. The tests on main effects (which in the presence of interactions may be regarded as the effects averaged over the levels of the other factors) indicate significance for effects A and B. The signs of the effects are also important. An increase in the level from low

```
                        The GLM Procedure
Dependent Variable: y
                                Sum of
 Source             DF         Squares  Mean Square  F Value  Pr > F
 Model              10     1689.237462   168.923746     9.37  0.0117
 Error               5       90.180831    18.036166
 Corrected Total    15     1779.418294
 R-Square     Coeff Var      Root MSE        y Mean
 0.949320      5.308667      4.246901      79.99938
 Source      DF    Type III SS   Mean Square  F Value    Pr > F
 A            1    446.1600062   446.1600062    24.74    0.0042
 B            1    619.6365563   619.6365563    34.36    0.0020
 C            1     23.3047563    23.3047563     1.29    0.3072
 D            1     18.3826563    18.3826563     1.02    0.3590
 A*B          1    520.1820562   520.1820562    28.84    0.0030
 A*C          1      6.3630063     6.3630063     0.35    0.5784
 A*D          1     13.3042562    13.3042562     0.74    0.4297
 B*C          1     13.1950562    13.1950562     0.73    0.4314
 B*D          1     20.7708062    20.7708062     1.15    0.3322
 C*D          1      7.9383063     7.9383063     0.44    0.5364
                                    Standard
 Parameter      Estimate             Error   t Value    Pr > |t|
 Intercept   79.99937500        1.06172520     75.35      <.0001
 A            5.28062500        1.06172520      4.97      0.0042
 B            6.22312500        1.06172520      5.86      0.0020
 C            1.20687500        1.06172520      1.14      0.3072
 D            1.07187500        1.06172520      1.01      0.3590
 A*B          5.70187500        1.06172520      5.37      0.0030
 A*C          0.63062500        1.06172520      0.59      0.5784
 A*D         -0.91187500        1.06172520     -0.86      0.4297
 B*C          0.90812500        1.06172520      0.86      0.4314
 B*D         -1.13937500        1.06172520     -1.07      0.3322
 C*D          0.70437500        1.06172520      0.66      0.5364
```

Figure 15.7: *SAS* printout for data of Case Study 15.1.

to high of A, injection velocity, results in increased shrinkage. The same is true for B. However, because of the significant interaction AB, main effect interpretations may be viewed as trends across the levels of the other factors. The impact of the significant AB interaction is better understood by using a two-way table of means.

Exercises

15.1 The following data are obtained from a 2^3 factorial experiment replicated three times. Evaluate the sums of squares for all factorial effects by the contrast method. Draw conclusions.

Treatment Combination	**Rep 1**	**Rep 2**	**Rep 3**
(1)	12	19	10
a	15	20	16
b	24	16	17
ab	23	17	27
c	17	25	21
ac	16	19	19
bc	24	23	29
abc	28	25	20

15.2 In an experiment conducted by the Mining Engineering Department at Virginia Tech to study a particular filtering system for coal, a coagulant was added to a solution in a tank containing coal and sludge, which was then placed in a recirculation system in order that the coal could be washed. Three factors were varied in the experimental process:

Factor A: percent solids circulated initially in the overflow
Factor B: flow rate of the polymer
Factor C: pH of the tank

The amount of solids in the underflow of the cleansing system determines how clean the coal has become. Two levels of each factor were used and two experimental runs were made for each of the $2^3 = 8$ combinations. The response measurements in percent solids by weight in the underflow of the circulation system are as specified in the following table:

Treatment	**Response**	
Combination	**Replication 1**	**Replication 2**
(1)	4.65	5.81
a	21.42	21.35
b	12.66	12.56
ab	18.27	16.62
c	7.93	7.88
ac	13.18	12.87
bc	6.51	6.26
abc	18.23	17.83

Assuming that all interactions are potentially important, do a complete analysis of the data. Use P-values in your conclusion.

15.3 In a metallurgy experiment, it is desired to test the effect of four factors and their interactions on the concentration (percent by weight) of a particular phosphorus compound in casting material. The variables are A, percent phosphorus in the refinement; B, percent remelted material; C, fluxing time; and D, holding time. The four factors are varied in a 2^4 factorial experiment with two castings taken at each factor combination. The 32 castings were made in random order. The following table shows the data and an ANOVA table is given in Figure 15.8 on page 610. Discuss the effects of the factors and their interactions on the concentration of the phosphorus compound.

Treatment	**Weight % of Phosphorus Compound**		
Combination	**Rep 1**	**Rep 2**	**Total**
(1)	30.3	28.6	58.9
a	28.5	31.4	59.9
b	24.5	25.6	50.1
ab	25.9	27.2	53.1
c	24.8	23.4	48.2
ac	26.9	23.8	50.7
bc	24.8	27.8	52.6
abc	22.2	24.9	47.1
d	31.7	33.5	65.2
ad	24.6	26.2	50.8
bd	27.6	30.6	58.2
abd	26.3	27.8	54.1
cd	29.9	27.7	57.6
acd	26.8	24.2	51.0
bcd	26.4	24.9	51.3
$abcd$	26.9	29.3	56.2
Total	428.1	436.9	865.0

15.4 A preliminary experiment is conducted to study the effects of four factors and their interactions on the output of a certain machining operation. Two runs are made at each of the treatment combinations in order to supply a measure of pure experimental error. Two levels of each factor are used, resulting in the data shown next page. Make tests on all main effects and interactions at the 0.05 level of significance. Draw conclusions.

Source of Variation	Effects	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-Value
Main effect :						
A	−1.2000	11.52	1	11.52	4.68	0.0459
B	−1.2250	12.01	1	12.01	4.88	0.0421
C	−2.2250	39.61	1	39.61	16.10	0.0010
D	1.4875	17.70	1	17.70	7.20	0.0163
Two-factor interaction :						
AB	0.9875	7.80	1	7.80	3.17	0.0939
AC	0.6125	3.00	1	3.00	1.22	0.2857
AD	−1.3250	14.05	1	14.05	5.71	0.0295
BC	1.1875	11.28	1	11.28	4.59	0.0480
BD	0.6250	3.13	1	3.13	1.27	0.2763
CD	0.7000	3.92	1	3.92	1.59	0.2249
Three-factor interaction :						
ABC	−0.5500	2.42	1	2.42	0.98	0.3360
ABD	1.7375	24.15	1	24.15	9.82	0.0064
ACD	1.4875	17.70	1	17.70	7.20	0.0163
BCD	−0.8625	5.95	1	5.95	2.42	0.1394
Four-factor interaction :						
$ABCD$	0.7000	3.92	1	3.92	1.59	0.2249
Error		39.36	16	2.46		
Total		217.51	31			

Figure 15.8: ANOVA table for Exercise 15.3.

Treatment Combination	Replicate 1	Replicate 2
(1)	7.9	9.6
a	9.1	10.2
b	8.6	5.8
c	10.4	12.0
d	7.1	8.3
ab	11.1	12.3
ac	16.4	15.5
ad	7.1	8.7
bc	12.6	15.2
bd	4.7	5.8
cd	7.4	10.9
abc	21.9	21.9
abd	9.8	7.8
acd	13.8	11.2
bcd	10.2	11.1
$abcd$	12.8	14.3

15.5 In the study *An X-Ray Fluorescence Method for Analyzing Polybutadiene-Acrylic Acid (PBAA) Propellants* (Quarterly Reports, RK-TR-62-1, Army Ordnance Missile Command), an experiment was conducted to determine whether or not there was a significant difference in the amount of aluminum obtained in an analysis with certain levels of certain processing variables. The data are shown below.

Obs.	Phys. State	Mixing Time	Blade Speed	Nitrogen Condition	Aluminum
1	1	1	2	2	16.3
2	1	2	2	2	16.0
3	1	1	1	1	16.2
4	1	2	1	2	16.1
5	1	1	1	2	16.0
6	1	2	1	1	16.0
7	1	2	2	1	15.5
8	1	1	2	1	15.9
9	2	1	2	2	16.7
10	2	2	2	2	16.1
11	2	1	1	1	16.3
12	2	2	1	2	15.8
13	2	1	1	2	15.9
14	2	2	1	1	15.9
15	2	2	2	1	15.6
16	2	1	2	1	15.8

The variables in the data are given as below.

A: mixing time

level 1: 2 hours
level 2: 4 hours

B: blade speed
  level 1: 36 rpm
  level 2: 78 rpm
C: condition of nitrogen passed over propellant
  level 1: dry
  level 2: 72% relative humidity
D: physical state of propellant
  level 1: uncured
  level 2: cured

Assuming all three- and four-factor interactions to be negligible, analyze the data. Use a 0.05 level of significance. Write a brief report summarizing the findings.

15.6 It is important to study the effect of the concentration of the reactant and the feed rate on the viscosity of the product from a chemical process. Let the reactant concentration be factor *A*, at levels 15% and 25%. Let the feed rate be factor *B*, with levels 20 lb/hr and 30 lb/hr. The experiment involves two experimental runs at each of the four combinations (L = low and H = high). The viscosity readings are as follows.

B	H	132 137	149 152
	L	145 147	154 150
		L	H
		A	

(a) Assuming a model containing two main effects and an interaction, calculate the three effects. Do you have any interpretation at this point?

(b) Do an analysis of variance and test for interaction. Give conclusions.

(c) Test for main effects and give final conclusions regarding the importance of all these effects.

15.7 Consider Exercise 15.3. It is of interest to the researcher to learn not only that *AD*, *BC*, and possibly *AB* are important, but also what they mean scientifically. Show two-dimensional interaction plots for all three and give an interpretation.

15.8 Consider Exercise 15.3 once again. Three-factor interactions are often not significant, and even if they are, they are difficult to interpret. The interaction *ABD* appears to be important. To gain some sense of interpretation, show two *AD* interaction plots, one for $B = -1$ and the other for $B = +1$. From the appearance of these, give an interpretation of the *ABD* interaction.

15.9 Consider Exercise 15.6. Use a $+1$ and -1 scaling for "high" and "low," respectively, and do a multiple linear regression with the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i,$$

with x_{1i} = reactant concentration $(-1, +1)$ and x_{2i} = feed rate $(-1, +1)$.

(a) Compute regression coefficients.

(b) How do the coefficients b_1, b_2, and b_{12} relate to the effects you found in Exercise 15.6(a)?

(c) In your regression analysis, do *t*-tests on b_1, b_2, and b_{12}. How do these test results relate to those in Exercise 15.6(b) and (c)?

15.10 Consider Exercise 15.5. Compute all 15 effects and do normal probability plots of the effects.

(a) Does it appear as if your assumption of negligible three- and four-factor interactions has merit?

(b) Are the results of the effect plots consistent with what you communicated about the importance of main effects and two-factor interactions in your summary report?

15.11 In Myers, Montgomery, and Anderson-Cook (2009), a data set is discussed in which a 2^3 factorial is used by an engineer to study the effects of cutting speed (*A*), tool geometry (*B*), and cutting angle (*C*) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and duplicates are run at each design point with the order of the runs being random. The data are presented here.

	A	***B***	***C***	**Life**
(1)	−	−	−	22, 31
a	+	−	−	32, 43
b	−	+	−	35, 34
ab	+	+	−	35, 47
c	−	−	+	44, 45
ac	+	−	+	40, 37
bc	−	+	+	60, 50
abc	+	+	+	39, 41

(a) Calculate all seven effects. Which appear, based on their magnitude, to be important?

(b) Do an analysis of variance and observe *P*-values.

(c) Do your results in (a) and (b) agree?

(d) The engineer felt confident that cutting speed and cutting angle should interact. If this interaction is significant, draw an interaction plot and discuss the engineering meaning of the interaction.

15.12 Consider Exercise 15.11. Suppose there was some experimental difficulty in making the runs. In fact, the total experiment had to be halted after only 4 runs. As a result, the abbreviated experiment is given by

	Life
a	43
b	35
c	44
abc	39

With only these runs, we have the signs for contrasts given by

	A	B	C	AB	AC	BC	ABC
a	+	−	−	−	−	+	+
b	−	+	−	−	+	−	+
c	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+

Comment. In your comments, determine whether or not the contrasts are orthogonal. Which are and which are not? Are main effects orthogonal to each other? In this abbreviated experiment (called a *fractional factorial*), can we study interactions independent of main effects? Is it a useful experiment if we are convinced that interactions are negligible? Explain.

15.4 Factorial Experiments in a Regression Setting

Thus far in this chapter, we have mostly confined our discussion of analysis of the data for a 2^k factorial to the method of analysis of variance. The only reference to an alternative analysis resides in Exercise 15.9. Indeed, this exercise introduces much of what motivates the present section. There are situations in which model fitting is important **and** the factors under study **can be controlled**. For example, a biologist may wish to study the growth of a certain type of algae in the water, and so a model that looks at units of algae as a function of the *amount of a pollutant* and, say, *time* would be very helpful. Thus, the study involves a factorial experiment in a laboratory setting in which concentration of the pollutant and time are the factors. As we shall discuss later in this section, a more precise model can be fitted if the factors are controlled in a factorial array, with the 2^k factorial often being a useful choice. In many biological and chemical processes, the levels of the regressor variables can and should be controlled.

Recall that the regression model employed in Chapter 12 can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

The **X** matrix is referred to as the **model matrix**. Suppose, for example, that a 2^3 factorial experiment is employed with the variables

Temperature:	150°C	200°C
Humidity:	15%	20%
Pressure (psi):	1000	1500

The familiar +1, −1 levels can be generated through the following centering and scaling to *design units*:

$$x_1 = \frac{\text{temperature} - 175}{25}, \quad x_2 = \frac{\text{humidity} - 17.5}{2.5}, \quad x_3 = \frac{\text{pressure} - 1250}{250}.$$

As a result, the **X** matrix becomes

$$\mathbf{X} = \begin{array}{cccc|c} & x_1 & x_2 & x_3 & \text{Design Identification} \\ \left[\begin{matrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{matrix}\right. & \begin{matrix} -1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ 1 \end{matrix} & \begin{matrix} -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{matrix} & \left.\begin{matrix} -1 \\ -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \end{matrix}\right] & \begin{matrix} (1) \\ a \\ b \\ c \\ ab \\ ac \\ bc \\ abc \end{matrix} \end{array}$$

It is now seen that the contrasts illustrated and discussed in Section 15.2 are directly related to regression coefficients. Notice that all the columns of the **X** matrix in our 2^3 example are *orthogonal.* As a result, the computation of regression coefficients as described in Section 12.3 becomes

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left(\frac{1}{8}\mathbf{I}\right)\mathbf{X}'\mathbf{y}$$

$$= \frac{1}{8}\begin{bmatrix} a + ab + ac + abc + (1) + b + c + bc \\ a + ab + ac + abc - (1) - b - c - bc \\ b + ab + bc + abc - (1) - a - c - ac \\ c + ac + bc + abc - (1) - a - b - ab \end{bmatrix},$$

where a, ab, and so on, are response measures.

One can now see that the notion of *calculated main effects*, which has been emphasized throughout this chapter with 2^k factorials, is related to coefficients in a fitted regression model when factors are quantitative. In fact, for a 2^k with, say, n experimental runs per design point, the relationships between effects and regression coefficients are as follows:

$$\text{Effect} = \frac{\text{contrast}}{2^{k-1}(n)}$$

$$\text{Regression coefficient} = \frac{\text{contrast}}{2^k(n)} = \frac{\text{effect}}{2}.$$

This relationship should make sense to the reader, since a regression coefficient b_j is an average rate of change in response *per unit change* in x_j. Of course, as one goes from -1 to $+1$ in x_j (low to high), the design variable changes by 2 units.

Example 15.2: Consider an experiment where an engineer desires to fit a linear regression of yield y against holding time x_1 and flexing time x_2 in a certain chemical system. All other factors are held fixed. The data in the natural units are given in Table 15.8. Estimate the multiple linear regression model.

***Solution*:** The fitted regression model is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

Table 15.8: Data for Example 15.2

Holding Time (hr)	Flexing Time (hr)	Yield (%)
0.5	0.10	28
0.8	0.10	39
0.5	0.20	32
0.8	0.20	46

The design units are

$$x_1 = \frac{\text{holding time} - 0.65}{0.15}, \qquad x_2 = \frac{\text{flexing time} - 0.15}{0.05}$$

and the $\mathbf{X}$ matrix is

$$\begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{ccc} & x_1 & x_2 \end{array}$$
$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

with the regression coefficients

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \frac{(1)+a+b+ab}{4} \\ \frac{a+ab-(1)-b}{4} \\ \frac{b+ab-(1)-a}{4} \end{bmatrix} = \begin{bmatrix} 36.25 \\ 6.25 \\ 2.75 \end{bmatrix}.$$

Thus, the least squares regression equation is

$$\hat{y} = 36.25 + 6.25x_1 + 2.75x_2.$$

This example provides an illustration of the use of the two-level factorial experiment in a regression setting. The four experimental runs in the 2^2 design were used to calculate a regression equation, with the obvious interpretation of the regression coefficients. The value $b_1 = 6.25$ represents the estimated increase in response (percent yield) per *design unit* change (0.15 hour) in holding time. The value $b_2 = 2.75$ represents a similar rate of change for flexing time. ■

Interaction in the Regression Model

The interaction contrasts discussed in Section 15.2 have definite interpretations in the regression context. In fact, interactions are accounted for in regression models by product terms. For example, in Example 15.2, the model with interaction is

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$$

with b_0, b_1, b_2 as before and

$$b_{12} = \frac{ab + (1) - a - b}{4} = \frac{46 + 28 - 39 - 32}{4} = 0.75.$$

Thus, the regression equation expressing two *linear main effects* and interaction is

$$\hat{y} = 36.25 + 6.25x_1 + 2.75x_2 + 0.75x_1x_2.$$

The regression context provides a framework in which the reader should better understand the advantage of orthogonality that is enjoyed by the 2^k factorial. In Section 15.2, the merits of orthogonality were discussed from the point of view of *analysis of variance* of the data in a 2^k factorial experiment. It was pointed out that orthogonality among effects leads to independence among the sums of squares. Of course, the presence of regression variables certainly does not rule out the use of analysis of variance. In fact, f-tests are conducted just as they were described in Section 15.2. Of course, a distinction must be made. In the case of ANOVA, the hypotheses evolve from population means, while in the regression case, the hypotheses involve regression coefficients.

For instance, consider the experimental design in Exercise 15.2 on page 609. Each factor is continuous. Suppose that the levels are

A (x_1):	20%	40%
B (x_2):	5 lb/sec	10 lb/sec
C (x_3):	5	5.5

and we have, for design levels,

$$x_1 = \frac{\%\text{ solids} - 30}{10}, \qquad x_2 = \frac{\text{flow rate} - 7.5}{2.5}, \qquad x_3 = \frac{\text{pH} - 5.25}{0.25}.$$

Suppose that it is of interest to fit a multiple regression model in which all linear coefficients and available interactions are to be considered. In addition, the engineer wants to obtain some insight into what levels of the factor will *maximize* cleansing (i.e., maximize the response). This problem will be the subject of Case Study 15.2.

Case Study 15.2: **Coal Cleansing Experiment**[1]: Figure 15.9 represents annotated computer printout for the regression analysis for the fitted model

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3,$$

where x_1, x_2, and x_3 are percent solids, flow rate, and pH of the system, respectively. The computer system used is *SAS* PROC REG.

Note the parameter estimates, standard error, and P-values in the printout. The parameter estimates represent coefficients in the model. All model coefficients are significant except the x_2x_3 term (BC interaction). Note also that residuals, confidence intervals, and prediction intervals appear as discussed in the regression material in Chapters 11 and 12.

The reader can use the values of the model coefficients and predicted values from the printout to ascertain what combination of the factors results in **maximum cleansing efficiency**. Factor A (percent solids circulated) has a large positive coefficient, suggesting a high value for percent solids. In addition, a low value for factor C (pH of the tank) is suggested. Though the B main effect (flow rate of the polymer) coefficient is positive, the rather large positive coefficient of

---

[1] See Exercise 15.2.

```
Dependent Variable: Y
                   Analysis of Variance
                        Sum of      Mean
Source           DF    Squares    Square  F Value  Pr > F
Model             7  490.23499  70.03357   254.43  <.0001
Error             8    2.20205   0.27526
Corrected Total  15  492.43704
Root MSE         0.52465    R-Square     0.9955
Dependent Mean  12.75188    Adj R-Sq     0.9916
Coeff Var        4.11429
                  Parameter Estimates
               Parameter  Standard
Variable   DF   Estimate     Error   t Value  Pr > |t|
Intercept   1   12.75188   0.13116     97.22    <.0001
A           1    4.71938   0.13116     35.98    <.0001
B           1    0.86563   0.13116      6.60    0.0002
C           1   -1.41563   0.13116    -10.79    <.0001
AB          1   -0.59938   0.13116     -4.57    0.0018
AC          1   -0.52813   0.13116     -4.03    0.0038
BC          1    0.00562   0.13116      0.04    0.9668
ABC         1    2.23063   0.13116     17.01    <.0001
    Dependent Predicted    Std Error
Obs  Variable     Value Mean Predict    95% CL Mean       95% CL Predict Residual
  1    4.6500    5.2300       0.3710   4.3745   6.0855   3.7483   6.7117  -0.5800
  2   21.4200   21.3850       0.3710  20.5295  22.2405  19.9033  22.8667   0.0350
  3   12.6600   12.6100       0.3710  11.7545  13.4655  11.1283  14.0917   0.0500
  4   18.2700   17.4450       0.3710  16.5895  18.3005  15.9633  18.9267   0.8250
  5    7.9300    7.9050       0.3710   7.0495   8.7605   6.4233   9.3867   0.0250
  6   13.1800   13.0250       0.3710  12.1695  13.8805  11.5433  14.5067   0.1550
  7    6.5100    6.3850       0.3710   5.5295   7.2405   4.9033   7.8667   0.1250
  8   18.2300   18.0300       0.3710  17.1745  18.8855  16.5483  19.5117   0.2000
  9    5.8100    5.2300       0.3710   4.3745   6.0855   3.7483   6.7117   0.5800
 10   21.3500   21.3850       0.3710  20.5295  22.2405  19.9033  22.8667  -0.0350
 11   12.5600   12.6100       0.3710  11.7545  13.4655  11.1283  14.0917  -0.0500
 12   16.6200   17.4450       0.3710  16.5895  18.3005  15.9633  18.9267  -0.8250
 13    7.8800    7.9050       0.3710   7.0495   8.7605   6.4233   9.3867  -0.0250
 14   12.8700   13.0250       0.3710  12.1695  13.8805  11.5433  14.5067  -0.1550
 15    6.2600    6.3850       0.3710   5.5295   7.2405   4.9033   7.8667  -0.1250
 16   17.8300   18.0300       0.3710  17.1745  18.8855  16.5483  19.5117  -0.2000
```

Figure 15.9: *SAS* printout for data of Case Study 15.2.

$x_1x_2x_3$ (ABC) suggests that flow rate should be at the low level to enhance efficiency. Indeed, the regression model generated in the *SAS* printout suggests that the combination of factors that may produce optimum results, or perhaps suggest direction for further experimentation, is given by

A: high level
B: low level
C: low level

15.5 The Orthogonal Design

In experimental situations where it is appropriate to fit models that are linear in the design variables and possibly should involve interactions or product terms, there are advantages gained from the two-level *orthogonal design*, or orthogonal array. By an orthogonal design we mean one in which there is orthogonality among the columns of the **X** matrix. For example, consider the **X** matrix for the 2^2 factorial of Example 15.2. Notice that all three columns are mutually orthogonal. The **X** matrix for the 2^3 factorial also contains orthogonal columns. The 2^3 factorial with interactions would yield an **X** matrix of the type

$$\mathbf{X} = \begin{array}{c} \\ \end{array}\begin{array}{cccccccc} & x_1 & x_2 & x_3 & x_1x_2 & x_1x_3 & x_2x_3 & x_1x_2x_3 \\ \end{array}$$

$$\mathbf{X} = \left[\begin{array}{rrrrrrrr} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$$

The outline of degrees of freedom is

Source	**d.f.**	
Regression	3	
Lack of fit	4	$(x_1x_2,\ x_1x_3,\ x_2x_3,\ x_1x_2x_3)$
Error (pure)	8	
Total	15	

The 8 degrees of freedom for pure error are obtained from the *duplicate runs* at each design point. Lack-of-fit degrees of freedom may be viewed as the difference between the number of distinct design points and the number of total model terms; in this case, there are 8 points and 4 model terms.

Standard Error of Coefficients and *T*-Tests

In previous sections, we showed how the designer of an experiment may exploit the notion of orthogonality to design a regression experiment with coefficients that attain minimum variance on a per cost basis. We should be able to make use of our exposure to regression in Section 12.4 to compute estimates of variances of coefficients and hence their standard errors. It is also of interest to note the relationship between the t-statistic on a coefficient and the F-statistic described and illustrated in previous chapters.

Recall from Section 12.4 that the variances and covariances of coefficients appear in A^{-1}, or, in terms of present notation, the *variance-covariance matrix* of coefficients is

$$\sigma^2 A^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

In the case of the 2^k factorial experiment, the columns of **X** are mutually orthogonal, imposing a very special structure. In general, for the 2^k we can write

$$\begin{array}{cccccccc} & x_1 & x_2 & \cdots & x_k & x_1x_2 & \cdots & \\ \mathbf{X} = [\mathbf{1} & \pm\mathbf{1} & \pm\mathbf{1} & \cdots & \pm\mathbf{1} & \pm\mathbf{1} & \cdots], \end{array}$$

where each column contains 2^k or $2^k n$ entries, where n is the number of replicate runs at each design point. Thus, formation of $\mathbf{X}'\mathbf{X}$ yields

$$\mathbf{X}'\mathbf{X} = 2^k n \mathbf{I}_p,$$

where $\mathbf{I}$ is the identity matrix of dimension p, the number of model parameters.

Example 15.3: Consider a 2^3 factorial design with duplicated runs fitted to the model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3.$$

Give expressions for the standard errors of the least squares estimates of b_0, b_1, b_2, b_3, b_{12}, b_{13}, and b_{23}.

Solution:

$$\mathbf{X} = \begin{array}{c} \\ \left[\begin{array}{rrrrrrr} & x_1 & x_2 & x_3 & x_1x_2 & x_1x_3 & x_2x_3 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right] \end{array}$$

with each unit viewed as being *repeated* (i.e., each observation is duplicated). As a result,

$$\mathbf{X}'\mathbf{X} = 16\mathbf{I}_7.$$

Thus,

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{16}\mathbf{I}_7.$$

From the foregoing it should be clear that the variances of all coefficients for a 2^k factorial with n runs at each design point are

$$\text{Var}(b_j) = \frac{\sigma^2}{2^k n},$$

and, of course, all covariances are zero. As a result, standard errors of coefficients are calculated as

$$s_{b_j} = s\sqrt{\frac{1}{2^k n}},$$

where s is found from the square root of the mean square error (hopefully obtained from adequate replication). Thus, in our case with the 2^3,

$$s_{b_j} = s\left(\frac{1}{4}\right).$$

Example 15.4: Consider the metallurgy experiment in Exercise 15.3 on page 609. Suppose that the fitted model is

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 \\ + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4.$$

What are the standard errors of the least squares regression coefficients?

Solution: Standard errors of all coefficients for the 2^k factorial are equal and are

$$s_{b_j} = s\sqrt{\frac{1}{2^k n}},$$

which in this illustration is

$$s_{b_j} = s\sqrt{\frac{1}{(16)(2)}}.$$

In this case, the pure mean square error is given by $s^2 = 2.46$ (16 degrees of freedom). Thus,

$$s_{b_j} = 0.28.$$

The standard errors of coefficients can be used to construct t-statistics on all coefficients. These t-values are related to the F-statistics in the analysis of variance. We have already demonstrated that an F-statistic on a coefficient, using the 2^k factorial, is

$$f = \frac{(\text{contrast})^2}{(2^k n)s^2}.$$

This is the form of the F-statistics on page 610 for the metallurgy experiment (Exercise 15.3). It is easy to verify that if we write

$$t = \frac{b_j}{s_{b_j}}, \qquad \text{where} \qquad b_j = \frac{\text{contrast}}{2^k n},$$

then

$$t^2 = \frac{(\text{contrast})^2}{s^2 2^k n} = f.$$

As a result, the usual relationship holds between t-statistics on coefficients and the F-values. As we might expect, the only difference between the use of t and F in assessing significance lies in the fact that the t-statistic indicates the sign, or direction, of the effect of the coefficient.

It would appear that the 2^k factorial plan would handle many practical situations in which regression models are fitted. It can accommodate linear and interaction terms, providing optimal estimates of all coefficients (from a variance point of view). However, when k is large, the number of design points required is very large. Often, portions of the total design can be used and still allow orthogonality with all its advantages. These designs are discussed in Section 15.6.

A More Thorough Look at the Orthogonality Property in the 2^k Factorial

We have learned that for the case of the 2^k factorial all the information that is delivered to the analyst about the main effects and interactions is in the form of contrasts. These "$2^k - 1$ pieces of information" carry a single degree of freedom apiece and they are independent of each other. In an analysis of variance, they manifest themselves as *effects*, whereas if a regression model is being constructed, the effects turn out to be regression coefficients, apart from a factor of 2. With either form of analysis, significance tests can be carried out and the t-test for a given effect is numerically the same as that for the corresponding regression coefficient. In the case of ANOVA, variable screening and scientific interpretation of interactions are important, whereas in the case of a regression analysis, a model may be used to predict response and/or determine which factor/level combinations are optimum (e.g. maximize yield or maximize cleaning efficiency, as in the case of Case Study 15.2).

It turns out that the orthogonality property is important whether the analysis is to be ANOVA or regression. The orthogonality among the columns of $\mathbf{X}$, the model matrix in, say, Example 15.3, provides special conditions that have an important impact on the **variance of effects** or **regression coefficients**. In fact, it has already become apparent that the orthogonal design results in equality of variance for all effects or coefficients. Thus, in this way, the precision, for purposes of estimation or testing, is the same for all coefficients, main effects, or interactions. In addition, if the regression model contains only linear terms and thus only main effects are of interest, the following conditions result in the minimization of variances of all effects (or, correspondingly, first-order regression coefficients).

Conditions for Minimum Variances of Coefficients

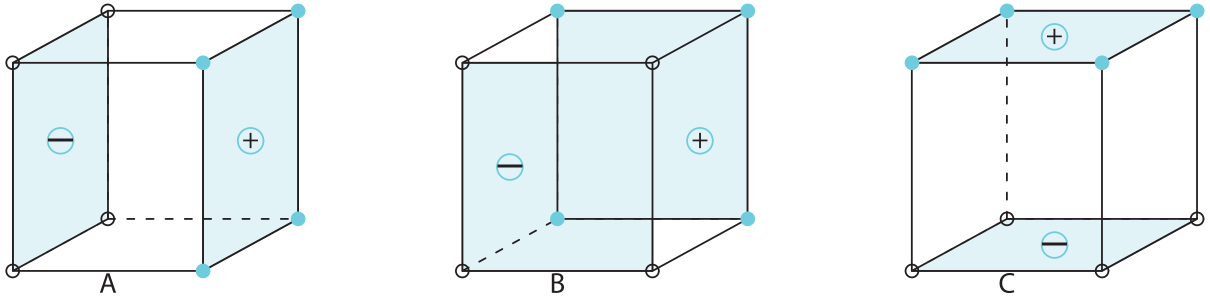
If the regression model contains terms no higher than first order, and if the ranges on the variables are given by $x_j \in [-1, +1]$ for $j = 1, 2, \ldots, k$, then $\text{Var}(b_j)/\sigma^2$, for $j = 1, 2, \ldots, k$, is minimized if the design is orthogonal and all x_i levels in the design are at ± 1 for $i = 1, 2, \ldots, k$.

Thus, in terms of coefficients of model terms or main effects, orthogonality in the 2^k is a very desirable property.

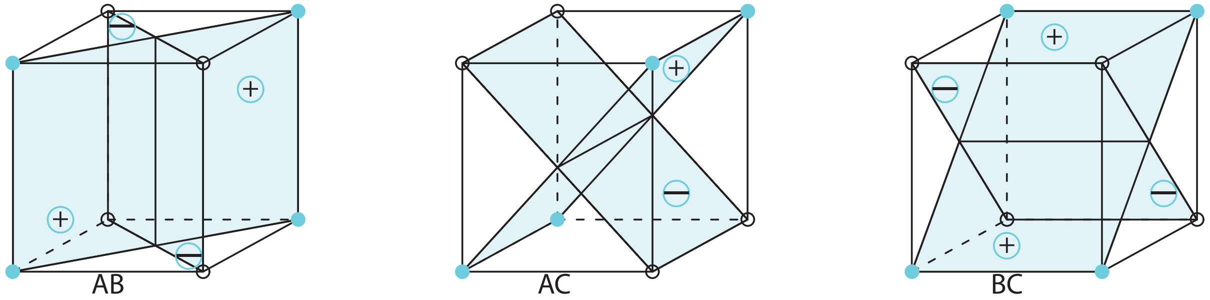
Another approach to a better understanding of the "balance" provided by the 2^3 is to look at the situation graphically. All of the contrasts that are orthogonal and thus mutually independent are shown graphically in Figure 15.10. In the graphs, the planes of the squares whose vertices contain the responses labeled "+" are compared to those containing the responses labeled "−." Those given in (a) show contrasts for main effects and should be obvious to the reader. Those in (b) show the planes representing "+" vertices and "−" vertices for the three two-factor interaction contrasts. In (c), we see the geometric representation of the contrasts for the three-factor (ABC) interaction.

Center Runs with 2^k Designs

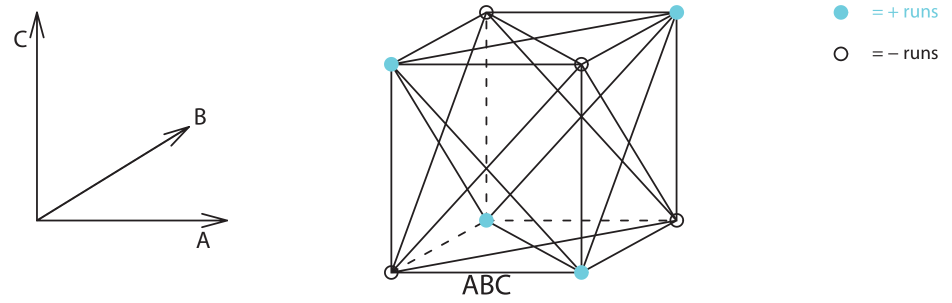
In the situation in which the 2^k design is implemented with **continuous** design variables and one is seeking to fit a linear regression model, the use of replicated runs in the **design center** can be extremely useful. In fact, quite apart from the advantages that will be discussed in what follows, a majority of scientists and



(a) Main effects



(b) Two–factor interactions



(c) Three–factor interaction

Figure 15.10: Geometric presentation of contrasts for the 2^3 factorial design.

engineers would consider center runs (i.e., the runs at $x_i = 0$ for $i = 1, 2, \ldots, k$) as not only a reasonable practice but something that was intuitively appealing. In many areas of application of the 2^k design, the scientist desires to determine if he or she might benefit from moving to a different region of interest in the factors. In many cases, the center (i.e., the point $(0, 0, \ldots, 0)$ in the coded factors) is often either the current operating conditions of the process or at least those conditions that are considered "currently optimum." So it is often the case that the scientist will require data on the response at the center.

Center Runs and Lack of Fit

In addition to the intuitive appeal of the augmentation of the 2^k with center runs, a second advantage is enjoyed that relates to the kind of model that is fitted to the data. Consider, for example, the case with $k = 2$, illustrated in Figure 15.11.

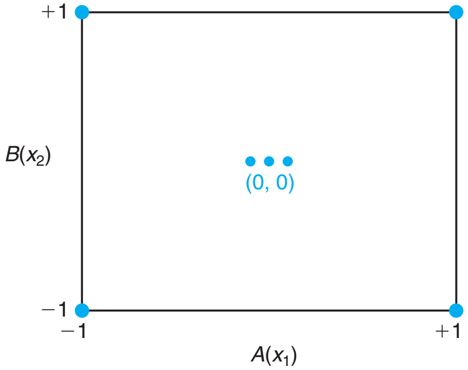


Figure 15.11: A 2^2 design with center runs.

It is clear that *without the center runs* the model terms are the intercept, x_1, x_2, x_1x_2. These account for the four model degrees of freedom delivered by the four design points, apart from any replication. Since each factor has response information available *only at two locations* $\{-1, +1\}$, no "pure" second-order curvature terms can be accommodated in the model (i.e, x_1^2 or x_2^2). But the information at $(0, 0)$ produces an additional model degree of freedom. While this important degree of freedom does not allow both x_1^2 and x_2^2 to be used in the model, it does allow for testing the significance of a linear combination of x_1^2 and x_2^2. For n_c center runs, there are then $n_c - 1$ degrees of freedom available for replication or "pure" error. This allows an estimate of σ^2 for testing the model terms and significance of the 1 d.f. for **quadratic lack of fit**. The concept here is very much like that discussed in the lack-of-fit material in Chapter 11.

In order to gain a complete understanding of how the lack-of-fit test works, assume that for $k = 2$ the **true model** contains the full second-order complement of terms, including x_1^2 and x_2^2. In other words,

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2.$$

Now, consider the contrast

$$\bar{y}_f - \bar{y}_0,$$

where $\bar{y}_f$ is the average response at the factorial locations and $\bar{y}_0$ is the average response at the center point. It can be shown easily (see Review Exercise 15.46) that

$$E(\bar{y}_f - \bar{y}_0) = \beta_{11} + \beta_{22},$$

and, in fact, for the general case with k factors,

$$E(\bar{y}_f - \bar{y}_0) = \sum_{i=1}^{k} \beta_{ii}.$$

As a result, the lack-of-fit test is a simple t-test (or $F = t^2$) with

$$t_{n_c-1} = \frac{\bar{y}_f - \bar{y}_0}{s_{\bar{y}_f - \bar{y}_0}} = \frac{\bar{y}_f - \bar{y}_0}{\sqrt{MSE(1/n_f + 1/n_c)}},$$

where n_f is the number of factorial points and MSE is simply the sample variance of the response values at $(0, 0, \ldots, 0)$.

Example 15.5: This example is taken from Myers, Montgomery, and Anderson-Cook (2009). A chemical engineer is attempting to model the percent conversion in a process. There are two variables of interest, reaction time and reaction temperature. In an attempt to arrive at the appropriate model, a preliminary experiment was conducted in a 2^2 factorial using the current region of interest in reaction time and temperature. Single runs were made at each of the four factorial points and five runs were made at the design center in order that a lack-of-fit test for curvature could be conducted. Figure 15.12 shows the design region and the experimental runs on yield.

The time and temperature readings at the center are, of course, 35 minutes and 145°C. The estimates of the main effects and single interaction coefficient are computed through contrasts, just as before. The center runs **play no role in the computation of b_1, b_2, and b_{12}**. This should be intuitively reasonable to the reader. The intercept is merely $\bar{y}$ for the entire experiment. This value is $\bar{y} = 40.4444$. The standard errors are found through the use of diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$, as discussed earlier. For this case,

$$\mathbf{X} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cccc} & x_1 & x_2 & x_1x_2 \end{array}$$

$$\mathbf{X} = \left[\begin{array}{rrrr} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$$

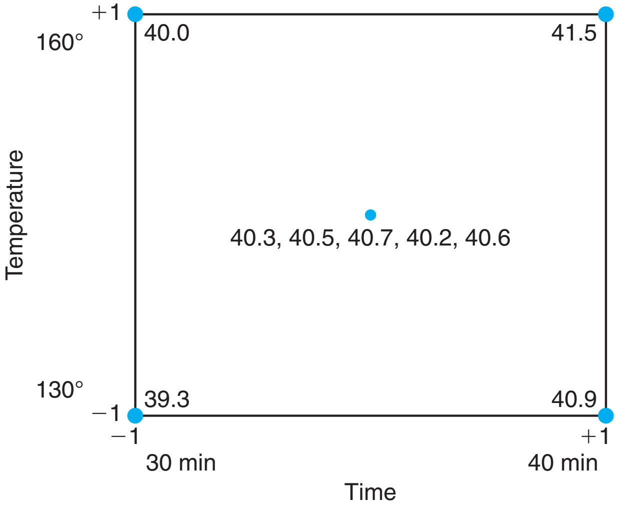


Figure 15.12: 2^2 factorial with 5 center runs.

After the computations, we have

$$\begin{array}{llll} b_0 = 40.4444, & b_1 = 0.7750, & b_2 = 0.3250, & b_{12} = -0.0250, \\ s_{b_0} = 0.06231, & s_{b_1} = 0.09347, & s_{b_2} = 0.09347, & s_{b_{12}} = 0.09347, \\ t_{b_0} = 649.07 & t_{b_1} = 8.29 & t_{b_2} = 3.48 & t_{b_{12}} = -0.27 \quad (P = 0.800). \end{array}$$

The contrast $\bar{y}_f - \bar{y}_0 = 40.425 - 40.46 = -0.035$, and the t-statistic that *tests for curvature* is given by

$$t = \frac{40.425 - 40.46}{\sqrt{0.0430(1/4 + 1/5)}} = 0.251 \qquad (P = 0.814).$$

As a result, it appears as if the appropriate model should contain only first-order terms (apart from the intercept).

An Intuitive Look at the Test on Curvature

If one considers the simple case of a single design variable with runs at -1 and $+1$, it should seem clear that the average response at -1 and $+1$ should be close to the response at 0, the center, if the model is first order in nature. Any deviation would certainly suggest curvature. This is simple to extend to two variables. Consider Figure 15.13.

The figure shows the plane on y that passes through the y values of the factorial points. This is the plane that would represent the perfect fit for the model containing x_1, x_2, and x_1x_2. If the model contains no quadratic curvature (i.e., $\beta_{11} = \beta_{22} = 0$), we would expect the response at $(0, 0)$ to be at or near the plane. If the response is far away from the plane, as in the case of Figure 15.13, then it can be seen graphically that quadratic curvature is present.

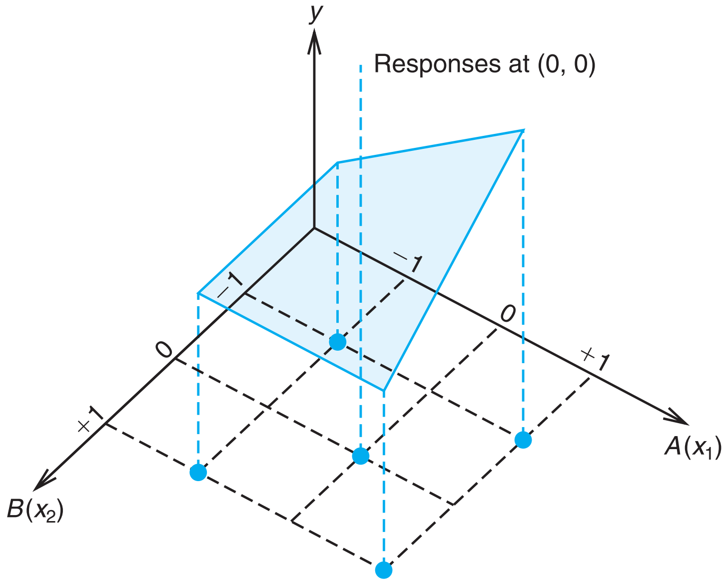


Figure 15.13: 2^2 factorial with runs at $(0,0)$.

Exercises

15.13 Consider a 2^5 experiment where the experimental runs are on 4 different machines. Use the machines as blocks, and assume that all main effects and two-factor interactions may be important.

(a) Which runs would be made on each of the 4 machines?

(b) Which effects are confounded with blocks?

15.14 An experiment is described in Myers, Montgomery, and Anderson-Cook (2009) in which optimum conditions are sought for storing bovine semen to obtain maximum survival. The variables are percent sodium citrate, percent glycerol, and equilibration time in hours. The response is percent survival of the motile spermatozoa. The natural levels are found in the above reference. The data with coded levels for the factorial portion of the design and the center runs are given.

(a) Fit a linear regression model to the data and determine which linear and interaction terms are significant. Assume that the $x_1x_2x_3$ interaction is negligible.

(b) Test for quadratic lack of fit and comment.

x_1, % Sodium Citrate	x_2, % Glycerol	x_3 Equilibration Time	% Survival
−1	−1	−1	57
1	−1	−1	40
−1	1	1	19
1	1	1	40
−1	−1	−1	54
1	−1	−1	41
−1	1	1	21
1	1	1	43
0	0	0	63
0	0	0	61

15.15 Oil producers are interested in nickel alloys that are strong and corrosion resistant. An experiment was conducted in which yield strengths were compared for nickel alloy tensile specimens charged in solutions of sulfuric acid saturated with carbon disulfide. Two alloys were compared: a 75% nickel composition and a 30% nickel composition. The alloys were tested under two different charging times, 25 and 50 days. A 2^3

factorial was conducted with the following factors:

% sulfuric acid: 4%, 6% (x_1)
charging time: 25 days, 50 days (x_2)
nickel composition: 30%, 75% (x_3)

A specimen was prepared for each of the 8 conditions. Since the engineers were not certain of the nature of the model (i.e., whether or not quadratic terms would be needed), a third level (middle level) was incorporated, and 4 center runs were employed using 4 specimens at 5% sulfuric acid, 37.5 days, and 52.5% nickel composition. The following are the yield strengths in kilograms per square inch.

	Charging Time			
	25 Days		**50 Days**	
Nickel	**Sulfuric Acid**		**Sulfuric Acid**	
Comp.	**4%**	**6%**	**4%**	**6%**
75%	52.5	56.5	47.9	47.2
30%	50.2	50.8	47.4	41.7

The center runs gave the following strengths:

51.6, 51.4, 52.4, 52.9

(a) Test to determine which main effects and interactions should be involved in the fitted model.

(b) Test for quadratic curvature.

(c) If quadratic curvature is significant, how many additional design points are needed to determine which quadratic terms should be included in the model?

15.16 Suppose a second replicate of the experiment in Exercise 15.13 could be performed.

(a) Would a second replication of the blocking scheme of Exercise 15.13 be the best choice?

(b) If the answer to part (a) is no, give the layout for a better choice for the second replicate.

(c) What concept did you use in your design selection?

15.17 Consider Figure 15.14, which represents a 2^2 factorial with 3 center runs. If quadratic curvature is significant, what additional design points would you select that might allow the estimation of the terms x_1^2, x_2^2? Explain.

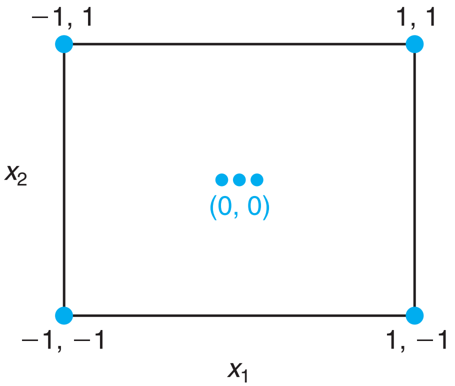


Figure 15.14: Graph for Exercise 15.17.

15.6 Fractional Factorial Experiments

The 2^k factorial experiment can become quite demanding, in terms of the number of experimental units required, when k is large. One of the real advantages of this experimental plan is that it allows a degree of freedom for each interaction. However, in many experimental situations, it is known that certain interactions are negligible, and thus it would be a waste of experimental effort to use the complete factorial experiment. In fact, the experimenter may have an economic constraint that disallows taking observations at all of the 2^k treatment combinations. When k is large, we can often make use of a **fractional factorial experiment** where perhaps one-half, one-fourth, or even one-eighth of the total factorial plan is actually carried out.

Construction of $\frac{1}{2}$ Fraction

The construction of the half-replicate design is identical to the allocation of the 2^k factorial experiment into two blocks. We begin by selecting a defining contrast that is to be completely sacrificed. We then construct the two blocks accordingly and choose either of them as the experimental plan.

A $\frac{1}{2}$ fraction of a 2^k factorial is often referred to as a 2^{k-1} design, the latter indicating the number of design points. Our first illustration of a 2^{k-1} will be a $\frac{1}{2}$ of 2^3, or a 2^{3-1}, design. In other words, the scientist or engineer cannot use the full complement (i.e., the full 2^3 with 8 design points) and hence must settle for a design with only four design points. The question is, of the design points (1), a, b, ab, ac, c, bc, and abc, which four design points would result in the most useful design? The answer, along with the important concepts involved, appears in the table of + and − signs displaying contrasts for the full 2^3. Consider Table 15.9.

Table 15.9: Contrasts for the Seven Available Effects for a 2^3 Factorial Experiment

	Treatment Combination	**Effects**							
		$\boldsymbol{I}$	$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C}$	$\boldsymbol{AB}$	$\boldsymbol{AC}$	$\boldsymbol{BC}$	$\boldsymbol{ABC}$
2^{3-1}	a	+	+	−	−	−	−	+	+
	b	+	−	+	−	−	+	−	+
	c	+	−	−	+	+	−	−	+
	abc	+	+	+	+	+	+	+	+
2^{3-1}	ab	+	+	+	−	+	−	−	−
	ac	+	+	−	+	−	+	−	−
	bc	+	−	+	+	−	−	+	−
	(1)	+	−	−	−	+	+	+	−

Note that the two $\frac{1}{2}$ fractions are $\{a, b, c, abc\}$ and $\{ab, ac, bc, (1)\}$. Note also from Table 15.9 that in both designs ABC has no contrast but all other effects do have contrasts. In one of the fractions we have ABC containing all + signs, and in the other fraction the ABC effect contains all − signs. As a result, we say that the top design in the table is described by $\boldsymbol{ABC = I}$ and the bottom design by $\boldsymbol{ABC = -I}$. The interaction $\boldsymbol{ABC}$ is called the **design generator**, and $ABC = I$ (or $ABC = -I$ for the second design) is called the **defining relation**.

Aliases in the 2^{3-1}

If we focus on the $ABC = I$ design (the upper 2^{3-1}), it becomes apparent that six effects contain contrasts. This produces the initial appearance that all *effects* can be studied apart from ABC. However, the reader can certainly recall that with only four design points, even if points are replicated, the degrees of freedom available (apart from experimental error) are

Regression model terms	3
Intercept	1
	4

A closer look suggests that the seven effects are not orthogonal, and each contrast is represented in another effect. In fact, using $\equiv$ to signify **identical contrasts**, we have

$$A \equiv BC; \qquad B \equiv AC; \qquad C \equiv AB.$$

As a result, within a pair an effect cannot be estimated independently of its alias "partner." The effects

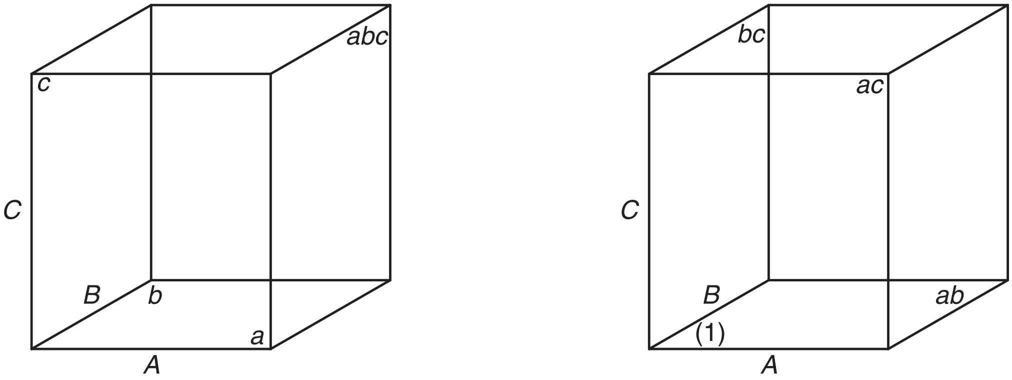
$$A = \frac{a + abc - b - c}{2} \qquad \text{and} \qquad BC = \frac{a + abc - b - c}{2}$$

will produce the same numerical result and thus contain the same information. In fact, it is often said that they **share a degree of freedom**. In truth, the estimated effect actually estimates the sum, namely $A + BC$. We say that A and BC are aliases, B and AC are aliases, and C and AB are aliases.

For the $ABC = -I$ fraction we can observe that the aliases are the same as those for the $ABC = I$ fraction, apart from sign. Thus, we have

$$A \equiv -BC; \qquad B \equiv -AC; \qquad C \equiv -AB.$$

The two fractions appear on corners of the cubes in Figures 15.15(a) and 15.15(b).



(a) The $ABC = I$ fraction

(b) The $ABC = -I$ fraction

Figure 15.15: The $\frac{1}{2}$ fractions of the 2^3 factorial.

How Aliases Are Determined in General

In general, for a 2^{k-1}, each effect, apart from that defined by the generator, will have a *single alias partner*. The effect defined by the generator will not be aliased by another effect but rather will be aliased with the mean since the least squares estimator will be the mean. To determine the alias for each effect, one merely begins with the defining relation, say $ABC = I$ for the 2^{3-1}. Then to find, say, the alias for effect A, multiply A by both sides of the equation $ABC = I$ and reduce any exponent by modulo 2. For example,

$$A \cdot ABC = A, \qquad \text{thus,} \quad BC \equiv A.$$

In a similar fashion,

$$B \equiv B \cdot ABC \equiv AB^2C \equiv AC,$$

and, of course,

$$C \equiv C \cdot ABC \equiv ABC^2 \equiv AB.$$

Now for the second fraction (i.e., defined by the relation $ABC = -I$),

$$A \equiv -BC; \qquad B \equiv -AC; \qquad C \equiv -AB.$$

As a result, the numerical value of effect A is actually estimating $A - BC$. Similarly, the value of B estimates $B - AC$, and the value of C estimates $C - AB$.

Formal Construction of the 2^{k-1}

A clear understanding of the concept of aliasing makes it very simple to understand the construction of the 2^{k-1}. We begin with investigation of the 2^{3-1}. There are three factors and four design points required. The procedure begins with a **full factorial** in $k - 1 = 2$ factors A and B. Then a third factor is added according to the desired alias structures. For example, with ABC as the generator, clearly $C = \pm AB$. Thus, $C = AB$ or $C = -AB$ is found to supplement the full factorial in A and B. Table 15.10 illustrates what is a very simple procedure.

Table 15.10: Construction of the Two 2^{3-1} Designs

Basic 2^2		**2^{3-1}; $ABC = I$**			**2^{3-1}; $ABC = -I$**		
A	B	A	B	$C = AB$	A	B	$C = -AB$
$-$	$-$	$-$	$-$	$+$	$-$	$-$	$-$
$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
$-$	$+$	$-$	$+$	$-$	$-$	$+$	$+$
$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$

Note that we saw earlier that $ABC = I$ gives the design points a, b, c, and abc while $ABC = -I$ gives (1), ac, bc, and ab. Earlier we were able to construct the same designs using the table of contrasts in Table 15.9. However, as the design becomes more complicated with higher fractions, these contrast tables become more difficult to deal with.

Consider now a 2^{4-1} (i.e., a $\frac{1}{2}$ of a 2^4 factorial design) involving factors A, B, C, and D. As in the case of the 2^{3-1}, the highest-order interaction, in this case

$ABCD$, is used as the generator. We must keep in mind that $ABCD = I$; the defining relation suggests that the information on $ABCD$ is sacrificed. Here we begin with the full 2^3 in A, B, and C and form $D = \pm ABC$ to generate the two 2^{4-1} designs. Table 15.11 illustrates the construction of both designs.

Table 15.11: Construction of the Two 2^{4-1} Designs

Basic 2^3			**2^{4-1}; $ABCD = I$**				**2^{4-1}; $ABCD = -I$**			
A	B	C	A	B	C	$D = ABC$	A	B	C	$D = -ABC$
−	−	−	−	−	−	−	−	−	−	+
+	−	−	+	−	−	+	+	−	−	−
−	+	−	−	+	−	+	−	+	−	−
+	+	−	+	+	−	−	+	+	−	+
−	−	+	−	−	+	+	−	−	+	−
+	−	+	+	−	+	−	+	−	+	+
−	+	+	−	+	+	−	−	+	+	+
+	+	+	+	+	+	+	+	+	+	−

Here, using the notation a, b, c, and so on, we have the following designs:

$$ABCD = I,\ (1),\ ad,\ bd,\ ab,\ cd,\ ac,\ bc,\ abcd$$
$$ABCD = -I,\ d,\ a,\ b,\ abd,\ c,\ acd,\ bcd,\ abc.$$

The aliases in the case of the 2^{4-1} are found as illustrated earlier for the 2^{3-1}. Each effect has a single alias partner and is found by multiplication via the use of the defining relation. For example, the alias for A for the $ABCD = I$ design is given by

$$A = A \cdot ABCD = A^2BCD = BCD.$$

The alias for AB is given by

$$AB = AB \cdot ABCD = A^2B^2CD = CD.$$

As we can observe easily, main effects are aliased with three-factor interactions and two-factor interactions are aliased with other two-factor interactions. A complete listing is given by

$$\begin{array}{ll} A = BCD & AB = CD \\ B = ACD & AC = BD \\ C = ABD & AD = BC \\ D = ABC. & \end{array}$$

Construction of the $\frac{1}{4}$ Fraction

In the case of the $\frac{1}{4}$ fraction, two interactions are selected to be sacrificed rather than one, and the third results from finding the generalized interaction of the selected two. Note that this is very much like the construction of four blocks. The fraction used is simply one of the blocks. A simple example aids a great deal in seeing the connection to the construction of the $\frac{1}{2}$ fraction. Consider the construction of $\frac{1}{4}$ of a 2^5 factorial (i.e., a 2^{5-2}), with factors A, B, C, D, and E. One procedure that **avoids the confounding of two main effects** is the choice of ABD and ACE as the interactions that correspond to the two generators, giving $ABD = I$ and $ACE = I$ as the defining relations. The third interaction sacrificed would then be $(ABD)(ACE) = A^2BCDE = BCDE$. For the construction of the design, we begin with a $2^{5-2} = 2^3$ factorial in A, B, and C. We use the interactions ABD and ACE to supply the generators, so the 2^3 factorial in A, B, and C is supplemented by factor $D = \pm AB$ and $E = \pm AC$. Thus, one of the fractions is given by

$$\begin{array}{c}
\begin{array}{ccccc} A & B & C & D=AB & E=AC \end{array} \\
\left[\begin{array}{ccccc}
- & - & - & + & + \\
+ & - & - & - & - \\
- & + & - & - & + \\
+ & + & - & + & - \\
- & - & + & + & - \\
+ & - & + & - & + \\
- & + & + & - & - \\
+ & + & + & + & +
\end{array}\right]
\end{array}
\quad
\begin{array}{l}
\\ de \\ a \\ be \\ abd \\ cd \\ ace \\ bc \\ abcde
\end{array}$$

The other three fractions are found by using the generators $\{D = -AB, E = AC\}$, $\{D = AB, E = -AC\}$, and $\{D = -AB, E = -AC\}$. Consider an analysis of the above 2^{5-2} design. It contains eight design points to study five factors. The aliases for main effects are given by

$$\begin{array}{rlrlrl}
A(ABD) & \equiv BD & A(ACE) & \equiv CE & A(BCDE) & \equiv ABCDE \\
B & \equiv AD & & \equiv ABCE & & \equiv CDE \\
C & \equiv ABCD & & \equiv AE & & \equiv BDE \\
D & \equiv AB & & \equiv ACDE & & \equiv BCE \\
E & \equiv ABDE & & \equiv AC & & \equiv BCD
\end{array}$$

Aliases for other effects can be found in the same fashion. The breakdown of degrees of freedom is given by (apart from replication)

$$\begin{array}{lcl}
\text{Main effects} & 5 & \\
\text{Lack of fit} & \underline{\;2\;} & (CD = BE,\ BC = DE) \\
\text{Total} & 7 &
\end{array}$$

We list interactions only through degree 2 in the lack of fit.

Consider now the case of a 2^{6-2}, which allows 16 design points to study six factors. Once again two design generators are chosen. A pragmatic choice to supplement a $2^{6-2} = 2^4$ full factorial in A, B, C, and D is to use $E = \pm ABC$ and $F = \pm BCD$. The construction is given in Table 15.12.

Obviously, with eight more design points than in the 2^{5-2}, the aliases for main effects will not present as difficult a problem. In fact, note that with defining relations $ABCE = \pm I$, $BCDF = \pm I$, and $(ABCE)(BCDF) = ADEF = \pm I$,

Table 15.12: A 2^{6-2} Design

A	***B***	***C***	***D***	$\boldsymbol{E = ABC}$	$\boldsymbol{F = BCD}$	**Treatment Combination**
−	−	−	−	−	−	(1)
+	−	−	−	+	−	ae
−	+	−	−	+	+	bef
+	+	−	−	−	+	abf
−	−	+	−	+	+	cef
+	−	+	−	−	+	acf
−	+	+	−	−	−	bc
+	+	+	−	+	−	$abce$
−	−	−	+	−	+	df
+	−	−	+	+	+	$adef$
−	+	−	+	+	−	bde
+	+	−	+	−	−	abd
−	−	+	+	+	−	cde
+	−	+	+	−	−	acd
−	+	+	+	−	+	$bcdf$
+	+	+	+	+	+	$abcdef$

main effects will be aliased with interactions that are no less complex than those of third order. The alias structure for main effects is written

$$\begin{aligned} &A \equiv BCE \equiv ABCDF \equiv DEF, && D \equiv ABCDE \equiv BCF \equiv AEF,\\ &B \equiv ACE \equiv CDF \equiv ABDEF, && E \equiv ABC \equiv BCDEF \equiv ADF,\\ &C \equiv ABE \equiv BDF \equiv ACDEF, && F \equiv ABCEF \equiv BCD \equiv ADE, \end{aligned}$$

each with a single degree of freedom. For the two-factor interactions,

$$\begin{aligned} &AB \equiv CE \equiv ACDF \equiv BDEF, && AF \equiv BCEF \equiv ABCD \equiv DE,\\ &AC \equiv BE \equiv ABDF \equiv CDEF, && BD \equiv ACDE \equiv CF \equiv ABEF,\\ &AD \equiv BCDE \equiv ABCF \equiv EF, && BF \equiv ACEF \equiv CD \equiv ABDE,\\ &AE \equiv BC \equiv ABCDEF \equiv DF. \end{aligned}$$

Here, of course, there is some aliasing among the two-factor interactions. The remaining 2 degrees of freedom are accounted for by the following groups:

$$ABD \equiv CDE \equiv ACF \equiv BEF, \qquad ACD \equiv BDE \equiv ABF \equiv CEF.$$

It becomes evident that we should always be aware of what the alias structure is for a fractional experiment before we finally recommend the experimental plan. Proper choice in defining contrasts is important, since it dictates the alias structure.

15.7 Analysis of Fractional Factorial Experiments

The difficulty of making formal significance tests using data from fractional factorial experiments lies in the determination of the proper error term. Unless there are data available from prior experiments, the error must come from a pooling of contrasts representing effects that are presumed to be negligible.

Sums of squares for individual effects are found by using essentially the same procedures given for the complete factorial. We can form a contrast in the treatment combinations by constructing the table of positive and negative signs. For example, for a half-replicate of a 2^3 factorial experiment with ABC the defining contrast, one possible set of treatment combinations, along with the appropriate algebraic sign for each contrast used in computing effects and the sums of squares for the various effects, is presented in Table 15.13.

Table 15.13: Signs for Contrasts in a Half-Replicate of a 2^3 Factorial Experiment

Treatment	**Factorial Effect**						
Combination	$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C}$	$\boldsymbol{AB}$	$\boldsymbol{AC}$	$\boldsymbol{BC}$	$\boldsymbol{ABC}$
a	+	−	−	−	−	+	+
b	−	+	−	−	+	−	+
c	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+

Note that in Table 15.13 the A and BC contrasts are identical, illustrating the aliasing. Also, $B \equiv AC$ and $C \equiv AB$. In this situation, we have three orthogonal contrasts representing the 3 degrees of freedom available. If two observations were obtained for each of the four treatment combinations, we would then have an estimate of the error variance with 4 degrees of freedom. Assuming the interaction effects to be negligible, we could test all the main effects for significance.

An example effect and corresponding sum of squares is

$$A = \frac{a - b - c + abc}{2n}, \qquad SSA = \frac{(a - b - c + abc)^2}{2^2 n}.$$

In general, the single-degree-of-freedom sum of squares for any effect in a 2^{-p} fraction of a 2^k factorial experiment $(p < k)$ is obtained by squaring contrasts in the treatment totals selected and dividing by $2^{k-p}n$, where n is the number of replications of these treatment combinations.

Example 15.6: Suppose that we wish to use a half-replicate to study the effects of five factors, each at two levels, on some response, and it is known that whatever the effect of each factor, it will be constant for each level of the other factors. In other words, there are no interactions. Let the defining contrast be $ABCDE$, causing main effects to be aliased with four-factor interactions. The pooling of contrasts involving interactions provides $15 - 5 = 10$ degrees of freedom for error. Perform an analysis of variance on the data in Table 15.14, testing all main effects for significance at the 0.05 level.

Solution: The sums of squares and effects for the main effects are

$$SSA = \frac{(11.3 - 15.6 - \cdots - 14.7 + 13.2)^2}{2^{5-1}} = \frac{(-17.5)^2}{16} = 19.14,$$

Table 15.14: Data for Example 15.6

Treatment	Response	Treatment	Response
a	11.3	bcd	14.1
b	15.6	abe	14.2
c	12.7	ace	11.7
d	10.4	ade	9.4
e	9.2	bce	16.2
abc	11.0	bde	13.9
abd	8.9	cde	14.7
acd	9.6	$abcde$	13.2

$A = -\frac{17.5}{8} = -2.19,$

$$SSB = \frac{(-11.3 + 15.6 - \cdots - 14.7 + 13.2)^2}{2^{5-1}} = \frac{(18.1)^2}{16} = 20.48,$$

$B = \frac{18.1}{8} = 2.26,$

$$SSC = \frac{(-11.3 - 15.6 + \cdots + 14.7 + 13.2)^2}{2^{5-1}} = \frac{(10.3)^2}{16} = 6.63,$$

$C = \frac{10.3}{8} = 1.21,$

$$SSD = \frac{(-11.3 - 15.6 - \cdots + 14.7 + 13.2)^2}{2^{5-1}} = \frac{(-7.7)^2}{16} = 3.71,$$

$D = \frac{-7.7}{8} = -0.96,$

$$SSE = \frac{(-11.3 - 15.6 - \cdots + 14.7 + 13.2)^2}{2^{5-1}} = \frac{(8.9)^2}{16} = 4.95,$$

$E = \frac{8.9}{8} = 1.11.$

All other calculations and tests of significance are summarized in Table 15.15. The tests indicate that factor A has a significant negative effect on the response, whereas factor B has a significant positive effect. Factors C, D, and E are not significant at the 0.05 level. ▟

Exercises

15.18 List the aliases for the various effects in a 2^5 factorial experiment when the defining contrast is $ACDE$.

15.19 (a) Obtain a $\frac{1}{2}$ fraction of a 2^4 factorial design using BCD as the defining contrast.

(b) Divide the $\frac{1}{2}$ fraction into 2 blocks of 4 units each by confounding ABC.

(c) Show the analysis-of-variance table (sources of variation and degrees of freedom) for testing all unconfounded main effects, assuming that all interaction effects are negligible.

15.20 Construct a $\frac{1}{4}$ fraction of a 2^6 factorial design using $ABCD$ and $BDEF$ as the defining contrasts. Show what effects are aliased with the six main effects.

Table 15.15: Analysis of Variance for the Data of a Half-Replicate of a 2^5 Factorial Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Main effect:				
A	19.14	1	19.14	6.21
B	20.48	1	20.48	6.65
C	6.63	1	6.63	2.15
D	3.71	1	3.71	1.20
E	4.95	1	4.95	1.61
Error	30.83	10	3.08	
Total	85.74	15		

15.21 (a) Using the defining contrasts $ABCE$ and $ABDF$, obtain a $\frac{1}{4}$ fraction of a 2^6 design.

(b) Show the analysis-of-variance table (sources of variation and degrees of freedom) for all appropriate tests, assuming that E and F do not interact and all three-factor and higher interactions are negligible.

15.22 Seven factors are varied at two levels in an experiment involving only 16 trials. A $\frac{1}{8}$ fraction of a 2^7 factorial experiment is used, with the defining contrasts being ACD, BEF, and CEG. The data are as follows:

Treat. Comb.	Response	Treat. Comb.	Response
(1)	31.6	*acg*	31.1
ad	28.7	*cdg*	32.0
abce	33.1	*beg*	32.8
cdef	33.6	*adefg*	35.3
acef	33.7	*efg*	32.4
bcde	34.2	*abdeg*	35.3
abdf	32.5	*bcdfg*	35.6
bf	27.8	*abcfg*	35.1

Perform an analysis of variance on all seven main effects, assuming that interactions are negligible. Use a 0.05 level of significance.

15.23 An experiment is conducted so that an engineer can gain insight into the influence of sealing temperature A, cooling bar temperature B, percent polyethylene additive C, and pressure D on the seal strength (in grams per inch) of a bread-wrapper stock. A $\frac{1}{2}$ fraction of a 2^4 factorial experiment is used, with the defining contrast being $ABCD$. The data are presented here. Perform an analysis of variance on main effects only. Use $\alpha = 0.05$.

A	B	C	D	Response
−1	−1	−1	−1	6.6
1	−1	−1	1	6.9
−1	1	−1	1	7.9
1	1	−1	−1	6.1
−1	−1	1	1	9.2
1	−1	1	−1	6.8
−1	1	1	−1	10.4
1	1	1	1	7.3

15.24 In an experiment conducted at the Department of Mechanical Engineering and analyzed by the Statistics Consulting Center at Virginia Tech, a sensor detects an electrical charge each time a turbine blade makes one rotation. The sensor then measures the amplitude of the electrical current. Six factors are rpm A, temperature B, gap between blades C, gap between blade and casing D, location of input E, and location of detection F. A $\frac{1}{4}$ fraction of a 2^6 factorial experiment is used, with defining contrasts being $ABCE$ and $BCDF$. The data are as follows:

A	B	C	D	E	F	Response
−1	−1	−1	−1	−1	−1	3.89
1	−1	−1	−1	1	−1	10.46
−1	1	−1	−1	1	1	25.98
1	1	−1	−1	−1	1	39.88
−1	−1	1	−1	1	1	61.88
1	−1	1	−1	−1	1	3.22
−1	1	1	−1	−1	−1	8.94
1	1	1	−1	1	−1	20.29
−1	−1	−1	1	−1	1	32.07
1	−1	−1	1	1	1	50.76
−1	1	−1	1	1	−1	2.80
1	1	−1	1	−1	−1	8.15
−1	−1	1	1	1	−1	16.80
1	−1	1	1	−1	−1	25.47
−1	1	1	1	−1	1	44.44
1	1	1	1	1	1	2.45

Perform an analysis of variance on main effects and two-factor interactions, assuming that all three-factor and higher interactions are negligible. Use $\alpha = 0.05$.

15.25 In the study *Durability of Rubber to Steel Adhesively Bonded Joints*, conducted at the Department of Environmental Science and Mechanics and analyzed by the Statistics Consulting Center at Virginia Tech, an experimenter measured the number of breakdowns in an adhesive seal. It was postulated that concentration of seawater A, temperature B, pH C, voltage D, and stress E influence the breakdown of an adhesive seal. A $\frac{1}{2}$ fraction of a 2^5 factorial experiment was used, with the defining contrast being $ABCDE$. The data are as follows:

A	***B***	***C***	***D***	***E***	**Response**
−1	−1	−1	−1	1	462
1	−1	−1	−1	−1	746
−1	1	−1	−1	−1	714
1	1	−1	−1	1	1070
−1	−1	1	−1	−1	474
1	−1	1	−1	1	832
−1	1	1	−1	1	764
1	1	1	−1	−1	1087
−1	−1	−1	1	−1	522
1	−1	−1	1	1	854
−1	1	−1	1	1	773
1	1	−1	1	−1	1068
−1	−1	1	1	1	572
1	−1	1	1	−1	831
−1	1	1	1	−1	819
1	1	1	1	1	1104

Perform an analysis of variance on main effects and two factor interactions AD, AE, BD, BE, assuming that all three-factor and higher interactions are negligible. Use $\alpha = 0.05$.

15.26 Consider a 2^{5-1} design with factors A, B, C, D, and E. Construct the design by beginning with a 2^4 and use $E = ABCD$ as the generator. Show all aliases.

15.27 There are six factors and only eight design points can be used. Construct a 2^{6-3} by beginning with a 2^3 and use $D = AB$, $E = -AC$, and $F = BC$ as the generators.

15.28 Consider Exercise 15.27. Construct another 2^{6-3} that is different from the design chosen in Exercise 15.27.

15.29 For Exercise 15.27, give all aliases for the six main effects.

15.30 In Myers, Montgomery, and Anderson-Cook (2009), an application is discussed in which an engineer is concerned with the effects on the cracking of a titanium alloy. The three factors are A, temperature; B, titanium content; and C, amount of grain refiner. The following table gives a portion of the design and the response, crack length induced in the sample of the alloy.

A	***B***	***C***	**Response**
−1	−1	−1	0.5269
1	1	−1	2.3380
1	−1	1	4.0060
−1	1	1	3.3640

(a) What is the defining relation?

(b) Give aliases for all three main effects assuming that two-factor interactions may be real.

(c) Assuming that interactions are negligible, which main factor is most important?

(d) At what level would you suggest the factor named in (c) be for final production, high or low?

(e) At what levels would you suggest the other factors be for final production?

(f) What hazards lie in the recommendations you made in (d) and (e)? Be thorough in your answer.

15.8 Higher Fractions and Screening Designs

Some industrial situations require the analyst to determine which of a large number of controllable factors have an impact on some important response. The factors may be qualitative or class variables, regression variables, or a mixture of both. The analytical procedure may involve analysis of variance, regression, or both. Often the regression model used involves only linear main effects, although a few interactions may be estimated. The situation calls for variable screening and the resulting experimental designs are known as **screening designs**. Clearly, two-level orthogonal designs that are saturated or nearly saturated are viable candidates.

Design Resolution

Two-level orthogonal designs are often classified according to their **resolution**, the latter determined through the following definition.

Definition 15.1: The **resolution** of a two-level orthogonal design is the length of the smallest (least complex) interaction among the set of defining contrasts.

If the design is constructed as a full or fractional factorial (i.e., either a 2^k or a 2^{k-p} design, $p = 1, 2, \ldots, k-1$), the notion of design resolution is an aid in categorizing the impact of the aliasing. For example, a resolution II design would have little use, since there would be at least one instance of aliasing of one main effect with another. A resolution III design will have all main effects (linear effects) orthogonal to each other. However, there will be some aliasing among linear effects and two-factor interactions. Clearly, then, if the analyst is interested in studying main effects (linear effects in the case of regression) and there are no two-factor interactions, a design of resolution at least III is required.

15.9 Construction of Resolution III and IV Designs with 8, 16, and 32 Design Points

Useful designs of resolution III and IV can be constructed for 2 to 7 variables with 8 design points. We begin with a 2^3 factorial that has been symbolically saturated with interactions.

$$\begin{array}{ccccccc}
x_1 & x_2 & x_3 & x_1x_2 & x_1x_3 & x_2x_3 & x_1x_2x_3 \\
\end{array}$$
$$\begin{bmatrix}
-1 & -1 & -1 & 1 & 1 & 1 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 \\
-1 & 1 & -1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & -1 \\
-1 & 1 & 1 & -1 & -1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}$$

It is clear that a resolution III design can be constructed merely by replacing interaction columns by new main effects through 7 variables. For example, we may define

$$\begin{aligned}
x_4 &= x_1x_2 && \text{(defining contrast } ABD\text{)} \\
x_5 &= x_1x_3 && \text{(defining contrast } ACE\text{)} \\
x_6 &= x_2x_3 && \text{(defining contrast } BCF\text{)} \\
x_7 &= x_1x_2x_3 && \text{(defining contrast } ABCG\text{)}
\end{aligned}$$

and obtain a 2^{-4} fraction of a 2^7 factorial. The preceding expressions identify the chosen defining contrasts. Eleven additional defining contrasts result, and all defining contrasts contain at least three letters. Thus, the design is a resolution III design. Clearly, if we begin with a *subset* of the augmented columns and conclude

Table 15.16: Some Resolution III, IV, V, VI and VII 2^{k-p} Designs

Number of Factors	Design	Number of Points	Generators
3	2_{III}^{3-1}	4	$C = \pm AB$
4	2_{IV}^{4-1}	8	$D = \pm ABC$
5	2_{III}^{5-2}	8	$D = \pm AB;\ E = \pm AC$
6	2_{VI}^{6-1}	32	$F = \pm ABCDE$
	2_{IV}^{6-2}	16	$E = \pm ABC;\ F = \pm BCD$
	2_{III}^{6-3}	8	$D = \pm AB;\ F = \pm BC;\ E = \pm AC$
7	2_{VII}^{7-1}	64	$G = \pm ABCDEF$
	2_{IV}^{7-2}	32	$E = \pm ABC;\ G = \pm ABDE$
	2_{IV}^{7-3}	16	$E = \pm ABC;\ F = \pm BCD;\ G = \pm ACD$
	2_{III}^{7-4}	8	$D = \pm AB;\ E = \pm AC;\ F = \pm BC;\ G = \pm ABC$
8	2_{V}^{8-2}	64	$G = \pm ABCD;\ H = \pm ABEF$
	2_{IV}^{8-3}	32	$F = \pm ABC;\ G = \pm ABD;\ H = \pm BCDE$
	2_{IV}^{8-4}	16	$E = \pm BCD;\ F = \pm ACD;\ G = \pm ABC;\ H = \pm ABD$

with a design involving fewer than 7 design variables, the result is a resolution III design in fewer than 7 variables.

A similar set of possible designs can be constructed for 16 design points by beginning with a 2^4 saturated with interactions. Definitions of variables that correspond to these interactions produce resolution III designs through 15 variables. In a similar fashion, designs containing 32 runs can be constructed by beginning with a 2^5.

Table 15.16 provides guidelines for constructing 8, 16, 32, and 64 point designs that are resolution III, IV and even V. The table gives the number of factors, the number of runs, and the generators that are used to produce the 2^{k-p} designs. The generator given is used to **augment the full factorial** containing $k - p$ factors.

15.10 Other Two-Level Resolution III Designs; The Plackett-Burman Designs

A family of designs developed by Plackett and Burman (1946; see the Bibliography) fills sample size voids that exist with the fractional factorials. The latter are useful with sample sizes 2^r (i.e., they involve sample sizes $4, 8, 16, 32, 64, \ldots$). The Plackett-Burman designs involve $4r$ design points, and thus designs of sizes 12, 20, 24, 28, and so on, are available. These two-level Plackett-Burman designs are resolution III designs and are very simple to construct. “Basic lines” are given for each sample size. These lines of + and − signs are $n - 1$ in number. To construct the columns of the design matrix, we begin with the basic line and do a cyclic permutation on the columns until k (the desired number of variables) columns are formed. Then we fill in the last row with negative signs. The result will be a resolution III design with k variables ($k = 1, 2, \ldots, N$). The basic lines are as follows:

$N = 12$ + + − + + + − − − + −

$N = 16$ + + + + − + − + + − − + − − −

$N = 20$ + + − − + + + + − + − + − − − − + + −

$N = 24$ + + + + + − + − + + − − + + − − + − + − − − −

Example 15.7: Construct a two-level screening design with 6 variables containing 12 design points.

***Solution*:** Begin with the basic line in the initial column. The second column is formed by bringing the bottom entry of the first column to the top of the second column and repeating the first column. The third column is formed in the same fashion, using entries in the second column. When there is a sufficient number of columns, **simply fill in the last row with negative signs**. The resulting design is as follows:

$$\begin{array}{cccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
\end{array}$$
$$\begin{bmatrix}
+ & - & + & - & - & - \\
+ & + & - & + & - & - \\
- & + & + & - & + & - \\
+ & - & + & + & - & + \\
+ & + & - & + & + & - \\
+ & + & + & - & + & + \\
- & + & + & + & - & + \\
- & - & + & + & + & - \\
- & - & - & + & + & + \\
+ & - & - & - & + & + \\
- & + & - & - & - & + \\
- & - & - & - & - & -
\end{bmatrix}$$

The Plackett-Burman designs are popular in industry for screening situations. Because they are resolution III designs, all linear effects are orthogonal. For any sample size, the user has available a design for $k = 2, 3, \ldots, N - 1$ variables.

The alias structure for the Plackett-Burman design is very complicated, and thus the user cannot construct the design with complete control over the alias structure, as in the case of 2^k or 2^{k-p} designs. However, in the case of regression models, the Plackett-Burman design can accommodate interactions (although they will not be orthogonal) when sufficient degrees of freedom are available. ▟

15.11 Introduction to Response Surface Methodology

In Case Study 15.2, a regression model was fitted to a set of data with the specific goal of finding conditions on those design variables that optimize (maximize) the cleansing efficiency of coal. The model contained three linear main effects, three two-factor interaction terms, and one three-factor interaction term. The model response was the cleansing efficiency, and the optimum conditions on x_1, x_2, and x_3

were found by using the signs and the magnitude of the model coefficients. In this example, a two-level design was employed for process improvement or process optimization. In many areas of science and engineering, the application is expanded to involve more complicated models and designs, and this collection of techniques is called **response surface methodology (RSM)**. It encompasses both graphical and analytical approaches. The term *response surface* is derived from the appearance of the multidimensional surface of constant estimated response from a second-order model, i.e., a model with first- and second-order terms. An example will follow.

The Second-Order Response Surface Model

In many industrial examples of process optimization, a *second-order response surface model* is used. For the case of, say, $k = 2$ process variables, or design variables, and a single response y, the model is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon.$$

Here we have $k = 2$ first-order terms, two pure second-order, or quadratic, terms, and one interaction term given by $\beta_{12} x_1 x_2$. The terms x_1 and x_2 are taken to be in the familiar ±1 coded form. The ϵ term designates the usual model error. In general, for k design variables the model will contain $1 + k + k + \binom{k}{2}$ model terms, and hence the experimental design must contain at least a like number of design points. In addition, the quadratic terms require that the design variables be fixed in the design with at least three levels. The resulting design is referred to as a *second-order design*. Illustrations will follow.

The following **central composite design (CCD)** and example is taken from Myers, Montgomery, and Anderson-Cook (2009). Perhaps the most popular class of second-order designs is the class of central composite designs. The example given in Table 15.17 involves a chemical process in which reaction temperature, ξ_1, and reactant concentration, ξ_2, are shown at their natural levels. They also appear in coded form. There are five levels of each of the two factors. In addition, we have the order in which the observations on x_1 and x_2 were run. The column on the right gives values of the response y, the percent conversion of the process. The first four design points represent the familiar factorial points at levels ±1. The next four points are called axial points. They are followed by the center runs that were discussed and illustrated earlier in this chapter. Thus, the five levels of each of the two factors are -1, $+1$, -1.414, $+1.414$, and 0. A clear picture of the geometry of the central composite design for this $k = 2$ example appears in Figure 15.16. This figure illustrates the source of the term **axial points**.These four points are on the factor axes at an axial distance of $\alpha = \sqrt{2} = 1.414$ from the design center. In fact, for this particular CCD, the perimeter points, axial and factorial, are all at the distance $\sqrt{2}$ from the design center, and as a result we have eight equally spaced points on a circle plus four replications at the design center.

Example 15.8: **Response Surface Analysis:** An analysis of the data in the two-variable example may involve the fitting of a second-order response function. The resulting response surface can be used analytically or graphically to determine the impact that x_1

Table 15.17: Central Composite Design for Example 15.8

		Temperature (°C)	**Concentration** (%)			
Observation	**Run**	ξ_1	ξ_2	x_1	x_2	y
1	4	200	15	−1	−1	43
2	12	250	15	1	−1	78
3	11	200	25	−1	1	69
4	5	250	25	1	1	73
5	6	189.65	20	−1.414	0	48
6	7	260.35	20	1.414	0	78
7	1	225	12.93	0	−1.414	65
8	3	225	27.07	0	1.414	74
9	8	225	20	0	0	76
10	10	225	20	0	0	79
11	9	225	20	0	0	83
12	2	225	20	0	0	81

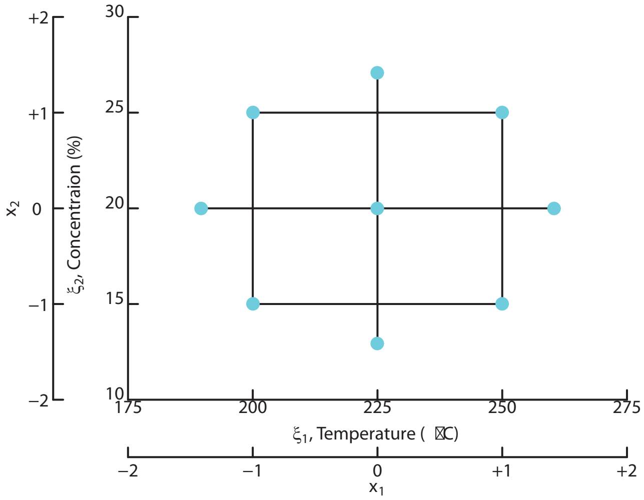


Figure 15.16: Central composite design for Example 15.8.

and x_2 have on percent conversion of the process. The coefficients in the response function are determined by the method of least squares developed in Chapter 12 and illustrated throughout this chapter. The resulting second-order response model is given in the coded variables as

$$\hat{y} = 79.75 + 10.18x_1 + 4.22x_2 - 8.50x_1^2 - 5.25x_2^2 - 7.75x_1x_2,$$

whereas in the natural variables it is given by

$$\hat{y} = -1080.22 + 7.7671\xi_1 + 23.1932\xi_2 - 0.0136\xi_1^2 - 0.2100\xi_2^2 - 0.0620\xi_1\xi_2.$$

Since the current example contains only two design variables, the most illuminating approach to determining the nature of the response surface in the design region is through two- or three-dimensional graphics. It is of interest to determine what levels of temperature x_1 and concentration x_2 produce a desirable estimated percent conversion, $\hat{y}$. The estimated response function above was plotted in three dimensions, and the resulting *response surface* is shown in Figure 15.17. The height of the surface is $\hat{y}$ in percent. It is readily seen from this figure why the term **response surface** is employed. In cases where only two design variables are used, two-dimensional contour plotting can be useful. Thus, make note of Figure 15.18. Contours of constant estimated conversion are seen as slices from the response surface. Note that the viewer of either figure can readily observe which coordinates of temperature and concentration produce the largest estimated percent conversion. In the plots, the coordinates are given in both coded units and natural units. Notice that the largest estimated conversion is at approximately 240°C and 20% concentration. The maximum estimated (or predicted) response at that location is 82.47%.

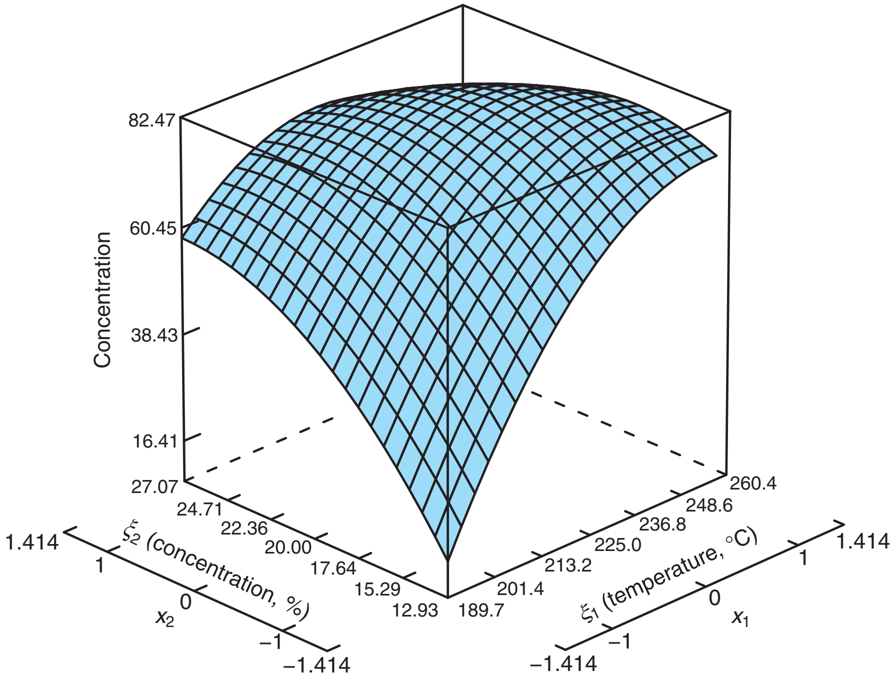


Figure 15.17: Plot for the response surface prediction conversion for Example 15.8.

Other Comments Concerning Response Surface Analysis

The book by Myers, Montgomery, and Anderson-Cook (2009) provides a great deal of information concerning both design and analysis of RSM. The graphical illustration we have used here can be augmented by analytical results that provide information about the nature of the response surface inside the design region.

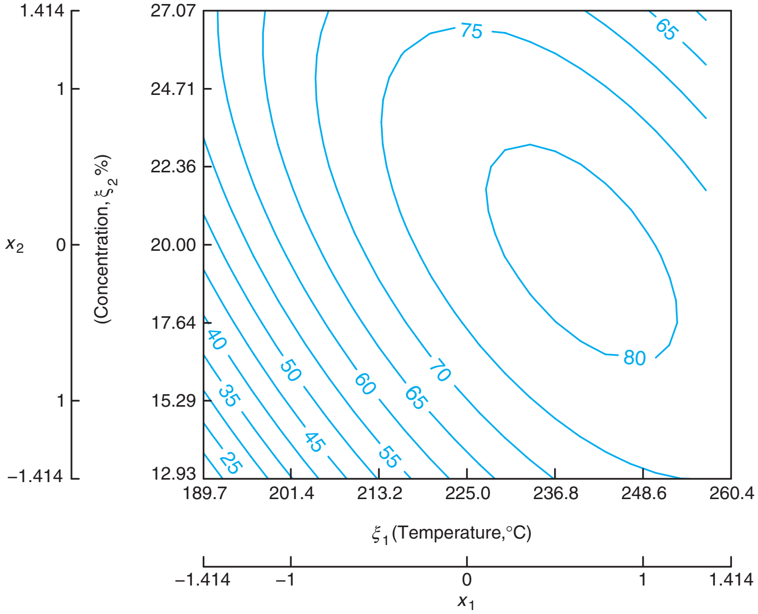


Figure 15.18: Contour plot of predicted conversion for Example 15.8.

Other computations can be used to determine whether the location of the optimum conditions is, in fact, inside or remote from the experimental design region. There are many important considerations when one is required to determine appropriate conditions for future operation of a process.

Other material in Myers, Montgomery, and Anderson-Cook (2009) deals with further experimental design issues. For example, the CCD, while the most generally useful design, is not the only class of design used in RSM. Many others are discussed in the aforementioned text. Also, the CCD discussed here is a special case in which $k = 2$. The more general $k > 2$ case is discussed in Myers, Montgomery, and Anderson-Cook (2009).

15.12 Robust Parameter Design

In this chapter, we have emphasized the notion of using design of experiments (DOE) to learn about engineering and scientific processes. In the case where the process involves a product, DOE can be used to provide product improvement or quality improvement. As we pointed out in Chapter 1, much importance has been attached to the use of statistical methods in product improvement. An important aspect of this quality improvement effort that surfaced in the 1980s and continued through the 1990s is to design quality into processes and products at the research stage or the process design stage. One often requires DOE in the development of processes that have the following properties:

1. Insensitive (robust) to environmental conditions

2. Insensitive (robust) to factors difficult to control
3. Provide minimum variation in performance

The methods used to attain the desirable characteristics in 1, 2, and 3 are a part of what is referred to as *robust parameter design*, or RPD (see Taguchi, 1991; Taguchi and Wu, 1985; and Kackar, 1985, in the Bibliography). The term *design* in this context refers to the design of the process or system; *parameter* refers to the parameters in the system. These are what we have been calling *factors* or *variables*.

It is very clear that goals 1, 2, and 3 above are quite noble. For example, a petroleum engineer may have a fine gasoline blend that performs quite well as long as conditions are ideal and stable. However, the performance may deteriorate because of changes in environmental conditions, such as type of driver, weather conditions, type of engine, and so forth. A scientist at a food company may have a cake mix that is quite good unless the user does not exactly follow directions on the box, directions that deal with oven temperature, baking time, and so forth. A product or process whose performance is consistent when exposed to these changing environmental conditions is called a **robust product** or **robust process**. (See Myers, Montgomery, and Anderson-Cook, 2009, in the Bibliography.)

Control and Noise Variables

Taguchi (1991) emphasized the notion of using two classes of design variables in a study involving RPD: *control factors* and *noise factors*.

Definition 15.2: **Control factors** are variables that can be controlled both in the experiment and in the process. **Noise factors** are variables that may or may not be controlled in the experiment but cannot be controlled in the process (or not controlled well in the process).

An important approach is to use control variables and noise variables in the same experiment as fixed effects. Orthogonal designs or orthogonal arrays are popular designs to use in this effort.

Goal of Robust Parameter Design

The goal of robust parameter design is to choose the levels of the control variables (i.e., the design of the process) that are most robust (insensitive) to changes in the noise variables.

It should be noted that *changes in the noise variables* actually imply changes during the process, changes in the field, changes in the environment, changes in handling or usage by the consumer, and so forth.

The Product Array

One approach to the design of experiments involving both control and noise variables is to use an experimental plan that calls for an orthogonal design for both the control and the noise variables separately. The complete experiment, then, is merely the product or crossing of these two orthogonal designs. The following is a simple example of a product array with two control and two noise variables.

Example 15.9: In the article "The Taguchi Approach to Parameter Design" in *Quality Progress*, December 1987, D. M. Byrne and S. Taguchi discuss an interesting example in which a method is sought for attaching an electrometric connector to a nylon tube so as to deliver the pull-off performance required for an automotive engine application. The objective is to find controllable conditions that maximize pull-off force. Among the controllable variables are A, connector wall thickness, and B, insertion depth. During routine operation there are several variables that cannot be controlled, although they will be controlled during the experiment. Among them are C, conditioning time, and D, conditioning temperature. Three levels are taken for each control variable and two for each noise variable. As a result, the crossed array is as follows. The control array is a 3×3 array, and the noise array is a familiar 2^2 factorial with (1), c, d, and cd representing the four factor combinations. The purpose of the noise factor is to create the *kind of variability in the response, pull-off force, that might be expected in day-to-day operation with the process.* The design is shown in Table 15.18.

Table 15.18: Design for Example 15.9

		B (depth)		
		Shallow	**Medium**	**Deep**
A (wall thickness)	**Thin**	(1)	(1)	(1)
		c	c	c
		d	d	d
		cd	cd	cd
	Medium	(1)	(1)	(1)
		c	c	c
		d	d	d
		cd	cd	cd
	Thick	(1)	(1)	(1)
		c	c	c
		d	d	d
		cd	cd	cd

Case Study 15.3: **Solder Process Optimization:** In an experiment described in *Understanding Industrial Designed Experiments* by Schmidt and Launsby (1991; see the Bibliography), solder process optimization is accomplished by a printed circuit-board assembly plant. Parts are inserted either manually or automatically into a bare board with a circuit printed on it. After the parts are inserted, the board is put through a wave solder machine, which is used to connect all the parts into the circuit. Boards are placed on a conveyor and taken through a series of steps. They are bathed in a flux mixture to remove oxide. To minimize warpage, they are preheated before the solder is applied. Soldering takes place as the boards move across the wave of solder. The object of the experiment is to minimize the number of solder defects per million joints. The control factors and levels are as given in Table 15.19.

Table 15.19: Control Factors for Case Study 15.3

Factor	$(-\mathbf{1})$	$(+\mathbf{1})$
A, solder pot temperature (°F)	480	510
B, conveyor speed (ft/min)	7.2	10
C, flux density	0.9°	1.0°
D, preheat temperature	150	200
E, wave height (in.)	0.5	0.6

These factors are easy to control at the experimental level but are more formidable at the plant or process level.

Noise Factors: Tolerances on Control Factors

Often in processes such as this one, the natural noise factors are tolerances on the control factors. For example, in the actual on-line process, solder pot temperature and conveyor-belt speed are difficult to control. It is known that the control of temperature is within ±5°F and the control of conveyor-belt speed is within ±0.2 ft/min. It is certainly conceivable that variability in the product response (soldering performance) is increased because of an inability to control these two factors at some nominal levels. The third noise factor is the type of assembly involved. In practice, one of two types of assemblies will be used. Thus, we have the noise factors given in Table 15.20.

Table 15.20: Noise Factors for Case Study 15.3

Factor	$(-\mathbf{1})$	$(+\mathbf{1})$
$A*$, solder pot temperature tolerance (°F) (deviation from nominal)	-5	$+5$
$B*$, conveyor speed tolerance (ft/min) (deviation from ideal)	-0.2	$+0.2$
$C*$, assembly type	1	2

Both the control array (inner array) and the noise array (outer array) were chosen to be fractional factorials, the former a $\frac{1}{4}$ of a 2^5 and the latter a $\frac{1}{2}$ of a 2^3. The crossed array and the response values are shown in Table 15.21. The first three columns of the inner array represent a 2^3. The fourth and fifth columns are formed by $D = -AC$ and $E = -BC$. Thus, the defining interactions for the inner array are ACD, BCE, and $ABDE$. The outer array is a standard resolution III fraction of a 2^3. Notice that each inner array point contains runs from the outer array. Thus, four response values are observed at each combination of the control array. Figure 15.19 displays plots which reveal the effect of temperature and density on the mean response.

Table 15.21: Crossed Arrays and Response Values for Case Study 15.3

Inner Array					**Outer Array**					
A	B	C	D	E	(1)	$a{*}b{*}$	$a{*}c{*}$	$b{*}c{*}$	$\bar{y}$	s_y
1	1	1	−1	−1	194	197	193	275	214.75	40.20
1	1	−1	1	1	136	136	132	136	135.00	2.00
1	−1	1	−1	1	185	261	264	264	243.50	39.03
1	−1	−1	1	−1	47	125	127	42	85.25	47.11
−1	1	1	1	−1	295	216	204	293	252.00	48.75
−1	1	−1	−1	1	234	159	231	157	195.25	43.04
−1	−1	1	1	1	328	326	247	322	305.75	39.25
−1	−1	−1	−1	−1	186	187	105	104	145.50	47.35

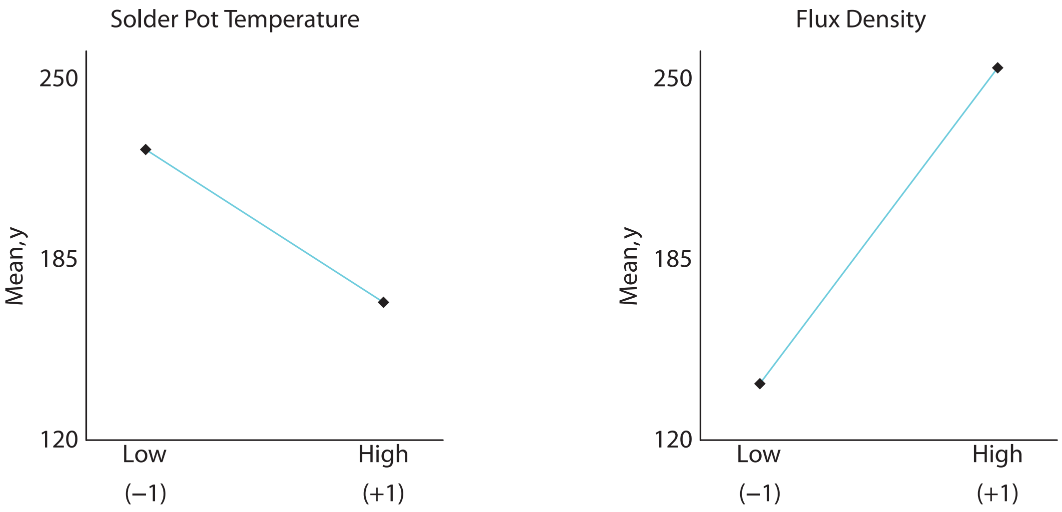


Figure 15.19: Plot showing the influence of factors on the mean response.

Simultaneous Analysis of Process Mean and Variance

In most examples using RPD, the analyst is interested in finding conditions on the control variables that give suitable values for the mean response $\bar{y}$. However, varying the noise variables produces information on the process variance σ_y^2 that might be anticipated in the process. Obviously a robust product is one for which the process is consistent and thus has a small process variance. RPD may involve the simultaneous analysis of $\bar{y}$ and s_y.

It turns out that temperature and flux density are the most important factors in Case Study 15.3. They seem to influence both s_y and $\bar{y}$. Fortunately, *high temperature* and *low flux density* are preferable for both. From Figure 15.19, the "optimum" conditions are

$$\text{solder temperature} = 510^\circ\text{F}, \qquad \text{flux density} = 0.9^\circ.$$

Alternative Approaches to Robust Parameter Design

One approach suggested by many is to model the sample mean and sample variance separately. Separate modeling often helps the experimenter to obtain a better understanding of the process involved. In the following example, we illustrate this approach with the solder process experiment.

Case Study 15.4: Consider the data set of Case Study 15.3. An alternative approach is to fit separate models for the mean $\bar{y}$ and the sample standard deviation. Suppose that we use the usual +1 and −1 coding for the control factors. Based on the apparent importance of solder pot temperature x_1 and flux density x_2, linear regression on the response (number of errors per million joints) produces

$$\hat{y} = 197.125 - 27.5x_1 + 57.875x_2.$$

To find the most robust levels of temperature and flux density, it is important to procure a compromise between the mean response and variability, which requires a modeling of the variability. An important tool in this regard is the log transformation (see Bartlett and Kendall, 1946, or Carroll and Ruppert, 1988):

$$\ln s^2 = \gamma_0 + \gamma_1(x_1) + \gamma_2(x_2).$$

This modeling process produces the following result:

$$\widehat{\ln s^2} = 6.6975 - 0.7458x_1 + 0.6150x_2.$$

The *log linear* model finds extensive use for modeling sample variance, since the log transformation on the sample variance lends itself to use of the method of least squares. This results from the fact that normality and homogeneous variance assumptions are often quite good when one uses $\ln s^2$ rather than s^2 as the model response.

The analysis that is important to the scientist or engineer makes use of the two models simultaneously. A graphical approach can be very useful. Figure 15.20 shows simple plots of the mean and standard deviation models simultaneously. As one would expect, the location of temperature and flux density that minimizes the mean number of errors is the same as that which minimizes variability, namely high temperature and low flux density. The graphical *multiple response surface* approach allows the user to see tradeoffs between process mean and process variability. For this example, the engineer may be dissatisfied with the extreme conditions in solder temperature and flux density. The figure offers estimates of how much is lost as one moves away from the optimum mean and variability conditions to any intermediate conditions. ▟

In Case Study 15.4, values for control variables were chosen that gave desirable conditions for both the mean and the variance of the process. The mean and variance were taken across the distribution of noise variables in the process and were modeled separately, and appropriate conditions were found through a dual response surface approach. Since Case Study 15.4 involved two models (mean and variance), this can be viewed as a dual response surface analysis. Fortunately, in this example the same conditions on the two relevant control variables, solder

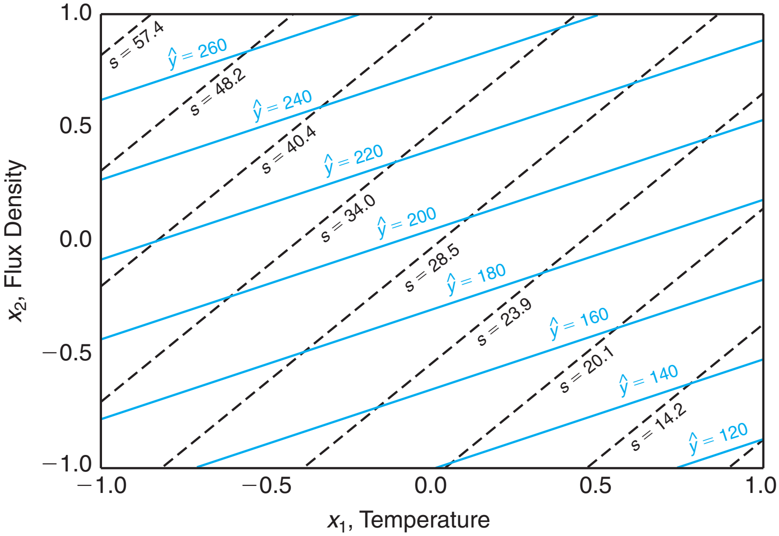


Figure 15.20: Mean and standard deviation for Case Study 15.4.

temperature and flux density, were optimal for both the process mean and the variance. Much of the time in practice some type of compromise between the mean and variance would need to be invoked.

The approach illustrated in Case Study 15.4 involves finding optimal process conditions when the data used are from a product array (or crossed array) type of experimental design. Often, using the product array, a cross between two designs, can be very costly. However, the development of dual response surface models, i.e., a model for the mean and a model for the variance, can be accomplished without a product array. A design that involves both control and noise variables is often called a *combined array*. This type of design and the resulting analysis can be used to determine what conditions on the control variables are most robust (insensitive) to variation in the noise variables. This can be viewed as tantamount to finding control levels that minimize the process variance produced by movement in the noise variables.

The Role of the Control-by-Noise Interaction

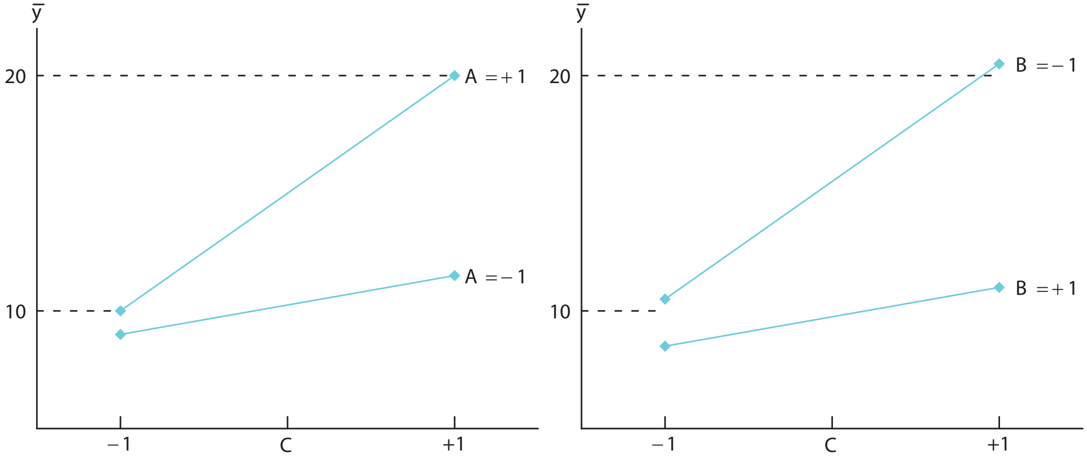
The structure of the process variance is greatly determined by the nature of the control-by-noise interaction. The nature of the nonhomogeneity of process variance is a function of which control variables interact with which noise variables. Specifically, as we will illustrate, those control variables that interact with one or more noise variables can be the object of the analysis. For example, let us consider an illustration used in Myers, Montgomery, and Anderson-Cook (2009) involving two control variables and a single noise variable with the data given in Table 15.22. A and B are control variables and C is a noise variable.

One can illustrate the interactions AC and BC with plots, as given in Figure

Table 15.22: Experimental Data in a Crossed Array

Inner Array		**Outer Array**		
$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C = -1}$	$\boldsymbol{C = +1}$	**Response Mean**
−1	−1	11	15	13.0
−1	1	7	8	7.5
1	−1	10	26	18.0
1	1	10	14	12.0

15.21. One must understand that while A and B are held constant in the process C follows a probability distribution during the process. Given this information, it becomes clear that $A = -1$ and $B = +1$ are levels that produce smaller values for the process variance, while $A = +1$ and $B = -1$ give larger values. Thus, we say that $A = -1$ and $B = +1$ are robust values, i.e., insensitive to inevitable changes in the noise variable C during the process.



(a) AC interaction plot. (b) BC interaction plot.

Figure 15.21: Interaction plots for the data in Table 15.22.

In the above example, we say that both A and B are dispersion effects (i.e. both factors impact the process variance). In addition, both factors are location effects since the mean of y changes as both factors move from -1 to $+1$.

Analysis Involving the Model Containing Both Control and Noise Variables

While it has been emphasized that noise variables are not constant during the working of the process, analysis that results in desirable or even optimal conditions on the control variables is best accomplished through an experiment in which both control and noise variables are fixed effects. Thus, both main effects in the control and noise variables and all the important control-by-noise interactions can be evaluated. This model in x and z, often called a response model, can both directly and indirectly provide useful information regarding the process. The response model is actually a response surface model in vector $\mathbf{x}$ and vector $\mathbf{z}$, where $\mathbf{x}$ contains control variables and $\mathbf{z}$ the noise variables. Certain operations allow models to be generated for the process mean and variance much as in Case Study 15.4. Details are supplied in Myers, Montgomery, and Anderson-Cook (2009); we will illustrate with a very simple example. Consider the data of Table 15.22 on page 650 with control variables A and B and noise variable C. There are eight experimental runs in a $2^2 \times 2$, or 2^3, factorial. Thus, the response model can be written

$$y(x, z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z + \beta_{12} x_1 x_2 + \beta_{1z} x_1 z + \beta_{2z} x_2 z + \epsilon.$$

We will not include the three-factor interaction in the regression model. A, B, and C in Table 15.22 are represented by x_1, x_2, and z, respectively, in the model. We assume that the error term ϵ has the usual independence and constant variance properties.

The Mean and Variance Response Surfaces

The process mean and variance response surfaces are best understood by considering the expectation and variance of z across the process. We assume that the noise variable C [denoted by z in $y(x, z)$] is continuous with mean 0 and variance σ_z^2. The process mean and variance models may be viewed as

$$\begin{aligned} E_z[y(x, z)] &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2, \\ \text{Var}_z[y(x, z)] &= \sigma^2 + \sigma_z^2(\beta_3 + \beta_{1z} x_1 + \beta_{2z} x_2)^2 = \sigma^2 + \sigma_z^2 l_x^2, \end{aligned}$$

where l_x is the slope $\frac{\partial y(x,z)}{\partial z}$ in the direction of z. As we indicated earlier, note how the interactions of factors A and B with the noise variable C are key components of the process variance.

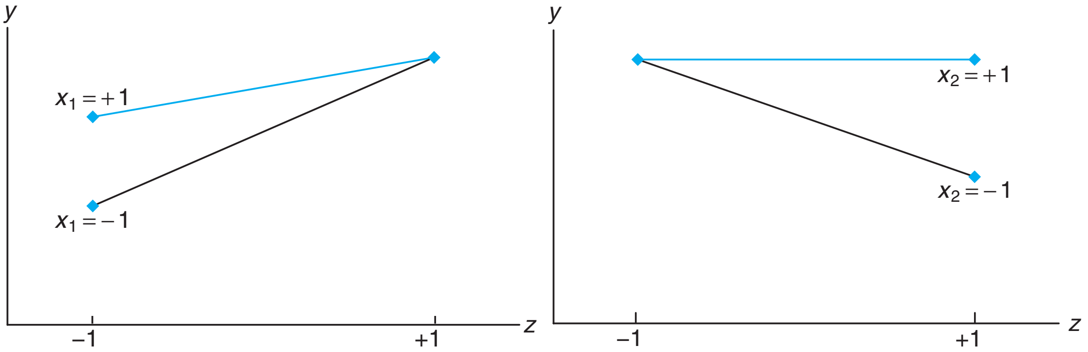
Though we have already analyzed the current example through plots in Figure 15.21, which displayed the role of AB and AC interactions, it is instructive to look at the analysis in light of $E_z[y(x, z)]$ and $\text{Var}_z[y(x, z)]$ above. In this example, the reader can easily verify the estimate b_{1z} for β_{1z} is 15/8 while the estimate b_{2z} for β_{2z} is $-15/8$. The coefficient $b_3 = 25/8$. Thus, the condition $x_1 = +1$ and $x_2 = -1$ results in a process variance estimate of

$$\begin{aligned} \widehat{\text{Var}}_z[y(x, z)] &= \sigma^2 + \sigma_z^2(b_3 + b_{1z} x_1 + b_{2z} x_2)^2 \\ &= \sigma^2 + \sigma_z^2 \left[\frac{25}{8} + \left(\frac{15}{8}\right)(1) + \left(\frac{-15}{8}\right)(-1)\right]^2 = \sigma^2 + \sigma_z^2 \left(\frac{55}{8}\right)^2, \end{aligned}$$

whereas for $x_1 = -1$ and $x_2 = 1$, we have

$$\begin{aligned} \widehat{\text{Var}}_z[y(x, z)] &= \sigma^2 + \sigma_z^2(b_3 + b_{1z} x_1 + b_{2z} x_2)^2 \\ &= \sigma^2 + \sigma_z^2 \left[\frac{25}{8} + \left(\frac{15}{8}\right)(-1) + \left(\frac{15}{8}\right)(-1)\right]^2 = \sigma^2 + \sigma_z^2 \left(\frac{-5}{8}\right)^2. \end{aligned}$$

Thus, for the most desirable (robust) condition of $x_1 = -1$ and $x_2 = 1$, the estimated process variance due to the noise variable C (or z) is $(25/64)\sigma_z^2$. The most undesirable condition, the condition of maximum process variance (i.e., $x_1 = +1$ and $x_2 = -1$), produces an estimated process variance of $(3025/64)\sigma_z^2$. As far as the mean response is concerned, Figure 15.21 indicates that if maximum response is desired $x_1 = +1$ and $x_2 = -1$ produce the best result.



(a) x_1z interaction plot. (b) x_2z interaction plot.

Figure 15.22: Interaction plots for the data in Exercise 15.31.

Exercises

15.31 Consider an example in which there are two control variables x_1 and x_2 and a single noise variable z. The goal is to determine the levels of x_1 and x_2 that are robust to changes in z, i.e., levels of x_1 and x_2 that minimize the variance produced in the response y as z moves between -1 and $+1$. The variables x_1 and x_2 are at two levels, -1 and $+1$, in the experiment. The data produce the plots in Figure 15.22 above. Note that x_1 and x_2 interact with the noise variable z. What settings on x_1 and x_2 (-1 or $+1$ for each) result in minimum variance in y? Explain.

15.32 Consider the following 2^3 factorial with control variables x_1 and x_2 and noise variable z. Can x_1 and x_2 be chosen at levels for which $\text{Var}(y)$ is minimized? Explain why or why not.

	$z = -1$		$z = +1$	
	$x_2 = -1$	$x_2 = +1$	$x_2 = -1$	$x_2 = +1$
$x_1 = -1$	4	6	8	10
$x_1 = +1$	1	3	3	5

15.33 Consider Case Study 15.1 involving the injection molding data. Suppose mold temperature is difficult to control and thus it can be assumed that in the process it follows a normal distribution with mean 0 and variance σ_z^2. Of concern is the variance of the shrinkage response in the process itself. In the analysis of Figure 15.7, it is clear that mold temperature, injection velocity, and the interaction between the two are the only important factors.

(a) Can the setting on velocity be used to create some type of control on the process variance in shrinkage which arises due to the inability to control temperature? Explain.

(b) Using parameter estimates from Figure 15.7, give an estimate of the following models:

(i) mean shrinkage across the distribution of temperature;

(ii) shrinkage variance as a function of σ_z^2.

(c) Use the estimated variance model to determine the level of velocity that minimizes the shrinkage variance.

(d) Use the mean shrinkage model to determine what value of velocity minimizes mean shrinkage.

(e) Are your results above consistent with your analysis from the interaction plot in Figure 15.6? Explain.

15.34 In Case Study 15.2 involving the coal cleansing data, the percent solids in the process system is known to vary uncontrollably during the process and is viewed as a noise factor with mean 0 and variance σ_z^2. The response, cleansing efficiency, has a mean and variance that change behavior during the process. Use only significant terms in the following parts.

(a) Use the estimates in Figure 15.9 to develop the process mean efficiency and variance models.

(b) What factor (or factors) might be controlled at certain levels to control or otherwise minimize the process variance?

(c) What conditions of factors B and C within the design region maximize the estimated mean?

(d) What level of C would you suggest for minimization of process variance when $B = 1$? When $B = -1$?

15.35 Use the coal cleansing data of Exercise 15.2 on page 609 to fit a model of the type

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where the levels are

x_1, percent solids: 8, 12

x_2, flow rate: 150, 250 gal/min

x_3, pH: 5, 6

Center and scale the variables to design units. Also conduct a test for lack of fit, and comment concerning the adequacy of the linear regression model.

15.36 A 2^5 factorial plan is used to build a regression model containing first-order coefficients and model terms for all two-factor interactions. Duplicate runs are made for each factor. Outline the analysis-of-variance table, showing degrees of freedom for regression, lack of fit, and pure error.

15.37 Consider the $\frac{1}{16}$ of the 2^7 factorial discussed in Section 15.9. List the additional 11 defining contrasts.

15.38 Construct a Plackett-Burman design for 10 variables containing 24 experimental runs.

Review Exercises

15.39 A Plackett-Burman design was used to study the rheological properties of high-molecular-weight copolymers. Two levels of each of six variables were fixed in the experiment. The viscosity of the polymer is the response. The data were analyzed by the Statistics Consulting Center at Virginia Tech for personnel in the Chemical Engineering Department at the University. The variables are as follows: hard block chemistry x_1, nitrogen flow rate x_2, heat-up time x_3, percent compression x_4, scans (high and low) x_5, percent strain x_6. The data are presented here.

Obs.	x_1	x_2	x_3	x_4	x_5	x_6	y
1	1	−1	1	−1	−1	−1	194,700
2	1	1	−1	1	−1	−1	588,400
3	−1	1	1	−1	1	−1	7533
4	1	−1	1	1	−1	1	514,100
5	1	1	−1	1	1	−1	277,300
6	1	1	1	−1	1	1	493,500
7	−1	1	1	1	−1	1	8969
8	−1	−1	1	1	1	−1	18,340
9	−1	−1	−1	1	1	1	6793
10	1	−1	−1	−1	1	1	160,400
11	−1	1	−1	−1	−1	1	7008
12	−1	−1	−1	−1	−1	−1	3637

Build a regression equation relating viscosity to the levels of the six variables. Conduct t-tests for all main effects. Recommend factors that should be retained for future studies and those that should not. Use the residual mean square (5 degrees of freedom) as a measure of experimental error.

15.40 A large petroleum company in the Southwest regularly conducts experiments to test additives to drilling fluids. Plastic viscosity is a rheological measure reflecting the thickness of the fluid. Various polymers are added to the fluid to increase viscosity. The following is a data set in which two polymers are used at two levels each and the viscosity measured. The concentration of the polymers is indicated as "low" or "high." Conduct an analysis of the 2^2 factorial experiment. Test for effects for the two polymers and interaction.

	Polymer 1			
Polymer 2	**Low**		**High**	
Low	3.0	3.5	11.3	12.0
High	11.7	12.0	21.7	22.4

15.41 A 2^2 factorial experiment is analyzed by the Statistics Consulting Center at Virginia Tech. The client is a member of the Department of Housing, Interior Design, and Resource Management. The client is interested in comparing cold start to preheating ovens in terms of total energy delivered to the product. In addition, convection is being compared to regular mode. Four experimental runs are made at each of the four factor combinations. Following are the data from the experiment:

	Preheat		Cold	
Convection	618	619.3	575	573.7
Mode	629	611	574	572
Regular	581	585.7	558	562
Mode	581	595	562	566

Do an analysis of variance to study main effects and interaction. Draw conclusions.

15.42 In the study "The Use of Regression Analysis for Correcting Matrix Effects in the X-Ray Fluorescence Analysis of Pyrotechnic Compositions," published in the *Proceedings of the Tenth Conference on the Design of Experiments in Army Research Development and Testing*, ARO-D Report 65-3 (1965), an experiment was conducted in which the concentrations of four components of a propellant mixture and the weights of fine and coarse particles in the slurry were each allowed to vary. Factors A, B, C, and D, each at two levels, represent the concentrations of the four components, and factors E and F, also at two levels, represent the weights of the fine and coarse particles present in the slurry. The goal of the analysis was to determine if the X-ray intensity ratios associated with component 1 of the propellant were significantly influenced by varying the concentrations of the various components and the weights of the particles in the mixture. A $\frac{1}{8}$ fraction of a 2^6 factorial experiment was used, with the defining contrasts being ADE, BCE, and ACF. The data shown here represent the total of a pair of intensity readings.
The pooled mean square error with 8 degrees of freedom is given by 0.02005. Analyze the data using a 0.05 level of significance to determine if the concentrations of the components and the weights of the fine and coarse particles present in the slurry have a significant influence on the intensity ratios associated with component 1. Assume that no interaction exists among the six factors.

Batch	Treatment Combination	Intensity Ratio Total
1	$abef$	2.2480
2	$cdef$	1.8570
3	(1)	2.2428
4	ace	2.3270
5	bde	1.8830
6	$abcd$	1.8078
7	adf	2.1424
8	bcf	1.9122

15.43 Use Table 15.16 to construct a 16-run design with 8 factors that is resolution IV.

15.44 Verify that your design in Review Exercise 15.43 is indeed resolution IV.

15.45 Construct a design that contains 9 design points, is orthogonal, contains 12 total runs and 3 degrees of freedom for replication error, and allows for a lack-of-fit test for pure quadratic curvature.

15.46 Consider a design which is a 2_{III}^{3-1} with 2 center runs. Consider $\bar{y}_f$ as the average response at the design parameter and $\bar{y}_0$ as the average response at the design center. Suppose the true regression model is

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2.$$

(a) Give (and verify) $E(\bar{y}_f - \bar{y}_0)$.

(b) Explain what you have learned from the result in (a).

15.13 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

In the use of fractional factorial experiments, one of the most important considerations that the analyst must be aware of is the *design resolution*. A design of low resolution is smaller (and hence cheaper) than one of higher resolution. However, a price is paid for the cheaper design. The design of lower resolution has heavier aliasing than one of higher resolution. For example, if the researcher has expectations that two-factor interactions may be important, then resolution III should not be used. A resolution III design is strictly a **main effects plan**.

Chapter 16

Nonparametric Statistics

16.1 Nonparametric Tests

Most of the hypothesis-testing procedures discussed in previous chapters are based on the assumption that the random samples are selected from normal populations. Fortunately, most of these tests are still reliable when we experience slight departures from normality, particularly when the sample size is large. Traditionally, these testing procedures have been referred to as **parametric methods**. In this chapter, we consider a number of alternative test procedures, called **nonparametric** or **distribution-free methods**, that often assume no knowledge whatsoever about the distributions of the underlying populations, except perhaps that they are continuous.

Nonparametric, or distribution-free procedures, are used with increasing frequency by data analysts. There are many applications in science and engineering where the data are reported as values not on a continuum but rather on an **ordinal scale** such that it is quite natural to assign ranks to the data. In fact, the reader may notice quite early in this chapter that the distribution-free methods described here involve an *analysis of ranks*. Most analysts find the computations involved in nonparametric methods to be very appealing and intuitive.

For an example where a nonparametric test is applicable, consider the situation in which two judges rank five brands of premium beer by assigning a rank of 1 to the brand believed to have the best overall quality, a rank of 2 to the second best, and so forth. A nonparametric test could then be used to determine whether there is any agreement between the two judges.

We should also point out that there are a number of disadvantages associated with nonparametric tests. Primarily, they do not utilize all the information provided by the sample, and thus a nonparametric test will be less efficient than the corresponding parametric procedure when both methods are applicable. Consequently, to achieve the same power, a nonparametric test will require a larger sample size than will the corresponding parametric test.

As we indicated earlier, slight departures from normality result in minor deviations from the ideal for the standard parametric tests. This is particularly true for the t-test and the F-test. In the case of the t-test and the F-test, the P-value quoted may be slightly in error if there is a moderate violation of the normality assumption.

In summary, if a parametric and a nonparametric test are both applicable to the same set of data, we should carry out the more efficient parametric technique. However, we should recognize that the assumptions of normality often cannot be justified and that we do not always have quantitative measurements. It is fortunate that statisticians have provided us with a number of useful nonparametric procedures. Armed with nonparametric techniques, the data analyst has more ammunition to accommodate a wider variety of experimental situations. It should be pointed out that even under the standard normal theory assumptions, the efficiencies of the nonparametric techniques are remarkably close to those of the corresponding parametric procedure. On the other hand, serious departures from normality will render the nonparametric method much more efficient than the parametric procedure.

Sign Test

The reader should recall that the procedures discussed in Section 10.4 for testing the null hypothesis that $\mu = \mu_0$ are valid only if the population is approximately normal or if the sample is large. If $n < 30$ and the population is decidedly nonnormal, we must resort to a nonparametric test.

The sign test is used to test hypotheses on a population *median*. In the case of many of the nonparametric procedures, the mean is replaced by the median as the pertinent **location parameter** under test. Recall that the sample median was defined in Section 1.3. The population counterpart, denoted by $\tilde{\mu}$, has an analogous definition. Given a random variable X, $\tilde{\mu}$ is defined such that $P(X > \tilde{\mu}) \leq 0.5$ and $P(X < \tilde{\mu}) \leq 0.5$. In the continuous case,

$$P(X > \tilde{\mu}) = P(X < \tilde{\mu}) = 0.5.$$

Of course, if the distribution is symmetric, the population mean and median are equal. In testing the null hypothesis H_0 that $\tilde{\mu} = \tilde{\mu}_0$ against an appropriate alternative, on the basis of a random sample of size n, we replace each sample value exceeding $\tilde{\mu}_0$ with a *plus* sign and each sample value less than $\tilde{\mu}_0$ with a *minus* sign. If the null hypothesis is true and the population is symmetric, the sum of the plus signs should be approximately equal to the sum of the minus signs. When one sign appears more frequently than it should based on chance alone, we reject the hypothesis that the population median $\tilde{\mu}$ is equal to $\tilde{\mu}_0$.

In theory, the sign test is applicable only in situations where $\tilde{\mu}_0$ cannot equal the value of any of the observations. Although there is a zero probability of obtaining a sample observation exactly equal to $\tilde{\mu}_0$ when the population is continuous, nevertheless, in practice a sample value equal to $\tilde{\mu}_0$ will often occur from a lack of precision in recording the data. When sample values equal to $\tilde{\mu}_0$ are observed, they are excluded from the analysis and the sample size is correspondingly reduced.

The appropriate test statistic for the sign test is the binomial random variable X, representing the number of plus signs in our random sample. If the null hypothesis that $\tilde{\mu} = \tilde{\mu}_0$ is true, the probability that a sample value results in either a plus or a minus sign is equal to 1/2. Therefore, to test the null hypothesis that

$\tilde{\mu} = \tilde{\mu}_0$, we actually test the null hypothesis that the number of plus signs is a value of a random variable having the binomial distribution with the parameter $p = 1/2$. P-values for both one-sided and two-sided alternatives can then be calculated using this binomial distribution. For example, in testing

$$H_0\colon\ \tilde{\mu} = \tilde{\mu}_0,$$
$$H_1\colon\ \tilde{\mu} < \tilde{\mu}_0,$$

we shall reject H_0 in favor of H_1 only if the proportion of plus signs is sufficiently less than 1/2, that is, when the value x of our random variable is small. Hence, if the computed P-value

$$P = P(X \le x \text{ when } p = 1/2)$$

is less than or equal to some preselected significance level α, we reject H_0 in favor of H_1. For example, when $n = 15$ and $x = 3$, we find from Table A.1 that

$$P = P(X \le 3 \text{ when } p = 1/2) = \sum_{x=0}^{3} b\left(x; 15, \frac{1}{2}\right) = 0.0176,$$

so the null hypothesis $\tilde{\mu} = \tilde{\mu}_0$ can certainly be rejected at the 0.05 level of significance but not at the 0.01 level.

To test the hypothesis

$$H_0\colon\ \tilde{\mu} = \tilde{\mu}_0,$$
$$H_1\colon\ \tilde{\mu} > \tilde{\mu}_0,$$

we reject H_0 in favor of H_1 only if the proportion of plus signs is sufficiently greater than 1/2, that is, when x is large. Hence, if the computed P-value

$$P = P(X \ge x \text{ when } p = 1/2)$$

is less than α, we reject H_0 in favor of H_1. Finally, to test the hypothesis

$$H_0\colon\ \tilde{\mu} = \tilde{\mu}_0,$$
$$H_1\colon\ \tilde{\mu} \ne \tilde{\mu}_0,$$

we reject H_0 in favor of H_1 when the proportion of plus signs is significantly less than or greater than 1/2. This, of course, is equivalent to x being sufficiently small or sufficiently large. Therefore, if $x < n/2$ and the computed P-value

$$P = 2P(X \le x \text{ when } p = 1/2)$$

is less than or equal to α, or if $x > n/2$ and the computed P-value

$$P = 2P(X \ge x \text{ when } p = 1/2)$$

is less than or equal to α, we reject H_0 in favor of H_1.

Whenever $n > 10$, binomial probabilities with $p = 1/2$ can be approximated from the normal curve, since $np = nq > 5$. Suppose, for example, that we wish to test the hypothesis

$$H_0\colon\ \tilde{\mu} = \tilde{\mu}_0,$$
$$H_1\colon\ \tilde{\mu} < \tilde{\mu}_0,$$

at the $\alpha = 0.05$ level of significance, for a random sample of size $n = 20$ that yields $x = 6$ plus signs. Using the normal curve approximation with

$$\tilde{\mu} = np = (20)(0.5) = 10$$

and

$$\sigma = \sqrt{npq} = \sqrt{(20)(0.5)(0.5)} = 2.236,$$

we find that

$$z = \frac{6.5 - 10}{2.236} = -1.57.$$

Therefore,

$$P = P(X \leq 6) \approx P(Z < -1.57) = 0.0582,$$

which leads to the nonrejection of the null hypothesis.

Example 16.1: The following data represent the number of hours that a rechargeable hedge trimmer operates before a recharge is required:

$$1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2, 1.7.$$

Use the sign test to test the hypothesis, at the 0.05 level of significance, that this particular trimmer operates a median of 1.8 hours before requiring a recharge.

Solution:

1. $H_0\colon\ \tilde{\mu} = 1.8$.
2. $H_1\colon\ \tilde{\mu} \neq 1.8$.
3. $\alpha = 0.05$.
4. Test statistic: Binomial variable X with $p = \frac{1}{2}$.
5. Computations: Replacing each value by the symbol "+" if it exceeds 1.8 and by the symbol "−" if it is less than 1.8 and discarding the one measurement that equals 1.8, we obtain the sequence

$$- \quad + \quad - \quad - \quad + \quad - \quad - \quad + \quad - \quad -$$

for which $n = 10$, $x = 3$, and $n/2 = 5$. Therefore, from Table A.1 the computed P-value is

$$P = 2P\left(X \leq 3 \text{ when } p = \frac{1}{2}\right) = 2\sum_{x=0}^{3} b\left(x; 10, \frac{1}{2}\right) = 0.3438 > 0.05.$$

6. Decision: Do not reject the null hypothesis and conclude that the median operating time is not significantly different from 1.8 hours.

We can also use the sign test to test the null hypothesis $\tilde{\mu}_1 - \tilde{\mu}_2 = d_0$ for paired observations. Here we replace each difference, d_i, with a plus or minus sign depending on whether the adjusted difference, $d_i - d_0$, is positive or negative. Throughout this section, we have assumed that the populations are symmetric. However, even if populations are skewed, we can carry out the same test procedure, but the hypotheses refer to the population medians rather than the means.

Example 16.2: A taxi company is trying to decide whether the use of radial tires instead of regular belted tires improves fuel economy. Sixteen cars are equipped with radial tires and driven over a prescribed test course. Without changing drivers, the same cars are then equipped with the regular belted tires and driven once again over the test course. The gasoline consumption, in kilometers per liter, is given in Table 16.1. Can we conclude at the 0.05 level of significance that cars equipped with radial tires obtain better fuel economy than those equipped with regular belted tires?

Table 16.1: Data for Example 16.2

Car	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Radial Tires	4.2	4.7	6.6	7.0	6.7	4.5	5.7	6.0
Belted Tires	4.1	4.9	6.2	6.9	6.8	4.4	5.7	5.8
Car	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Radial Tires	7.4	4.9	6.1	5.2	5.7	6.9	6.8	4.9
Belted Tires	6.9	4.9	6.0	4.9	5.3	6.5	7.1	4.8

***Solution*:** Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ represent the median kilometers per liter for cars equipped with radial and belted tires, respectively.

1. H_0: $\tilde{\mu}_1 - \tilde{\mu}_2 = 0$.
2. H_1: $\tilde{\mu}_1 - \tilde{\mu}_2 > 0$.
3. $\alpha = 0.05$.
4. Test statistic: Binomial variable X with $p = 1/2$.
5. Computations: After replacing each positive difference by a "+" symbol and each negative difference by a "−" symbol and then discarding the two zero differences, we obtain the sequence

$$+ \; - \; + \; + \; - \; + \; + \; + \; + \; + \; + \; + \; - \; +$$

for which $n = 14$ and $x = 11$. Using the normal curve approximation, we find

$$z = \frac{10.5 - 7}{\sqrt{(14)(0.5)(0.5)}} = 1.87,$$

and then

$$P = P(X \geq 11) \approx P(Z > 1.87) = 0.0307.$$

6. Decision: Reject H_0 and conclude that, on the average, radial tires do improve fuel economy.

Not only is the sign test one of the simplest nonparametric procedures to apply; it has the additional advantage of being applicable to dichotomous data that cannot be recorded on a numerical scale but can be represented by positive and negative responses. For example, the sign test is applicable in experiments where a qualitative response such as "hit" or "miss" is recorded, and in sensory-type experiments where a plus or minus sign is recorded depending on whether the taste tester correctly or incorrectly identifies the desired ingredient.

We shall attempt to make comparisons between many of the nonparametric procedures and the corresponding parametric tests. In the case of the sign test the competition is, of course, the t-test. If we are sampling from a normal distribution, the use of the t-test will result in a larger power for the test. If the distribution is merely symmetric, though not normal, the t-test is preferred in terms of power unless the distribution has extremely "heavy tails" compared to the normal distribution.

16.2 Signed-Rank Test

The reader should note that the sign test utilizes only the plus and minus signs of the differences between the observations and $\tilde{\mu}_0$ in the one-sample case, or the plus and minus signs of the differences between the pairs of observations in the paired-sample case; it does not take into consideration the magnitudes of these differences. A test utilizing both direction and magnitude, proposed in 1945 by Frank Wilcoxon, is now commonly referred to as the **Wilcoxon signed-rank test**.

The analyst can extract more information from the data in a nonparametric fashion if it is reasonable to invoke an additional restriction on the distribution from which the data were taken. The Wilcoxon signed-rank test applies in the case of a **symmetric continuous distribution**. Under this condition, we can test the null hypothesis $\tilde{\mu} = \tilde{\mu}_0$. We first subtract $\tilde{\mu}_0$ from each sample value, discarding all differences equal to zero. The remaining differences are then ranked without regard to sign. A rank of 1 is assigned to the smallest absolute difference (i.e., without sign), a rank of 2 to the next smallest, and so on. When the absolute value of two or more differences is the same, assign to each the average of the ranks that would have been assigned if the differences were distinguishable. For example, if the fifth and sixth smallest differences are equal in absolute value, each is assigned a rank of 5.5. If the hypothesis $\tilde{\mu} = \tilde{\mu}_0$ is true, the total of the ranks corresponding to the positive differences should nearly equal the total of the ranks corresponding to the negative differences. Let us represent these totals by w_+ and w_-, respectively. We designate the smaller of w_+ and w_- by w.

In selecting repeated samples, we would expect w_+ and w_-, and therefore w, to vary. Thus, we may think of w_+, w_-, and w as values of the corresponding random variables W_+, W_-, and W. The null hypothesis $\tilde{\mu} = \tilde{\mu}_0$ can be rejected in favor of the alternative $\tilde{\mu} < \tilde{\mu}_0$ only if w_+ is small and w_- is large. Likewise, the alternative $\tilde{\mu} > \tilde{\mu}_0$ can be accepted only if w_+ is large and w_- is small. For a two-sided alternative, we may reject H_0 in favor of H_1 if either w_+ or w_-, and hence w, is sufficiently small. Therefore, no matter what the alternative hypothesis may be, we reject the null hypothesis when the value of the appropriate statistic W_+, W_-, or W is sufficiently small.

Two Samples with Paired Observations

To test the null hypothesis that we are sampling two continuous symmetric populations with $\tilde{\mu}_1 = \tilde{\mu}_2$ for the paired-sample case, we rank the differences of the paired observations without regard to sign and proceed as in the single-sample case. The various test procedures for both the single- and paired-sample cases are summarized in Table 16.2.

Table 16.2: Signed-Rank Test

H_0	H_1	Compute
$\tilde{\mu} = \tilde{\mu}_0$	$\tilde{\mu} < \tilde{\mu}_0$	w_+
	$\tilde{\mu} > \tilde{\mu}_0$	w_-
	$\tilde{\mu} \neq \tilde{\mu}_0$	w
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 < \tilde{\mu}_2$	w_+
	$\tilde{\mu}_1 > \tilde{\mu}_2$	w_-
	$\tilde{\mu}_1 \neq \tilde{\mu}_2$	w

It is not difficult to show that whenever $n < 5$ and the level of significance does not exceed 0.05 for a one-tailed test or 0.10 for a two-tailed test, all possible values of w_+, w_-, or w will lead to the acceptance of the null hypothesis. However, when $5 \leq n \leq 30$, Table A.16 shows approximate critical values of W_+ and W_- for levels of significance equal to 0.01, 0.025, and 0.05 for a one-tailed test and critical values of W for levels of significance equal to 0.02, 0.05, and 0.10 for a two-tailed test. The null hypothesis is rejected if the computed value w_+, w_-, or w is **less than or equal to** the appropriate tabled value. For example, when $n = 12$, Table A.16 shows that a value of $w_+ \leq 17$ is required for the one-sided alternative $\tilde{\mu} < \tilde{\mu}_0$ to be significant at the 0.05 level.

Example 16.3: Rework Example 16.1 by using the signed-rank test.

Solution:

1. H_0: $\tilde{\mu} = 1.8$.
2. H_1: $\tilde{\mu} \neq 1.8$.
3. $\alpha = 0.05$.
4. Critical region: Since $n = 10$ after discarding the one measurement that equals 1.8, Table A.16 shows the critical region to be $w \leq 8$.
5. Computations: Subtracting 1.8 from each measurement and then ranking the differences without regard to sign, we have

d_i	−0.3	0.4	−0.9	−0.5	0.2	−0.2	−0.3	0.2	−0.6	−0.1
Ranks	5.5	7	10	8	3	3	5.5	3	9	1

Now $w_+ = 13$ and $w_- = 42$, so $w = 13$, the smaller of w_+ and w_-.

6. Decision: As before, do not reject H_0 and conclude that the median operating time is not significantly different from 1.8 hours.

The signed-rank test can also be used to test the null hypothesis that $\tilde{\mu}_1 - \tilde{\mu}_2 = d_0$. In this case, the populations need not be symmetric. As with the sign test, we subtract d_0 from each difference, rank the adjusted differences without regard to sign, and apply the same procedure as above.

Example 16.4: It is claimed that a college senior can increase his or her score in the major field area of the graduate record examination by at least 50 points if he or she is provided with sample problems in advance. To test this claim, 20 college seniors are divided into 10 pairs such that the students in each matched pair have almost the same overall grade-point averages for their first 3 years in college. Sample problems and answers are provided at random to one member of each pair 1 week prior to the examination. The examination scores are given in Table 16.3.

Table 16.3: Data for Example 16.4

	Pair									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
With Sample Problems	531	621	663	579	451	660	591	719	543	575
Without Sample Problems	509	540	688	502	424	683	568	748	530	524

Test the null hypothesis, at the 0.05 level of significance, that sample problems increase scores by 50 points against the alternative hypothesis that the increase is less than 50 points.

***Solution*:** Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ represent the median scores of all students taking the test in question with and without sample problems, respectively.

1. H_0: $\tilde{\mu}_1 - \tilde{\mu}_2 = 50$.
2. H_1: $\tilde{\mu}_1 - \tilde{\mu}_2 < 50$.
3. $\alpha = 0.05$.
4. Critical region: Since $n = 10$, Table A.16 shows the critical region to be $w_+ \leq 11$.
5. Computations:

	Pair									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
d_i	22	81	−25	77	27	−23	23	−29	13	51
$d_i - d_0$	−28	31	−75	27	−23	−73	−27	−79	−37	1
Ranks	5	6	9	3.5	2	8	3.5	10	7	1

Now we find that $w_+ = 6 + 3.5 + 1 = 10.5$.

6. Decision: Reject H_0 and conclude that sample problems do not, on average, increase one's graduate record score by as much as 50 points.

Normal Approximation for Large Samples

When $n \geq 15$, the sampling distribution of W_+ (or W_-) approaches the normal distribution with mean and variance given by

$$\mu_{W_+} = \frac{n(n+1)}{4} \text{ and } \sigma^2_{W_+} = \frac{n(n+1)(2n+1)}{24}.$$

Therefore, when n exceeds the largest value in Table A.16, the statistic

$$Z = \frac{W_+ - \mu_{W_+}}{\sigma_{W_+}}$$

can be used to determine the critical region for the test.

Exercises

16.1 The following data represent the time, in minutes, that a patient has to wait during 12 visits to a doctor's office before being seen by the doctor:

17	15	20	20	32	28
12	26	25	25	35	24

Use the sign test at the 0.05 level of significance to test the doctor's claim that the median waiting time for her patients is not more than 20 minutes.

16.2 The following data represent the number of hours of flight training received by 18 student pilots from a certain instructor prior to their first solo flight:

9	12	18	14	12	14	12	10	16
11	9	11	13	11	13	15	13	14

Using binomial probabilities from Table A.1, perform a sign test at the 0.02 level of significance to test the instructor's claim that the median time required before his students' solo is 12 hours of flight training.

16.3 A food inspector examined 16 jars of a certain brand of jam to determine the percent of foreign impurities. The following data were recorded:

2.4	2.3	3.1	2.2	2.3	1.2	1.0	2.4
1.7	1.1	4.2	1.9	1.7	3.6	1.6	2.3

Using the normal approximation to the binomial distribution, perform a sign test at the 0.05 level of significance to test the null hypothesis that the median percent of impurities in this brand of jam is 2.5% against the alternative that the median percent of impurities is not 2.5%.

16.4 A paint supplier claims that a new additive will reduce the drying time of its acrylic paint. To test this claim, 12 panels of wood were painted, one-half of each panel with paint containing the regular additive and the other half with paint containing the new additive. The drying times, in hours, were recorded as follows:

	Drying Time (hours)	
Panel	**New Additive**	**Regular Additive**
1	6.4	6.6
2	5.8	5.8
3	7.4	7.8
4	5.5	5.7
5	6.3	6.0
6	7.8	8.4
7	8.6	8.8
8	8.2	8.4
9	7.0	7.3
10	4.9	5.8
11	5.9	5.8
12	6.5	6.5

Use the sign test at the 0.05 level to test the null hypothesis that the new additive is no better than the regular additive in reducing the drying time of this kind of paint.

16.5 It is claimed that a new diet will reduce a person's weight by 4.5 kilograms, on average, in a period of 2 weeks. The weights of 10 women were recorded before and after a 2-week period during which they followed this diet, yielding the following data:

Woman	**Weight Before**	**Weight After**
1	58.5	60.0
2	60.3	54.9
3	61.7	58.1
4	69.0	62.1
5	64.0	58.5
6	62.6	59.9
7	56.7	54.4
8	63.6	60.2
9	68.2	62.3
10	59.4	58.7

Use the sign test at the 0.05 level of significance to test the hypothesis that the diet reduces the median weight by 4.5 kilograms against the alternative hypothesis that the median weight loss is less than 4.5 kilograms.

16.6 Two types of instruments for measuring the amount of sulfur monoxide in the atmosphere are being compared in an air-pollution experiment. The following readings were recorded daily for a period of 2 weeks:

	Sulfur Monoxide	
Day	**Instrument *A***	**Instrument *B***
1	0.96	0.87
2	0.82	0.74
3	0.75	0.63
4	0.61	0.55
5	0.89	0.76
6	0.64	0.70
7	0.81	0.69
8	0.68	0.57
9	0.65	0.53
10	0.84	0.88
11	0.59	0.51
12	0.94	0.79
13	0.91	0.84
14	0.77	0.63

Using the normal approximation to the binomial distribution, perform a sign test to determine whether the different instruments lead to different results. Use a 0.05 level of significance.

16.7 The following figures give the systolic blood pressure of 16 joggers before and after an 8-kilometer run:

Jogger	**Before**	**After**
1	158	164
2	149	158
3	160	163
4	155	160
5	164	172
6	138	147
7	163	167
8	159	169
9	165	173
10	145	147
11	150	156
12	161	164
13	132	133
14	155	161
15	146	154
16	159	170

Use the sign test at the 0.05 level of significance to test the null hypothesis that jogging 8 kilometers increases the median systolic blood pressure by 8 points against the alternative that the increase in the median is less than 8 points.

16.8 Analyze the data of Exercise 16.1 by using the signed-rank test.

16.9 Analyze the data of Exercise 16.2 by using the signed-rank test.

16.10 The weights of 5 people before they stopped smoking and 5 weeks after they stopped smoking, in kilograms, are as follows:

	Individual				
	1	**2**	**3**	**4**	**5**
Before	66	80	69	52	75
After	71	82	68	56	73

Use the signed-rank test for paired observations to test the hypothesis, at the 0.05 level of significance, that giving up smoking has no effect on a person's weight against the alternative that one's weight increases if he or she quits smoking.

16.11 Rework Exercise 16.5 by using the signed-rank test.

16.12 The following are the numbers of prescriptions filled by two pharmacies over a 20-day period:

Day	**Pharmacy *A***	**Pharmacy *B***
1	19	17
2	21	15
3	15	12
4	17	12
5	24	16
6	12	15
7	19	11
8	14	13
9	20	14
10	18	21
11	23	19
12	21	15
13	17	11
14	12	10
15	16	20
16	15	12
17	20	13
18	18	17
19	14	16
20	22	18

Use the signed-rank test at the 0.01 level of significance to determine whether the two pharmacies, on average, fill the same number of prescriptions against the alternative that pharmacy *A* fills more prescriptions than pharmacy *B*.

16.13 Rework Exercise 16.7 by using the signed-rank test.

16.14 Rework Exercise 16.6 by using the signed-rank test.

16.3 Wilcoxon Rank-Sum Test

As we indicated earlier, the nonparametric procedure is generally an appropriate alternative to the normal theory test when the normality assumption does not hold. When we are interested in testing equality of means of two continuous distributions that are obviously nonnormal, and samples are independent (i.e., there is no pairing of observations), the **Wilcoxon rank-sum test** or **Wilcoxon two-sample test** is an appropriate alternative to the two-sample t-test described in Chapter 10.

We shall test the null hypothesis H_0 that $\tilde{\mu}_1 = \tilde{\mu}_2$ against some suitable alternative. First we select a random sample from each of the populations. Let n_1 be the number of observations in the smaller sample, and n_2 the number of observations in the larger sample. When the samples are of equal size, n_1 and n_2 may be randomly assigned. Arrange the $n_1 + n_2$ observations of the combined samples in ascending order and substitute a rank of $1, 2, \ldots, n_1 + n_2$ for each observation. In the case of ties (identical observations), we replace the observations by the mean of the ranks that the observations would have if they were distinguishable. For example, if the seventh and eighth observations were identical, we would assign a rank of 7.5 to each of the two observations.

The sum of the ranks corresponding to the n_1 observations in the smaller sample is denoted by w_1. Similarly, the value w_2 represents the sum of the n_2 ranks corresponding to the larger sample. The total $w_1 + w_2$ depends only on the number of observations in the two samples and is in no way affected by the results of the experiment. Hence, if $n_1 = 3$ and $n_2 = 4$, then $w_1 + w_2 = 1 + 2 + \cdots + 7 = 28$, regardless of the numerical values of the observations. In general,

$$w_1 + w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2},$$

the arithmetic sum of the integers $1, 2, \ldots, n_1 + n_2$. Once we have determined w_1, it may be easier to find w_2 by the formula

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1.$$

In choosing repeated samples of sizes n_1 and n_2, we would expect w_1, and therefore w_2, to vary. Thus, we may think of w_1 and w_2 as values of the random variables W_1 and W_2, respectively. The null hypothesis $\tilde{\mu}_1 = \tilde{\mu}_2$ will be rejected in favor of the alternative $\tilde{\mu}_1 < \tilde{\mu}_2$ only if w_1 is small and w_2 is large. Likewise, the alternative $\tilde{\mu}_1 > \tilde{\mu}_2$ can be accepted only if w_1 is large and w_2 is small. For a two-tailed test, we may reject H_0 in favor of H_1 if w_1 is small and w_2 is large or if w_1 is large and w_2 is small. In other words, the alternative $\tilde{\mu}_1 < \tilde{\mu}_2$ is accepted if w_1 is sufficiently small; the alternative $\tilde{\mu}_1 > \tilde{\mu}_2$ is accepted if w_2 is sufficiently small; and the alternative $\tilde{\mu}_1 \neq \tilde{\mu}_2$ is accepted if the minimum of w_1 and w_2 is sufficiently small. In actual practice, we usually base our decision on the value

$$u_1 = w_1 - \frac{n_1(n_1 + 1)}{2} \qquad \text{or} \qquad u_2 = w_2 - \frac{n_2(n_2 + 1)}{2}$$

of the related statistic U_1 or U_2 or on the value u of the statistic U, the minimum of U_1 and U_2. These statistics simplify the construction of tables of critical values, since both U_1 and U_2 have symmetric sampling distributions and assume values in the interval from 0 to n_1n_2 such that $u_1 + u_2 = n_1n_2$.

From the formulas for u_1 and u_2 we see that u_1 will be small when w_1 is small and u_2 will be small when w_2 is small. Consequently, the null hypothesis will be rejected whenever the appropriate statistic U_1, U_2, or U assumes a value less than or equal to the desired critical value given in Table A.17. The various test procedures are summarized in Table 16.4.

Table 16.4: Rank-Sum Test

H_0	H_1	Compute
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 < \tilde{\mu}_2$	u_1
	$\tilde{\mu}_1 > \tilde{\mu}_2$	u_2
	$\tilde{\mu}_1 \neq \tilde{\mu}_2$	u

Table A.17 gives critical values of U_1 and U_2 for levels of significance equal to 0.001, 0.01, 0.025, and 0.05 for a one-tailed test, and critical values of U for levels of significance equal to 0.002, 0.02, 0.05, and 0.10 for a two-tailed test. If the observed value of u_1, u_2, or u is **less than or equal** to the tabled critical value, the null hypothesis is rejected at the level of significance indicated by the table. Suppose, for example, that we wish to test the null hypothesis that $\tilde{\mu}_1 = \tilde{\mu}_2$ against the one-sided alternative that $\tilde{\mu}_1 < \tilde{\mu}_2$ at the 0.05 level of significance for random samples of sizes $n_1 = 3$ and $n_2 = 5$ that yield the value $w_1 = 8$. It follows that

$$u_1 = 8 - \frac{(3)(4)}{2} = 2.$$

Our one-tailed test is based on the statistic U_1. Using Table A.17, we reject the null hypothesis of equal means when $u_1 \leq 1$. Since $u_1 = 2$ does not fall in the rejection region, the null hypothesis cannot be rejected.

Example 16.5: The nicotine content of two brands of cigarettes, measured in milligrams, was found to be as follows:

Brand *A*	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3		
Brand *B*	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	5.4

Test the hypothesis, at the 0.05 level of significance, that the median nicotine contents of the two brands are equal against the alternative that they are unequal.

Solution:

1. H_0: $\tilde{\mu}_1 = \tilde{\mu}_2$.
2. H_1: $\tilde{\mu}_1 \neq \tilde{\mu}_2$.
3. $\alpha = 0.05$.
4. Critical region: $u \leq 17$ (from Table A.17).
5. Computations: The observations are arranged in ascending order and ranks from 1 to 18 assigned.

Original Data	**Ranks**	**Original Data**	**Ranks**
0.6	1	4.0	10.5*
1.6	2	4.0	10.5
1.9	3	4.1	12
2.1	4*	4.8	13*
2.2	5	5.4	14.5*
2.5	6	5.4	14.5
3.1	7	6.1	16*
3.3	8*	6.2	17
3.7	9*	6.3	18*

*The ranks marked with an asterisk belong to sample A.

Now

$$w_1 = 4 + 8 + 9 + 10.5 + 13 + 14.5 + 16 + 18 = 93$$

and

$$w_2 = \frac{(18)(19)}{2} - 93 = 78.$$

Therefore,

$$u_1 = 93 - \frac{(8)(9)}{2} = 57, \qquad u_2 = 78 - \frac{(10)(11)}{2} = 23.$$

6. Decision: Do not reject the null hypothesis H_0 and conclude that there is no significant difference in the median nicotine contents of the two brands of cigarettes.

Normal Theory Approximation for Two Samples

When both n_1 and n_2 exceed 8, the sampling distribution of U_1 (or U_2) approaches the normal distribution with mean and variance given by

$$\mu_{U_1} = \frac{n_1 n_2}{2} \text{ and } \sigma^2_{U_1} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

Consequently, when n_2 is greater than 20, the maximum value in Table A.17, and n_1 is at least 9, we can use the statistic

$$Z = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}}$$

for our test, with the critical region falling in either or both tails of the standard normal distribution, depending on the form of H_1.

The use of the Wilcoxon rank-sum test is not restricted to nonnormal populations. It can be used in place of the two-sample t-test when the populations are normal, although the power will be smaller. The Wilcoxon rank-sum test is always superior to the t-test for decidedly nonnormal populations.

16.4 Kruskal-Wallis Test

In Chapters 13, 14, and 15, the technique of analysis of variance was prominent as an analytical technique for testing equality of $k \geq 2$ population means. Again, however, the reader should recall that normality must be assumed in order for the F-test to be theoretically correct. In this section, we investigate a nonparametric alternative to analysis of variance.

The **Kruskal-Wallis test**, also called the **Kruskal-Wallis H test**, is a generalization of the rank-sum test to the case of $k > 2$ samples. It is used to test the null hypothesis H_0 that k independent samples are from identical populations. Introduced in 1952 by W. H. Kruskal and W. A. Wallis, the test is a nonparametric procedure for testing the equality of means in the one-factor analysis of variance when the experimenter wishes to avoid the assumption that the samples were selected from normal populations.

Let n_i $(i = 1, 2, \ldots, k)$ be the number of observations in the ith sample. First, we combine all k samples and arrange the $n = n_1 + n_2 + \cdots + n_k$ observations in ascending order, substituting the appropriate rank from $1, 2, \ldots, n$ for each observation. In the case of ties (identical observations), we follow the usual procedure of replacing the observations by the mean of the ranks that the observations would have if they were distinguishable. The sum of the ranks corresponding to the n_i observations in the ith sample is denoted by the random variable R_i. Now let us consider the statistic

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1),$$

which is approximated very well by a chi-squared distribution with $k-1$ degrees of freedom when H_0 is true, provided each sample consists of at least 5 observations. The fact that h, the assumed value of H, is large when the independent samples come from populations that are not identical allows us to establish the following decision criterion for testing H_0:

Kruskal-Wallis Test

To test the null hypothesis H_0 that k independent samples are from identical populations, compute

$$h = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(n+1),$$

where r_i is the assumed value of R_i, for $i = 1, 2, \ldots, k$. If h falls in the critical region $H > \chi^2_\alpha$ with $v = k - 1$ degrees of freedom, reject H_0 at the α-level of significance; otherwise, fail to reject H_0.

Example 16.6: In an experiment to determine which of three different missile systems is preferable, the propellant burning rate is measured. The data, after coding, are given in Table 16.5. Use the Kruskal-Wallis test and a significance level of $\alpha = 0.05$ to test the hypothesis that the propellant burning rates are the same for the three missile systems.

Table 16.5: Propellant Burning Rates

Missile System								
1			**2**			**3**		
24.0	16.7	22.8	23.2	19.8	18.1	18.4	19.1	17.3
19.8	18.9		17.6	20.2	17.8	17.3	19.7	18.9
						18.8	19.3	

Solution:

1. H_0: $\mu_1 = \mu_2 = \mu_3$.
2. H_1: The three means are not all equal.
3. $\alpha = 0.05$.
4. Critical region: $h > \chi^2_{0.05} = 5.991$, for $v = 2$ degrees of freedom.
5. Computations: In Table 16.6, we convert the 19 observations to ranks and sum the ranks for each missile system.

Table 16.6: Ranks for Propellant Burning Rates

Missile System		
1	**2**	**3**
19	18	7
1	14.5	11
17	6	2.5
14.5	4	2.5
9.5	16	13
$r_1 = 61.0$	5	9.5
	$r_2 = 63.5$	8
		12
		$r_3 = 65.5$

Now, substituting $n_1 = 5$, $n_2 = 6$, $n_3 = 8$ and $r_1 = 61.0$, $r_2 = 63.5$, $r_3 = 65.5$, our test statistic H assumes the value

$$h = \frac{12}{(19)(20)}\left(\frac{61.0^2}{5} + \frac{63.5^2}{6} + \frac{65.5^2}{8}\right) - (3)(20) = 1.66.$$

6. Decision: Since $h = 1.66$ does not fall in the critical region $h > 5.991$, we have insufficient evidence to reject the hypothesis that the propellant burning rates are the same for the three missile systems.

Exercises

16.15 A cigarette manufacturer claims that the tar content of brand B cigarettes is lower than that of brand A cigarettes. To test this claim, the following determinations of tar content, in milligrams, were recorded:

Brand A	1	12	9	13	11	14
Brand B	8	10	7			

Use the rank-sum test with $\alpha = 0.05$ to test whether the claim is valid.

16.16 To find out whether a new serum will arrest leukemia, nine patients, who have all reached an advanced stage of the disease, are selected. Five patients receive the treatment and four do not. The survival times, in years, from the time the experiment commenced are

Treatment	2.1	5.3	1.4	4.6	0.9
No treatment	1.9	0.5	2.8	3.1	

Use the rank-sum test, at the 0.05 level of significance, to determine if the serum is effective.

16.17 The following data represent the number of hours that two different types of scientific pocket calculators operate before a recharge is required.

Calculator A	5.5	5.6	6.3	4.6	5.3	5.0	6.2	5.8	5.1
Calculator B	3.8	4.8	4.3	4.2	4.0	4.9	4.5	5.2	4.5

Use the rank-sum test with $\alpha = 0.01$ to determine if calculator A operates longer than calculator B on a full battery charge.

16.18 A fishing line is being manufactured by two processes. To determine if there is a difference in the mean breaking strength of the lines, 10 pieces manufactured by each process are selected and then tested for breaking strength. The results are as follows:

Process 1	10.4	9.8	11.5	10.0	9.9
	9.6	10.9	11.8	9.3	10.7
Process 2	8.7	11.2	9.8	10.1	10.8
	9.5	11.0	9.8	10.5	9.9

Use the rank-sum test with $\alpha = 0.1$ to determine if there is a difference between the mean breaking strengths of the lines manufactured by the two processes.

16.19 From a mathematics class of 12 equally capable students using programmed materials, 5 students are selected at random and given additional instruction by the teacher. The results on the final examination are as follows:

	Grade						
Additional Instruction	87	69	78	91	80		
No Additional Instruction	75	88	64	82	93	79	67

Use the rank-sum test with $\alpha = 0.05$ to determine if the additional instruction affects the average grade.

16.20 The following data represent the weights, in kilograms, of personal luggage carried on various flights by a member of a baseball team and a member of a basketball team.

Luggage Weight (kilograms)

Baseball Player			**Basketball Player**	
16.3	20.0	18.6	15.4	16.3
18.1	15.0	15.4	17.7	18.1
15.9	18.6	15.6	18.6	16.8
14.1	14.5	18.3	12.7	14.1
17.7	19.1	17.4	15.0	13.6
16.3	13.6	14.8	15.9	16.3
13.2	17.2	16.5		

Use the rank-sum test with $\alpha = 0.05$ to test the null hypothesis that the two athletes carry the same amount of luggage on the average against the alternative hypothesis that the average weights of luggage for the two athletes are different.

16.21 The following data represent the operating times in hours for three types of scientific pocket calculators before a recharge is required:

Calculator

A			***B***			***C***		
4.9	6.1	4.3	5.5	5.4	6.2	6.4	6.8	5.6
4.6	5.2		5.8	5.5	5.2	6.5	6.3	6.6
				4.8				

Use the Kruskal-Wallis test, at the 0.01 level of significance, to test the hypothesis that the operating times for all three calculators are equal.

16.22 In Exercise 13.6 on page 519, use the Kruskal-Wallis test at the 0.05 level of significance to determine if the organic chemical solvents differ significantly in sorption rate.

16.5 Runs Test

In applying the many statistical concepts discussed throughout this book, it was always assumed that the sample data had been collected by some randomization procedure. The **runs test**, based on the order in which the sample observations are obtained, is a useful technique for testing the null hypothesis H_0 that the observations have indeed been drawn at random.

To illustrate the runs test, let us suppose that 12 people are polled to find out if they use a certain product. We would seriously question the assumed randomness of the sample if all 12 people were of the same sex. We shall designate a male and a female by the symbols M and F, respectively, and record the outcomes according to their sex in the order in which they occur. A typical sequence for the experiment might be

$$\underbrace{M\ M}\ \underbrace{F\ F\ F}\ \underbrace{M}\ \underbrace{F\ F}\ \underbrace{M\ M\ M\ M},$$

where we have grouped subsequences of identical symbols. Such groupings are called **runs**.

Definition 16.1: A **run** is a subsequence of one or more identical symbols representing a common property of the data.

Regardless of whether the sample measurements represent qualitative or quantitative data, the runs test divides the data into two mutually exclusive categories: male or female; defective or nondefective; heads or tails; above or below the median; and so forth. Consequently, a sequence will always be limited to two distinct symbols. Let n_1 be the number of symbols associated with the category that occurs the least and n_2 be the number of symbols that belong to the other category. Then the sample size $n = n_1 + n_2$.

For the $n = 12$ symbols in our poll, we have five runs, with the first containing two M's, the second containing three F's, and so on. If the number of runs is larger or smaller than what we would expect by chance, the hypothesis that the sample was drawn at random should be rejected. Certainly, a sample resulting in only two runs,

$$M\ M\ M\ M\ M\ M\ M\ F\ F\ F\ F\ F$$

or the reverse, is most unlikely to occur from a random selection process. Such a result indicates that the first 7 people interviewed were all males, followed by 5 females. Likewise, if the sample resulted in the maximum number of 12 runs, as in the alternating sequence

$$M\ F\ M\ F\ M\ F\ M\ F\ M\ F\ M\ F,$$

we would again be suspicious of the order in which the individuals were selected for the poll.

The runs test for randomness is based on the random variable V, the total number of runs that occur in the complete sequence of the experiment. In Table A.18, values of $P(V \leq v^*$ when H_0 is true) are given for $v^* = 2, 3, \ldots, 20$ runs and values of n_1 and n_2 less than or equal to 10. The P-values for both one-tailed and two-tailed tests can be obtained using these tabled values.

For the poll taken previously, we exhibit a total of 5 F's and 7 M's. Hence, with $n_1 = 5$, $n_2 = 7$, and $v = 5$, we note from Table A.18 that the P-value for a two-tailed test is

$$P = 2P(V \leq 5 \text{ when } H_0 \text{ is true}) = 0.394 > 0.05.$$

That is, the value $v = 5$ is reasonable at the 0.05 level of significance when H_0 is true, and therefore we have insufficient evidence to reject the hypothesis of randomness in our sample.

When the number of runs is large (for example, if $v = 11$ while $n_1 = 5$ and $n_2 = 7$), the P-value for a two-tailed test is

$$\begin{aligned} P &= 2P(V \geq 11 \text{ when } H_0 \text{ is true}) = 2[1 - P(V \leq 10 \text{ when } H_0 \text{ is true})] \\ &= 2(1 - 0.992) = 0.016 < 0.05, \end{aligned}$$

which leads us to reject the hypothesis that the sample values occurred at random.

The runs test can also be used to detect departures from randomness of a sequence of quantitative measurements over time, caused by trends or periodicities. Replacing each measurement, in the order in which it was collected, by a *plus* symbol if it falls above the median or by a *minus* symbol if it falls below the median and omitting all measurements that are exactly equal to the median, we generate a sequence of plus and minus symbols that is tested for randomness as illustrated in the following example.

Example 16.7: A machine dispenses acrylic paint thinner into containers. Would you say that the amount of paint thinner being dispensed by this machine varies randomly if the contents of the next 15 containers are measured and found to be 3.6, 3.9, 4.1, 3.6, 3.8, 3.7, 3.4, 4.0, 3.8, 4.1, 3.9, 4.0, 3.8, 4.2, and 4.1 liters? Use a 0.1 level of significance.

Solution:

1. H_0: Sequence is random.
2. H_1: Sequence is not random.
3. $\alpha = 0.1$.
4. Test statistic: V, the total number of runs.
5. Computations: For the given sample, we find $\tilde{x} = 3.9$. Replacing each measurement by the symbol "+" if it falls above 3.9 or by the symbol "−" if it falls below 3.9 and omitting the two measurements that equal 3.9, we obtain the sequence

$$- \; + \; - \; - \; - \; - \; + \; - \; + \; + \; - \; + \; +$$

for which $n_1 = 6$, $n_2 = 7$, and $v = 8$. Therefore, from Table A.18, the computed P-value is

$$\begin{aligned} P &= 2P(V \geq 8 \text{ when } H_0 \text{ is true}) \\ &= 2[1 - P(V \leq 8 \text{ when } H_0 \text{ is true})] = 2(0.5) = 1. \end{aligned}$$

6. Decision: Do not reject the hypothesis that the sequence of measurements varies randomly.

The runs test, although less powerful, can also be used as an alternative to the Wilcoxon two-sample test to test the claim that two random samples come from populations having the same distributions and therefore equal means. If the populations are symmetric, rejection of the claim of equal distributions is equivalent to accepting the alternative hypothesis that the means are not equal. In performing the test, we first combine the observations from both samples and arrange them in ascending order. Now assign the letter A to each observation taken from one of the populations and the letter B to each observation from the other population, thereby generating a sequence consisting of the symbols A and B. If observations from one population are tied with observations from the other population, the sequence of A and B symbols generated will not be unique and consequently the number of runs is unlikely to be unique. Procedures for breaking ties usually result in additional tedious computations, and for this reason we might prefer to apply the Wilcoxon rank-sum test whenever these situations occur.

To illustrate the use of runs in testing for equal means, consider the survival times of the leukemia patients of Exercise 16.16 on page 670, for which we have

$$\begin{array}{ccccccccc} 0.5 & 0.9 & 1.4 & 1.9 & 2.1 & 2.8 & 3.1 & 4.6 & 5.3 \\ B & A & A & B & A & B & B & A & A \end{array}$$

resulting in $v = 6$ runs. If the two symmetric populations have equal means, the observations from the two samples will be intermingled, resulting in many runs. However, if the population means are significantly different, we would expect most of the observations for one of the two samples to be smaller than those for the other sample. In the extreme case where the populations do not overlap, we would obtain a sequence of the form

$$A\ A\ A\ A\ A\ B\ B\ B\ B \quad \text{or} \quad B\ B\ B\ B\ A\ A\ A\ A\ A$$

and in either case there would be only two runs. Consequently, the hypothesis of equal population means will be rejected at the α-level of significance only when v is small enough so that

$$P = P(V \leq v \text{ when } H_0 \text{ is true}) \leq \alpha,$$

implying a one-tailed test.

Returning to the data of Exercise 16.16 on page 670, for which $n_1 = 4$, $n_2 = 5$, and $v = 6$, we find from Table A.18 that

$$P = P(V \leq 6 \text{ when } H_0 \text{ is true}) = 0.786 > 0.05$$

and therefore fail to reject the null hypothesis of equal means. Hence, we conclude that the new serum does not prolong life by arresting leukemia.

When n_1 and n_2 increase in size, the sampling distribution of V approaches the normal distribution with mean and variance given by

$$\mu_V = \frac{2n_1n_2}{n_1 + n_2} + 1 \text{ and } \sigma_V^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

Consequently, when n_1 and n_2 are both greater than 10, we can use the statistic

$$Z = \frac{V - \mu_V}{\sigma_V}$$

to establish the critical region for the runs test.

16.6 Tolerance Limits

Tolerance limits for a normal distribution of measurements were discussed in Chapter 9. In this section, we consider a method for constructing tolerance intervals that is independent of the shape of the underlying distribution. As we might suspect, for a reasonable degree of confidence they will be substantially longer than those constructed assuming normality, and the sample size required is generally very large. Nonparametric tolerance limits are stated in terms of the smallest and largest observations in our sample.

Two-Sided Tolerance Limits

For any distribution of measurements, two-sided tolerance limits are indicated by the smallest and largest observations in a sample of size n, where n is determined so that one can assert with $100(1-\gamma)\%$ confidence that **at least** the proportion $1-\alpha$ of the distribution is included between the sample extremes.

Table A.19 gives required sample sizes for selected values of γ and $1-\alpha$. For example, when $\gamma = 0.01$ and $1-\alpha = 0.95$, we must choose a random sample of size $n = 130$ in order to be 99% confident that at least 95% of the distribution of measurements is included between the sample extremes.

Instead of determining the sample size n such that a specified proportion of measurements is contained between the sample extremes, it is desirable in many industrial processes to determine the sample size such that a fixed proportion of the population falls below the largest (or above the smallest) observation in the sample. Such limits are called one-sided tolerance limits.

One-Sided Tolerance Limits

For any distribution of measurements, a one-sided tolerance limit is determined by the smallest (largest) observation in a sample of size n, where n is determined so that one can assert with $100(1-\gamma)\%$ confidence that **at least** the proportion $1-\alpha$ of the distribution will exceed the smallest (be less than the largest) observation in the sample.

Table A.20 shows required sample sizes corresponding to selected values of γ and $1-\alpha$. Hence, when $\gamma = 0.05$ and $1-\alpha = 0.70$, we must choose a sample of size $n = 9$ in order to be 95% confident that 70% of our distribution of measurements will exceed the smallest observation in the sample.

16.7 Rank Correlation Coefficient

In Chapter 11, we used the sample correlation coefficient r to measure the population correlation coefficient ρ, the linear relationship between two continuous variables X and Y. If ranks $1, 2, \ldots, n$ are assigned to the x observations in order of magnitude and similarly to the y observations, and if these ranks are then substituted for the actual numerical values in the formula for the correlation coefficient in Chapter 11, we obtain the nonparametric counterpart of the conventional correlation coefficient. A correlation coefficient calculated in this manner is known as the **Spearman rank correlation coefficient** and is denoted by r_s. When there are no ties among either set of measurements, the formula for r_s reduces to a much simpler expression involving the differences d_i between the ranks assigned to the n pairs of x's and y's, which we now state.

Rank Correlation Coefficient

A nonparametric measure of association between two variables X and Y is given by the **rank correlation coefficient**

$$r_s = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2,$$

where d_i is the difference between the ranks assigned to x_i and y_i and n is the number of pairs of data.

In practice, the preceding formula is also used when there are ties among either the x or y observations. The ranks for tied observations are assigned as in the signed-rank test by averaging the ranks that would have been assigned if the observations were distinguishable.

The value of r_s will usually be close to the value obtained by finding r based on numerical measurements and is interpreted in much the same way. As before, the value of r_s will range from -1 to $+1$. A value of $+1$ or -1 indicates perfect association between X and Y, the plus sign occurring for identical rankings and the minus sign occurring for reverse rankings. When r_s is close to zero, we conclude that the variables are uncorrelated.

Example 16.8: The figures listed in Table 16.7, released by the Federal Trade Commission, show the milligrams of tar and nicotine found in 10 brands of cigarettes. Calculate the rank correlation coefficient to measure the degree of relationship between tar and nicotine content in cigarettes.

Table 16.7: Tar and Nicotine Contents

Cigarette Brand	Tar Content	Nicotine Content
Viceroy	14	0.9
Marlboro	17	1.1
Chesterfield	28	1.6
Kool	17	1.3
Kent	16	1.0
Raleigh	13	0.8
Old Gold	24	1.5
Philip Morris	25	1.4
Oasis	18	1.2
Players	31	2.0

***Solution*:** Let X and Y represent the tar and nicotine contents, respectively. First we assign ranks to each set of measurements, with the rank of 1 assigned to the lowest number in each set, the rank of 2 to the second lowest number in each set, and so forth, until the rank of 10 is assigned to the largest number. Table 16.8 shows the individual rankings of the measurements and the differences in ranks for the 10 pairs of observations.

Table 16.8: Rankings for Tar and Nicotine Content

Cigarette Brand	x_i	y_i	d_i
Viceroy	2.0	2.0	0.0
Marlboro	4.5	4.0	0.5
Chesterfield	9.0	9.0	0.0
Kool	4.5	6.0	−1.5
Kent	3.0	3.0	0.0
Raleigh	1.0	1.0	0.0
Old Gold	7.0	8.0	−1.0
Philip Morris	8.0	7.0	1.0
Oasis	6.0	5.0	1.0
Players	10.0	10.0	0.0

Substituting into the formula for r_s, we find that

$$r_s = 1 - \frac{(6)(5.50)}{(10)(100-1)} = 0.967,$$

indicating a high positive correlation between the amounts of tar and nicotine found in cigarettes. ▟

Some advantages to using r_s rather than r do exist. For instance, we no longer assume the underlying relationship between X and Y to be linear and therefore, when the data possess a distinct curvilinear relationship, the rank correlation coefficient will likely be more reliable than the conventional measure. A second advantage to using the rank correlation coefficient is the fact that no assumptions of normality are made concerning the distributions of X and Y. Perhaps the greatest advantage occurs when we are unable to make meaningful numerical measurements but nevertheless can establish rankings. Such is the case, for example, when different judges rank a group of individuals according to some attribute. The rank correlation coefficient can be used in this situation as a measure of the consistency of the two judges.

To test the hypothesis that $\rho = 0$ by using a rank correlation coefficient, one needs to consider the sampling distribution of the r_s-values under the assumption of no correlation. Critical values for $\alpha = 0.05, 0.025, 0.01$, and 0.005 have been calculated and appear in Table A.21. The setup of this table is similar to that of the table of critical values for the t-distribution except for the left column, which now gives the number of pairs of observations rather than the degrees of freedom. Since the distribution of the r_s-values is symmetric about zero when $\rho = 0$, the r_s-value that leaves an area of α to the left is equal to the negative of the r_s-value that leaves an area of α to the right. For a two-sided alternative hypothesis, the critical region of size α falls equally in the two tails of the distribution. For a test in which the alternative hypothesis is negative, the critical region is entirely in the left tail of the distribution, and when the alternative is positive, the critical region is placed entirely in the right tail.

Example 16.9: Refer to Example 16.8 and test the hypothesis that the correlation between the amounts of tar and nicotine found in cigarettes is zero against the alternative that it is greater than zero. Use a 0.01 level of significance.

Solution:

1. H_0: $\rho = 0$.
2. H_1: $\rho > 0$.
3. $\alpha = 0.01$.
4. Critical region: $r_s > 0.745$ from Table A.21.
5. Computations: From Example 16.8, $r_s = 0.967$.
6. Decision: Reject H_0 and conclude that there is a significant correlation between the amounts of tar and nicotine found in cigarettes.

Under the assumption of no correlation, it can be shown that the distribution of the r_s-values approaches a normal distribution with a mean of 0 and a standard deviation of $1/\sqrt{n-1}$ as n increases. Consequently, when n exceeds the values given in Table A.21, one can test for a significant correlation by computing

$$z = \frac{r_s - 0}{1/\sqrt{n-1}} = r_s\sqrt{n-1}$$

and comparing with critical values of the standard normal distribution shown in Table A.3.

Exercises

16.23 A random sample of 15 adults living in a small town were selected to estimate the proportion of voters favoring a certain candidate for mayor. Each individual was also asked if he or she was a college graduate. By letting Y and N designate the responses of "yes" and "no" to the education question, the following sequence was obtained:

$$N\ N\ N\ N\ Y\ Y\ N\ Y\ Y\ N\ Y\ N\ N\ N\ N$$

Use the runs test at the 0.1 level of significance to determine if the sequence supports the contention that the sample was selected at random.

16.24 A silver-plating process is used to coat a certain type of serving tray. When the process is in control, the thickness of the silver on the trays will vary randomly following a normal distribution with a mean of 0.02 millimeter and a standard deviation of 0.005 millimeter. Suppose that the next 12 trays examined show the following thicknesses of silver: 0.019, 0.021, 0.020, 0.019, 0.020, 0.018, 0.023, 0.021, 0.024, 0.022, 0.023, 0.022. Use the runs test to determine if the fluctuations in thickness from one tray to another are random. Let $\alpha = 0.05$.

16.25 Use the runs test to test, at level 0.01, whether there is a difference in the average operating time for the two calculators of Exercise 16.17 on page 670.

16.26 In an industrial production line, items are inspected periodically for defectives. The following is a sequence of defective items, D, and nondefective items, N, produced by this production line:

$$D\ D\ N\ N\ N\ D\ N\ N\ D\ D\ N\ N\ N\ N$$
$$N\ D\ D\ D\ N\ N\ D\ N\ N\ N\ N\ D\ N\ D$$

Use the large-sample theory for the runs test, with a significance level of 0.05, to determine whether the defectives are occurring at random.

16.27 Assuming that the measurements of Exercise 1.14 on page 30 were recorded successively from left to right as they were collected, use the runs test, with $\alpha = 0.05$, to test the hypothesis that the data represent a random sequence.

16.28 How large a sample is required to be 95% confident that at least 85% of the distribution of measurements is included between the sample extremes?

16.29 What is the probability that the range of a random sample of size 24 includes at least 90% of the population?

16.30 How large a sample is required to be 99% confident that at least 80% of the population will be less than the largest observation in the sample?

16.31 What is the probability that at least 95% of a population will exceed the smallest value in a random sample of size $n = 135$?

16.32 The following table gives the recorded grades for 10 students on a midterm test and the final examination in a calculus course:

Student	Midterm Test	Final Examination
L.S.A.	84	73
W.P.B.	98	63
R.W.K.	91	87
J.R.L.	72	66
J.K.L.	86	78
D.L.P.	93	78
B.L.P.	80	91
D.W.M.	0	0
M.N.M.	92	88
R.H.S.	87	77

(a) Calculate the rank correlation coefficient.

(b) Test the null hypothesis that $\rho = 0$ against the alternative that $\rho > 0$. Use $\alpha = 0.025$.

16.33 With reference to the data of Exercise 11.1 on page 398,

(a) calculate the rank correlation coefficient;

(b) test the null hypothesis, at the 0.05 level of significance, that $\rho = 0$ against the alternative that $\rho \neq 0$. Compare your results with those obtained in Exercise 11.44 on page 435.

16.34 Calculate the rank correlation coefficient for the daily rainfall and amount of particulate removed in Exercise 11.13 on page 400.

16.35 With reference to the weights and chest sizes of infants in Exercise 11.47 on page 436,

(a) calculate the rank correlation coefficient;

(b) test the hypothesis, at the 0.025 level of significance, that $\rho = 0$ against the alternative that $\rho > 0$.

16.36 A consumer panel tests nine brands of microwave ovens for overall quality. The ranks assigned by the panel and the suggested retail prices are as follows:

Manufacturer	Panel Rating	Suggested Price
A	6	\$480
B	9	395
C	2	575
D	8	550
E	5	510
F	1	545
G	7	400
H	4	465
I	3	420

Is there a significant relationship between the quality and the price of a microwave oven? Use a 0.05 level of significance.

16.37 Two judges at a college homecoming parade rank eight floats in the following order:

	Float							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Judge *A*	5	8	4	3	6	2	7	1
Judge *B*	7	5	4	2	8	1	6	3

(a) Calculate the rank correlation coefficient.

(b) Test the null hypothesis that $\rho = 0$ against the alternative that $\rho > 0$. Use $\alpha = 0.05$.

16.38 In the article called "Risky Assumptions" by Paul Slovic, Baruch Fischoff, and Sarah Lichtenstein, published in *Psychology Today* (June 1980), the risk of dying in the United States from 30 activities and technologies is ranked by members of the League of Women Voters and also by experts who are professionally involved in assessing risks. The rankings are as shown in Table 16.9.

(a) Calculate the rank correlation coefficient.

(b) Test the null hypothesis of zero correlation between the rankings of the League of Women Voters and the experts against the alternative that the correlation is not zero. Use a 0.05 level of significance.

Table 16.9: The Ranking Data for Exercise 16.38

Activity or Technology Risk	**Voters**	**Experts**	**Activity or Technology Risk**	**Voters**	**Experts**
Nuclear power	1	20	Motor vehicles	2	1
Handguns	3	4	Smoking	4	2
Motorcycles	5	6	Alcoholic beverages	6	3
Private aviation	7	12	Police work	8	17
Pesticides	9	8	Surgery	10	5
Fire fighting	11	18	Large construction	12	13
Hunting	13	23	Spray cans	14	26
Mountain climing	15	29	Bicycles	16	15
Commercial aviation	17	16	Electric power	18	9
Swimming	19	10	Contraceptives	20	11
Skiing	21	30	X-rays	22	7
Football	23	27	Railroads	24	19
Food preservatives	25	14	Food coloring	26	21
Power mowers	27	28	Antibiotics	28	24
Home appliances	29	22	Vaccinations	30	25

Review Exercises

16.39 A study by a chemical company compared the drainage properties of two different polymers. Ten different sludges were used, and both polymers were allowed to drain in each sludge. The free drainage was measured in mL/min.

Sludge Type	**Polymer *A***	**Polymer *B***
1	12.7	12.0
2	14.6	15.0
3	18.6	19.2
4	17.5	17.3
5	11.8	12.2
6	16.9	16.6
7	19.9	20.1
8	17.6	17.6
9	15.6	16.0
10	16.0	16.1

(a) Use the sign test at the 0.05 level to test the null hypothesis that polymer A has the same median drainage as polymer B.

(b) Use the signed-rank test to test the hypotheses of part (a).

16.40 In Review Exercise 13.45 on page 555, use the Kruskal-Wallis test, at the 0.05 level of significance, to determine if the chemical analyses performed by the four laboratories give, on average, the same results.

16.41 Use the data from Exercise 13.14 on page 530 to see if the median amount of nitrogen lost in perspiration is different for the three levels of dietary protein.

Chapter 17

Statistical Quality Control

17.1 Introduction

The notion of using sampling and statistical analysis techniques in a production setting had its beginning in the 1920s. The objective of this highly successful concept is the systematic reduction of variability and the accompanying isolation of sources of difficulties *during production.* In 1924, Walter A. Shewhart of the Bell Telephone Laboratories developed the concept of a control chart. However, it was not until World War II that the use of control charts became widespread. This was due to the importance of maintaining quality in production processes during that period. In the 1950s and 1960s, the development of quality control and the general area of quality assurance grew rapidly, particularly with the emergence of the space program in the United States. There has been widespread and successful use of quality control in Japan thanks to the efforts of W. Edwards Deming, who served as a consultant in Japan following World War II. Quality control has been, and is, an important ingredient in the development of Japan's industry and economy.

Quality control is receiving increasing attention as a management tool in which important characteristics of a product are observed, assessed, and compared with some type of standard. The various procedures in quality control involve considerable use of sampling procedures and statistical principles that have been presented in previous chapters. The primary users of quality control are, of course, industrial corporations. It has become clear that an effective quality control program enhances the quality of the product being produced and increases profits. This is particularly true today since products are produced in such high volume. Before the movement toward quality control methods, quality often suffered because of lack of efficiency, which, of course, increases cost.

The Control Chart

The purpose of a control chart is to determine if the performance of a process is maintaining an acceptable level of quality. It is expected, of course, that any process will experience natural variability, that is, variability due to essentially unimportant and uncontrollable sources of variation. On the other hand, a process may experience more serious types of variability in key performance measures.

These sources of variability may arise from one of several types of nonrandom "assignable causes," such as operator errors or improperly adjusted dials on a machine. A process operating in this state is called **out of control**. A process experiencing only chance variation is said to be in **statistical control**. Of course, a successful production process may operate in an in-control state for a long period. It is presumed that during this period, the process is producing an acceptable product. However, there may be either a gradual or a sudden "shift" that requires detection.

A control chart is intended as a device to detect the nonrandom or out-of-control state of a process. Typically, the control chart takes the form indicated in Figure 17.1. It is important that the shift be detected quickly so that the problem can be corrected. Obviously, if detection is slow, many defective or nonconforming items are produced, resulting in considerable waste and increased cost.

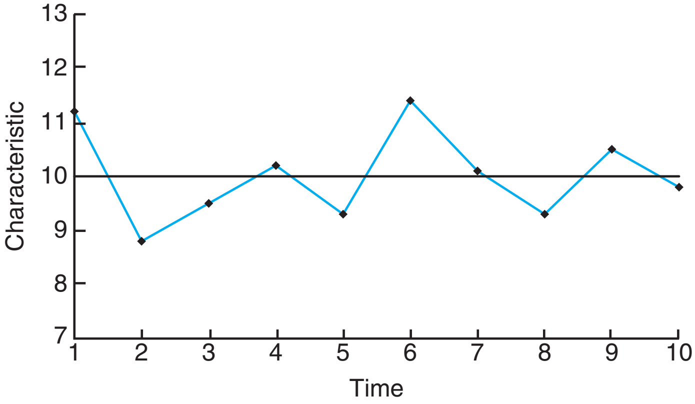


Figure 17.1: Typical control chart.

Some type of quality characteristic must be under consideration, and units of the process must be sampled over time. Say, for example, the characteristic is the circumference of an engine bearing. The centerline represents the average value of the characteristic when the process is in control. The points depicted in the figure represent results of, say, sample averages of this characteristic, with the samples taken over time. The upper control limit and the lower control limit are chosen in such a way that one would expect all sample points to be covered by these boundaries if the process is in control. As a result, the general complexion of the plotted points over time determines whether or not the process is concluded to be in control. The "in control" evidence is produced by a random pattern of points, with all plotted values being inside the control limits. When a point falls outside the control limits, this is taken to be evidence of a process that is out of control, and a search for the assignable cause is suggested. In addition, a nonrandom pattern of points may be considered suspicious and certainly an indication that an investigation for the appropriate corrective action is needed.

17.2 Nature of the Control Limits

The fundamental ideas on which control charts are based are similar in structure to those of hypothesis testing. Control limits are established to control the probability of making the error of concluding that the process is out of control when in fact it is not. This corresponds to the probability of making a type I error if we were testing the null hypothesis that the process is in control. On the other hand, we must be attentive to an error of the second kind, namely, not finding the process out of control when in fact it is (type II error). Thus, the choice of control limits is similar to the choice of a critical region.

As in the case of hypothesis testing, the sample size at each point is important. The choice of sample size depends to a large extent on the sensitivity or power of detection of the out-of-control state. In this application, the notion of *power* is very similar to that of the hypothesis-testing situation. Clearly, the larger the sample at each time period, the quicker the detection of an out-of-control process. In a sense, the control limits actually define what the user considers as being *in control.* In other words, the latitude given by the control limits must depend in some sense on the process variability. As a result, the computation of the control limits will naturally depend on data taken from the process results. Thus, any quality control application must have its beginning with computation from a preliminary sample or set of samples which will establish both the centerline and the quality control limits.

17.3 Purposes of the Control Chart

One obvious purpose of the control chart is mere surveillance of the process, that is, to determine if changes need to be made. In addition, the constant systematic gathering of data often allows management to assess process capability. Clearly, if a single performance characteristic is important, continual sampling and estimation of the mean and standard deviation of that performance characteristic provide an update on what the process can do in terms of mean performance and random variation. This is valuable even if the process stays in control for long periods. The systematic and formal structure of the control chart can often prevent overreaction to changes that represent only random fluctuations. Obviously, in many situations, changes brought about by overreaction can create serious problems that are difficult to solve.

Quality characteristics of control charts fall generally into *two* categories, **variables** and **attributes**. As a result, types of control charts often take the same classifications. In the case of the variables type of chart, the characteristic is usually a measurement on a continuum, such as diameter or weight. For the attribute chart, the characteristic reflects whether the individual product *conforms* (defective or not). Applications for these two distinct situations are obvious.

In the case of the variables chart, control must be exerted on both central tendency and variability. A quality control analyst must be concerned about whether there has been a shift in values of the performance characteristic *on average.* In addition, there will always be a concern about whether some change in process conditions results in a decrease in precision (i.e., an increase in variability). Separate control charts are essential for dealing with these two concepts. Central tendency is controlled by the $\bar{X}$-*chart*, where means of relatively small samples are plotted on a control chart. Variability around the mean is controlled by the *range* in the sample, or the sample *standard deviation*. In the case of attribute sampling, the *proportion defective* from a sample is often the quantity plotted on the chart. In the following section, we discuss the development of control charts for the variables type of performance characteristic.

17.4 Control Charts for Variables

Providing an example is a relatively easy way to explain the rudiments of the $\bar{X}$-chart for variables. Suppose that quality control charts are to be used on a process for manufacturing a certain engine part. Suppose the process mean is $\mu = 50$ mm and the standard deviation is $\sigma = 0.01$ mm. Suppose that groups of 5 are sampled every hour and the values of the *sample mean* $\bar{X}$ are recorded and plotted on a chart like the one in Figure 17.2. The limits for the $\bar{X}$-charts are based on the standard deviation of the random variable $\bar{X}$. We know from material in Chapter 8 that for the average of independent observations in a sample of size n,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}},$$

where σ is the standard deviation of an individual observation. The control limits are designed to result in a small probability that a given value of $\bar{X}$ is outside the limits given that, indeed, the process is in control (i.e., $\mu = 50$). If we invoke the Central Limit Theorem, we have that under the condition that the process is in control,

$$\bar{X} \sim N\left(50, \frac{0.01}{\sqrt{5}}\right).$$

As a result, $100(1-\alpha)\%$ of the $\bar{X}$-values fall inside the limits when the process is in control if we use the limits

$$\text{LCL} = \mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 50 - z_{\alpha/2}(0.0045), \qquad \text{UCL} = \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 50 + z_{\alpha/2}(0.0045).$$

Here LCL and UCL stand for lower control limit and upper control limit, respectively. Often the $\bar{X}$-charts are based on limits that are referred to as “three-sigma” limits, referring, of course, to $z_{\alpha/2} = 3$ and limits that become

$$\mu \ \pm \ 3\frac{\sigma}{\sqrt{n}}.$$

In our illustration, the upper and lower limits become

$$\text{LCL} = 50 - 3(0.0045) = 49.9865, \qquad \text{UCL} = 50 + 3(0.0045) = 50.0135.$$

Thus, if we view the structure of the 3σ limits from the point of view of hypothesis testing, for a given sample point, the probability is 0.0026 that the $\bar{X}$-value falls outside control limits, given that the process is in control. This is the probability

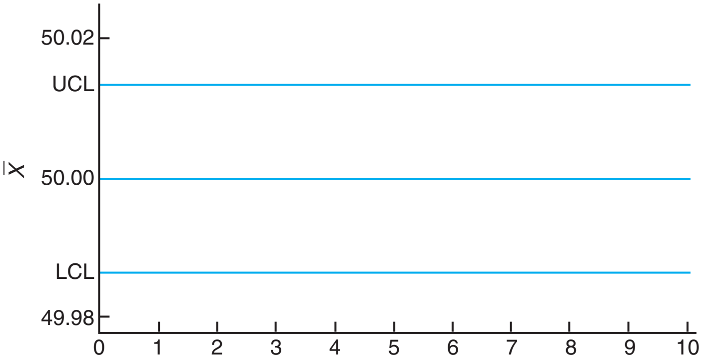


Figure 17.2: The 3σ control limits for the engine part example.

of the analyst *erroneously* determining that the process is out of control (see Table A.3).

The example above not only illustrates the $\bar{X}$-chart for variables, but also should provide the reader with insight into the nature of control charts in general. The centerline generally reflects the ideal value of an important parameter. Control limits are established from knowledge of the sampling properties of the statistic that estimates the parameter in question. They very often involve a multiple of the standard deviation of the statistic. It has become general practice to use 3σ limits. In the case of the $\bar{X}$-chart provided here, the Central Limit Theorem provides the user with a good approximation of the probability of falsely ruling that the process is out of control. In general, though, the user may not be able to rely on the normality of the statistic on the centerline. As a result, the exact probability of "type I error" may not be known. Despite this, it has become fairly standard to use the $k\sigma$ limits. While use of the 3σ limits is widespread, at times the user may wish to deviate from this approach. A smaller multiple of σ may be appropriate when it is important to quickly detect an out-of-control situation. Because of economic considerations, it may prove costly to allow a process to continue to run out of control for even short periods, while the cost of the search and correction of assignable causes may be relatively small. Clearly, in this case, control limits that are tighter than 3σ limits are appropriate.

Rational Subgroups

The sample values to be used in a quality control effort are divided into subgroups, with a *sample* representing a subgroup. As we indicated earlier, time order of production is certainly a natural basis for selection of the subgroups. We may view the quality control effort very simply as (1) sampling, (2) detection of an out-of-control state, and (3) a search for assignable causes that may be occurring over time. The selection of the basis for these sample groups would appear to be straightforward, but the choice of these subgroups of sampling information can have an important effect on the success of the quality control program. These subgroups are often called **rational subgroups**. Generally, if the analyst is interested in detecting a *shift in location*, the subgroups should be chosen so that within-subgroup variability is small and assignable causes, if they are present, have the greatest chance of being detected. Thus, we want to choose the subgroups in such a way as to maximize the between-subgroup variability. Choosing units in a subgroup that are produced close together in time, for example, is a reasonable approach. On the other hand, control charts are often used to control variability, in which case the performance statistic is *variability within the sample.* Thus, it is more important to choose the rational subgroups to maximize the within-sample variability. In this case, the observations in the subgroups should behave more like a random sample and the variability within samples needs to be a depiction of the variability of the process.

It is important to note that control charts on variability should be established before the development of charts on center of location (say, $\bar{X}$-charts). Any control chart on center of location will certainly depend on variability. For example, we have seen an illustration of the central tendency chart and it depends on σ. In the sections that follow, an estimate of σ from the data will be discussed.

$\bar{X}$-Chart with Estimated Parameters

In the foregoing, we have illustrated notions of the $\bar{X}$-chart that make use of the Central Limit Theorem and employ *known* values of the process mean and standard deviation. As we indicated earlier, the control limits

$$\text{LCL} = \mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad \text{UCL} = \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

are used, and an $\bar{X}$-value falling outside these limits is viewed as evidence that the mean μ has changed and thus the process may be out of control.

In many practical situations, it is unreasonable to assume that we know μ and σ. As a result, estimates must be supplied from data taken when the process is in control. Typically, the estimates are determined during a period in which *background information* or *start-up information* is gathered. A basis for rational subgroups is chosen, and data are gathered with samples of size n in each subgroup. The sample sizes are usually small, say 4, 5, or 6, and k samples are taken, with k being at least 20. During this period in which it is assumed that the process is in control, the user establishes estimates of μ and σ on which the control chart is based. The important information gathered during this period includes the sample means in the subgroup, the overall mean, and the sample range in each subgroup. In the following paragraphs, we outline how this information is used to develop the control chart.

A portion of the sample information from these k samples takes the form $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$, where the random variable $\bar{X}_i$ is the average of the values in the ith sample. Obviously, the overall average is the random variable

$$\bar{\bar{X}} = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_i.$$

This is the appropriate estimator of the process mean and, as a result, is the centerline in the $\bar{X}$ control chart. In quality control applications, it is often convenient to estimate σ from the information related to the *ranges* in the samples rather than sample standard deviations. Let us define

$$R_i = X_{\max,i} - X_{\min,i}$$

as the range for the data in the ith sample. Here $X_{\max,i}$ and $X_{\min,i}$ are the largest and smallest observations, respectively, in the sample. The appropriate estimate of σ is a function of the average range

$$\bar{R} = \frac{1}{k}\sum_{i=1}^{k} R_i.$$

An estimate of σ, say $\hat{\sigma}$, is obtained by

$$\hat{\sigma} = \frac{\bar{R}}{d_2},$$

where d_2 is a constant depending on the sample size. Values of d_2 are shown in Table A.22.

Use of the range in producing an estimate of σ has roots in quality-control-type applications, particularly since the range was so easy to compute, compared to other variability estimates, in the era when efficient computation was still an issue. The assumption of normality of the individual observations is implicit in the $\bar{X}$-chart. Of course, the existence of the Central Limit Theorem is certainly helpful in this regard. Under the assumption of normality, we make use of a random variable called the relative range, given by

$$W = \frac{R}{\sigma}.$$

It turns out that the moments of W are simple functions of the sample size n (see the reference to Montgomery, 2000b, in the Bibliography). The expected value of W is often referred to as d_2. Thus, by taking the expected value of W above, we have

$$\frac{E(R)}{\sigma} = d_2.$$

As a result, the rationale for the estimate $\hat{\sigma} = \bar{R}/d_2$ is readily understood. It is well known that the range method produces an efficient estimator of σ in relatively small samples. This makes the estimator particularly attractive in quality control applications, since the sample sizes in the subgroups are generally small. Using the range method for estimation of σ results in control charts with the following parameters:

$$\text{UCL} = \bar{\bar{X}} + \frac{3\bar{R}}{d_2\sqrt{n}}, \qquad \text{centerline} = \bar{\bar{X}}, \qquad \text{LCL} = \bar{\bar{X}} - \frac{3\bar{R}}{d_2\sqrt{n}}.$$

Defining the quantity

$$A_2 = \frac{3}{d_2\sqrt{n}},$$

we have that

$$\text{UCL} = \bar{\bar{X}} + A_2\bar{R}, \qquad \text{LCL} = \bar{\bar{X}} - A_2\bar{R}.$$

To simplify the structure, the user of $\bar{X}$-charts often finds values of A_2 tabulated. Values of A_2 are given for various sample sizes in Table A.22.

R-Charts to Control Variation

Up to this point, all illustrations and details have dealt with the quality control analysts' attempts at detection of out-of-control conditions produced by a *shift in the mean*. The control limits are based on the distribution of the random variable $\bar{X}$ and depend on the assumption of normality of the individual observations. It is important for control to be applied to variability as well as center of location. In fact, many experts believe that control of variability of the performance characteristic is more important and should be established before center of location is considered. Process variability can be controlled through the use of *plots of the sample range*. A plot over time of the sample ranges is called an ***R*-chart**. The same general structure can be used as in the case of the $\bar{X}$-chart, with $\bar{R}$ *being the centerline* and the control limits depending on an estimate of the standard deviation of the random variable R. Thus, as in the case of the $\bar{X}$-chart, 3σ limits are established where "3σ" implies $3\sigma_R$. The quantity σ_R must be estimated from the data just as $\sigma_{\bar{X}}$ is estimated.

The estimate of σ_R, the standard deviation, is also based on the distribution of the relative range

$$W = \frac{R}{\sigma}.$$

The standard deviation of W is a known function of the sample size and is generally denoted by d_3. As a result,

$$\sigma_R = \sigma d_3.$$

We can now replace σ by $\hat{\sigma} = \bar{R}/d_2$, and thus the estimator of σ_R is

$$\hat{\sigma}_R = \frac{\bar{R}d_3}{d_2}.$$

Thus, the quantities that define the R-chart are

$$\text{UCL} = \bar{R}D_4, \qquad \text{centerline} = \bar{R}, \qquad \text{LCL} = \bar{R}D_3,$$

where the constants D_4 and D_3 (depending only on n) are

$$D_4 = 1 + 3\frac{d_3}{d_2}, \qquad D_3 = 1 - 3\frac{d_3}{d_2}.$$

The constants D_4 and D_3 are tabulated in Table A.22.

$\bar{X}$- and R-Charts for Variables

A process manufacturing missile component parts is being controlled, with the performance characteristic being the tensile strength in pounds per square inch. Samples of size 5 each are taken every hour and 25 samples are reported. The data are shown in Table 17.1.

Table 17.1: Sample Information on Tensile Strength Data

Sample Number	**Observations**					$\bar{X}_i$	R_i
1	1515	1518	1512	1498	1511	1510.8	20
2	1504	1511	1507	1499	1502	1504.6	12
3	1517	1513	1504	1521	1520	1515.0	17
4	1497	1503	1510	1508	1502	1504.0	13
5	1507	1502	1497	1509	1512	1505.4	15
6	1519	1522	1523	1517	1511	1518.4	12
7	1498	1497	1507	1511	1508	1504.2	14
8	1511	1518	1507	1503	1509	1509.6	15
9	1506	1503	1498	1508	1506	1504.2	10
10	1503	1506	1511	1501	1500	1504.2	11
11	1499	1503	1507	1503	1501	1502.6	8
12	1507	1503	1502	1500	1501	1502.6	7
13	1500	1506	1501	1498	1507	1502.4	9
14	1501	1509	1503	1508	1503	1504.8	8
15	1507	1508	1502	1509	1501	1505.4	8
16	1511	1509	1503	1510	1507	1508.0	8
17	1508	1511	1513	1509	1506	1509.4	7
18	1508	1509	1512	1515	1519	1512.6	11
19	1520	1517	1519	1522	1516	1518.8	6
20	1506	1511	1517	1516	1508	1511.6	11
21	1500	1498	1503	1504	1508	1502.6	10
22	1511	1514	1509	1508	1506	1509.6	8
23	1505	1508	1500	1509	1503	1505.0	9
24	1501	1498	1505	1502	1505	1502.2	7
25	1509	1511	1507	1500	1499	1505.2	12

As we indicated earlier, it is important initially to establish "in control" conditions on variability. The calculated centerline for the R-chart is

$$\bar{R} = \frac{1}{25}\sum_{i=1}^{25} R_i = 10.72.$$

We find from Table A.22 that for $n = 5$, $D_3 = 0$ and $D_4 = 2.114$. As a result, the control limits for the R-chart are

$$\begin{aligned}\text{LCL} &= \bar{R}D_3 = (10.72)(0) = 0,\\ \text{UCL} &= \bar{R}D_4 = (10.72)(2.114) = 22.6621.\end{aligned}$$

The R-chart is shown in Figure 17.3. None of the plotted ranges fall outside the control limits. As a result, there is no indication of an out-of-control situation.

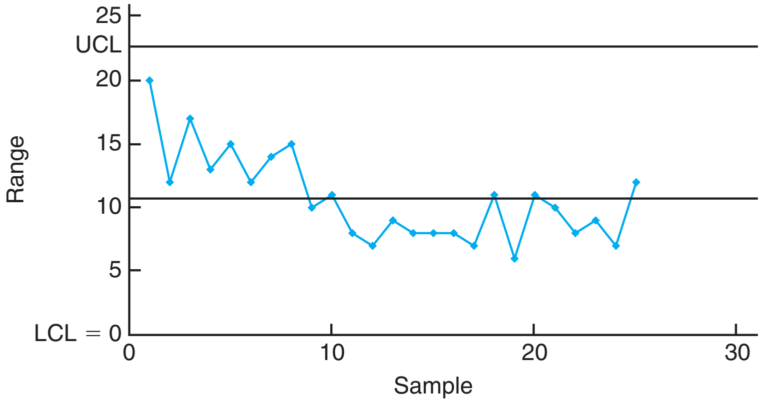


Figure 17.3: R-chart for the tensile strength example.

The $\bar{X}$-chart can now be constructed for the tensile strength readings. The centerline is

$$\bar{\bar{X}} = \frac{1}{25}\sum_{i=1}^{25}\bar{X}_i = 1507.328.$$

For samples of size 5, we find $A_2 = 0.577$ from Table A.22. Thus, the control limits are

$$\text{UCL} = \bar{\bar{X}} + A_2\bar{R} = 1507.328 + (0.577)(10.72) = 1513.5134,$$
$$\text{LCL} = \bar{\bar{X}} - A_2\bar{R} = 1507.328 - (0.577)(10.72) = 1501.1426.$$

The $\bar{X}$-chart is shown in Figure 17.4. As the reader can observe, three values fall outside the control limits. As a result, the control limits for $\bar{X}$ should not be used for line quality control.

Further Comments about Control Charts for Variables

A process may appear to be in control and, in fact, may stay in control for a long period. Does this necessarily mean that the process is operating successfully? A process that is operating *in control* is merely one in which the process mean and variability are stable. Apparently, no serious changes have occurred. “In control” implies that the process remains consistent with natural variability. Quality control charts may be viewed as a method in which the inherent natural variability governs the width of the control limits. There is no implication, however, to what extent an in-control process satisfies predetermined *specifications* required of the process. Specifications are limits that are established by the consumer. If the current natural

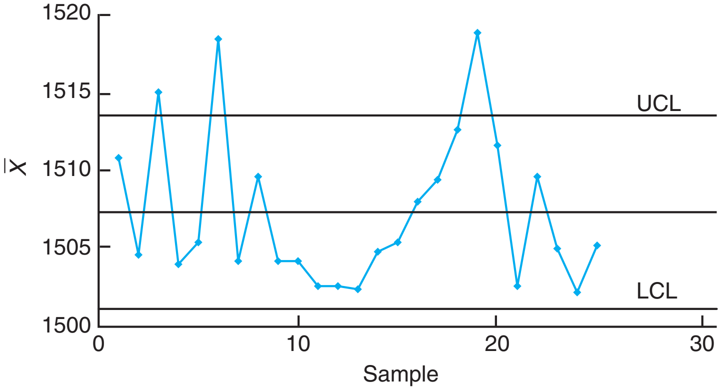


Figure 17.4: $\bar{X}$-chart for the tensile strength example.

variability of the process is larger than that dictated by the specifications, the process will not produce items that meet specifications with high frequency, even though the process is stable and in control.

We have alluded to the normality assumption on the individual observations in a variables control chart. For the $\bar{X}$-chart, if the individual observations are normal, the statistic $\bar{X}$ is normal. As a result, the quality control analyst has control over the probability of type I error in this case. If the individual X's are not normal, $\bar{X}$ is approximately normal and thus there is approximate control over the probability of type I error for the case in which σ is known. However, the use of the range method for estimating the standard deviation also depends on the normality assumption. Studies regarding the robustness of the $\bar{X}$-chart to departures from normality indicate that for samples of size $k \geq 4$ the $\bar{X}$ chart results in an α-risk close to that advertised (see the work by Montgomery, 2000b, and Schilling and Nelson, 1976, in the Bibliography). We indicated earlier that the $\pm k\sigma_R$ approach to the R-chart is a matter of convenience and tradition. Even if the distribution of individual observations is normal, the distribution of R is not normal. In fact, the distribution of R is not even symmetric. The symmetric control limits of $\pm k\sigma_R$ only give an approximation to the α-risk, and in some cases the approximation is not particularly good.

Choice of Sample Size (Operating Characteristic Function) in the Case of the $\bar{X}$-Chart

Scientists and engineers dealing in quality control often refer to factors that affect the *design of the control chart.* Components that determine the design of the chart include the sample size taken in each subgroup, the width of the control limits, and the frequency of sampling. All of these factors depend to a large extent on economic and practical considerations. Frequency of sampling obviously depends on the cost of sampling and the cost incurred if the process continues out of control for a long period. These same factors affect the width of the "in-control" region. The cost that is associated with investigation and search for assignable causes has an impact on the width of the region and on frequency of sampling. A considerable amount of attention has been devoted to optimal design of control charts, and extensive details will not be given here. The reader should refer to the work by Montgomery (2000b) cited in the Bibliography for an excellent historical account of much of this research.

Choice of sample size and frequency of sampling involves balancing available resources allocated to these two efforts. In many cases, the analyst may need to make changes in the strategy until the proper balance is achieved. The analyst should always be aware that if the cost of producing nonconforming items is great, a high sampling frequency with relatively small sample size is a proper strategy.

Many factors must be taken into consideration in the choice of a sample size. In the illustrations and discussion, we have emphasized the use of $n = 4$, 5, or 6. These values are considered relatively small for general problems in statistical inference but perhaps proper sample sizes for quality control. One justification, of course, is that quality control is a continuing process and the results produced by one sample or set of units will be followed by results from many more. Thus, the "effective" sample size of the entire quality control effort is many times larger than that used in a subgroup. It is generally considered to be more effective to *sample frequently* with a small sample size.

The analyst can make use of the notion of the power of a test to gain some insight into the effectiveness of the sample size chosen. This is particularly important since small sample sizes are usually used in each subgroup. Refer to Chapters 10 and 13 for a discussion of the power of formal tests on means and the analysis of variance. Although formal tests of hypotheses are not actually being conducted in quality control, one can treat the sampling information as if the strategy at each subgroup were to test a hypothesis, either on the population mean μ or on the standard deviation σ. Of interest is the *probability of detection* of an out-of-control condition for a given sample and, perhaps more important, the expected number of runs required for detection. The probability of detection of a specified out-of-control condition corresponds to the power of a test. It is not our intention to show development of the power for all of the types of control charts presented here, but rather to show the development for the $\bar{X}$-chart and present power results for the R-chart.

Consider the $\bar{X}$-chart for σ known. Suppose that the in-control state has $\mu = \mu_0$. A study of the role of the subgroup sample size is tantamount to investigating the β-risk, that is, the probability that an $\bar{X}$-value remains inside the control limits given that, indeed, a shift in the mean has occurred. Suppose that the form the shift takes is

$$\mu = \mu_0 + r\sigma.$$

Again, making use of the normality of $\bar{X}$, we have

$$\beta = P(\text{LCL} \le \bar{X} \le \text{UCL} \mid \mu = \mu_0 + r\sigma).$$

For the case of $k\sigma$ limits,

$$\text{LCL} = \mu_0 - \frac{k\sigma}{\sqrt{n}} \quad \text{and} \quad \text{UCL} = \mu_0 + \frac{k\sigma}{\sqrt{n}}.$$

As a result, if we denote by Z the standard normal random variable,

$$\begin{aligned}\beta &= P\left[Z < \left(\frac{\mu_0 + k\sigma/\sqrt{n} - \mu}{\sigma/\sqrt{n}}\right)\right] - P\left[Z < \left(\frac{\mu_0 - k\sigma/\sqrt{n} - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= P\left\{Z < \left[\frac{\mu_0 + k\sigma/\sqrt{n} - (\mu + r\sigma)}{\sigma/\sqrt{n}}\right]\right\} - P\left\{Z < \left[\frac{\mu_0 - k\sigma/\sqrt{n} - (\mu + r\sigma)}{\sigma/\sqrt{n}}\right]\right\} \\ &= P(Z < k - r\sqrt{n}) - P(Z < -k - r\sqrt{n}).\end{aligned}$$

Notice the role of n, r, and k in the expression for the β-risk. The probability of not detecting a specific shift clearly increases with an increase in k, as expected. β decreases with an increase in r, the magnitude of the shift, and decreases with an increase in the sample size n.

It should be emphasized that the expression above results in the β-risk (probability of type II error) for the case of a *single sample.* For example, suppose that in the case of a sample of size 4, a shift of σ occurs in the mean. The probability of detecting the shift (power) *in the first sample following the shift* is (assuming 3σ limits)

$$1 - \beta = 1 - [P(Z < 1) - P(Z < -5)] = 0.1587.$$

On the other hand, the probability of detecting a shift of 2σ is

$$1 - \beta = 1 - [P(Z < -1) - P(Z < -7)] = 0.8413.$$

The results above illustrate a fairly modest probability of detecting a shift of magnitude σ and a fairly high probability of detecting a shift of magnitude 2σ. The complete picture of how, say, 3σ control limits perform for the $\bar{X}$-chart described here is depicted in Figure 17.5. Rather than plotting the power functions, a plot is given of β against r, where the shift in the mean is of magnitude $r\sigma$. Of course, the sample sizes of $n = 4, 5, 6$ result in a small probability of detecting a shift of 1.0σ or even 1.5σ on the first sample after the shift.

But if sampling is done frequently, the probability may not be as important as the average or expected number of runs required before detection of the shift. Quick detection is important and is certainly possible even though the probability of detection on the first sample is not high. It turns out that $\bar{X}$-charts with these small samples will result in relatively rapid detection. If β is the probability of not detecting a shift on the first sample following the shift, then the probability of detecting the shift on the sth sample after the shift is (assuming independent samples)

$$P_s = (1 - \beta)\beta^{s-1}.$$

The reader should recognize this as an application of the geometric distribution. The average or expected value of the number of samples required for detection is

$$\sum_{s=1}^{\infty} s\beta^{s-1}(1 - \beta) = \frac{1}{1 - \beta}.$$

Thus, the expected number of samples required to detect the shift in the mean is the *reciprocal of the power* (i.e., the probability of detection on the first sample following the shift).

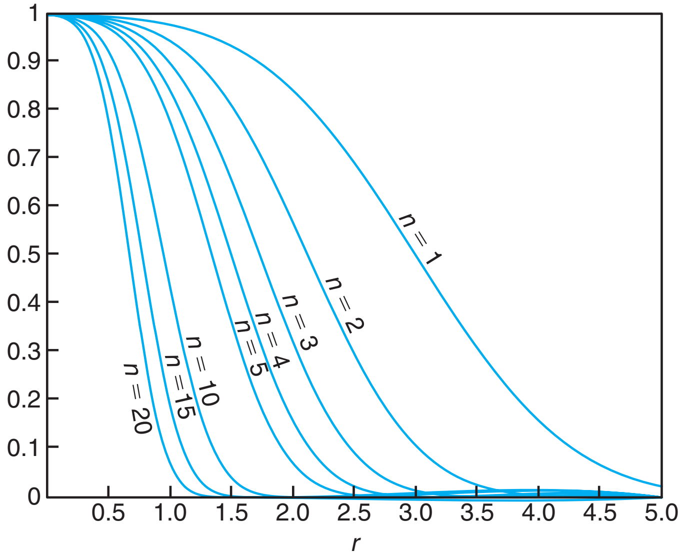


Figure 17.5: Operating characteristic curves for the $\bar{X}$-chart with 3σ limits. Here β is the type II probability error on the first sample after a shift in the mean of $r\sigma$.

Example 17.1: In a certain quality control effort, it is important for the quality control analyst to quickly detect shifts in the mean of $\pm\sigma$ while using a 3σ control chart with a sample size $n = 4$. The expected number of samples that are required following the shift for the detection of the out-of-control state can be an aid in the assessment of the quality control procedure.

From Figure 17.5, for $n = 4$ and $r = 1$, it can be seen that $\beta \approx 0.84$. If we allow s to denote the number of samples required to detect the shift, the mean of s is

$$E(s) = \frac{1}{1-\beta} = \frac{1}{0.16} = 6.25.$$

Thus, on the average, seven subgroups are required before detection of a shift of $\pm\sigma$.

Choice of Sample Size for the *R*-Chart

The OC curve for the R-chart is shown in Figure 17.6. Since the R-chart is used for control of the process standard deviation, the β-risk is plotted as a function of the in-control standard deviation, σ_0, and the standard deviation after the process goes out of control. The latter standard deviation will be denoted σ_1. Let

$$\lambda = \frac{\sigma_1}{\sigma_0}.$$

For various sample sizes, β is plotted against λ.

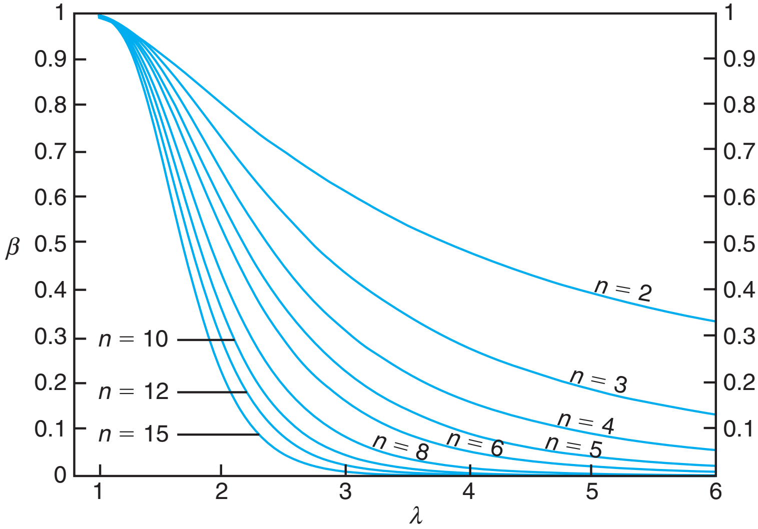


Figure 17.6: Operating characteristic curve for the R-charts with 3σ limits.

$\bar{X}$- and S-Charts for Variables

It is natural for the student of statistics to anticipate use of the sample variance in the $\bar{X}$-chart and in a chart to control variability. The range is efficient as an estimator for σ, but this efficiency decreases as the sample size gets larger. For n as large as 10, the familiar statistic

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

should be used in the control chart for both the mean and the variability. The reader should recall from Chapter 9 that S^2 is an unbiased estimator for σ^2 but that S is not unbiased for σ. It has become customary to correct S for bias in control chart applications. We know, in general, that

$$E(S) \neq \sigma.$$

In the case in which the X_i are independent and normally distributed with mean μ and variance σ^2,

$$E(S) = c_4\sigma, \qquad \text{where} \qquad c_4 = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}$$

and $\Gamma(\cdot)$ refers to the gamma function (see Chapter 6). For example, for $n = 5$, $c_4 = (3/8)\sqrt{2\pi}$. In addition, the variance of the estimator S is

$$\text{Var}(S) = \sigma^2(1 - c_4^2).$$

We have established the properties of S that will allow us to write control limits for both $\bar{X}$ and S. To build a proper structure, we begin by assuming that σ is known. Later we discuss estimating σ from a preliminary set of samples.

If the statistic S is plotted, the obvious control chart parameters are

$$\text{UCL} = c_4\sigma + 3\sigma\sqrt{1 - c_4^2}, \quad \text{centerline} = c_4\sigma, \quad \text{LCL} = c_4\sigma - 3\sigma\sqrt{1 - c_4^2}.$$

As usual, the control limits are defined more succinctly through use of tabulated constants. Let

$$B_5 = c_4 - 3\sqrt{1 - c_4^2}, \qquad B_6 = c_4 + 3\sqrt{1 - c_4^2},$$

and thus we have

$$\text{UCL} = B_6\sigma, \qquad \text{centerline} = c_4\sigma, \qquad \text{LCL} = B_5\sigma.$$

The values of B_5 and B_6 for various sample sizes are tabulated in Table A.22.

Now, of course, the control limits above serve as a basis for the development of the quality control parameters for the situation that is most often seen in practice, namely, that in which σ is unknown. We must once again assume that a set of *base samples* or preliminary samples is taken to produce an estimate of σ during what is assumed to be an "in-control" period. Sample standard deviations $S_1, S_2, \ldots, S_m$ are obtained from samples that are each of size n. An unbiased estimator of the type

$$\frac{\bar{S}}{c_4} = \left(\frac{1}{m}\sum_{i=1}^{m} S_i\right) \Big/ c_4$$

is often used for σ. Here, of course, $\bar{S}$, the average value of the sample standard deviation in the preliminary sample, is the logical centerline in the control chart to control variability. The upper and lower control limits are unbiased estimators of the control limits that are appropriate for the case where σ is known. Since

$$E\left(\frac{\bar{S}}{c_4}\right) = \sigma,$$

the statistic $\bar{S}$ is an appropriate centerline (as an unbiased estimator of $c_4\sigma$) and the quantities

$$\bar{S} - 3\frac{\bar{S}}{c_4}\sqrt{1 - c_4^2} \qquad \text{and} \qquad \bar{S} + 3\frac{\bar{S}}{c_4}\sqrt{1 - c_4^2}$$

are the appropriate lower and upper 3σ control limits, respectively. As a result, the centerline and limits for the S-chart to control variability are

$$\text{LCL} = B_3\bar{S}, \qquad \text{centerline} = \bar{S}, \qquad \text{UCL} = B_4\bar{S},$$

where

$$B_3 = 1 - \frac{3}{c_4}\sqrt{1 - c_4^2}, \qquad B_4 = 1 + \frac{3}{c_4}\sqrt{1 - c_4^2}.$$

The constants B_3 and B_4 appear in Table A.22.

We can now write the parameters of the corresponding $\bar{X}$-chart involving the use of the sample standard deviation. Let us assume that $\bar{S}$ and $\bar{\bar{X}}$ are available from the base preliminary sample. The centerline remains $\bar{\bar{X}}$ and the 3σ limits are merely of the form $\bar{\bar{X}} \pm 3\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is an unbiased estimator. We simply supply $\bar{S}/c_4$ as an estimator for σ, and thus we have

$$\text{LCL} = \bar{\bar{X}} - A_3\bar{S}, \qquad \text{centerline} = \bar{\bar{X}}, \qquad \text{UCL} = \bar{\bar{X}} + A_3\bar{S},$$

where

$$A_3 = \frac{3}{c_4\sqrt{n}}.$$

The constant A_3 appears for various sample sizes in Table A.22.

Example 17.2: Containers are produced by a process where the volume of the containers is subject to quality control. Twenty-five samples of size 5 each were used to establish the quality control parameters. Information from these samples is documented in Table 17.2.

From Table A.22, $B_3 = 0$, $B_4 = 2.089$, and $A_3 = 1.427$. As a result, the control limits for $\bar{X}$ are given by

$$\text{UCL} = \bar{\bar{X}} + A_3\bar{S} = 62.3771, \qquad \text{LCL} = \bar{\bar{X}} - A_3\bar{S} = 62.2741,$$

and the control limits for the S-chart are

$$\text{LCL} = B_3\bar{S} = 0, \qquad \text{UCL} = B_4\bar{S} = 0.0754.$$

Figures 17.7 and 17.8 show the $\bar{X}$ and S control charts, respectively, for this example. Sample information for all 25 samples in the preliminary data set is plotted on the charts. Control seems to have been established after the first few samples.

17.5 Control Charts for Attributes

As we indicated earlier in this chapter, many industrial applications of quality control require that the quality characteristic indicate no more than that the item "conforms." In other words, there is no continuous measurement that is crucial to the performance of the item. An obvious illustration of this type of sampling, called **sampling for attributes**, is the performance of a light bulb, which either performs satisfactorily or does not. The item is either **defective or not defective**. Manufactured metal pieces may contain deformities. Containers from a production line may leak. In both of these cases, a defective item negates usage by the customer. The standard control chart for this situation is the p-chart, or *chart for fraction defective*. As we might expect, the probability distribution involved is the binomial distribution. The reader is referred to Chapter 5 for background on the binomial distribution.

Table 17.2: Volume of Containers for 25 Samples in a Preliminary Sample (in cubic centimeters)

Sample	**Observations**					$\bar{X}_i$	S_i
1	62.255	62.301	62.289	62.189	62.311	62.269	0.0495
2	62.187	62.225	62.337	62.297	62.307	62.271	0.0622
3	62.421	62.377	62.257	62.295	62.222	62.314	0.0829
4	62.301	62.315	62.293	62.317	62.409	62.327	0.0469
5	62.400	62.375	62.295	62.272	62.372	62.343	0.0558
6	62.372	62.275	62.315	62.372	62.302	62.327	0.0434
7	62.297	62.303	62.337	62.392	62.344	62.335	0.0381
8	62.325	62.362	62.351	62.371	62.397	62.361	0.0264
9	62.327	62.297	62.318	62.342	62.318	62.320	0.0163
10	62.297	62.325	62.303	62.307	62.333	62.313	0.0153
11	62.315	62.366	62.308	62.318	62.319	62.325	0.0232
12	62.297	62.322	62.344	62.342	62.313	62.324	0.0198
13	62.375	62.287	62.362	62.319	62.382	62.345	0.0406
14	62.317	62.321	62.297	62.372	62.319	62.325	0.0279
15	62.299	62.307	62.383	62.341	62.394	62.345	0.0431
16	62.308	62.319	62.344	62.319	62.378	62.334	0.0281
17	62.319	62.357	62.277	62.315	62.295	62.313	0.0300
18	62.333	62.362	62.292	62.327	62.314	62.326	0.0257
19	62.313	62.387	62.315	62.318	62.341	62.335	0.0313
20	62.375	62.321	62.354	62.342	62.375	62.353	0.0230
21	62.399	62.308	62.292	62.372	62.299	62.334	0.0483
22	62.309	62.403	62.318	62.295	62.317	62.328	0.0427
23	62.293	62.293	62.342	62.315	62.349	62.318	0.0264
24	62.388	62.308	62.315	62.392	62.303	62.341	0.0448
25	62.324	62.318	62.315	62.295	62.319	62.314	0.0111
						$\bar{\bar{X}} = 62.3256$	
						$\bar{S} = 0.0361$	

The p-Chart for Fraction Defective

Any manufactured item may have several characteristics that are important and should be examined by an inspector. However, the entire development here focuses on a single characteristic. Suppose that for all items the probability of a defective item is p, and that all items are being produced independently. Then, in a random sample of n items produced, allowing X to be the number of defective items, we have

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

As one might suspect, the mean and variance of the binomial random variable will play an important role in the development of the control chart. The reader should recall that

$$E(X) = np \qquad \text{and} \qquad \mathrm{Var}(X) = np(1-p).$$

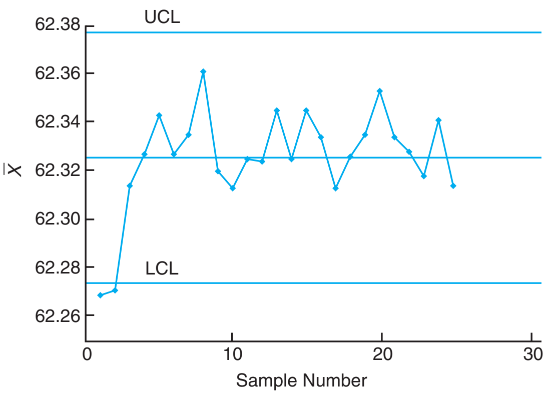


Figure 17.7: The $\bar{X}$-chart with control limits established by the data of Example 17.2.

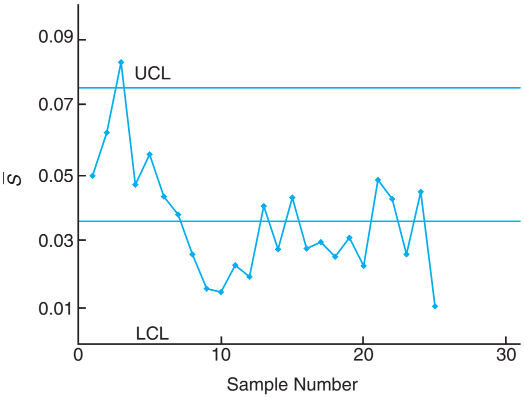


Figure 17.8: The S-chart with control limits established by the data of Example 17.2.

An unbiased estimator of p is the **fraction defective** or the **proportion defective**, $\hat{p}$, where

$$\hat{p} = \frac{\text{number of defectives in the sample of size } n}{n}.$$

As in the case of the variables control charts, the distributional properties of p are important in the development of the control chart. We know that

$$E(\hat{p}) = p, \qquad \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}.$$

Here we apply the same 3σ principles that we use for the variables charts. Let us assume initially that p is known. The structure, then, of the control charts involves the use of 3σ limits with

$$\hat{\sigma} = \sqrt{\frac{p(1-p)}{n}}.$$

Thus, the limits are

$$\mathrm{LCL} = p - 3\sqrt{\frac{p(1-p)}{n}}, \qquad \mathrm{UCL} = p + 3\sqrt{\frac{p(1-p)}{n}},$$

with the process considered in control when the $\hat{p}$-values from the sample lie inside the control limits.

Generally, of course, the value of p is not known and must be estimated from a base set of samples very much like the case of μ and σ in the variables charts. Assume that there are m preliminary samples of size n. For a given sample, each of the n observations is reported as either "defective" or "not defective." The obvious unbiased estimator for p to use in the control chart is

$$\bar{p} = \frac{1}{m}\sum_{i=1}^{m} \hat{p}_i,$$

where $\hat{p}_i$ is the proportion defective in the ith sample. As a result, the control limits are

$$\text{LCL} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \qquad \text{centerline} = \bar{p}, \qquad \text{UCL} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}.$$

Example 17.3: Consider the data shown in Table 17.3 on the number of defective electronic components in samples of size 50. Twenty samples were taken in order to establish preliminary control chart values. The control charts determined by this preliminary period will have centerline $\bar{p} = 0.088$ and control limits

$$\text{LCL} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{50}} = -0.0322 \quad \text{and} \quad \text{UCL} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{50}} = 0.2082.$$

Table 17.3: Data for Example 17.3 to Establish Control Limits for p-Charts, Samples of Size 50

Sample	**Number of Defective Components**	**Fraction Defective $\hat{p}_i$**
1	8	0.16
2	6	0.12
3	5	0.10
4	7	0.14
5	2	0.04
6	5	0.10
7	3	0.06
8	8	0.16
9	4	0.08
10	4	0.08
11	3	0.06
12	1	0.02
13	5	0.10
14	4	0.08
15	4	0.08
16	2	0.04
17	3	0.06
18	5	0.10
19	6	0.12
20	3	0.06
		$\bar{p} = 0.088$

Obviously, with a computed value that is negative, the LCL will be set to zero. It is apparent from the values of the control limits that the process is in control during this preliminary period. ▟

Choice of Sample Size for the p-Chart

The choice of sample size for the p-chart for attributes involves the same general types of considerations as that of the chart for variables. A sample size is required that is sufficiently large to have a high probability of detection of an out-of-control condition when, in fact, a specified change in p has occurred. There is *no best method* for choice of sample size. However, one reasonable approach, suggested by Duncan (1986; see the Bibliography), is to choose n so that there is probability 0.5 that we detect a shift in p of a particular amount. The resulting solution for n is quite simple. Suppose that the normal approximation to the binomial distribution applies. We wish, under the condition that p has shifted to, say, $p_1 > p_0$, that

$$P(\hat{p} \geq \text{UCL}) = P\left[Z \geq \frac{\text{UCL} - p_1}{\sqrt{p_1(1-p_1)/n}}\right] = 0.5.$$

Since $P(Z > 0) = 0.5$, we set

$$\frac{\text{UCL} - p_1}{\sqrt{p_1(1-p_1)/n}} = 0.$$

Substituting

$$p + 3\sqrt{\frac{p(1-p)}{n}} = \text{UCL},$$

we have

$$(p - p_1) + 3\sqrt{\frac{p(1-p)}{n}} = 0.$$

We can now solve for n, the size of each sample:

$$n = \frac{9}{\Delta^2}p(1-p),$$

where, of course, Δ is the "shift" in the value of p, and p is the probability of a defective on which the control limits are based. However, if the control charts are based on $k\sigma$ limits, then

$$n = \frac{k^2}{\Delta^2}p(1-p).$$

Example 17.4: Suppose that an attribute quality control chart is being designed with a value of $p = 0.01$ for the in-control probability of a defective. What is the sample size per subgroup producing a probability of 0.5 that a process shift to $p = p_1 = 0.05$ will be detected? The resulting p-chart will involve 3σ limits.

***Solution*:** Here we have $\Delta = 0.04$. The appropriate sample size is

$$n = \frac{9}{(0.04)^2}(0.01)(0.99) = 55.69 \approx 56.$$

Control Charts for Defects (Use of the Poisson Model)

In the preceding development, we assumed that the item under consideration is one that is either defective (i.e., nonfunctional) or not defective. In the latter case, it is functional and thus acceptable to the consumer. In many situations, this "defective or not" approach is too simplistic. Units may contain defects or nonconformities but still function quite well for the consumer. Indeed, in this case, it may be important to exert control on the *number of defects* or *number of nonconformities*. This type of quality control effort finds application when the units are either not simplistic or large. For example, the number of defects may be quite useful as the object of control when the single item or unit is, say, a personal computer. Another example is a unit defined by 50 feet of manufactured pipeline, where the number of defective welds is the object of quality control; the number of defects in 50 feet of manufactured carpeting; or the number of "bubbles" in a large manufactured sheet of glass.

It is clear from what we describe here that the binomial distribution is not appropriate. The total number of nonconformities in a unit or the average number per unit can be used as the measure for the control chart. Often it is assumed that the number of nonconformities in a sample of items follows the Poisson distribution. This type of chart is often called a ***C*-chart**.

Suppose that the number of defects X in one unit of product follows the Poisson distribution with parameter λ. (Here $t = 1$ for the Poisson model.) Recall that for the Poisson distribution,

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots.$$

Here, the random variable X is the number of nonconformities. In Chapter 5, we learned that the mean and variance of the Poisson random variable are both λ. Thus, if the quality control chart were to be structured according to the usual 3σ limits, we could have, for λ known,

$$\text{UCL} = \lambda + 3\sqrt{\lambda}, \qquad \text{centerline} = \lambda, \qquad \text{LCL} = \lambda - 3\sqrt{\lambda}.$$

As usual, λ often must come from an estimator from the data. An unbiased estimate of λ is the *average* number of nonconformities per sample. Denote this estimate by $\hat{\lambda}$. Thus, the control chart has the limits

$$\text{UCL} = \hat{\lambda} + 3\sqrt{\hat{\lambda}}, \qquad \text{centerline} = \hat{\lambda}, \qquad \text{LCL} = \hat{\lambda} - 3\sqrt{\hat{\lambda}}.$$

Example 17.5: Table 17.4 represents the number of defects in 20 successive samples of sheet metal rolls each 100 feet long. A control chart is to be developed from these preliminary data for the purpose of controlling the number of defects in such samples. The estimate of the Poisson parameter λ is given by $\hat{\lambda} = 5.95$. As a result, the control limits suggested by these preliminary data are

$$\text{UCL} = \hat{\lambda} + 3\sqrt{\hat{\lambda}} = 13.2678 \quad \text{and} \quad \text{LCL} = \hat{\lambda} - 3\sqrt{\hat{\lambda}} = -1.3678,$$

with LCL being set to zero.

Table 17.4: Data for Example 17.5; Control Involves Number of Defects in Sheet Metal Rolls

Sample Number	**Number of Defects**	**Sample Number**	**Number of Defects**
1	8	11	3
2	7	12	7
3	5	13	5
4	4	14	9
5	4	15	7
6	7	16	7
7	6	17	8
8	4	18	6
9	5	19	7
10	6	20	4
			Ave. 5.95

Figure 17.9 shows a plot of the preliminary data with the control limits revealed.

Table 17.5 shows additional data taken from the production process. For each sample, the unit on which the chart was based, namely 100 feet of the metal, was inspected. The information on 20 samples is included. Figure 17.10 shows a plot of the additional production data. It is clear that the process is in control, at least through the period for which the data were taken.

Table 17.5: Additional Data from the Production Process of Example 17.5

Sample Number	**Number of Defects**	**Sample Number**	**Number of Defects**
1	3	11	7
2	5	12	5
3	8	13	9
4	5	14	4
5	8	15	6
6	4	16	5
7	3	17	3
8	6	18	2
9	5	19	1
10	2	20	6

In Example 17.5, we have made very clear what the sampling or inspection unit is, namely, 100 feet of metal. In many cases where the item is a specific one (e.g., a personal computer or a specific type of electronic device), the inspection unit may be a *set of items*. For example, the analyst may decide to use 10 computers in each subgroup and observe a count of the total number of defects found. Thus, the preliminary sample for construction of the control chart would involve several samples, each containing 10 computers. The choice of the sample size may depend on many factors. Often, we may want a sample size that will ensure an LCL that is positive.

The analyst may wish to use the average number of defects per sampling unit as the basic measure in the control chart. For example, for the case of the personal

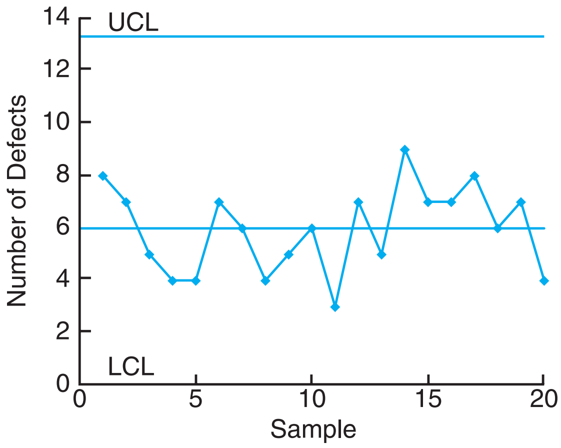


Figure 17.9: Preliminary data plotted on the control chart for Example 17.5.

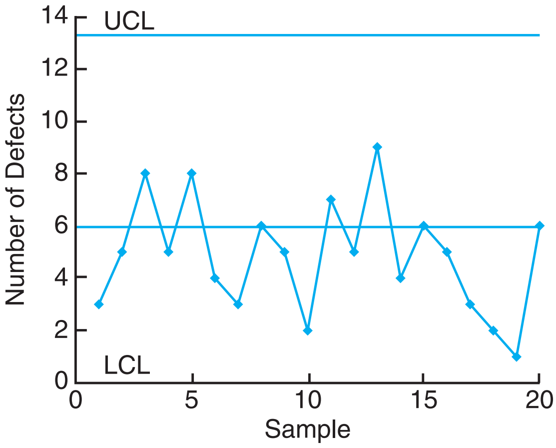


Figure 17.10: Additional production data for Example 17.5.

computer, let the random variable total number of defects

$$U = \frac{\text{total number of defects}}{n}$$

be measured for each sample of, say, $n = 10$. We can use the method of moment-generating functions to show that U is a Poisson random variable (see Review Exercise 17.1) if we assume that the number of defects per sampling unit is Poisson with parameter λ. Thus, the control chart for this situation is characterized by the following:

$$\text{UCL} = \bar{U} + 3\sqrt{\frac{\bar{U}}{n}}, \qquad \text{centerline} = \bar{U}, \qquad \text{LCL} = \bar{U} - 3\sqrt{\frac{\bar{U}}{n}}.$$

Here, of course, $\bar{U}$ is the average of the U-values in the preliminary or base data set. The term $\bar{U}/n$ is derived from the result that

$$E(U) = \lambda, \qquad \text{Var}(U) = \frac{\lambda}{n},$$

and thus $\bar{U}$ is an unbiased estimate of $E(U) = \lambda$ and $\bar{U}/n$ is an unbiased estimate of $\text{Var}(U) = \lambda/n$. This type of control chart is often called a ***U*-chart**.

In this section, we based our entire development of control charts on the Poisson probability model. This model has been used in combination with the 3σ concept. As we implied earlier in this chapter, the notion of 3σ limits has its roots in the normal approximation, although many users feel that the concept works well as a pragmatic tool even if normality is not even approximately correct. The difficulty, of course, is that in the absence of normality, we cannot control the probability of incorrect specification of an out-of-control state. In the case of the Poisson model, when λ is small the distribution is quite asymmetric, a condition that may produce undesirable results if we hold to the 3σ approach.

17.6 Cusum Control Charts

The disadvantage of the Shewhart-type control charts, developed and illustrated in the preceding sections, lies in their inability to detect small changes in the mean. A quality control mechanism that has received considerable attention in the statistics literature and usage in industry is the **cumulative sum (cusum) chart**. The method for the cusum chart is simple and its appeal is intuitive. It should become obvious to the reader why it is more responsive to small changes in the mean. Consider a control chart for the mean with a reference level established at value W. Consider particular observations $X_1, X_2, \ldots, X_r$. The first r cusums are

$$\begin{aligned} S_1 &= X_1 - W \\ S_2 &= S_1 + (X_2 - W) \\ S_3 &= S_2 + (X_3 - W) \\ &\vdots \\ S_r &= S_{r-1} + (X_r - W). \end{aligned}$$

It becomes clear that the cusum is merely the accumulation of differences from the reference level. That is,

$$S_k = \sum_{i=1}^{k} (X_i - W), \qquad k = 1, 2, \ldots.$$

The cusum chart is, then, a plot of S_k against time.

Suppose that we consider the reference level W to be an acceptable value of the mean μ. Clearly, if there is no shift in μ, the cusum chart should be approximately horizontal, with some minor fluctuations balanced around zero. Now, if there is only a moderate change in the mean, a relatively large change in the *slope* of the cusum chart should result, since each new observation has a chance of contributing a shift and the measure being plotted is accumulating these shifts. Of course, the signal that the mean has shifted lies in the nature of the slope of the cusum chart. The purpose of the chart is to detect changes that are moving away from the reference level. A nonzero slope (in either direction) represents a change away from the reference level. A positive slope indicates an increase in the mean above the reference level, while a negative slope signals a decrease.

Cusum charts are often devised with a defined *acceptable quality level* (AQL) and *rejectable quality level* (RQL) preestablished by the user. Both represent values of the mean. These may be viewed as playing roles somewhat similar to those of the null and alternative mean of hypothesis testing. Consider a situation where the analyst hopes to detect an increase in the value of the process mean. We shall use the notation μ_0 for AQL and μ_1 for RQL and let $\mu_1 > \mu_0$. The reference level is now set at

$$W = \frac{\mu_0 + \mu_1}{2}.$$

The values of S_r $(r = 1, 2, \ldots.)$ will have a negative slope if the process mean is at μ_0 and a positive slope if the process mean is at μ_1.

Decision Rule for Cusum Charts

As indicated earlier, the slope of the cusum chart provides the signal for action by the quality control analyst. The decision rule calls for action if, at the rth sampling period,

$$d_r > h,$$

where h is a prespecified value called the **length of the decision interval** and

$$d_r = S_r - \min_{1 \leq i \leq r-1} S_i.$$

In other words, action is taken if the data reveal that the current cusum value exceeds by a specified amount the previous smallest cusum value.

A modification in the mechanics described above makes employing the method easier. We have described a procedure that plots the cusums and computes differences. A simple modification involves plotting the differences directly and allows for checking against the decision interval. The general expression for d_r is quite simple. For the cusum procedure where we are detecting increases in the mean,

$$d_r = \max[0, d_{r-1} + (X_r - W)].$$

The choice of the value of h is, of course, very important. We do not choose in this book to provide the many details in the literature dealing with this choice. The reader is referred to Ewan and Kemp, 1960, and Montgomery, 2000b (see the Bibliography) for a thorough discussion. One important consideration is the **expected run length**. Ideally, the expected run length is quite large under $\mu = \mu_0$ and quite small when $\mu = \mu_1$.

Review Exercises

17.1 Consider $X_1, X_2, \ldots, X_n$ independent Poisson random variables with parameters $\mu_1, \mu_2, \ldots, \mu_n$. Use the properties of moment-generating functions to show that the random variable $\sum_{i=1}^{n} X_i$ is a Poisson random variable with mean $\sum_{i=1}^{n} \mu_i$ and variance $\sum_{i=1}^{n} \mu_i$.

17.2 Consider the following data taken on subgroups of size 5. The data contain 20 averages and ranges on the diameter (in millimeters) of an important component part of an engine. Display $\bar{X}$- and R-charts. Does the process appear to be in control?

Sample	$\bar{X}$	R
1	2.3972	0.0052
2	2.4191	0.0117
3	2.4215	0.0062
4	2.3917	0.0089
5	2.4151	0.0095
6	2.4027	0.0101
7	2.3921	0.0091
8	2.4171	0.0059
9	2.3951	0.0068
10	2.4215	0.0048
11	2.3887	0.0082
12	2.4107	0.0032
13	2.4009	0.0077
14	2.3992	0.0107
15	2.3889	0.0025
16	2.4107	0.0138
17	2.4109	0.0037
18	2.3944	0.0052
19	2.3951	0.0038
20	2.4015	0.0017

17.3 Suppose for Review Exercise 17.2 that the buyer has set specifications for the part. The specifications require that the diameter fall in the range covered by 2.40000 ± 0.0100 mm. What proportion of units produced by this process will not conform to specifications?

17.4 For the situation of Review Exercise 17.2, give numerical estimates of the mean and standard deviation of the diameter for the part being manufactured in the process.

17.5 Consider the data of Table 17.1. Suppose that additional samples of size 5 are taken and tensile strength recorded. The sampling produces the following results (in pounds per square inch).

Sample	$\bar{X}$	R
1	1511	22
2	1508	14
3	1522	11
4	1488	18
5	1519	6
6	1524	11
7	1519	8
8	1504	7
9	1500	8
10	1519	14

(a) Plot the data, using the $\bar{X}$- and R-charts for the preliminary data of Table 17.1.

(b) Does the process appear to be in control? If not, explain why.

17.6 Consider an in-control process with mean $\mu = 25$ and $\sigma = 1.0$. Suppose that subgroups of size 5 are used with control limits $\mu \pm 3\sigma/\sqrt{n}$, and centerline at μ. Suppose that a shift occurs in the mean, and the new mean is $\mu = 26.5$.

(a) What is the average number of samples required (following the shift) to detect the out-of-control situation?

(b) What is the standard deviation of the number of runs required?

17.7 Consider the situation of Example 17.2. The following data are taken on additional samples of size 5. Plot the $\bar{X}$- and S-values on the $\bar{X}$- and S-charts that were produced with the data in the preliminary sample. Does the process appear to be in control? Explain why or why not.

Sample	$\bar{X}$	S_i
1	62.280	0.062
2	62.319	0.049
3	62.297	0.077
4	62.318	0.042
5	62.315	0.038
6	62.389	0.052
7	62.401	0.059
8	62.315	0.042
9	62.298	0.036
10	62.337	0.068

17.8 Samples of size 50 are taken every hour from a process producing a certain type of item that is considered either defective or not defective. Twenty samples are taken.

(a) Construct a control chart for control of proportion defective.

(b) Does the process appear to be in control? Explain.

Sample	**Number of Defective Items**	**Sample**	**Number of Defective Items**
1	4	11	2
2	3	12	4
3	5	13	1
4	3	14	2
5	2	15	3
6	2	16	1
7	2	17	1
8	1	18	2
9	4	19	3
10	3	20	1

17.9 For the situation of Review Exercise 17.8, suppose that additional data are collected as follows:

Sample	**Number of Defective Items**
1	3
2	4
3	2
4	2
5	3
6	1
7	3
8	5
9	7
10	7

Does the process appear to be in control? Explain.

17.10 A quality control effort is being undertaken for a process where large steel plates are manufactured and surface defects are of concern. The goal is to set up a quality control chart for the number of defects per plate. The data are given below. Set up the appropriate control chart, using this sample information. Does the process appear to be in control?

Sample	**Number of Defects**	**Sample**	**Number of Defects**
1	4	11	1
2	2	12	2
3	1	13	2
4	3	14	3
5	0	15	1
6	4	16	4
7	5	17	3
8	3	18	2
9	2	19	1
10	2	20	3

This page intentionally left blank

Chapter 18

Bayesian Statistics

18.1 Bayesian Concepts

The classical methods of estimation that we have studied in this text are based solely on information provided by the random sample. These methods essentially interpret probabilities as relative frequencies. For example, in arriving at a 95% confidence interval for μ, we interpret the statement

$$P(-1.96 < Z < 1.96) = 0.95$$

to mean that 95% of the time in repeated experiments Z will fall between -1.96 and 1.96. Since

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

for a normal sample with known variance, the probability statement here means that 95% of the random intervals $(\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n})$ contain the true mean μ. Another approach to statistical methods of estimation is called **Bayesian methodology**. The main idea of the method comes from Bayes' rule, described in Section 2.7. The key difference between the Bayesian approach and the classical or frequentist approach is that in Bayesian concepts, the parameters are viewed as random variables.

Subjective Probability

Subjective probability is the foundation of Bayesian concepts. In Chapter 2, we discussed two possible approaches to probability, namely the relative frequency and the indifference approaches. The first one determines a probability as a consequence of repeated experiments. For instance, to decide the free-throw percentage of a basketball player, we can record the number of shots made and the total number of attempts this player has made. The probability of hitting a free-throw for this player can be calculated as the ratio of these two numbers. On the other hand, if we have no knowledge of any bias in a die, the probability that a 3 will appear in the next throw will be 1/6. Such an approach to probability interpretation is based on the indifference rule.

However, in many situations, the preceding probability interpretations cannot be applied. For instance, consider the questions “What is the probability that it will rain tomorrow?” “How likely is it that this stock will go up by the end of the month?” and “What is the likelihood that two companies will be merged together?” They can hardly be interpreted by the aforementioned approaches, and the answers to these questions may be different for different people. Yet these questions are constantly asked in daily life, and the approach used to explain these probabilities is called *subjective probability*, which reflects one’s subjective opinion.

Conditional Perspective

Recall that in Chapters 9 through 17, all statistical inferences were based on the fact that the parameters are unknown but fixed quantities, apart from those in Section 9.14, in which the parameters were treated as variables and the maximum likelihood estimates (MLEs) were calculated conditioning on the observed sample data. In Bayesian statistics, not only are the parameters treated as variables as in MLE calculation, but also they are treated as random.

Because the observed data are the only experimental results for the practitioner, statistical inference is based on the actual observed data from a given experiment. Such a view is called a *conditional perspective.* Furthermore, in Bayesian concepts, since the parameters are treated as random, a probability distribution can be specified, generally by using the *subjective probability* for the parameter. Such a distribution is called a *prior distribution* and it usually reflects the experimenter’s prior belief about the parameter. In the Bayesian perspective, once an experiment is conducted and data are observed, all knowledge about the parameter is contained in the actual observed data and in the prior information.

Bayesian Applications

Although Bayes’ rule is credited to Thomas Bayes, Bayesian applications were first introduced by French scientist Pierre Simon Laplace, who published a paper on using Bayesian inference on the unknown binomial proportions (for binomial distribution, see Section 5.2).

Since the introduction of the Markov chain Monte Carlo (MCMC) computational tools for Bayesian analysis in the early 1990s, Bayesian statistics has become more and more popular in statistical modeling and data analysis. Meanwhile, methodology developments using Bayesian concepts have progressed dramatically, and they are applied in fields such as bioinformatics, biology, business, engineering, environmental and ecology science, life science and health, medicine, and many others.

18.2 Bayesian Inferences

Consider the problem of finding a point estimate of the parameter θ for the population with distribution $f(x|\ \theta)$, given θ. Denote by $\pi(\theta)$ the prior distribution of θ. Suppose that a random sample of size n, denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, is observed.

Definition 18.1: The distribution of θ, given $\mathbf{x}$, which is called the posterior distribution, is given by

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{g(\mathbf{x})},$$

where $g(\mathbf{x})$ is the marginal distribution of $\mathbf{x}$.

The marginal distribution of $\mathbf{x}$ in the above definition can be calculated using the following formula:

$$g(\mathbf{x}) = \begin{cases} \sum\limits_{\theta} f(\mathbf{x}|\theta)\pi(\theta), & \theta \text{ is discrete,} \\ \int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)\ d\theta, & \theta \text{ is continuous.} \end{cases}$$

Example 18.1: Assume that the prior distribution for the proportion of defectives produced by a machine is

p	0.1	0.2
$\pi(p)$	0.6	0.4

Denote by x the number of defectives among a random sample of size 2. Find the posterior probability distribution of p, given that x is observed.

***Solution*:** The random variable X follows a binomial distribution

$$f(x|p) = b(x; 2, p) = \binom{2}{x} p^x q^{2-x}, \quad x = 0, 1, 2.$$

The marginal distribution of x can be calculated as

$$\begin{aligned} g(x) &= f(x|0.1)\pi(0.1) + f(x|0.2)\pi(0.2) \\ &= \binom{2}{x}[(0.1)^x(0.9)^{2-x}(0.6) + (0.2)^x(0.8)^{2-x}(0.4)]. \end{aligned}$$

Hence, for $x = 0, 1, 2$, we obtain the marginal probabilities as

x	0	1	2
$g(x)$	0.742	0.236	0.022

The posterior probability of $p = 0.1$, given x, is

$$\pi(0.1|x) = \frac{f(x|0.1)\pi(0.1)}{g(x)} = \frac{(0.1)^x(0.9)^{2-x}(0.6)}{(0.1)^x(0.9)^{2-x}(0.6) + (0.2)^x(0.8)^{2-x}(0.4)},$$

and $\pi(0.2|x) = 1 - \pi(0.1|x)$.

Suppose that $x = 0$ is observed.

$$\pi(0.1|0) = \frac{f(0 \mid 0.1)\pi(0.1)}{g(0)} = \frac{(0.1)^0(0.9)^{2-0}(0.6)}{0.742} = 0.6550,$$

and $\pi(0.2|0) = 0.3450$. If $x = 1$ is observed, $\pi(0.1|1) = 0.4576$, and $\pi(0.2|1) = 0.5424$. Finally, $\pi(0.1|2) = 0.2727$, and $\pi(0.2|2) = 0.7273$. ▟

The prior distribution for Example 18.1 is discrete, although the natural range of p is from 0 to 1. Consider the following example, where we have a prior distribution covering the whole space for p.

Example 18.2: Suppose that the prior distribution of p is uniform (i.e., $\pi(p) = 1$, for $0 < p < 1$). Use the same random variable X as in Example 18.1 to find the posterior distribution of p.

***Solution*:** As in Example 18.1, we have

$$f(x|p) = b(x; 2, p) = \binom{2}{x} p^x q^{2-x}, \quad x = 0, 1, 2.$$

The marginal distribution of x can be calculated as

$$g(x) = \int_0^1 f(x|p)\pi(p)\ dp = \binom{2}{x} \int_0^1 p^x (1-p)^{2-x}\ dp.$$

The integral above can be evaluated at each x directly as $g(0) = 1/3$, $g(1) = 1/3$, and $g(2) = 1/3$. Therefore, the posterior distribution of p, given x, is

$$\pi(p|x) = \frac{\binom{2}{x} p^x (1-p)^{2-x}}{1/3} = 3\binom{2}{x} p^x (1-p)^{2-x}, \quad 0 < p < 1.$$

The posterior distribution above is actually a beta distribution (see Section 6.8) with parameters $\alpha = x + 1$ and $\beta = 3 - x$. So, if $x = 0$ is observed, the posterior distribution of p is a beta distribution with parameters $(1, 3)$. The posterior mean is $\mu = \frac{1}{1+3} = \frac{1}{4}$ and the posterior variance is $\sigma^2 = \frac{(1)(3)}{(1+3)^2(1+3+1)} = \frac{3}{80}$. ■

Using the posterior distribution, we can estimate the parameter(s) in a population in a straightforward fashion. In computing posterior distributions, it is very helpful if one is familiar with the distributions in Chapters 5 and 6. Note that in Definition 18.1, the *variable* in the posterior distribution is θ, while $\mathbf{x}$ is given. Thus, we can treat $g(\mathbf{x})$ as a constant as we calculate the posterior distribution of θ. Then the posterior distribution can be expressed as

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta),$$

where the symbol "$\propto$" stands for *is proportional to*. In the calculation of the posterior distribution above, we can leave the factors that do not depend on θ out of the normalization constant, i.e., the marginal density $g(\mathbf{x})$.

Example 18.3: Suppose that random variables $X_1, \ldots, X_n$ are independent and from a Poisson distribution with mean λ. Assume that the prior distribution of λ is exponential with mean 1. Find the posterior distribution of λ when $\bar{x} = 3$ with $n = 10$.

***Solution*:** The density function of $\mathbf{X} = (X_1, \ldots, X_n)$ is

$$f(\mathbf{x}|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!},$$

and the prior distribution is

$$\pi(\theta) = e^{-\lambda}, \text{ for } \lambda > 0.$$

Hence, using Definition 18.1 we obtain the posterior distribution of λ as

$$\pi(\lambda|\mathbf{x}) \propto f(\mathbf{x}|\lambda)\pi(\lambda) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-\lambda} \propto e^{-(n+1)\lambda}\lambda^{\sum_{i=1}^{n} x_i}.$$

Referring to the gamma distribution in Section 6.6, we conclude that the posterior distribution of λ follows a gamma distribution with parameters $1 + \sum_{i=1}^{n} x_i$ and $\frac{1}{n+1}$. Hence, we have the posterior mean and variance of λ as $\frac{\sum_{i=1}^n x_i + 1}{n+1}$ and $\frac{\sum_{i=1}^n x_i + 1}{(n+1)^2}$. So, when $\bar{x} = 3$ with $n = 10$, we have $\sum_{i=1}^{10} x_i = 30$. Hence, the posterior distribution of λ is a gamma distribution with parameters 31 and 1/11. ▟

From Example 18.3 we observe that sometimes it is quite convenient to use the "proportional to" technique in calculating the posterior distribution, especially when the result can be formed to a commonly used distribution as described in Chapters 5 and 6.

Point Estimation Using the Posterior Distribution

Once the posterior distribution is derived, we can easily use the summary of the posterior distribution to make inferences on the population parameters. For instance, the posterior mean, median, and mode can all be used to estimate the parameter.

Example 18.4: Suppose that $x = 1$ is observed for Example 18.2. Find the posterior mean and the posterior mode.

***Solution*:** When $x = 1$, the posterior distribution of p can be expressed as

$$\pi(p|1) = 6p(1-p), \qquad \text{for} \quad 0 < p < 1.$$

To calculate the mean of this distribution, we need to find

$$\int_0^1 6p^2(1-p)\,dp = 6\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{1}{2}.$$

To find the posterior mode, we need to obtain the value of p such that the posterior distribution is maximized. Taking derivative of $\pi(p)$ with respect to p, we obtain $6 - 12p$. Solving for p in $6 - 12p = 0$, we obtain $p = 1/2$. The second derivative is -12, which implies that the posterior mode is achieved at $p = 1/2$. ▟

Bayesian methods of estimation concerning the mean μ of a normal population are based on the following example.

Example 18.5: If $\bar{x}$ is the mean of a random sample of size n from a normal population with known variance σ^2, and the prior distribution of the population mean is a normal distribution with known mean μ_0 and known variance σ_0^2, then show that the posterior distribution of the population mean is also a normal distribution with mean μ^* and standard deviation σ^*, where

$$\mu^* = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\bar{x} + \frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\mu_0 \qquad \text{and} \qquad \sigma^* = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}.$$

***Solution*:** The density function of our sample is

$$f(x_1, x_2, \ldots, x_n \mid \mu) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right],$$

for $-\infty < x_i < \infty$ and $i = 1, 2, \ldots, n$, and the prior is

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right], \qquad -\infty < \mu < \infty.$$

Then the posterior distribution of μ is

$$\begin{aligned}\pi(\mu|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 + \left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{n(\bar{x} - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right]\right\},\end{aligned}$$

due to

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

from Section 8.5. Completing the squares for μ yields the posterior distribution

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu^*}{\sigma^*}\right)^2\right],$$

where

$$\mu^* = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \qquad \sigma^* = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}.$$

This is a normal distribution with mean μ^* and standard deviation σ^*. ▟

The Central Limit Theorem allows us to use Example 18.5 also when we select sufficiently large random samples ($n \geq 30$ for many engineering experimental cases) from nonnormal populations (the distribution is not very far from symmetric), and when the prior distribution of the mean is approximately normal.

Several comments need to be made about Example 18.5. The posterior mean μ^* can also be written as

$$\mu^* = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\bar{x} + \frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\mu_0,$$

which is a weighted average of the sample mean $\bar{x}$ and the prior mean μ_0. Since both coefficients are between 0 and 1 and they sum to 1, the posterior mean μ^* is always between $\bar{x}$ and μ_0. This means that the posterior estimation of μ is influenced by both $\bar{x}$ and μ_0. Furthermore, the weight of $\bar{x}$ depends on the prior variance as well as the variance of the sample mean. For a large sample problem ($n \to \infty$), the posterior mean $\mu^* \to \bar{x}$. This means that the prior mean does not play any role in estimating the population mean μ using the posterior distribution. This is very reasonable since it indicates that when the amount of data is substantial, information from the data will dominate the information on μ provided by the prior. On the other hand, when the prior variance is large ($\sigma_0^2 \to \infty$), the posterior mean μ^* also goes to $\bar{x}$. Note that for a normal distribution, the larger the variance, the flatter the density function. The flatness of the normal distribution in this case means that there is almost no subjective prior information available on the parameter μ before the data are collected. Thus, it is reasonable that the posterior estimation μ^* only depends on the data value $\bar{x}$.

Now consider the posterior standard deviation σ^*. This value can also be written as

$$\sigma^* = \sqrt{\frac{\sigma_0^2\sigma^2/n}{\sigma_0^2 + \sigma^2/n}}.$$

It is obvious that the value σ^* is smaller than both σ_0 and $\sigma/\sqrt{n}$, the prior standard deviation and the standard deviation of $\bar{x}$, respectively. This suggests that the posterior estimation is more accurate than both the prior and the sample data. Hence, incorporating both the data and prior information results in better posterior information than using any of the data or prior alone. This is a common phenomenon in Bayesian inference. Furthermore, to compute μ^* and σ^* by the formulas in Example 18.5, we have assumed that σ^2 is known. Since this is generally not the case, we shall replace σ^2 by the sample variance s^2 whenever $n \geq 30$.

Bayesian Interval Estimation

Similar to the classical confidence interval, in Bayesian analysis we can calculate a $100(1-\alpha)\%$ Bayesian interval using the posterior distribution.

Definition 18.2: The interval $a < \theta < b$ will be called a $100(1-\alpha)\%$ **Bayesian interval** for θ if

$$\int_{-\infty}^{a} \pi(\theta|x)\ d\theta = \int_{b}^{\infty} \pi(\theta|x)\ d\theta = \frac{\alpha}{2}.$$

Recall that under the frequentist approach, the probability of a confidence interval, say 95%, is interpreted as a coverage probability, which means that if an experiment is repeated again and again (with considerable unobserved data), the probability that the intervals calculated according to the rule will cover the true parameter is 95%. However, in Bayesian interval interpretation, say for a 95% interval, we can state that the probability of the unknown parameter falling into the calculated interval (which only depends on the observed data) is 95%.

Example 18.6: Supposing that $X \sim b(x; n, p)$, with known $n = 2$, and the prior distribution of p is uniform $\pi(p) = 1$, for $0 < p < 1$, find a 95% Bayesian interval for p.

Solution: As in Example 18.2, when $x = 0$, the posterior distribution is a beta distribution with parameters 1 and 3, i.e., $\pi(p|0) = 3(1-p)^2$, for $0 < p < 1$. Thus, we need to solve for a and b using Definition 18.2, which yields the following:

$$0.025 = \int_0^a 3(1-p)^2 \, dp = 1 - (1-a)^3$$

and

$$0.025 = \int_b^1 3(1-p)^2 \, dp = (1-b)^3.$$

The solutions to the above equations result in $a = 0.0084$ and $b = 0.7076$. Therefore, the probability that p falls into $(0.0084, 0.7076)$ is 95%. ▟

For the normal population and normal prior case described in Example 18.5, the posterior mean μ^* is the Bayes estimate of the population mean μ, and a $100(1-\alpha)\%$ **Bayesian interval** for μ can be constructed by computing the interval

$$\mu^* - z_{\alpha/2}\sigma^* < \mu < \mu^* + z_{\alpha/2}\sigma^*,$$

which is centered at the posterior mean and contains $100(1-\alpha)\%$ of the posterior probability.

Example 18.7: An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 100 hours. Prior experience leads us to believe that μ is a value of a normal random variable with a mean $\mu_0 = 800$ hours and a standard deviation $\sigma_0 = 10$ hours. If a random sample of 25 bulbs has an average life of 780 hours, find a 95% Bayesian interval for μ.

Solution: According to Example 18.5, the posterior distribution of the mean is also a normal distribution with mean

$$\mu^* = \frac{(25)(780)(10)^2 + (800)(100)^2}{(25)(10)^2 + (100)^2} = 796$$

and standard deviation

$$\sigma^* = \sqrt{\frac{(10)^2(100)^2}{(25)(10)^2 + (100)^2}} = \sqrt{80}.$$

The 95% Bayesian interval for μ is then given by

$$796 - 1.96\sqrt{80} < \mu < 796 + 1.96\sqrt{80},$$

or

$$778.5 < \mu < 813.5.$$

Hence, we are 95% sure that μ will be between 778.5 and 813.5.

On the other hand, ignoring the prior information about μ, we could proceed as in Section 9.4 and construct the classical 95% confidence interval

$$780 - (1.96)\left(\frac{100}{\sqrt{25}}\right) < \mu < 780 + (1.96)\left(\frac{100}{\sqrt{25}}\right),$$

or $740.8 < \mu < 819.2$, which is seen to be wider than the corresponding Bayesian interval. ▟

18.3 Bayes Estimates Using Decision Theory Framework

Using Bayesian methodology, the posterior distribution of a parameter can be obtained. Bayes estimates can also be derived using the posterior distribution and a loss function when a loss is incurred. A loss function is a function that describes the cost of a decision associated with an event of interest. Here we only list a few commonly used loss functions and their associated Bayes estimates.

Squared-Error Loss

Definition 18.3: The **squared-error loss function** is

$$L(\theta, a) = (\theta - a)^2,$$

where θ is the parameter (or state of nature) and a an action (or estimate).

A Bayes estimate minimizes the posterior expected loss, given on the observed sample data.

Theorem 18.1: The mean of the posterior distribution $\pi(\theta|x)$, denoted by θ^*, is the **Bayes estimate of $\boldsymbol{\theta}$** under the squared-error loss function.

Example 18.8: Find the Bayes estimates of p, for all the values of x, for Example 18.1 when the squared-error loss function is used.

***Solution*:** When $x = 0$, $p^* = (0.1)(0.6550) + (0.2)(0.3450) = 0.1345$.
When $x = 1$, $p^* = (0.1)(0.4576) + (0.2)(0.5424) = 0.1542$.
When $x = 2$, $p^* = (0.1)(0.2727) + (0.2)(0.7273) = 0.1727$.

Note that the classical estimate of p is $\hat{p} = x/n = 0$, $1/2$, and 1, respectively, for the x values at 0, 1, and 2. These classical estimates are very different from the corresponding Bayes estimates. ▟

Example 18.9: Repeat Example 18.8 in the situation of Example 18.2.

***Solution*:** Since the posterior distribution of p is a $B(x + 1, 3 - x)$ distribution (see Section 6.8 on page 201), the Bayes estimate of p is

$$p^* = E^{\pi(p|x)}(p) = 3\binom{2}{x}\int_0^1 p^{x+1}(1-p)^{2-x}\,dp,$$

which yields $p^* = 1/4$ for $x = 0$, $p^* = 1/2$ for $x = 1$, and $p^* = 3/4$ for $x = 2$, respectively. Notice that when $x = 1$ is observed, the Bayes estimate and the classical estimate $\hat{p}$ are equivalent. ▟

For the normal situation as described in Example 18.5, the Bayes estimate of μ under the squared-error loss will be the posterior mean μ^*.

Example 18.10: Suppose that the sampling distribution of a random variable, X, is Poisson with parameter λ. Assume that the prior distribution of λ follows a gamma distribution with parameters (α, β). Find the Bayes estimate of λ under the squared-error loss function.

***Solution*:** Using Example 18.3, we conclude that the posterior distribution of λ follows a gamma distribution with parameters $(x+\alpha, (1+1/\beta)^{-1})$. Using Theorem 6.4, we obtain the posterior mean

$$\hat{\lambda} = \frac{x+\alpha}{1+1/\beta}.$$

Since the posterior mean is the Bayes estimate under the squared-error loss, $\hat{\lambda}$ is our Bayes estimate.

Absolute-Error Loss

The squared-error loss described above is similar to the least-squares concept we discussed in connection with regression in Chapters 11 and 12. In this section, we introduce another loss function as follows.

Definition 18.4: The **absolute-error loss function** is defined as

$$L(\theta, a) = |\theta - a|,$$

where θ is the parameter and a an action.

Theorem 18.2: The median of the posterior distribution $\pi(\theta|x)$, denoted by θ^*, is the **Bayes estimate of** θ under the absolute-error loss function.

Example 18.11: Under the absolute-error loss, find the Bayes estimator for Example 18.9 when $x = 1$ is observed.

***Solution*:** Again, the posterior distribution of p is a $B(x+1, 3-x)$. When $x = 1$, it is a beta distribution with density $\pi(p \mid x = 1) = 6x(1-x)$ for $0 < x < 1$ and 0 otherwise. The median of this distribution is the value of p^* such that

$$\frac{1}{2} = \int_0^{p^*} 6p(1-p)\, dp = 3p^{*2} - 2p^{*3},$$

which yields the answer $p^* = \frac{1}{2}$. Hence, the Bayes estimate in this case is 0.5.

Exercises

18.1 Estimate the proportion of defectives being produced by the machine in Example 18.1 if the random sample of size 2 yields 2 defectives.

18.2 Let us assume that the prior distribution for the proportion p of drinks from a vending machine that overflow is

p	0.05	0.10	0.15
$\pi(p)$	0.3	0.5	0.2

If 2 of the next 9 drinks from this machine overflow, find

(a) the posterior distribution for the proportion p;

(b) the Bayes estimate of p.

18.3 Repeat Exercise 18.2 when 1 of the next 4 drinks overflows and the uniform prior distribution is

$$\pi(p) = 10, \quad 0.05 < p < 0.15.$$

18.4 Service calls come to a maintenance center according to a Poisson process with λ calls per minute. A data set of 20 one-minute periods yields an average of 1.8 calls. If the prior for λ follows an exponential distribution with mean 2, determine the posterior distribution of λ.

18.5 A previous study indicates that the percentage of chain smokers, p, who have lung cancer follows a beta distribution (see Section 6.8) with mean 70% and standard deviation 10%. Suppose a new data set collected shows that 81 out of 120 chain smokers have lung cancer.

(a) Determine the posterior distribution of the percentage of chain smokers who have lung cancer by combining the new data and the prior information.

(b) What is the posterior probability that p is larger than 50%?

18.6 The developer of a new condominium complex claims that 3 out of 5 buyers will prefer a two-bedroom unit, while his banker claims that it would be more correct to say that 7 out of 10 buyers will prefer a two-bedroom unit. In previous predictions of this type, the banker has been twice as reliable as the developer. If 12 of the next 15 condominiums sold in this complex are two-bedroom units, find

(a) the posterior probabilities associated with the claims of the developer and banker;

(b) a point estimate of the proportion of buyers who prefer a two-bedroom unit.

18.7 The burn time for the first stage of a rocket is a normal random variable with a standard deviation of 0.8 minute. Assume a normal prior distribution for μ with a mean of 8 minutes and a standard deviation of 0.2 minute. If 10 of these rockets are fired and the first stage has an average burn time of 9 minutes, find a 95% Bayesian interval for μ.

18.8 The daily profit from a juice vending machine placed in an office building is a value of a normal random variable with unknown mean μ and variance σ^2. Of course, the mean will vary somewhat from building to building, and the distributor feels that these average daily profits can best be described by a normal distribution with mean $\mu_0 = \$30.00$ and standard deviation $\sigma_0 = \$1.75$. If one of these juice machines, placed in a certain building, showed an average daily profit of $\bar{x} = \$24.90$ during the first 30 days with a standard deviation of $s = \$2.10$, find

(a) a Bayes estimate of the true average daily profit for this building;

(b) a 95% Bayesian interval of μ for this building;

(c) the probability that the average daily profit from the machine in this building is between \$24.00 and \$26.00.

18.9 The mathematics department of a large university is designing a placement test to be given to incoming freshman classes. Members of the department feel that the average grade for this test will vary from one freshman class to another. This variation of the average class grade is expressed subjectively by a normal distribution with mean $\mu_0 = 72$ and variance $\sigma_0^2 = 5.76$.

(a) What prior probability does the department assign to the actual average grade being somewhere between 71.8 and 73.4 for next year's freshman class?

(b) If the test is tried on a random sample of 100 students from the next incoming freshman class, resulting in an average grade of 70 with a variance of 64, construct a 95% Bayesian interval for μ.

(c) What posterior probability should the department assign to the event of part (a)?

18.10 Suppose that in Example 18.7 the electrical firm does not have enough prior information regarding the population mean length of life to be able to assume a normal distribution for μ. The firm believes, however, that μ is surely between 770 and 830 hours, and it is thought that a more realistic Bayesian approach would be to assume the prior distribution

$$\pi(\mu) = \frac{1}{60}, \quad 770 < \mu < 830.$$

If a random sample of 25 bulbs gives an average life of 780 hours, follow the steps of the proof for Example 18.5 to find the posterior distribution

$$\pi(\mu \mid x_1, x_2, \ldots, x_{25}).$$

18.11 Suppose that the time to failure T of a certain hinge is an exponential random variable with probability density

$$f(t) = \theta e^{-\theta t}, \quad t > 0.$$

From prior experience we are led to believe that θ is a value of an exponential random variable with probability density

$$\pi(\theta) = 2e^{-2\theta}, \quad \theta > 0.$$

If we have a sample of n observations on T, show that the posterior distribution of Θ is a gamma distribution with parameters

$$\alpha = n + 1 \quad \text{and} \quad \beta = \left(\sum_{i=1}^{n} t_i + 2 \right)^{-1}.$$

18.12 Suppose that a sample consisting of 5, 6, 6, 7, 5, 6, 4, 9, 3, and 6 comes from a Poisson population with mean λ. Assume that the parameter λ follows a gamma distribution with parameters $(3, 2)$. Under the squared-error loss function, find the Bayes estimate of λ.

18.13 A random variable X follows a negative binomial distribution with parameters $k = 5$ and p [i.e., $b^*(x; 5, p)$]. Furthermore, we know that p follows a uniform distribution on the interval $(0, 1)$. Find the Bayes estimate of p under the squared-error loss function.

18.14 A random variable X follows an exponential distribution with mean $1/\beta$. Assume the prior distribution of β is another exponential distribution with mean 2.5. Determine the Bayes estimate of β under the absolute-error loss function.

18.15 A random sample $X_1, \ldots, X_n$ comes from a uniform distribution (see Section 6.1) population $U(0, \theta)$ with unknown θ. The data are given below:

$$0.13, 1.06, 1.65, 1.73, 0.95, 0.56, 2.14, 0.33, 1.22, 0.20,$$
$$1.55, 1.18, 0.71, 0.01, 0.42, 1.03, 0.43, 1.02, 0.83, 0.88$$

Suppose the prior distribution of θ has the density

$$\pi(\theta) = \begin{cases} \frac{1}{\theta^2}, & \theta > 1, \\ 0, & \theta \le 1. \end{cases}$$

Determine the Bayes estimator under the absolute-error loss function.

Bibliography

[1] Bartlett, M. S., and Kendall, D. G. (1946). "The Statistical Analysis of Variance Heterogeneity and Logarithmic Transformation," *Journal of the Royal Statistical Society*, Ser. B. **8**, 128–138.

[2] Bowker, A. H., and Lieberman, G. J. (1972). *Engineering Statistics*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall.

[3] Box, G. E. P. (1988). "Signal to Noise Ratios, Performance Criteria and Transformations (with discussion)," *Technometrics*, **30**, 1–17.

[4] Box, G. E. P., and Fung, C. A. (1986). "Studies in Quality Improvement: Minimizing Transmitted Variation by Parameter Design," Report 8. University of Wisconsin-Madison, Center for Quality and Productivity Improvement.

[5] Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: John Wiley & Sons.

[6] Brownlee, K. A. (1984). *Statistical Theory and Methodology: In Science and Engineering*, 2nd ed. New York: John Wiley & Sons.

[7] Carroll, R. J., and Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall.

[8] Chatterjee, S., Hadi, A. S., and Price, B. (1999). *Regression Analysis by Example*, 3rd ed. New York: John Wiley & Sons.

[9] Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.

[10] Daniel, C. and Wood, F. S. (1999). *Fitting Equations to Data: Computer Analysis of Multifactor Data*, 2nd ed. New York: John Wiley & Sons.

[11] Daniel, W. W. (1989). *Applied Nonparametric Statistics*, 2nd ed. Belmont, Calif.: Wadsworth Publishing Company.

[12] Devore, J. L. (2003). *Probability and Statistics for Engineering and the Sciences*, 6th ed. Belmont, Calif: Duxbury Press.

[13] Dixon, W. J. (1983). *Introduction to Statistical Analysis*, 4th ed. New York: McGraw-Hill.

[14] Draper, N. R., and Smith, H. (1998). *Applied Regression Analysis*, 3rd ed. New York: John Wiley & Sons.

[15] Duncan, A. (1986). *Quality Control and Industrial Statistics*, 5th ed. Homewood, Ill.: Irwin.

[16] Dyer, D. D., and Keating, J. P. (1980). "On the Determination of Critical Values for Bartlett's Test," *Journal of the American Statistical Association*, **75**, 313–319.

[17] Ewan, W. D., and Kemp, K. W. (1960). "Sampling Inspection of Continuous Processes with No Autocorrelation between Successive Results," *Biometrika*, **47**, 363–380.

[18] Geary, R. C. (1947). "Testing for Normality," *Biometrika*, **34**, 209–242.

[19] Gunst, R. F., and Mason, R. L. (1980). *Regression Analysis and Its Application: A Data-Oriented Approach*. New York: Marcel Dekker.

[20] Guttman, I., Wilks, S. S., and Hunter, J. S. (1971). *Introductory Engineering Statistics*. New York: John Wiley & Sons.

[21] Harville, D. A. (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, **72**, 320--338.

[22] Hicks, C. R., and Turner, K. V. (1999). *Fundamental Concepts in the Design of Experiments*, 5th ed. Oxford: Oxford University Press.

[23] Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1991). *Fundamentals of Exploratory Analysis of Variance*. New York: John Wiley & Sons.

[24] Hocking, R. R. (1976). "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, **32**, 1–49.

[25] Hodges, J. L., and Lehmann, E. L. (2005). *Basic Concepts of Probability and Statistics*, 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics.

[26] Hoerl, A. E., and Wennard, R. W. (1970). "Ridge Regression: Applications to Nonorthogonal Problems," *Technometrics*, **12**, 55–67.

[27] Hogg, R. V., and Ledolter, J. (1992). *Applied Statistics for Engineers and Physical Scientists*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall.

[28] Hogg, R. V., McKean, J. W., and Craig, A. (2005). *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N.J.: Prentice Hall.

[29] Hollander, M., and Wolfe, D. (1999). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.

[30] Johnson, N. L., and Leone, F. C. (1977). *Statistics and Experimental Design: In Engineering and the Physical Sciences*, 2nd ed. Vols. I and II, New York: John Wiley & Sons.

[31] Kackar, R. (1985). "Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology*, **17**, 176–188.

[32] Koopmans, L. H. (1987). *An Introduction to Contemporary Statistics*, 2nd ed. Boston: Duxbury Press.

[33] Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Regression Models*, 5th ed. New York: McGraw-Hill/Irwin.

[34] Larsen, R. J., and Morris, M. L. (2000). *An Introduction to Mathematical Statistics and Its Applications*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall.

[35] Lehmann, E. L., and D'Abrera, H. J. M. (1998). *Nonparametrics: Statistical Methods Based on Ranks*, rev. ed. Upper Saddle River, N.J.: Prentice Hall.

[36] Lentner, M., and Bishop, T. (1986). *Design and Analysis of Experiments*, 2nd ed. Blacksburg, Va.: Valley Book Co.

[37] Mallows, C. L. (1973). "Some Comments on C_p," *Technometrics*, **15**, 661–675.

[38] McClave, J. T., Dietrich, F. H., and Sincich, T. (1997). *Statistics*, 7th ed. Upper Saddle River, N.J.: Prentice Hall.

[39] Montgomery, D. C. (2008a). *Design and Analysis of Experiments*, 7th ed. New York: John Wiley & Sons.

[40] Montgomery, D. C. (2008b). *Introduction to Statistical Quality Control*, 6th ed. New York: John Wiley & Sons.

[41] Mosteller, F., and Tukey, J. (1977). *Data Analysis and Regression.* Reading, Mass.: Addison-Wesley Publishing Co.

[42] Myers, R. H. (1990). *Classical and Modern Regression with Applications*, 2nd ed. Boston: Duxbury Press.

[43] Myers, R. H., Khuri, A. I., and Vining, G. G. (1992). "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, **46**, 131–139.

[44] Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed. New York: John Wiley & Sons.

[45] Myers, R. H., Montgomery, D. C., Vining, G. G., and Robinson, T. J. (2008). *Generalized Linear Models with Applications in Engineering and the Sciences*, 2nd ed., New York: John Wiley & Sons.

[46] Noether, G. E. (1976). *Introduction to Statistics: A Nonparametric Approach*, 2nd ed. Boston: Houghton Mifflin Company.

[47] Olkin, I., Gleser, L. J., and Derman, C. (1994). *Probability Models and Applications*, 2nd ed. New York: Prentice Hall.

[48] Ott, R. L., and Longnecker, M. T. (2000). *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Boston: Duxbury Press.

[49] Pacansky, J., England, C. D., and Wattman, R. (1986). "Infrared Spectroscopic Studies of Poly (perfluoropropyleneoxide) on Gold Substrate: A Classical Dispersion Analysis for the Refractive Index." *Applied Spectroscopy*, **40**, 8–16.

[50] Plackett, R. L., and Burman, J. P. (1946). "The Design of Multifactor Experiments," *Biometrika*, **33**, 305–325.

[51] Ross, S. M. (2002). *Introduction to Probability Models*, 9th ed. New York: Academic Press, Inc.

[52] Satterthwaite, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components," *Biometrics*, **2**, 110–114.

[53] Schilling, E. G., and Nelson, P. R. (1976). "The Effect of Nonnormality on the Control Limits of $\bar{X}$ Charts," *Journal of Quality Technology*, **8**, 347–373.

[54] Schmidt, S. R., and Launsby, R. G. (1991). *Understanding Industrial Designed Experiments.* Colorado Springs, Col. Air Academy Press.

[55] Shoemaker, A. C., Tsui, K.-L., and Wu, C. F. J. (1991). "Economical Experimentation Methods for Robust Parameter Design," *Technometrics*, **33**, 415–428.

[56] Snedecor, G. W., and Cochran, W. G. (1989). *Statistical Methods*, 8th ed. Allies, Iowa: The Iowa State University Press.

[57] Steel, R. G. D., Torrie, J. H., and Dickey, D. A. (1996). *Principles and Procedures of Statistics: A Biometrical Approach*, 3rd ed. New York: McGraw-Hill.

[58] Taguchi, G. (1991). *Introduction to Quality Engineering.* White Plains, N.Y.: Unipub/Kraus International.

[59] Taguchi, G., and Wu, Y. (1985). *Introduction to Off-Line Quality Control.* Nagoya, Japan: Central Japan Quality Control Association.

[60] Thompson, W. O., and Cady, F. B. (1973). *Proceedings of the University of Kentucky Conference on Regression with a Large Number of Predictor Variables.* Lexington, Ken.: University of Kentucky Press.

[61] Tukey, J. W. (1977). *Exploratory Data Analysis.* Reading, Mass.: Addison-Wesley Publishing Co.

[62] Vining, G. G., and Myers, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, **22**, 38–45.

[63] Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1990). "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, **22**, 15–22.

[64] Winer, B. J. (1991). *Statistical Principles in Experimental Design*, 3rd ed. New York: McGraw-Hill.

Appendix A
Statistical Tables and Proofs

Table A.1 Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
1	**0**	0.9000	0.8000	0.7500	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	**0**	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	**0**	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	**0**	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	**0**	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	**0**	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000
	1	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094	0.0410	0.0109	0.0016	0.0001
	2	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438	0.1792	0.0705	0.0170	0.0013
	3	0.9987	0.9830	0.9624	0.9295	0.8208	0.6563	0.4557	0.2557	0.0989	0.0159
	4	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7667	0.5798	0.3446	0.1143
	5	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9533	0.8824	0.7379	0.4686
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	**0**	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0016	0.0002	0.0000	
	1	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0188	0.0038	0.0004	0.0000
	2	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0963	0.0288	0.0047	0.0002
	3	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2898	0.1260	0.0333	0.0027
	4	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5801	0.3529	0.1480	0.0257
	5	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8414	0.6706	0.4233	0.1497
	6		1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9176	0.7903	0.5217
	7			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

						p					
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
8	**0**	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	
	1	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0085	0.0013	0.0001	
	2	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0498	0.0113	0.0012	0.0000
	3	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1737	0.0580	0.0104	0.0004
	4	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4059	0.1941	0.0563	0.0050
	5	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6846	0.4482	0.2031	0.0381
	6		0.9999	0.9996	0.9987	0.9915	0.9648	0.8936	0.7447	0.4967	0.1869
	7		1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8322	0.5695
	8				1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	**0**	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0003	0.0000		
	1	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0038	0.0004	0.0000	
	2	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0043	0.0003	0.0000
	3	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0994	0.0253	0.0031	0.0001
	4	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2666	0.0988	0.0196	0.0009
	5	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5174	0.2703	0.0856	0.0083
	6	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7682	0.5372	0.2618	0.0530
	7		1.0000	0.9999	0.9996	0.9962	0.9805	0.9295	0.8040	0.5638	0.2252
	8			1.0000	1.0000	0.9997	0.9980	0.9899	0.9596	0.8658	0.6126
	9					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	**0**	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000		
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7		0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8		1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9				1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	**0**	0.3138	0.0859	0.0422	0.0198	0.0036	0.0005	0.0000			
	1	0.6974	0.3221	0.1971	0.1130	0.0302	0.0059	0.0007	0.0000		
	2	0.9104	0.6174	0.4552	0.3127	0.1189	0.0327	0.0059	0.0006	0.0000	
	3	0.9815	0.8389	0.7133	0.5696	0.2963	0.1133	0.0293	0.0043	0.0002	
	4	0.9972	0.9496	0.8854	0.7897	0.5328	0.2744	0.0994	0.0216	0.0020	0.0000
	5	0.9997	0.9883	0.9657	0.9218	0.7535	0.5000	0.2465	0.0782	0.0117	0.0003
	6	1.0000	0.9980	0.9924	0.9784	0.9006	0.7256	0.4672	0.2103	0.0504	0.0028
	7		0.9998	0.9988	0.9957	0.9707	0.8867	0.7037	0.4304	0.1611	0.0185
	8		1.0000	0.9999	0.9994	0.9941	0.9673	0.8811	0.6873	0.3826	0.0896
	9			1.0000	1.0000	0.9993	0.9941	0.9698	0.8870	0.6779	0.3026
	10					1.0000	0.9995	0.9964	0.9802	0.9141	0.6862
	11						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
12	**0**	0.2824	0.0687	0.0317	0.0138	0.0022	0.0002	0.0000			
	1	0.6590	0.2749	0.1584	0.0850	0.0196	0.0032	0.0003	0.0000		
	2	0.8891	0.5583	0.3907	0.2528	0.0834	0.0193	0.0028	0.0002	0.0000	
	3	0.9744	0.7946	0.6488	0.4925	0.2253	0.0730	0.0153	0.0017	0.0001	
	4	0.9957	0.9274	0.8424	0.7237	0.4382	0.1938	0.0573	0.0095	0.0006	0.0000
	5	0.9995	0.9806	0.9456	0.8822	0.6652	0.3872	0.1582	0.0386	0.0039	0.0001
	6	0.9999	0.9961	0.9857	0.9614	0.8418	0.6128	0.3348	0.1178	0.0194	0.0005
	7	1.0000	0.9994	0.9972	0.9905	0.9427	0.8062	0.5618	0.2763	0.0726	0.0043
	8		0.9999	0.9996	0.9983	0.9847	0.9270	0.7747	0.5075	0.2054	0.0256
	9		1.0000	1.0000	0.9998	0.9972	0.9807	0.9166	0.7472	0.4417	0.1109
	10				1.0000	0.9997	0.9968	0.9804	0.9150	0.7251	0.3410
	11					1.0000	0.9998	0.9978	0.9862	0.9313	0.7176
	12						1.0000	1.0000	1.0000	1.0000	1.0000
13	**0**	0.2542	0.0550	0.0238	0.0097	0.0013	0.0001	0.0000			
	1	0.6213	0.2336	0.1267	0.0637	0.0126	0.0017	0.0001	0.0000		
	2	0.8661	0.5017	0.3326	0.2025	0.0579	0.0112	0.0013	0.0001		
	3	0.9658	0.7473	0.5843	0.4206	0.1686	0.0461	0.0078	0.0007	0.0000	
	4	0.9935	0.9009	0.7940	0.6543	0.3530	0.1334	0.0321	0.0040	0.0002	
	5	0.9991	0.9700	0.9198	0.8346	0.5744	0.2905	0.0977	0.0182	0.0012	0.0000
	6	0.9999	0.9930	0.9757	0.9376	0.7712	0.5000	0.2288	0.0624	0.0070	0.0001
	7	1.0000	0.9988	0.9944	0.9818	0.9023	0.7095	0.4256	0.1654	0.0300	0.0009
	8		0.9998	0.9990	0.9960	0.9679	0.8666	0.6470	0.3457	0.0991	0.0065
	9		1.0000	0.9999	0.9993	0.9922	0.9539	0.8314	0.5794	0.2527	0.0342
	10			1.0000	0.9999	0.9987	0.9888	0.9421	0.7975	0.4983	0.1339
	11				1.0000	0.9999	0.9983	0.9874	0.9363	0.7664	0.3787
	12					1.0000	0.9999	0.9987	0.9903	0.9450	0.7458
	13						1.0000	1.0000	1.0000	1.0000	1.0000
14	**0**	0.2288	0.0440	0.0178	0.0068	0.0008	0.0001	0.0000			
	1	0.5846	0.1979	0.1010	0.0475	0.0081	0.0009	0.0001			
	2	0.8416	0.4481	0.2811	0.1608	0.0398	0.0065	0.0006	0.0000		
	3	0.9559	0.6982	0.5213	0.3552	0.1243	0.0287	0.0039	0.0002		
	4	0.9908	0.8702	0.7415	0.5842	0.2793	0.0898	0.0175	0.0017	0.0000	
	5	0.9985	0.9561	0.8883	0.7805	0.4859	0.2120	0.0583	0.0083	0.0004	
	6	0.9998	0.9884	0.9617	0.9067	0.6925	0.3953	0.1501	0.0315	0.0024	0.0000
	7	1.0000	0.9976	0.9897	0.9685	0.8499	0.6047	0.3075	0.0933	0.0116	0.0002
	8		0.9996	0.9978	0.9917	0.9417	0.7880	0.5141	0.2195	0.0439	0.0015
	9		1.0000	0.9997	0.9983	0.9825	0.9102	0.7207	0.4158	0.1298	0.0092
	10			1.0000	0.9998	0.9961	0.9713	0.8757	0.6448	0.3018	0.0441
	11				1.0000	0.9994	0.9935	0.9602	0.8392	0.5519	0.1584
	12					0.9999	0.9991	0.9919	0.9525	0.8021	0.4154
	13					1.0000	0.9999	0.9992	0.9932	0.9560	0.7712
	14						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
15	**0**	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000				
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000			
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000		
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001		
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10		1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11			1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12				1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13					1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14						1.0000	0.9995	0.9953	0.9648	0.7941
	15							1.0000	1.0000	1.0000	1.0000
16	**0**	0.1853	0.0281	0.0100	0.0033	0.0003	0.0000				
	1	0.5147	0.1407	0.0635	0.0261	0.0033	0.0003	0.0000			
	2	0.7892	0.3518	0.1971	0.0994	0.0183	0.0021	0.0001			
	3	0.9316	0.5981	0.4050	0.2459	0.0651	0.0106	0.0009	0.0000		
	4	0.9830	0.7982	0.6302	0.4499	0.1666	0.0384	0.0049	0.0003		
	5	0.9967	0.9183	0.8103	0.6598	0.3288	0.1051	0.0191	0.0016	0.0000	
	6	0.9995	0.9733	0.9204	0.8247	0.5272	0.2272	0.0583	0.0071	0.0002	
	7	0.9999	0.9930	0.9729	0.9256	0.7161	0.4018	0.1423	0.0257	0.0015	0.0000
	8	1.0000	0.9985	0.9925	0.9743	0.8577	0.5982	0.2839	0.0744	0.0070	0.0001
	9		0.9998	0.9984	0.9929	0.9417	0.7728	0.4728	0.1753	0.0267	0.0005
	10		1.0000	0.9997	0.9984	0.9809	0.8949	0.6712	0.3402	0.0817	0.0033
	11			1.0000	0.9997	0.9951	0.9616	0.8334	0.5501	0.2018	0.0170
	12				1.0000	0.9991	0.9894	0.9349	0.7541	0.4019	0.0684
	13					0.9999	0.9979	0.9817	0.9006	0.6482	0.2108
	14					1.0000	0.9997	0.9967	0.9739	0.8593	0.4853
	15						1.0000	0.9997	0.9967	0.9719	0.8147
	16							1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
17	**0**	0.1668	0.0225	0.0075	0.0023	0.0002	0.0000				
	1	0.4818	0.1182	0.0501	0.0193	0.0021	0.0001	0.0000			
	2	0.7618	0.3096	0.1637	0.0774	0.0123	0.0012	0.0001			
	3	0.9174	0.5489	0.3530	0.2019	0.0464	0.0064	0.0005	0.0000		
	4	0.9779	0.7582	0.5739	0.3887	0.1260	0.0245	0.0025	0.0001		
	5	0.9953	0.8943	0.7653	0.5968	0.2639	0.0717	0.0106	0.0007	0.0000	
	6	0.9992	0.9623	0.8929	0.7752	0.4478	0.1662	0.0348	0.0032	0.0001	
	7	0.9999	0.9891	0.9598	0.8954	0.6405	0.3145	0.0919	0.0127	0.0005	
	8	1.0000	0.9974	0.9876	0.9597	0.8011	0.5000	0.1989	0.0403	0.0026	0.0000
	9		0.9995	0.9969	0.9873	0.9081	0.6855	0.3595	0.1046	0.0109	0.0001
	10		0.9999	0.9994	0.9968	0.9652	0.8338	0.5522	0.2248	0.0377	0.0008
	11		1.0000	0.9999	0.9993	0.9894	0.9283	0.7361	0.4032	0.1057	0.0047
	12			1.0000	0.9999	0.9975	0.9755	0.8740	0.6113	0.2418	0.0221
	13				1.0000	0.9995	0.9936	0.9536	0.7981	0.4511	0.0826
	14					0.9999	0.9988	0.9877	0.9226	0.6904	0.2382
	15					1.0000	0.9999	0.9979	0.9807	0.8818	0.5182
	16						1.0000	0.9998	0.9977	0.9775	0.8332
	17							1.0000	1.0000	1.0000	1.0000
18	**0**	0.1501	0.0180	0.0056	0.0016	0.0001	0.0000				
	1	0.4503	0.0991	0.0395	0.0142	0.0013	0.0001				
	2	0.7338	0.2713	0.1353	0.0600	0.0082	0.0007	0.0000			
	3	0.9018	0.5010	0.3057	0.1646	0.0328	0.0038	0.0002			
	4	0.9718	0.7164	0.5187	0.3327	0.0942	0.0154	0.0013	0.0000		
	5	0.9936	0.8671	0.7175	0.5344	0.2088	0.0481	0.0058	0.0003		
	6	0.9988	0.9487	0.8610	0.7217	0.3743	0.1189	0.0203	0.0014	0.0000	
	7	0.9998	0.9837	0.9431	0.8593	0.5634	0.2403	0.0576	0.0061	0.0002	
	8	1.0000	0.9957	0.9807	0.9404	0.7368	0.4073	0.1347	0.0210	0.0009	
	9		0.9991	0.9946	0.9790	0.8653	0.5927	0.2632	0.0596	0.0043	0.0000
	10		0.9998	0.9988	0.9939	0.9424	0.7597	0.4366	0.1407	0.0163	0.0002
	11		1.0000	0.9998	0.9986	0.9797	0.8811	0.6257	0.2783	0.0513	0.0012
	12			1.0000	0.9997	0.9942	0.9519	0.7912	0.4656	0.1329	0.0064
	13				1.0000	0.9987	0.9846	0.9058	0.6673	0.2836	0.0282
	14					0.9998	0.9962	0.9672	0.8354	0.4990	0.0982
	15					1.0000	0.9993	0.9918	0.9400	0.7287	0.2662
	16						0.9999	0.9987	0.9858	0.9009	0.5497
	17						1.0000	0.9999	0.9984	0.9820	0.8499
	18							1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	***r***	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
19	**0**	0.1351	0.0144	0.0042	0.0011	0.0001					
	1	0.4203	0.0829	0.0310	0.0104	0.0008	0.0000				
	2	0.7054	0.2369	0.1113	0.0462	0.0055	0.0004	0.0000			
	3	0.8850	0.4551	0.2631	0.1332	0.0230	0.0022	0.0001			
	4	0.9648	0.6733	0.4654	0.2822	0.0696	0.0096	0.0006	0.0000		
	5	0.9914	0.8369	0.6678	0.4739	0.1629	0.0318	0.0031	0.0001		
	6	0.9983	0.9324	0.8251	0.6655	0.3081	0.0835	0.0116	0.0006		
	7	0.9997	0.9767	0.9225	0.8180	0.4878	0.1796	0.0352	0.0028	0.0000	
	8	1.0000	0.9933	0.9713	0.9161	0.6675	0.3238	0.0885	0.0105	0.0003	
	9		0.9984	0.9911	0.9674	0.8139	0.5000	0.1861	0.0326	0.0016	
	10		0.9997	0.9977	0.9895	0.9115	0.6762	0.3325	0.0839	0.0067	0.0000
	11		1.0000	0.9995	0.9972	0.9648	0.8204	0.5122	0.1820	0.0233	0.0003
	12			0.9999	0.9994	0.9884	0.9165	0.6919	0.3345	0.0676	0.0017
	13			1.0000	0.9999	0.9969	0.9682	0.8371	0.5261	0.1631	0.0086
	14				1.0000	0.9994	0.9904	0.9304	0.7178	0.3267	0.0352
	15					0.9999	0.9978	0.9770	0.8668	0.5449	0.1150
	16					1.0000	0.9996	0.9945	0.9538	0.7631	0.2946
	17						1.0000	0.9992	0.9896	0.9171	0.5797
	18							0.9999	0.9989	0.9856	0.8649
	19							1.0000	1.0000	1.0000	1.0000
20	**0**	0.1216	0.0115	0.0032	0.0008	0.0000					
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000				
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002				
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000			
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003			
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000		
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003		
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	
	10		0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11		0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12		1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13			1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14				1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15					0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16					1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17						0.9998	0.9964	0.9645	0.7939	0.3231
	18						1.0000	0.9995	0.9924	0.9308	0.6083
	19							1.0000	0.9992	0.9885	0.8784
	20								1.0000	1.0000	1.0000

Table A.2 Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ								
r	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865
4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977
5				1.0000	1.0000	1.0000	0.9999	0.9998	0.9997
6							1.0000	1.0000	1.0000

	μ								
r	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8		1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9			1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10				0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11				1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12					1.0000	0.9999	0.9997	0.9992	0.9980
13						1.0000	0.9999	0.9997	0.9993
14							1.0000	0.9999	0.9998
15								1.0000	0.9999
16									1.0000

Table A.2 (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

					$\boldsymbol{\mu}$				
r	**5.5**	**6.0**	**6.5**	**7.0**	**7.5**	**8.0**	**8.5**	**9.0**	**9.5**
0	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885
6	0.6860	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649
7	0.8095	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687
8	0.8944	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918
9	0.9462	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218
10	0.9747	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453
11	0.9890	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520
12	0.9955	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364
13	0.9983	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981
14	0.9994	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400
15	0.9998	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665
16	0.9999	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823
17	1.0000	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911
18		1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957
19			1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980
20						0.9999	0.9998	0.9996	0.9991
21						1.0000	0.9999	0.9998	0.9996
22							1.0000	0.9999	0.9999
23								1.0000	0.9999
24									1.0000

Table A.2 (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

					μ				
r	**10.0**	**11.0**	**12.0**	**13.0**	**14.0**	**15.0**	**16.0**	**17.0**	**18.0**
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
3	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317
25		0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554
26		1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718
27			0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827
28			1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897
29				1.0000	0.9999	0.9996	0.9989	0.9973	0.9941
30					0.9999	0.9998	0.9994	0.9986	0.9967
31					1.0000	0.9999	0.9997	0.9993	0.9982
32						1.0000	0.9999	0.9996	0.9990
33							0.9999	0.9998	0.9995
34							1.0000	0.9999	0.9998
35								1.0000	0.9999
36									0.9999
37									1.0000

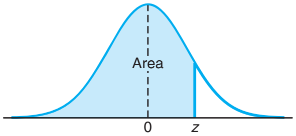


Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

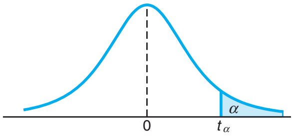


Table A.4 Critical Values of the t-Distribution

				α			
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Table A.4 (continued) Critical Values of the t-Distribution

	$\boldsymbol{\alpha}$						
v	**0.02**	**0.015**	**0.01**	**0.0075**	**0.005**	**0.0025**	**0.0005**
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290

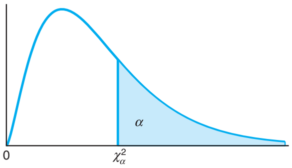


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	$0.0^{4}393$	$0.0^{3}157$	$0.0^{3}628$	$0.0^{3}982$	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872	39.335
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	49.335
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	59.335

Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

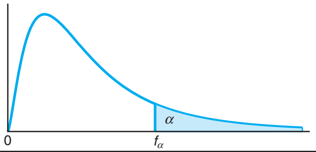


Table A.6 Critical Values of the *F*-Distribution

	$f_{0.05}(v_1, v_2)$								
	v_1								
v_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, by permission of E.S. Pearson and the Biometrika Trustees.

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.01}(v_1, v_2)$								
	v_1								
v_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.01}(v_1, v_2)$									
	v_1									
v_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	6055.85	6106.32	6157.28	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Table A.7 Tolerance Factors for Normal Distributions

	Two-Sided Intervals						**One-Sided Intervals**					
	$\gamma = 0.05$			$\gamma = 0.01$			$\gamma = 0.05$			$\gamma = 0.01$		
	$1-\alpha$			$1-\alpha$			$1-\alpha$			$1-\alpha$		
n	**0.90**	**0.95**	**0.99**	**0.90**	**0.95**	**0.99**	**0.90**	**0.95**	**0.99**	**0.90**	**0.95**	**0.99**
2	32.019	37.674	48.430	160.193	188.491	242.300	20.581	26.260	37.094	103.029	131.426	185.617
3	8.380	9.916	12.861	18.930	22.401	29.055	6.156	7.656	10.553	13.995	17.170	23.896
4	5.369	6.370	8.299	9.398	11.150	14.527	4.162	5.144	7.042	7.380	9.083	12.387
5	4.275	5.079	6.634	6.612	7.855	10.260	3.407	4.203	5.741	5.362	6.578	8.939
6	3.712	4.414	5.775	5.337	6.345	8.301	3.006	3.708	5.062	4.411	5.406	7.335
7	3.369	4.007	5.248	4.613	5.488	7.187	2.756	3.400	4.642	3.859	4.728	6.412
8	3.136	3.732	4.891	4.147	4.936	6.468	2.582	3.187	4.354	3.497	4.285	5.812
9	2.967	3.532	4.631	3.822	4.550	5.966	2.454	3.031	4.143	3.241	3.972	5.389
10	2.839	3.379	4.433	3.582	4.265	5.594	2.355	2.911	3.981	3.048	3.738	5.074
11	2.737	3.259	4.277	3.397	4.045	5.308	2.275	2.815	3.852	2.898	3.556	4.829
12	2.655	3.162	4.150	3.250	3.870	5.079	2.210	2.736	3.747	2.777	3.410	4.633
13	2.587	3.081	4.044	3.130	3.727	4.893	2.155	2.671	3.659	2.677	3.290	4.472
14	2.529	3.012	3.955	3.029	3.608	4.737	2.109	2.615	3.585	2.593	1.189	4.337
15	2.480	2.954	3.878	2.945	3.507	4.605	2.068	2.566	3.520	2.522	3.102	4.222
16	2.437	2.903	3.812	2.872	3.421	4.492	2.033	2.524	3.464	2.460	3.028	4.123
17	2.400	2.858	3.754	2.808	3.345	4.393	2.002	2.486	3.414	2.405	2.963	4.037
18	2.366	2.819	3.702	2.753	3.279	4.307	1.974	2.453	3.370	2.357	2.905	3.960
19	2.337	2.784	3.656	2.703	3.221	4.230	1.949	2.423	3.331	2.314	2.854	3.892
20	2.310	2.752	3.615	2.659	3.168	4.161	1.926	2.396	3.295	2.276	2.808	1.832
25	2.208	2.631	3.457	2.494	2.972	3.904	1.838	2.292	3.158	2.129	2.633	3.001
30	2.140	2.549	3.350	2.385	2.841	3.733	1.777	2.220	3.064	2.030	2.516	3.447
35	2.090	2.490	3.272	2.306	2.748	3.611	1.732	2.167	2.995	1.957	2.430	3.334
40	2.052	2.445	3.213	2.247	2.677	3.518	1.697	2.126	2.941	1.902	2.364	3.249
45	2.021	2.408	3.165	2.200	2.621	3.444	1.669	2.092	2.898	1.857	2.312	3.180
50	1.996	2.379	3.126	2.162	2.576	3.385	1.646	2.065	2.863	1.821	2.269	3.125
60	1.958	2.333	3.066	2.103	2.506	3.293	1.609	2.022	2.807	1.764	2.202	3.038
70	1.929	2.299	3.021	2.060	2.454	3.225	1.581	1.990	2.765	1.722	2.153	2.974
80	1.907	2.272	2.986	2.026	2.414	3.173	1.559	1.965	2.733	1.688	2.114	2.924
90	1.889	2.251	2.958	1.999	2.382	3.130	1.542	1.944	2.706	1.661	2.082	2.883
100	1.874	2.233	2.934	1.977	2.355	3.096	1.527	1.927	2.684	1.639	2.056	2.850
150	1.825	2.175	2.859	1.905	2.270	2.983	1.478	1.870	2.611	1.566	1.971	2.741
200	1.798	2.143	2.816	1.865	2.222	2.921	1.450	1.837	2.570	1.524	1.923	2.679
250	1.780	2.121	2.788	1.839	2.191	2.880	1.431	1.815	2.542	1.496	1.891	2.638
300	1.767	2.106	2.767	1.820	2.169	2.850	1.417	1.800	2.522	1.476	1.868	2.608
∞	1.645	1.960	2.576	1.645	1.960	2.576	1.282	1.645	2.326	1.282	1.645	2.326

Adapted from C. Eisenhart, M. W. Hastay, and W. A. Wallis, *Techniques of Statistical Analysis*, Chapter 2, McGraw-Hill Book Company, New York, 1947. Used with permission of McGraw-Hill Book Company.

Table A.8 Sample Size for the t-Test of the Mean

		Level of t-Test																			
	Single-Sided Test	$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$				
	Double-Sided Test	$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$				
	$\beta = 0.1$	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**
	0.05																				
	0.10																				
	0.15																				122
	0.20										139					99					70
	0.25					110					90				128	64			139	101	45
	0.30				134	78				115	63			119	90	45		122	97	71	32
	0.35			125	99	58			109	85	47		109	88	67	34		90	72	52	24
	0.40		115	97	77	45		101	85	66	37	117	84	68	51	26	101	70	55	40	19
	0.45		92	77	62	37	110	81	68	53	30	93	67	54	41	21	80	55	44	33	15
	0.50	100	75	63	51	30	90	66	55	43	25	76	54	44	34	18	65	45	36	27	13
	0.55	83	63	53	42	26	75	55	46	36	21	63	45	37	28	15	54	38	30	22	11
	0.60	71	53	45	36	22	63	47	39	31	18	53	38	32	24	13	46	32	26	19	9
	0.65	61	46	39	31	20	55	41	34	27	16	46	33	27	21	12	39	28	22	17	8
	0.70	53	40	34	28	17	47	35	30	24	14	40	29	24	19	10	34	24	19	15	8
	0.75	47	36	30	25	16	42	31	27	21	13	35	26	21	16	9	30	21	17	13	7
	0.80	41	32	27	22	14	37	28	24	19	12	31	22	19	15	9	27	19	15	12	6
	0.85	37	29	24	20	13	33	25	21	17	11	28	21	17	13	8	24	17	14	11	6
	0.90	34	26	22	18	12	29	23	19	16	10	25	19	16	12	7	21	15	13	10	5
Value of	**0.95**	31	24	20	17	11	27	21	18	14	9	23	17	14	11	7	19	14	11	9	5
$\Delta = \lvert\delta\rvert/\sigma$	**1.00**	28	22	19	16	10	25	19	16	13	9	21	16	13	10	6	18	13	11	8	5
	1.1	24	19	16	14	9	21	16	14	12	8	18	13	11	9	6		15	11	9	7
	1.2	21	16	14	12	8	18	14	12	10	7	15	12	10	8	5		13	10	8	6
	1.3	18	15	13	11	8	16	13	11	9	6		14	10	9	7		11	8	7	6
	1.4	16	13	12	10	7	14	11	10	9	6	12	9	8	7		10	8	7	5	
	1.5	15	12	11	9	7	13	10	9	8	6	11	8	7	6			9	7	6	
	1.6	13	11	10	8	6	12	10	9	7	5		10	8	7	6			8	6	6
	1.7	12	10	9	8	6		11	9	8	7		9	7	6	5			8	6	5
	1.8	12	10	9	8	6		10	8	7	7			8	7	6				7	6
	1.9	11	9	8	7	6		10	8	7	6			8	6	6				7	5
	2.0	10	8	8	7	5		9	7	7	6			7	6	5					6
	2.1		10	8	7	7		8	7	6	6				7	6					6
	2.2		9	8	7	6		8	7	6	5				7	6					6
	2.3		9	7	7	6			8	6	6				6	5					5
	2.4		8	7	7	6			7	6	6					6					
	2.5		8	7	6	6			7	6	6					6					
	3.0		7	6	6	5			6	5	5					5					
	3.5			6	5	5					5										
	4.0					6															

Reproduced with permission from O. L. Davies, ed., *Design and Analysis of Industrial Experiments*, Oliver & Boyd, Edinburgh, 1956.

Table A.9 Sample Size for the t-Test of the Difference between Two Means

		Level of t-Test																			
	Single-Sided Test	$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$				
	Double-Sided Test	$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$				
	$\beta = 0.1$	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**	**.01**	**.05**	**.1**	**.2**	**.5**
	0.05																				
	0.10																				
	0.15																				
	0.20																				137
	0.25															124					88
	0.30										123					87					61
	0.35					110					90					64				102	45
	0.40					85					70				100	50			108	78	35
	0.45				118	68				101	55			105	79	39		108	86	62	28
	0.50				96	55			106	82	45		106	86	64	32		88	70	51	23
	0.55			101	79	46		106	88	68	38		87	71	53	27	112	73	58	42	19
	0.60		101	85	67	39		90	74	58	32	104	74	60	45	23	89	61	49	36	16
	0.65		87	73	57	34	104	77	64	49	27	88	63	51	39	20	76	52	42	30	14
	0.70	100	75	63	50	29	90	66	55	43	24	76	55	44	34	17	66	45	36	26	12
	0.75	88	66	55	44	26	79	58	48	38	21	67	48	39	29	15	57	40	32	23	11
	0.80	77	58	49	39	23	70	51	43	33	19	59	42	34	26	14	50	35	28	21	10
	0.85	69	51	43	35	21	62	46	38	30	17	52	37	31	23	12	45	31	25	18	9
	0.90	62	46	39	31	19	55	41	34	27	15	47	34	27	21	11	40	28	22	16	8
Value of	**0.95**	55	42	35	28	17	50	37	31	24	14	42	30	25	19	10	36	25	20	15	7
$\Delta = \|\delta\|/\sigma$	**1.00**	50	38	32	26	15	45	33	28	22	13	38	27	23	17	9	33	23	18	14	7
	1.1	42	32	27	22	13	38	28	23	19	11	32	23	19	14	8	27	19	15	12	6
	1.2	36	27	23	18	11	32	24	20	16	9	27	20	16	12	7	23	16	13	10	5
	1.3	31	23	20	16	10	28	21	17	14	8	23	17	14	11	6	20	14	11	9	5
	1.4	27	20	17	14	9	24	18	15	12	8	20	15	12	10	6	17	12	10	8	4
	1.5	24	18	15	13	8	21	16	14	11	7	18	13	11	9	5	15	11	9	7	4
	1.6	21	16	14	11	7	19	14	12	10	6	16	12	10	8	5	14	10	8	6	4
	1.7	19	15	13	10	7	17	13	11	9	6	14	11	9	7	4	12	9	7	6	3
	1.8	17	13	71	10	6	15	12	10	8	5	13	10	8	6	4	11	8	7	5	
	1.9	16	12	11	9	6	14	11	9	8	5	12	9	7	6	4	10	7	6	5	
	2.0	14	11	10	8	6	13	10	9	7	5	11	8	7	6	4	9	7	6	4	
	2.1	13	10	9	8	5	12	9	8	7	5	10	8	6	5	3	8	6	5	4	
	2.2	12	10	8	7	5	11	9	7	6	4	9	7	6	5		8	6	5	4	
	2.3	11	9	8	7	5	10	8	7	6	4	9	7	6	5		7	5	5	4	
	2.4	11	9	8	6	5	10	8	7	6	48	6	5	4		7	5	4	4		
	2.5	10	8	7	6	4	9	7	6	5	4	8	6	5	4		6	5	4	3	
	3.0	8	6	6	5	4	7	6	5	4	3	6	5	4	4		5	4	3		
	3.5	6	5	5	4	3	6	5	4	4	5	4	4	3		4	3				
	4.0	6	5	4	4		5	4	4	3	4	4	3			4					

Reproduced with permission from O. L. Davies, ed., *Design and Analysis of Industrial Experiments*, Oliver & Boyd, Edinburgh, 1956.

Table A.10 Critical Values for Bartlett's Test

	$b_k(0.01; n)$								
	Number of Populations, k								
n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	0.1411	0.1672							
4	0.2843	0.3165	0.3475	0.3729	0.3937	0.4110			
5	0.3984	0.4304	0.4607	0.4850	0.5046	0.5207	0.5343	0.5458	0.5558
6	0.4850	0.5149	0.5430	0.5653	0.5832	O.5978	0.6100	0.6204	0.6293
7	0.5512	0.5787	0.6045	0.6248	0.6410	0.6542	0.6652	0.6744	0.6824
8	0.6031	0.6282	0.6518	0.6704	0.6851	0.6970	0.7069	0.7153	0.7225
9	0.6445	0.6676	0.6892	0.7062	0.7197	0.7305	0.7395	0.7471	0.7536
10	0.6783	0.6996	0.7195	0.7352	0.7475	0.7575	0.7657	0.7726	0.7786
11	0.7063	0.7260	0.7445	0.7590	0.7703	0.7795	0.7871	0.7935	0.7990
12	0.7299	0.7483	0.7654	0.7789	0.7894	0.7980	0.8050	0.8109	0.8160
13	0.7501	0.7672	0.7832	0.7958	0.8056	0.8135	0.8201	0.8256	0.8303
14	0.7674	0.7835	0.7985	0.8103	0.8195	0.8269	0.8330	0.8382	0.8426
15	0.7825	0.7977	0.8118	0.8229	0.8315	0.8385	0.8443	0.8491	0.8532
16	0.7958	0.8101	0.8235	0.8339	0.8421	0.8486	0.8541	0.8586	0.8625
17	0.8076	0.8211	0.8338	0.8436	0.8514	0.8576	0.8627	0.8670	0.8707
18	0.8181	0.8309	0.8429	0.8523	0.8596	0.8655	0.8704	0.8745	0.8780
19	0.8275	0.8397	0.8512	0.8601	0.8670	0.8727	0.8773	0.8811	0.8845
20	0.8360	0.8476	0.8586	0.8671	0.8737	0.8791	0.8835	0.8871	0.8903
21	0.8437	0.8548	0.8653	0.8734	0.8797	0.8848	0.8890	0.8926	0.8956
22	0.8507	0.8614	0.8714	0.8791	0.8852	0.8901	0.8941	0.8975	0.9004
23	0.8571	0.8673	0.8769	0.8844	0.8902	0.8949	0.8988	0.9020	0.9047
24	0.8630	0.8728	0.8820	0.8892	0.8948	0.8993	0.9030	0.9061	0.9087
25	0.8684	0.8779	0.8867	0.8936	0.8990	0.9034	0.9069	0.9099	0.9124
26	0.8734	0.8825	0.8911	0.8977	0.9029	0.9071	0.9105	0.9134	0.9158
27	0.8781	0.8869	0.8951	0.9015	0.9065	0.9105	0.9138	0.9166	0.9190
28	0.8824	0.8909	0.8988	0.9050	0.9099	0.9138	0.9169	0.9196	0.9219
29	0.8864	0.8946	0.9023	0.9083	0.9130	0.9167	0.9198	0.9224	0.9246
30	0.8902	0.8981	0.9056	0.9114	0.9159	0.9195	0.9225	0.9250	0.9271
40	0.9175	0.9235	0.9291	0.9335	0.9370	0.9397	0.9420	0.9439	0.9455
50	0.9339	0.9387	0.9433	0.9468	0.9496	0.9518	0.9536	0.9551	0.9564
60	0.9449	0.9489	0.9527	0.9557	0.9580	0.9599	0.9614	0.9626	0.9637
80	0.9586	0.9617	0.9646	0.9668	0.9685	0.9699	0.9711	0.9720	0.9728
100	0.9669	0.9693	0.9716	0.9734	0.9748	0.9759	0.9769	0.9776	0.9783

Reproduced from D. D. Dyer and J. P. Keating, "On the Determination of Critical Values for Bartlett's Test," *J. Am. Stat. Assoc.*, **75**, 1980, by permission of the Board of Directors.

Table A.10 (continued) Critical Values for Bartlett's Test

	$b_k(0.05; n)$								
	Number of Populations, k								
n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	0.3123	0.3058	0.3173	0.3299					
4	0.4780	0.4699	0.4803	0.4921	0.5028	0.5122	0.5204	0.5277	0.5341
5	0.5845	0.5762	0.5850	0.5952	0.6045	0.6126	0.6197	0.6260	0.6315
6	0.6563	0.6483	0.6559	0.6646	0.6727	0.6798	0.6860	0.6914	0.6961
7	0.7075	0.7000	0.7065	0.7142	0.7213	0.7275	0.7329	0.7376	0.7418
8	0.7456	0.7387	0.7444	0.7512	0.7574	0.7629	0.7677	0.7719	0.7757
9	0.7751	0.7686	0.7737	0.7798	0.7854	0.7903	0.7946	0.7984	0.8017
10	0.7984	0.7924	0.7970	0.8025	0.8076	0.8121	0.8160	0.8194	0.8224
11	0.8175	0.8118	0.8160	0.8210	0.8257	0.8298	0.8333	0.8365	0.8392
12	0.8332	0.8280	0.8317	0.8364	0.8407	0.8444	0.8477	0.8506	0.8531
13	0.8465	0.8415	0.8450	0.8493	0.8533	0.8568	0.8598	0.8625	0.8648
14	0.8578	0.8532	0.8564	0.8604	0.8641	0.8673	0.8701	0.8726	0.8748
15	0.8676	0.8632	0.8662	0.8699	0.8734	0.8764	0.8790	0.8814	0.8834
16	0.8761	0.8719	0.8747	0.8782	0.8815	0.8843	0.8868	0.8890	0.8909
17	0.8836	0.8796	0.8823	0.8856	0.8886	0.8913	0.8936	0.8957	0.8975
18	0.8902	0.8865	0.8890	0.8921	0.8949	0.8975	0.8997	0.9016	0.9033
19	0.8961	0.8926	0.8949	0.8979	0.9006	0.9030	0.9051	0.9069	0.9086
20	0.9015	0.8980	0.9003	0.9031	0.9057	0.9080	0.9100	0.9117	0.9132
21	0.9063	0.9030	0.9051	0.9078	0.9103	0.9124	0.9143	0.9160	0.9175
22	0.9106	0.9075	0.9095	0.9120	0.9144	0.9165	0.9183	0.9199	0.9213
23	0.9146	0.9116	0.9135	0.9159	0.9182	0.9202	0.9219	0.9235	0.9248
24	0.9182	0.9153	0.9172	0.9195	0.9217	0.9236	0.9253	0.9267	0.9280
25	0.9216	0.9187	0.9205	0.9228	0.9249	0.9267	0.9283	0.9297	0.9309
26	0.9246	0.9219	0.9236	0.9258	0.9278	0.9296	0.9311	0.9325	0.9336
27	0.9275	0.9249	0.9265	0.9286	0.9305	0.9322	0.9337	0.9350	0.9361
28	0.9301	0.9276	0.9292	0.9312	0.9330	0.9347	0.9361	0.9374	0.9385
29	0.9326	0.9301	0.9316	0.9336	0.9354	0.9370	0.9383	0.9396	0.9406
30	0.9348	0.9325	0.9340	0.9358	0.9376	0.9391	0.9404	0.9416	0.9426
40	0.9513	0.9495	0.9506	0.9520	0.9533	0.9545	0.9555	0.9564	0.9572
50	0.9612	0.9597	0.9606	0.9617	0.9628	0.9637	0.9645	0.9652	0.9658
60	0.9677	0.9665	0.9672	0.9681	0.9690	0.9698	0.9705	0.9710	0.9716
80	0.9758	0.9749	0.9754	0.9761	0.9768	0.9774	0.9779	0.9783	0.9787
100	0.9807	0.9799	0.9804	0.9809	0.9815	0.9819	0.9823	0.9827	0.9830

Table A.11 Critical Values for Cochran's Test

$\alpha = 0.01$														
	n													
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**17**	**37**	**145**	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0125	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Reproduced from C. Eisenhart, M. W. Hastay, and W. A. Wallis, *Techniques of Statistical Analysis*, Chapter 15, McGraw-Hill Book Company, New, York, 1947. Used with permission of McGraw-Hill Book Company.

Table A.11 (continued) Critical Values for Cochran's Test

$\alpha = 0.05$

	n													
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**17**	**37**	**145**	**∞**
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	6.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.12 Upper Percentage Points of the Studentized Range Distribution: Values of $q(0.05; k, v)$

Degrees of	**Number of Treatments k**								
Freedom, v	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	18.0	27.0	32.8	37.2	40.5	43.1	15.1	47.1	49.1
2	6.09	5.33	9.80	10.89	11.73	12.43	13.03	13.54	13.99
3	4.50	5.91	6.83	7.51	8.04	8.47	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.06	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.65	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.05
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47

Table A.13 Least Significant Studentized Ranges $r_p(0.05; p, v)$

$\alpha = 0.05$

					p				
v	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.501	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	3.927	4.013	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	3.635	3.749	3.797	3.814	3.814	3.814	3.814	3.814	3.814
6	3.461	3.587	3.649	3.68	3.694	3.697	3.697	3.697	3.697
7	3.344	3.477	3.548	3.588	3.611	3.622	3.626	3.626	3.626
8	3.261	3.399	3.475	3.521	3.549	3.566	3.575	3.579	3.579
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544	3.547
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522
11	3.113	3.256	3.342	3.397	3.435	3.462	3.48	3.493	3.501
12	3.082	3.225	3.313	3.370	3.410	3.439	3.459	3.474	3.484
13	3.055	3.200	3.289	3.348	3.389	3.419	3.442	3.458	3.470
14	3.033	3.178	3.268	3.329	3.372	3.403	3.426	3.444	3.457
15	3.014	3.160	3.25	3.312	3.356	3.389	3.413	3.432	3.446
16	2.998	3.144	3.235	3.298	3.343	3.376	3.402	3.422	3.437
17	2.984	3.130	3.222	3.285	3.331	3.366	3.392	3.412	3.429
18	2.971	3.118	3.210	3.274	3.321	3.356	3.383	3.405	3.421
19	2.960	3.107	3.199	3.264	3.311	3.347	3.375	3.397	3.415
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.391	3.409
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390
30	2.888	3.035	3.131	3.199	3.250	3.290	3.322	3.349	3.371
40	2.858	3.006	3.102	3.171	3.224	3.266	3.300	3.328	3.352
60	2.829	2.976	3.073	3.143	3.198	3.241	3.277	3.307	3.333
120	2.800	2.947	3.045	3.116	3.172	3.217	3.254	3.287	3.314
∞	2.772	2.918	3.017	3.089	3.146	3.193	3.232	3.265	3.294

Abridged from H. L. Harter, "Critical Values for Duncan's New Multiple Range Test," *Biometrics*, **16**, No. 4, 1960, by permission of the author and the editor.

Table A.13 (continued) Least Significant Studentized Ranges $r_p(0.01; p, v)$

$\alpha = 0.01$

	p								
v	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03
2	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04
3	8.261	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321
4	6.512	6.677	6.740	6.756	6.756	6.756	6.756	6.756	6.756
5	5.702	5.893	5.989	6.040	6.065	6.074	6.074	6.074	6.074
6	5.243	5.439	5.549	5.614	5.655	5.680	5.694	5.701	5.703
7	4.949	5.145	5.260	5.334	5.383	5.416	5.439	5.454	5.464
8	4.746	4.939	5.057	5.135	5.189	5.227	5.256	5.276	5.291
9	4.596	4.787	4.906	4.986	5.043	5.086	5.118	5.142	5.160
10	4.482	4.671	4.790	4.871	4.931	4.975	5.010	5.037	5.058
11	4.392	4.579	4.697	4.780	4.841	4.887	4.924	4.952	4.975
12	4.320	4.504	4.622	4.706	4.767	4.815	4.852	4.883	4.907
13	4.260	4.442	4.560	4.644	4.706	4.755	4.793	4.824	4.850
14	4.210	4.391	4.508	4.591	4.654	4.704	4.743	4.775	4.802
15	4.168	4.347	4.463	4.547	4.610	4.660	4.700	4.733	4.760
16	4.131	4.309	4.425	4.509	4.572	4.622	4.663	4.696	4.724
17	4.099	4.275	4.391	4.475	4.539	4.589	4.630	4.664	4.693
18	4.071	4.246	4.362	4.445	4.509	4.560	4.601	4.635	4.664
19	4.046	4.220	4.335	4.419	4.483	4.534	4.575	4.610	4.639
20	4.024	4.197	4.312	4.395	4.459	4.510	4.552	4.587	4.617
24	3.956	4.126	4.239	4.322	4.386	4.437	4.480	4.516	4.546
30	3.889	4.056	4.168	4.250	4.314	4.366	4.409	4.445	4.477
40	3.825	3.988	4.098	4.180	4.244	4.296	4.339	4.376	4.408
60	3.762	3.922	4.031	4.111	4.174	4.226	4.270	4.307	4.340
120	3.702	3.858	3.965	4.044	4.107	4.158	4.202	4.239	4.272
∞	3.643	3.796	3.900	3.978	4.040	4.091	4.135	4.172	4.205

Table A.14 Values of $d_{\alpha/2}(k, v)$ for Two-Sided Comparisons between k Treatments and a Control

	$\alpha = 0.05$								
	k = Number of Treatment Means (excluding control)								
v	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

Reproduced from Charles W. Dunnett, "New Tables for Multiple Comparison with a Control," *Biometrics*, **20**, No. 3, 1964, by permission of the author and the editor.

Table A.14 (continued) Values of $d_{\alpha/2}(k, v)$ for Two-Sided Comparisons between k Treatments and a Control

	$\alpha = 0.01$								
	k = Number of Treatment Means (excluding control)								
v	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	4.03	4.63	4.98	5.22	5.41	5.56	5.69	5.80	5.89
6	3.71	4.21	4.51	4.71	4.87	5.00	5.10	5.20	5.28
7	3.50	3.95	4.21	4.39	4.53	4.64	4.74	4.82	4.89
8	3.36	3.77	4.00	4.17	4.29	4.40	4.48	4.56	4.62
9	3.25	3.63	3.85	4.01	4.12	4.22	4.30	4.37	4.43
10	3.17	3.53	3.74	3.88	3.99	4.08	4.16	4.22	4.28
11	3.11	3.45	3.65	3.79	3.89	3.98	4.05	4.11	4.16
12	3.05	3.39	3.58	3.71	3.81	3.89	3.96	4.02	4.07
13	3.01	3.33	3.52	3.65	3.74	3.82	3.89	3.94	3.99
14	2.98	3.29	3.47	3.59	3.69	3.76	3.83	3.88	3.93
15	2.95	3.25	3.43	3.55	3.64	3.71	3.78	3.83	3.88
16	2.92	3.22	3.39	3.51	3.60	3.67	3.73	3.78	3.83
17	2.90	3.19	3.36	3.47	3.56	3.63	3.69	3.74	3.79
18	2.88	3.17	3.33	3.44	3.53	3.60	3.66	3.71	3.75
19	2.86	3.15	3.31	3.42	3.50	3.57	3.63	3.68	3.72
20	2.85	3.13	3.29	3.40	3.48	3.55	3.60	3.65	3.69
24	2.80	3.07	3.22	3.32	3.40	3.47	3.52	3.57	3.61
30	2.75	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52
40	2.70	2.95	3.09	3.19	3.26	3.32	3.37	3.41	3.44
60	2.66	2.90	3.03	3.12	3.19	3.25	3.29	3.33	3.37
120	2.62	2.85	2.97	3.06	3.12	3.18	3.22	3.26	3.29
∞	2.58	2.79	2.92	3.00	3.06	3.11	3.15	3.19	3.22

Table A.15 Values of $d_\alpha(k, v)$ for One-Sided Comparisons between k Treatments and a Control

$\alpha = 0.05$									
	k = Number of Treatment Means (excluding control)								
v	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

Reproduced from Charles W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *J. Am. Stat. Assoc.*, **50**, 1955, 1096–1121, by permission of the author and the editor.

Table A.15 (continued) Values of $d_\alpha(k, v)$ for One-Sided Comparisons between k Treatments and a Control

	$\alpha = 0.01$								
	k = Number of Treatment Means (excluding control)								
v	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

Table A.16 Critical Values for the Signed-Rank Test

n	**One-Sided $\alpha = 0.01$** **Two-Sided $\alpha = 0.02$**	**One-Sided $\alpha = 0.025$** **Two-Sided $\alpha = 0.05$**	**One-Sided $\alpha = 0.05$** **Two-Sided $\alpha = 0.1$**
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

Reproduced from F. Wilcoxon and R. A. Wilcox, *Some Rapid Approximate Statistical Procedures, American Cyanamid Company*, Pearl River, N.Y., 1964, by permission of the American Cyanamid Company.

Table A.17 Critical Values for the Wilcoxon Rank-Sum Test

One-Tailed Test at $\alpha = 0.001$ or Two-Tailed Test at $\alpha = 0.002$

	n_2														
n_1	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1															
2															
3												0	0	0	0
4					0	0	0	1	1	1	2	2	3	3	3
5		0	0	1	1	2	2	3	3	4	5	5	6	7	7
6	0	1	2	2	3	4	4	5	6	7	8	9	10	11	12
7		2	3	3	5	6	7	8	9	10	11	13	14	15	16
8			5	5	6	8	9	11	12	14	15	17	18	20	21
9				7	8	10	12	14	15	17	19	21	23	25	26
10					10	12	14	17	19	21	23	25	27	29	32
11						15	17	20	22	24	27	29	32	34	37
12							20	23	25	28	31	34	37	40	42
13								26	29	32	35	38	42	45	48
14									32	36	39	43	46	50	54
15										40	43	47	51	55	59
16											48	52	56	60	65
17												57	61	66	70
18													66	71	76
19														77	82
20															88

One-Tailed Test at $\alpha = 0.01$ or Two-Tailed Test at $\alpha = 0.02$

	n_2															
n_1	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1																
2									0	0	0	0	0	0	1	1
3			0	0	1	1	1	2	2	2	3	3	4	4	4	5
4	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6		3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
7			6	8	9	11	12	14	16	17	19	21	23	24	26	28
8				10	11	13	15	17	20	22	24	26	28	30	32	34
9					14	16	18	21	23	26	28	31	33	36	38	40
10						19	22	24	27	30	33	36	38	41	44	47
11							25	28	31	34	37	41	44	47	50	53
12								31	35	38	42	46	49	53	56	60
13									39	43	47	51	55	59	63	67
14										47	51	56	60	65	69	73
15											56	61	66	70	75	80
16												66	71	76	82	87
17													77	82	88	93
18														88	94	100
19															101	107
20																114

Based in part on Tables 1, 3, 5, and 7 of D. Auble, "Extended Tables for the Mann-Whitney Statistic," *Bulletin of the Institute of Educational Research at Indiana University*, **1**, No. 2, 1953, by permission of the director.

Table A.17 (continued) Critical Values for the Wilcoxon Rank-Sum Test

One-Tailed Test at $\alpha = 0.025$ or Two-Tailed Test at $\alpha = 0.05$

	n_2																
n_1	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1																	
2					0	0	0	0	1	1	1	1	1	2	2	2	2
3		0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5		2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7				8	10	12	14	16	18	20	22	24	26	28	30	32	34
8					13	15	17	19	22	24	26	29	31	34	36	38	41
9						17	20	23	26	28	31	34	37	39	42	45	48
10							23	26	29	33	36	39	42	45	48	52	55
11								30	33	37	40	44	47	51	55	58	62
12									37	41	45	49	53	57	61	65	69
13										45	50	54	59	63	67	72	76
14											55	59	64	67	74	78	83
15												64	70	75	80	85	90
16													75	81	86	92	98
17														87	93	99	105
18															99	106	112
19																113	119
20																	127

One-Tailed Test at $\alpha = 0.05$ or Two-Tailed Test at $\alpha = 0.1$

	n_2																	
n_1	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1																	0	0
2			0	0	0	1	1	1	1	2	2	3	3	3	3	4	4	4
3	0	0	1	2	2	3	4	4	5	5	6	7	7	8	9	9	10	11
4		1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5			4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6				7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7					11	13	15	17	19	21	24	26	28	30	33	35	37	39
8						15	18	20	23	26	28	31	33	36	39	41	44	47
9							21	24	27	30	33	36	39	42	45	48	51	54
10								27	31	34	37	41	44	48	51	55	58	62
11									34	38	42	46	50	54	57	61	65	69
12										42	47	51	55	60	64	68	72	77
13											51	56	61	65	70	75	80	84
14												61	66	71	77	82	87	92
15													72	77	83	88	94	100
16														83	89	95	101	107
17															96	102	109	115
18																109	116	123
19																	123	130
20																		138

Table A.18 $P(V \le v^*$ when H_0 is true) in the Runs Test

	v^*								
(n_1, n_2)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$(2, 3)$	0.200	0.500	0.900	1.000					
$(2, 4)$	0.133	0.400	0.800	1.000					
$(2, 5)$	0.095	0.333	0.714	1.000					
$(2, 6)$	0.071	0.286	0.643	1.000					
$(2, 7)$	0.056	0.250	0.583	1.000					
$(2, 8)$	0.044	0.222	0.533	1.000					
$(2, 9)$	0.036	0.200	0.491	1.000					
$(2, 10)$	0.030	0.182	0.455	1.000					
$(3, 3)$	0.100	0.300	0.700	0.900	1.000				
$(3, 4)$	0.057	0.200	0.543	0.800	0.971	1.000			
$(3, 5)$	0.036	0.143	0.429	0.714	0.929	1.000			
$(3, 6)$	0.024	0.107	0.345	0.643	0.881	1.000			
$(3, 7)$	0.017	0.083	0.283	0.583	0.833	1.000			
$(3, 8)$	0.012	0.067	0.236	0.533	0.788	1.000			
$(3, 9)$	0.009	0.055	0.200	0.491	0.745	1.000			
$(3, 10)$	0.007	0.045	0.171	0.455	0.706	1.000			
$(4, 4)$	0.029	0.114	0.371	0.629	0.886	0.971	1.000		
$(4, 5)$	0.016	0.071	0.262	0.500	0.786	0.929	0.992	1.000	
$(4, 6)$	0.010	0.048	0.190	0.405	0.690	0.881	0.976	1.000	
$(4, 7)$	0.006	0.033	0.142	0.333	0.606	0.833	0.954	1.000	
$(4, 8)$	0.004	0.024	0.109	0.279	0.533	0.788	0.929	1.000	
$(4, 9)$	0.003	0.018	0.085	0.236	0.471	0.745	0.902	1.000	
$(4, 10)$	0.002	0.014	0.068	0.203	0.419	0.706	0.874	1.000	
$(5, 5)$	0.008	0.040	0.167	0.357	0.643	0.833	0.960	0.992	1.000
$(5, 6)$	0.004	0.024	0.110	0.262	0.522	0.738	0.911	0.976	0.998
$(5, 7)$	0.003	0.015	0.076	0.197	0.424	0.652	0.854	0.955	0.992
$(5, 8)$	0.002	0.010	0.054	0.152	0.347	0.576	0.793	0.929	0.984
$(5, 9)$	0.001	0.007	0.039	0.119	0.287	0.510	0.734	0.902	0.972
$(5, 10)$	0.001	0.005	0.029	0.095	0.239	0.455	0.678	0.874	0.958
$(6, 6)$	0.002	0.013	0.067	0.175	0.392	0.608	0.825	0.933	0.987
$(6, 7)$	0.001	0.008	0.043	0.121	0.296	0.500	0.733	0.879	0.966
$(6, 8)$	0.001	0.005	0.028	0.086	0.226	0.413	0.646	0.821	0.937
$(6, 9)$	0.000	0.003	0.019	0.063	0.175	0.343	0.566	0.762	0.902
$(6, 10)$	0.000	0.002	0.013	0.047	0.137	0.288	0.497	0.706	0.864
$(7, 7)$	0.001	0.004	0.025	0.078	0.209	0.383	0.617	0.791	0.922
$(7, 8)$	0.000	0.002	0.015	0.051	0.149	0.296	0.514	0.704	0.867
$(7, 9)$	0.000	0.001	0.010	0.035	0.108	0.231	0.427	0.622	0.806
$(7, 10)$	0.000	0.001	0.006	0.024	0.080	0.182	0.355	0.549	0.743
$(8, 8)$	0.000	0.001	0.009	0.032	0.100	0.214	0.405	0.595	0.786
$(8, 9)$	0.000	0.001	0.005	0.020	0.069	0.157	0.319	0.500	0.702
$(8, 10)$	0.000	0.000	0.003	0.013	0.048	0.117	0.251	0.419	0.621
$(9, 9)$	0.000	0.000	0.003	0.012	0.044	0.109	0.238	0.399	0.601
$(9, 10)$	0.000	0.000	0.002	0.008	0.029	0.077	0.179	0.319	0.510
$(10, 10)$	0.000	0.000	0.001	0.004	0.019	0.051	0.128	0.242	0.414

Reproduced from C. Eisenhart and R. Swed, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," *Ann. Math. Stat.*, **14**, 1943, by permission of the editor.

Table A.18 (continued) $P(V \leq v^*$ when H_0 is true) in the Runs Test

	v^*									
(n_1, n_2)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
(2, 3)										
(2, 4)										
(2, 5)										
(2, 6)										
(2, 7)										
(2, 8)										
(2, 9)										
(2, 10)										
(3, 3)										
(3, 4)										
(3, 5)										
(3, 6)										
(3, 7)										
(3, 8)										
(3, 9)										
(3, 10)										
(4, 4)										
(4, 5)										
(4, 6)										
(4, 7)										
(4, 8)										
(4, 9)										
(4, 10)										
(5, 5)										
(5, 6)	1.000									
(5, 7)	1.000									
(5, 8)	1.000									
(5, 9)	1.000									
(5, 10)	1.000									
(6, 6)	0.998	1.000								
(6, 7)	0.992	0.999	1.000							
(6, 8)	0.984	0.998	1.000							
(6, 9)	0.972	0.994	1.000							
(6, 10)	0.958	0.990	1.000							
(7, 7)	0.975	0.996	0.999	1.000						
(7, 8)	0.949	0.988	0.998	1.000	1.000					
(7, 9)	0.916	0.975	0.994	0.999	1.000					
(7, 10)	0.879	0.957	0.990	0.998	1.000					
(8, 8)	0.900	0.968	0.991	0.999	1.000	1.000				
(8, 9)	0.843	0.939	0.980	0.996	0.999	1.000	1.000			
(8, 10)	0.782	0.903	0.964	0.990	0.998	1.000	1.000			
(9, 9)	0.762	0.891	0.956	0.988	0.997	1.000	1.000	1.000		
(9, 10)	0.681	0.834	0.923	0.974	0.992	0.999	1.000	1.000	1.000	
(10, 10)	0.586	0.758	0.872	0.949	0.981	0.996	0.999	1.000	1.000	1.000

Table A.19 Sample Size for Two-Sided Nonparametric Tolerance Limits

	$\mathbf{1-\gamma}$					
$\mathbf{1-\alpha}$	**0.50**	**0.70**	**0.90**	**0.95**	**0.99**	**0.995**
0.995	336	488	777	947	1325	1483
0.99	168	244	388	473	662	740
0.95	34	49	77	93	130	146
0.90	17	24	38	46	64	72
0.85	11	16	25	30	42	47
0.80	9	12	18	22	31	34
0.75	7	10	15	18	24	27
0.70	6	8	12	14	20	22
0.60	4	6	9	10	14	16
0.50	3	5	7	8	11	12

Reproduced from Table A–25d of Wilfrid J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis*, 3rd ed. McGraw-Hill, New York, 1969. Used with permission of McGraw-Hill Book Company.

Table A.20 Sample Size for One-Sided Nonparametric Tolerance Limits

	$\mathbf{1-\gamma}$				
$\mathbf{1-\alpha}$	**0.50**	**0.70**	**0.95**	**0.99**	**0.995**
0.995	139	241	598	919	1379
0.99	69	120	299	459	688
0.95	14	24	59	90	135
0.90	7	12	29	44	66
0.85	5	8	19	29	43
0.80	4	6	14	21	31
0.75	3	5	11	7	25
0.70	2	4	9	13	20
0.60	2	3	6	10	14
0.50	1	2	5	7	10

Reproduced from Table A–25e of Wilfrid J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis*, 3rd ed. McGraw-Hill, New York, 1969. Used with permission of McGraw-Hill Book Company.

Table A.21 Critical Values for Spearman's Rank Correlation Coefficients

n	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900			
6	0.829	0.886	0.943	
7	0.714	0.786	0.893	
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

Reproduced from E. G. Olds, "Distribution of Sums of Squares of Rank Differences for Small Samples," *Ann. Math. Stat.*, **9**, 1938, by permission of the editor.

Table A.22 Factors for Constructing Control Charts

	Chart for Averages		Chart for Standard Deviations						Chart for Ranges				
Obs. in Sample	**Factors for Control Limits**		**Factors for Centerline**		**Factors for Control Limits**				**Factors for Centerline**		**Factors for Control Limits**		
n	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6	d_2	$1/d_2$	d_3	D_3	D_4
2	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.267
3	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	2.574
4	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	2.282
5	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	2.114
6	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	2.004
7	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.076	1.924
8	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.136	1.864
9	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.184	1.816
10	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.223	1.777
11	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.256	1.744
12	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.283	1.717
13	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	0.307	1.693
14	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	0.328	1.672
15	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	0.347	1.653
16	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	0.363	1.637
17	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	0.378	1.622
18	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	0.391	1.608
19	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	0.403	1.597
20	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	0.415	1.585
21	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	0.425	1.575
22	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	0.434	1.566
23	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	0.443	1.557
24	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	0.451	1.548
25	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	0.459	4.541

Table A.23 The Incomplete Gamma Function: $F(x;\alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}\, dy$

	α									
x	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.6320	0.2640	0.0800	0.0190	0.0040	0.0010	0.0000	0.0000	0.0000	0.0000
2	0.8650	0.5940	0.3230	0.1430	0.0530	0.0170	0.0050	0.0010	0.0000	0.0000
3	0.9500	0.8010	0.5770	0.3530	0.1850	0.0840	0.0340	0.0120	0.0040	0.0010
4	0.9820	0.9080	0.7620	0.5670	0.3710	0.2150	0.1110	0.0510	0.0210	0.0080
5	0.9930	0.9600	0.8750	0.7350	0.5600	0.3840	0.2380	0.1330	0.0680	0.0320
6	0.9980	0.9830	0.9380	0.8490	0.7150	0.5540	0.3940	0.2560	0.1530	0.0840
7	0.9990	0.9930	0.9700	0.9180	0.8270	0.6990	0.5500	0.4010	0.2710	0.1700
8	1.0000	0.9970	0.9860	0.9580	0.9000	0.8090	0.6870	0.5470	0.4070	0.2830
9		0.9990	0.9940	0.9790	0.9450	0.8840	0.7930	0.6760	0.5440	0.4130
10		1.0000	0.9970	0.9900	0.9710	0.9330	0.8700	0.7800	0.6670	0.5420
11			0.9990	0.9950	0.9850	0.9620	0.9210	0.8570	0.7680	0.6590
12			1.0000	0.9980	0.9920	0.9800	0.9540	0.9110	0.8450	0.7580
13				0.9990	0.9960	0.9890	0.9740	0.9460	0.9000	0.8340
14				1.0000	0.9980	0.9940	0.9860	0.9680	0.9380	0.8910
15					0.9990	0.9970	0.9920	0.9820	0.9630	0.9300

A.24 Proof of Mean of the Hypergeometric Distribution

To find the mean of the hypergeometric distribution, we write

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = k \sum_{x=1}^{n} \frac{(k-1)!}{(x-1)!(k-x)!} \cdot \frac{\binom{N-k}{n-x}}{\binom{N}{n}} \\ &= k \sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}}. \end{aligned}$$

Since

$$\binom{N-k}{n-1-y} = \binom{(N-1)-(k-1)}{n-1-y} \quad \text{and} \quad \binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1},$$

letting $y = x - 1$, we obtain

$$\begin{aligned} E(X) &= k \sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{N-k}{n-1-y}}{\binom{N}{n}} \\ &= \frac{nk}{N} \sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{(N-1)-(k-1)}{n-1-y}}{\binom{N-1}{n-1}} = \frac{nk}{N}, \end{aligned}$$

since the summation represents the total of all probabilities in a hypergeometric experiment when $N-1$ items are selected at random from $N-1$, of which $k-1$ are labeled success.

A.25 Proof of Mean and Variance of the Poisson Distribution

Let $\mu = \lambda t$.

$$E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\mu}\mu^x}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{e^{-\mu}\mu^x}{x!} = \mu \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^{x-1}}{(x-1)!}.$$

Since the summation in the last term above is the total probability of a Poisson random variable with mean μ, which can be easily seen by letting $y = x - 1$, it equals 1. Therefore, $E(X) = \mu$. To calculate the variance of X, note that

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\mu}\mu^x}{x!} = \mu^2 \sum_{x=2}^{\infty} \frac{e^{-\mu}\mu^{x-2}}{(x_2)!}.$$

Again, letting $y = x - 2$, the summation in the last term above is the total probability of a Poisson random variable with mean μ. Hence, we obtain

$$\sigma^2 = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2 = \mu^2 + \mu - \mu^2 = \mu = \lambda t.$$

A.26 Proof of Mean and Variance of the Gamma Distribution

To find the mean and variance of the gamma distribution, we first calculate

$$E(X^k) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+k-1} e^{-x/\beta}\, dx = \frac{\beta^{k+\alpha}\Gamma(\alpha+k)}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{x^{\alpha+k-1}e^{-x/\beta}}{\beta^{k+\alpha}\Gamma(\alpha+k)}\, dx,$$

for $k = 0, 1, 2, \ldots$. Since the integrand in the last term above is a gamma density function with parameters $\alpha + k$ and β, it equals 1. Therefore,

$$E(X^k) = \beta^k \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)}.$$

Using the recursion formula of the gamma function from page 194, we obtain

$$\mu = \beta \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta \quad \text{and} \quad \sigma^2 = E(X^2) - \mu^2 = \beta^2 \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} - \mu^2 = \beta^2\alpha(\alpha+1) - (\alpha\beta)^2 = \alpha\beta^2.$$

Appendix B
Answers to Odd-Numbered Non-Review Exercises

Chapter 1

1.1 (a) Sample size = 15

(b) Sample mean = 3.787

(c) Sample median = 3.6

(e) $\bar{x}_{\text{tr}(20)} = 3.678$

(f) They are about the same.

1.3 (b) Yes, the aging process has reduced the tensile strength.

(c) $\bar{x}_{\text{Aging}} = 209.90$, $\bar{x}_{\text{No aging}} = 222.10$

(d) $\tilde{x}_{\text{Aging}} = 210.00$, $\tilde{x}_{\text{No aging}} = 221.50$. The means and medians are similar for each group.

1.5 (b) Control: $\bar{x} = 5.60$, $\tilde{x} = 5.00$, $\bar{x}_{\text{tr}(10)} = 5.13$. Treatment: $\bar{x} = 7.60$, $\tilde{x} = 4.50$, $\bar{x}_{\text{tr}(10)} = 5.63$.

(c) The extreme value of 37 in the treatment group plays a strong leverage role for the mean calculation.

1.7 Sample variance = 0.943
Sample standard deviation = 0.971

1.9 (a) No aging: sample variance = 23.66, sample standard deviation = 4.86. Aging: sample variance = 42.10, sample standard deviation = 6.49.

(b) Based on the numbers in (a), the variation in "Aging" is smaller than the variation in "No aging," although the difference is not so apparent in the plot.

1.11 Control: sample variance = 69.38, sample standard deviation = 8.33. Treatment: sample variance = 128.04, sample standard deviation = 11.32.

1.13 (a) Mean = 124.3, median = 120

(b) 175 is an extreme observation.

1.15 Yes, P-value = 0.03125, probability of obtaining $HHHHH$ with a fair coin.

1.17 (a) The sample means for nonsmokers and smokers are 30.32 and 43.70, respectively.

(b) The sample standard deviations for nonsmokers and smokers are 7.13 and 16.93, respectively.

(d) Smokers appear to take a longer time to fall asleep. For smokers the time to fall asleep is more variable.

1.19 (a)

Stem	Leaf	Frequency
0	22233457	8
1	023558	6
2	035	3
3	03	2
4	057	3
5	0569	4
6	0005	4

(b)

Class Interval	Class Midpoint	Freq.	Rel. Freq.
0.0−0.9	0.45	8	0.267
1.0−1.9	1.45	6	0.200
2.0−2.9	2.45	3	0.100
3.0−3.9	3.45	2	0.067
4.0−4.9	4.45	3	0.100
5.0−5.9	5.45	4	0.133
6.0−6.9	6.45	4	0.133

(c) Sample mean = 2.7967
Sample range = 6.3
Sample standard deviation = 2.2273

1.21 (a) $\bar{x} = 74.02$ and $\tilde{x} = 78$
(b) $s = 39.26$

1.23 (b) $\bar{x}_{1980} = 395.10$, $\bar{x}_{1990} = 160.15$
(c) The mean emissions dropped between 1980 and 1990; the variability also decreased because there were no longer extremely large emissions.

1.25 (a) Sample mean = 33.31
(b) Sample median = 26.35
(d) $\bar{x}_{\text{tr}(10)} = 30.97$

Chapter 2

2.1 (a) $S = \{8, 16, 24, 32, 40, 48\}$
(b) $S = \{-5, 1\}$
(c) $S = \{T, HT, HHT, HHH\}$
(d) S ={Africa, Antarctica, Asia, Australia, Europe, North America, South America}
(e) $S = \phi$

2.3 $A = C$

2.5 Using the tree diagram, we obtain
$S = \{1HH,\ 1HT,\ 1TH,\ 1TT,\ 2H,\ 2T,\ 3HH,\ 3HT,\ 3TH,\ 3TT,\ 4H,\ 4T,\ 5HH,\ 5HT,\ 5TH,\ 5TT,\ 6H,\ 6T\}$

2.7 $S_1 = \{MMMM, MMMF, MMFM, MFMM, FMMM, MMFF, MFMF, MFFM, FMFM, FFMM, FMMF, MFFF, FMFF, FFMF, FFFM, FFFF\}$;
$S_2 = \{0, 1, 2, 3, 4\}$

2.9 (a) $A = \{1HH, 1HT, 1TH, 1TT, 2H, 2T\}$
(b) $B = \{1TT, 3TT, 5TT\}$
(c) $A' = \{3HH, 3HT, 3TH, 3TT, 4H, 4T, 5HH, 5HT, 5TH, 5TT, 6H, 6T\}$
(d) $A' \cap B = \{3TT, 5TT\}$
(e) $A \cup B = \{1HH, 1HT, 1TH, 1TT, 2H, 2T, 3TT, 5TT\}$

2.11 (a) $S = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2, F_1M_1, F_1M_2, F_1F_2, F_2M_1, F_2M_2, F_2F_1\}$
(b) $A = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2\}$
(c) $B = \{M_1F_1, M_1F_2, M_2F_1, M_2F_2, F_1M_1, F_1M_2, F_2M_1, F_2M_2\}$
(d) $C = \{F_1F_2, F_2F_1\}$
(e) $A \cap B = \{M_1F_1, M_1F_2, M_2F_1, M_2F_2\}$
(f) $A \cup C = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2, F_1F_2, F_2F_1\}$

2.15 (a) {nitrogen, potassium, uranium, oxygen}
(b) {copper, sodium, zinc, oxygen}
(c) {copper, sodium, nitrogen, potassium, uranium, zinc}
(d) {copper, uranium, zinc}
(e) ϕ
(f) {oxygen}

2.19 (a) The family will experience mechanical problems but will receive no ticket for a traffic violation and will not arrive at a campsite that has no vacancies.
(b) The family will receive a traffic ticket and arrive at a campsite that has no vacancies but will not experience mechanical problems.
(c) The family will experience mechanical problems and will arrive at a campsite that has no vacancies.
(d) The family will receive a traffic ticket but will not arrive at a campsite that has no vacancies.
(e) The family will not experience mechanical problems.

2.21 18

2.23 156

2.25 20

2.27 48

2.29 210

2.31 72

2.33 (a) 1024; (b) 243

2.35 362,880

2.37 2880

2.39 (a) 40,320; (b) 336

2.41 360

2.43 24

2.45 3360

2.47 56

2.49 (a) Sum of the probabilities exceeds 1.
(b) Sum of the probabilities is less than 1.
(c) A negative probability
(d) Probability of both a heart and a black card is zero.

2.51 $S = \{\$10, \$25, \$100\}$; $P(10) = \frac{11}{20}$, $P(25) = \frac{3}{10}$, $P(100) = \frac{15}{100}$; $\frac{17}{20}$

2.53 (a) 0.3; (b) 0.2

2.55 10/117

2.57 (a) 5/26; (b) 9/26; (c) 19/26

2.59 (a) 94/54,145; (b) 143/39,984

2.61 (a) 22/25; (b) 3/25; (c) 17/50

2.63 (a) 0.32; (b) 0.68; (c) office or den

2.65 (a) 0.8; (b) 0.45; (c) 0.55

2.67 (a) 0.31; (b) 0.93; (c) 0.31

2.69 (a) 0.009; (b) 0.999; (c) 0.01

2.71 (a) 0.048; (b) \$50,000; (c) \$12,500

2.73 (a) The probability that a convict who pushed dope also committed armed robbery.
(b) The probability that a convict who committed armed robbery did not push dope.
(c) The probability that a convict who did not push dope also did not commit armed robbery.

2.75 (a) 14/39; (b) 95/112

2.77 (a) 5/34; (b) 3/8

2.79 (a) 0.018; (b) 0.614; (c) 0.166; (d) 0.479

2.81 (a) 0.35; (b) 0.875; (c) 0.55

2.83 (a) 9/28; (b) 3/4; (c) 0.91

2.85 0.27

2.87 5/8

2.89 (a) 0.0016; (b) 0.9984

2.91 (a) 91/323; (b) 91/323

2.93 (a) 0.75112; (b) 0.2045

2.95 0.0960

2.97 0.40625

2.99 0.1124

2.101 0.857

Chapter 3

3.1 Discrete; continuous; continuous; discrete; discrete; continuous

3.3

Sample Space	w
HHH	3
HHT	1
HTH	1
THH	1
HTT	−1
THT	−1
TTH	−1
TTT	−3

3.5 (a) 1/30; (b) 1/10

3.7 (a) 0.68; (b) 0.375

3.9 (b) 19/80

3.11

x	0	1	2
$f(x)$	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

3.13
$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ 0.41, & \text{for } 0 \le x < 1, \\ 0.78, & \text{for } 1 \le x < 2, \\ 0.94, & \text{for } 2 \le x < 3, \\ 0.99, & \text{for } 3 \le x < 4, \\ 1, & \text{for } x \ge 4 \end{cases}$$

3.15
$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ \frac{2}{7}, & \text{for } 0 \le x < 1, \\ \frac{6}{7}, & \text{for } 1 \le x < 2, \\ 1, & \text{for } x \ge 2 \end{cases}$$
(a) 4/7; (b) 5/7

3.17 (b) 1/4; (c) 0.3

3.19 $F(x) = \begin{cases} 0, & x < 1 \\ \frac{x-1}{2}, & 1 \le x < 3 \\ 1, & x \ge 3 \end{cases}$; 1/4

3.21 (a) 3/2; (b) $F(x) = \begin{cases} 0, & x < 0 \\ x^{3/2}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$; 0.3004

3.23

$$F(w) = \begin{cases} 0, & \text{for } w < -3, \\ \frac{1}{27}, & \text{for } -3 \le w < -1, \\ \frac{7}{27}, & \text{for } -1 \le w < 1, \\ \frac{19}{27}, & \text{for } 1 \le w < 3, \\ 1, & \text{for } w \ge 3 \end{cases}$$

(a) 20/27; (b) 2/3

3.25

t	20	25	30
$P(T = t)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

3.27 (a)

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-x/2000), & x \ge 0 \end{cases}$$

(b) 0.6065; (c) 0.6321

3.29 (b)

$$F(x) = \begin{cases} 0, & x < 1, \\ 1 - x^{-3}, & x \ge 1 \end{cases}$$

(c) 0.0156

3.31 (a) 0.2231; (b) 0.2212

3.33 (a) $k = 280$; (b) 0.3633; (c) 0.0563

3.35 (a) 0.1528; (b) 0.0446

3.37 (a) 1/36; (b) 1/15

3.39 (a)

		x			
	$f(x,y)$	0	1	2	3
y	0	0	$\frac{3}{70}$	$\frac{9}{70}$	$\frac{3}{70}$
	1	$\frac{2}{70}$	$\frac{18}{70}$	$\frac{18}{70}$	$\frac{2}{70}$
	2	$\frac{3}{70}$	$\frac{9}{70}$	$\frac{3}{70}$	0

(b) 1/2

3.41 (a) 1/16; (b) $g(x) = 12x(1-x)^2$, for $0 \le x \le 1$; (c) 1/4

3.43 (a) 3/64; (b) 1/2

3.45 0.6534

3.47 (a) Dependent; (b) 1/3

3.49 (a)

x	1	2	3
$g(x)$	0.10	0.35	0.55

(b)

y	1	2	3
$h(y)$	0.20	0.50	0.30

(c) 0.2857

3.51 (a)

		x			
	$f(x,y)$	0	1	2	3
y	0	$\frac{1}{55}$	$\frac{6}{55}$	$\frac{6}{55}$	$\frac{1}{55}$
	1	$\frac{6}{55}$	$\frac{16}{55}$	$\frac{6}{55}$	0
	2	$\frac{6}{55}$	$\frac{6}{55}$	0	0
	3	$\frac{1}{55}$	0	0	0

(b) 42/55

3.53 5/8

3.55 Independent

3.57 (a) 3; (b) 21/512

3.59 Dependent

Chapter 4

4.1 0.88

4.3 25¢

4.5 \$1.23

4.7 \$500

4.9 \$6900

4.11 $(\ln 4)/\pi$

4.13 100 hours

4.15 0

4.17 209

4.19 \$1855

4.21 \$833.33

4.23 (a) 35.2; (b) $\mu_X = 3.20$, $\mu_Y = 3.00$

4.25 2

4.27 2000 hours

4.29 (b) 3/2

4.31 (a) 1/6; (b) $(5/6)^5$

4.33 \$5,250,000

4.35 0.74

4.37 1/18; in terms of actual profit, the variance is $\frac{1}{18}(5000)^2$

4.39 1/6

4.41 118.9

4.43 $\mu_Y = 10$; $\sigma_Y^2 = 144$

4.45 0.01

4.47 -0.0062

4.49 $\sigma_X^2 = 0.8456$, $\sigma_X = 0.9196$

4.51 $-1/\sqrt{5}$

4.53 $\mu_{g(X)} = 10.33$, $\sigma_{g(X)} = 6.66$

4.55 \$0.80

4.57 209

4.59 $\mu = 7/2$, $\sigma^2 = 15/4$

4.61 3/14

4.63 52

4.65 (a) 7; (b) 0; (c) 12.25

4.67 46/63

4.69 (a) $E(X) = E(Y) = 1/3$ and $\text{Var}(X) = \text{Var}(Y) = 4/9$; (b) $E(Z) = 2/3$ and $\text{Var}(Z) = 8/9$

4.71 (a) 4; (b) 32; 16

4.73 By direct calculation, $E(e^Y) = 1884.32$. Using the second-order approximation, $E(e^Y) \approx 1883.38$, which is very close to the true value.

4.75 0.03125

4.77 (a) At most 4/9; (b) at least 5/9; (c) at least 21/25; (d) 10

Chapter 5

5.1 $\mu = \frac{1}{k}\sum_{i=1}^{k} x_i$, $\sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$

5.3 $f(x) = \frac{1}{10}$, for $x = 1, 2, \ldots, 10$, and $f(x) = 0$ elsewhere; 3/10

5.5 (a) 0.0480; (b) 0.2375; (c) $P(X = 5 \mid p = 0.3) = 0.1789$, $P = 0.3$ is reasonable.

5.7 (a) 0.0474; (b) 0.0171

5.9 (a) 0.7073; (b) 0.4613; (c) 0.1484

5.11 0.1240

5.13 0.8369

5.15 (a) 0.0778; (b) 0.3370; (c) 0.0870

5.17 $\mu = 3.5$, $\sigma^2 = 1.05$

5.19 $f(x_1, x_2, x_3) = \binom{n}{x_1, x_2, x_3} 0.35^{x_1} 0.05^{x_2} 0.60^{x_3}$

5.21 0.0095

5.23 0.0077

5.25 0.8670

5.27 (a) 0.2852; (b) 0.9887; (c) 0.6083

5.29 5/14

5.31 $h(x; 6, 3, 4) = \frac{\binom{4}{x}\binom{2}{3-x}}{\binom{6}{3}}$, for $x = 1, 2, 3$; $P(2 \le X \le 3) = 4/5$

5.33 (a) 0.3246; (b) 0.4496

5.35 0.9517

5.37 (a) 0.6815; (b) 0.1153

5.39 0.9453

5.41 0.6077

5.43 (a) 4/33; (b) 8/165

5.45 0.2315

5.47 (a) 0.3991; (b) 0.1316

5.49 0.0515

5.51 63/64

5.53 (a) 0.3840; (b) 0.0067

5.55 (a) 0.0630; (b) 0.9730

5.57 (a) 0.1429; (b) 0.1353

5.59 (a) 0.1638; (b) 0.032

5.61 0.2657

5.63 $\mu = 6$, $\sigma^2 = 6$

5.65 (a) 0.2650; (b) 0.9596

5.67 (a) 0.8243; (b) 14

5.69 4

5.71 5.53×10^{-4}; $\mu = 7.5$

5.73 (a) 0.0137; (b) 0.0830

5.75 0.4686

Chapter 6

6.3 (a) 0.6; (b) 0.7; (c) 0.5

6.5 (a) 0.0823; (b) 0.0250; (c) 0.2424;
(d) 0.9236; (e) 0.8133; (f) 0.6435

6.7 (a) 0.54; (b) -1.72; (c) 1.28

6.9 (a) 0.1151; (b) 16.1; (c) 20.275; (d) 0.5403

6.11 (a) 0.0548; (b) 0.4514; (c) 23 cups;
(d) 189.95 milliliters

6.13 (a) 0.8980; (b) 0.0287; (c) 0.6080

6.15 (a) 0.0571; (b) 99.11%; (c) 0.3974;
(d) 27.952 minutes; (e) 0.0092

6.17 6.24 years

6.19 (a) 51%; (b) $18.37

6.21 (a) 0.0401; (b) 0.0244

6.23 26 students

6.25 (a) 0.3085; (b) 0.0197

6.27 (a) 0.9514; (b) 0.0668

6.29 (a) 0.1171; (b) 0.2049

6.31 0.1357

6.33 (a) 0.0778; (b) 0.0571; (c) 0.6811

6.35 (a) 0.8749; (b) 0.0059

6.37 (a) 0.0228; (b) 0.3974

6.41 $2.8e^{-1.8} - 3.4e^{-2.4} = 0.1545$

6.43 (a) $\mu = 6$; $\sigma^2 = 18$;
(b) between 0 and 14.485 million liters

6.45 $\sum_{x=4}^{6} \binom{6}{x} (1 - e^{-3/4})^x (e^{-3/4})^{6-x} = 0.3968$

6.47 (a) $\sqrt{\pi/2} = 1.2533$ years; (b) e^{-2}

6.49 (a) Mean $= 0.25$, median $= 0.206$; (b) variance $= 0.0375$; (c) 0.2963

6.51 $e^{-4} = 0.0183$

6.53 (a) $\mu = \alpha\beta = 50$; (b) $\sigma^2 = \alpha\beta^2 = 500$; $\sigma = \sqrt{500}$; (c) 0.815

6.55 (a) 0.1889; (b) 0.0357

6.57 Mean $= e^6$, variance $= e^{12}(e^4 - 1)$

6.59 (a) e^{-5}; (b) $\beta = 0.2$

Chapter 7

7.1 $g(y) = 1/3$, for $y = 1, 3, 5$

7.3
$$g(y_1, y_2) = \binom{2}{\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}, 2 - y_1} \times \left(\frac{1}{4}\right)^{(y_1+y_2)/2} \left(\frac{1}{3}\right)^{(y_1-y_2)/2} \left(\frac{5}{12}\right)^{2-y_1};$$
for $y_1 = 0, 1, 2$; $y_2 = -2, -1, 0, 1, 2$;
$y_2 \leq y_1$; $y_1 + y_2 = 0, 2, 4$

7.7 Gamma distribution with $\alpha = 3/2$ and $\beta = m/2b$

7.9 (a) $g(y) = 32/y^3$, for $y > 4$; (b) 1/4

7.11 $h(z) = 2(1 - z)$, for $0 < z < 1$

7.13 $h(w) = 6 + 6w - 12w^{1/2}$, for $0 < w < 1$

7.15 $g(y) = \begin{cases} \frac{2}{9\sqrt{y}}, & 0 < y < 1, \\ \frac{\sqrt{y}+1}{9\sqrt{y}}, & 1 < y < 4 \end{cases}$

7.19 Both equal μ.

7.23 (a) Gamma(2, 1); (b) Uniform(0, 1)

Chapter 8

8.1 (a) Responses of all people in Richmond who have a telephone;
 (b) Outcomes for a large or infinite number of tosses of a coin;
 (c) Length of life of such tennis shoes when worn on the professional tour;
 (d) All possible time intervals for this lawyer to drive from her home to her office.

8.3 (a) $\bar{x} = 3.2$ seconds; (b) $\tilde{x} = 3.1$ seconds

8.5 (a) $\bar{x} = 2.4$; (b) $\tilde{x} = 2$; (c) $m = 3$

8.7 (a) 53.75; (b) 75 and 100

8.9 (a) Range is 10; (b) $s = 3.307$

8.11 (a) 2.971; (b) 2.971

8.13 $s = 0.585$

8.15 (a) 45.9; (b) 5.1

8.17 0.3159

8.19 (a) Variance is reduced from 0.49 to 0.16.
(b) Variance is increased from 0.04 to 0.64.

8.21 Yes

8.23 (a) $\mu = 5.3$; $\sigma^2 = 0.81$;
(b) $\mu_{\bar{X}} = 5.3$; $\sigma^2_{\bar{X}} = 0.0225$;
(c) 0.9082

8.25 (a) 0.6898; (b) 7.35

8.29 0.5596

8.31 (a) The chance that the difference in mean drying time is larger than 1.0 is 0.0013; (b) 13

8.33 (a) 1/2; (b) 0.3085

8.35 $P(\bar{X} \leq 775 \mid \mu = 760) = 0.9332$

8.37 (a) 27.488; (b) 18.475; (c) 36.415

8.39 (a) 0.297; (b) 32.852; (c) 46.928

8.41 (a) 0.05; (b) 0.94

8.45 (a) 0.975; (b) 0.10; (c) 0.875; (d) 0.99

8.47 (a) 2.500; (b) 1.319; (c) 1.714

8.49 No; $\mu > 20$

8.51 (a) 2.71; (b) 3.51; (c) 2.92;
(d) 0.47; (e) 0.34

8.53 The F-ratio is 1.44. The variances are not significantly different.

Chapter 9

9.1 56

9.3 $0.3097 < \mu < 0.3103$

9.5 (a) $22,496 < \mu < 24,504$; (b) error ≤ 1004

9.7 35

9.9 $10.15 < \mu < 12.45$

9.11 $0.978 < \mu < 1.033$

9.13 $47.722 < \mu < 49.278$

9.15 $(13,075, 33,925)$

9.17 $(6.05, 16.55)$

9.19 323.946 to 326.154

9.21 Upper prediction limit: 9.42;
upper tolerance limit: 11.72

9.25 Yes, the value of 6.9 is outside of the prediction interval.

9.27 (a) $(0.9876, 1.0174)$;
(b) $(0.9411, 1.0639)$;
(c) $(0.9334, 1.0716)$

9.35 $2.9 < \mu_1 - \mu_2 < 7.1$

9.37 $2.80 < \mu_1 - \mu_2 < 3.40$

9.39 $1.5 < \mu_1 - \mu_2 < 12.5$

9.41 $0.70 < \mu_1 - \mu_2 < 3.30$

9.43 $-6536 < \mu_1 - \mu_2 < 2936$

9.45 $(-0.74, 6.30)$

9.47 $(-6.92, 36.70)$

9.49 $0.54652 < \mu_B - \mu_A < 1.69348$

9.51 Method 1: $0.194 < p < 0.262$; method 2: $0.1957 < p < 0.2639$

9.53 (a) $0.498 < p < 0.642$; (b) error ≤ 0.072

9.55 (a) $0.739 < p < 0.961$; (b) no

9.57 (a) $0.644 < p < 0.690$; (b) error ≤ 0.023

9.59 2576

9.61 160

9.63 9604

9.65 $-0.0136 < p_F - p_M < 0.0636$

9.67 $0.0011 < p_1 - p_2 < 0.0869$

9.69 $(-0.0849, 0.0013)$; not significantly different

9.71 $0.293 < \sigma^2 < 6.736$; valid claim

9.73 $3.472 < \sigma^2 < 12.804$

9.75 $9.27 < \sigma < 34.16$

9.77 $0.549 < \sigma_1/\sigma_2 < 2.690$

9.79 $0.016 < \sigma_1^2/\sigma_2^2 < 0.454$; no

9.81 $\frac{1}{n}\sum_{i=1}^{n} x_i$

9.83 $\hat{\beta} = \bar{x}/5$

9.85 $\hat{\theta} = \max\{x_1, \ldots, x_n\}$

9.87 $x \ln p + (1-x)\ln(1-p)$. Set the derivative with respect to $p = 0$; $\hat{p} = x = 1.0$

Chapter 10

10.1 (a) Conclude that less than 30% of the public is allergic to some cheese products when, in fact, 30% or more is allergic.
(b) Conclude that at least 30% of the public is allergic to some cheese products when, in fact, less than 30% is allergic.

10.3 (a) The firm is not guilty;
(b) the firm is guilty.

10.5 (a) 0.0559;
(b) $\beta = 0.0017$; $\beta = 0.00968$; $\beta = 0.5557$

10.7 (a) 0.1286;
(b) $\beta = 0.0901$; $\beta = 0.0708$.
(c) The probability of a type I error is somewhat large.

10.9 (a) $\alpha = 0.0850$; (b) $\beta = 0.3410$

10.11 (a) $\alpha = 0.1357$; (b) $\beta = 0.2578$

10.13 $\alpha = 0.0094$; $\beta = 0.0122$

10.15 (a) $\alpha = 0.0718$; (b) $\beta = 0.1151$

10.17 (a) $\alpha = 0.0384$; (b) $\beta = 0.5$; $\beta = 0.2776$

10.19 $z = -2.76$; yes, $\mu < 40$ months; P-value $= 0.0029$

10.21 $z = -1.64$; P-value $= 0.10$

10.23 $t = 0.77$; fail to reject H_0.

10.25 $z = 8.97$; yes, $\mu > 20,000$ kilometers; P-value < 0.001

10.27 $t = 12.72$; P-value < 0.0005; reject H_0.

10.29 $t = -1.98$; P-value $= 0.0312$; reject H_0

10.31 $z = -2.60$; conclude $\mu_A - \mu_B \leq 12$ kilograms.

10.33 $t = 1.50$; there is not sufficient evidence to conclude that the increase in substrate concentration would cause an increase in the mean velocity of more than 0.5 micromole per 30 minutes.

10.35 $t = 0.70$; there is not sufficient evidence to support the conclusion that the serum is effective.

10.37 $t = 2.55$; reject H_0: $\mu_1 - \mu_2 > 4$ kilometers.

10.39 $t' = 0.22$; fail to reject H_0.

10.41 $t' = 2.76$; reject H_0.

10.43 $t = -2.53$; reject H_0; the claim is valid.

10.45 $t = 2.48$; P-value < 0.02; reject H_0.

10.47 $n = 6$

10.49 $78.28 \approx 79$

10.51 5

10.53 (a) H_0: $M_{\text{hot}} - M_{\text{cold}} = 0$,
H_1: $M_{\text{hot}} - M_{\text{cold}} \neq 0$;
(b) paired t, $t = 0.99$; P-value > 0.30; fail to reject H_0.

10.55 P-value $= 0.4044$ (with a one-tailed test); the claim is not refuted.

10.57 $z = 1.44$; fail to reject H_0.

10.59 $z = -5.06$ with P-value ≈ 0; conclude that fewer than one-fifth of the homes are heated by oil.

10.61 $z = 0.93$ with P-value $= P(Z > 0.93) = 0.1762$; there is not sufficient evidence to conclude that the new medicine is effective.

10.63 $z = 2.36$ with P-value $= 0.0182$; yes, the difference is significant.

10.65 $z = 1.10$ with P-value $= 0.1357$; we do not have sufficient evidence to conclude that breast cancer is more prevalent in the urban community.

10.67 $\chi^2 = 18.13$ with P-value $= 0.0676$ (from computer output); do not reject H_0: $\sigma^2 = 0.03$.

10.69 $\chi^2 = 63.75$ with P-value $= 0.8998$ (from computer output); do not reject H_0.

10.71 $\chi^2 = 42.37$ with P-value $= 0.0117$ (from computer output); machine is out of control.

10.73 $f = 1.33$ with P-value $= 0.3095$ (from computer output); fail to reject H_0: $\sigma_1 = \sigma_2$.

10.75 $f = 0.086$ with P-value $= 0.0328$ (from computer output); reject H_0: $\sigma_1 = \sigma_2$ at level greater than 0.0328.

10.77 $f = 19.67$ with P-value $= 0.0008$ (from computer output); reject H_0: $\sigma_1 = \sigma_2$.

10.79 $\chi^2 = 10.14$; reject H_0, the ratio is not 5:2:2:1.

10.81 $\chi^2 = 4.47$; there is not sufficient evidence to claim that the die is unbalanced.

10.83 $\chi^2 = 3.125$; do not reject H_0: geometric distribution.

10.85 $\chi^2 = 5.19$; do not reject H_0: normal distribution.

10.87 $\chi^2 = 5.47$; do not reject H_0.

10.89 $\chi^2 = 124.59$; yes, occurrence of these types of crime is dependent on the city district.

10.91 $\chi^2 = 5.92$ with P-value $= 0.4332$; do not reject H_0.

10.93 $\chi^2 = 31.17$ with P-value < 0.0001; attitudes are not homogeneous.

10.95 $\chi^2 = 1.84$; do not reject H_0.

Chapter 11

11.1 (a) $b_0 = 64.529$, $b_1 = 0.561$;
(b) $\hat{y} = 81.4$

11.3 (a) $\hat{y} = 5.8254 + 0.5676x$;
(c) $\hat{y} = 34.205$ at 50°C

11.5 (a) $\hat{y} = 6.4136 + 1.8091x$;
(b) $\hat{y} = 9.580$ at temperature 1.75

11.7 (b) $\hat{y} = 31.709 + 0.353x$

11.9 (b) $\hat{y} = 343.706 + 3.221x$;
(c) $\hat{y} = \$456$ at advertising costs $= \$35$

11.11 (b) $\hat{y} = -1847.633 + 3.653x$

11.13 (a) $\hat{y} = 153.175 - 6.324x$;
(b) $\hat{y} = 123$ at $x = 4.8$ units

11.15 (a) $s^2 = 176.4$;
(b) $t = 2.04$; fail to reject H_0: $\beta_1 = 0$.

11.17 (a) $s^2 = 0.40$;
(b) $4.324 < \beta_0 < 8.503$;
(c) $0.446 < \beta_1 < 3.172$

11.19 (a) $s^2 = 6.626$;
(b) $2.684 < \beta_0 < 8.968$;
(c) $0.498 < \beta_1 < 0.637$

11.21 $t = -2.24$; reject H_0 and conclude $\beta < 6$

11.23 (a) $24.438 < \mu_{Y|24.5} < 27.106$;
(b) $21.88 < y_0 < 29.66$

11.25 $7.81 < \mu_{Y|1.6} < 10.81$

11.27 (a) 17.1812 mpg;
(b) no, the 95% confidence interval on mean mpg is $(27.95, 29.60)$;
(c) miles per gallon will likely exceed 18.

11.29 (b) $\hat{y} = 3.4156x$

11.31 The f-value for testing the lack of fit is 1.58, and the conclusion is that H_0 is not rejected. Hence, the lack-of-fit test is insignificant.

11.33 (a) $\hat{y} = 2.003x$;
(b) $t = 1.40$, fail to reject H_0.

11.35 $f = 1.71$ and P-value $= 0.2517$; the regression is linear.

11.37 (a) $b_0 = 10.812$, $b_1 = -0.3437$;
(b) $f = 0.43$; the regression is linear.

11.39 (a) $\hat{P} = -11.3251 - 0.0449T$;
(b) yes;
(c) $R^2 = 0.9355$;
(d) yes

11.41 (b) $\hat{N} = -175.9025 + 0.0902Y$; $R^2 = 0.3322$

11.43 $r = 0.240$

11.45 (a) $r = -0.979$;
(b) P-value $= 0.0530$; do not reject H_0 at 0.025 level;
(c) 95.8%

11.47 (a) $r = 0.784$;
(b) reject H_0 and conclude that $\rho > 0$;
(c) 61.5%.

Chapter 12

12.1 $\hat{y} = 0.5800 + 2.7122x_1 + 2.0497x_2$

12.3 (a) $\hat{y} = 27.547 + 0.922x_1 + 0.284x_2$;
(b) $\hat{y} = 84$ at $x_1 = 64$ and $x_2 = 4$

12.5 (a) $\hat{y} = -102.7132 + 0.6054x_1 + 8.9236x_2 + 1.4374x_3 + 0.0136x_4$;
(b) $\hat{y} = 287.6$

12.7 $\hat{y} = 141.6118 - 0.2819x + 0.0003x^2$

12.9 (a) $\hat{y} = 56.4633 + 0.1525x - 0.00008x^2$;
(b) $\hat{y} = 86.7\%$ when temperature is at 225°C

12.11 $\hat{y} = -6.5122 + 1.9994x_1 - 3.6751x_2 + 2.5245x_3 + 5.1581x_4 + 14.4012x_5$

12.13 (a) $\hat{y} = 350.9943 - 1.2720x_1 - 0.1539x_2$;
(b) $\hat{y} = 140.9$

12.15 $\hat{y} = 3.3205 + 0.4210x_1 - 0.2958x_2 + 0.0164x_3 + 0.1247x_4$

12.17 0.1651

12.19 242.72

12.21 (a) $\hat{\sigma}^2_{B_2} = 28.0955$; (b) $\hat{\sigma}_{B_1B_2} = -0.0096$

12.23 $t = 5.91$ with P-value $= 0.0002$. Reject H_0 and claim that $\beta_1 \neq 0$.

12.25 $0.4516 < \mu_{Y|x_1=900,x_2=1} < 1.2083$ and $-0.1640 < y_0 < 1.8239$

12.27 $263.7879 < \mu_{Y|x_1=75,x_2=24,x_3=90,x_4=98} < 311.3357$ and $243.7175 < y_0 < 331.4062$

12.29 (a) $t = -1.09$ with P-value $= 0.3562$;
(b) $t = -1.72$ with P-value $= 0.1841$;
(c) yes; not sufficient evidence to show that x_1 and x_2 are significant

12.31 $R^2 = 0.9997$

12.33 $f = 5.106$ with P-value $= 0.0303$; the regression is not significant at level 0.01.

12.35 $f = 34.90$ with P-value $= 0.0002$; reject H_0 and conclude $\beta_1 > 0$.

12.37 $f = 10.18$ with P-value < 0.01; x_1 and x_2 are significant in the presence of x_3 and x_4.

12.39 The two-variable model is better.

12.41 First model: $R^2_{\text{adj}} = 92.7\%$, C.V. = 9.0385.
Second model: $R^2_{\text{adj}} = 98.1\%$, C.V. = 4.6287.
The partial F-test shows P-value $= 0.0002$; model 2 is better.

12.43 Using x_2 alone is not much different from using x_1 and x_2 together since the R^2_{adj} are 0.7696 versus 0.7591, respectively.

12.45 (a) $\widehat{\text{mpg}} = 5.9593 - 0.00003773$ odometer $+ 0.3374$ octane $- 12.6266z_1 - 12.9846z_2$;
(b) sedan;
(c) they are not significantly different.

12.47 (b) $\hat{y} = 4.690$ seconds;
(c) $4.450 < \mu_{Y|\{180,260\}} < 4.930$

12.49 $\hat{y} = 2.1833 + 0.9576x_2 + 3.3253x_3$

12.51 (a) $\hat{y} = -587.211 + 428.433x$;
(b) $\hat{y} = 1180 - 191.691x + 35.20945x^2$;
(c) quadratic model

12.53 $\hat{\sigma}^2_{B_1} = 20{,}588$; $\hat{\sigma}^2_{B_{11}} = 62.6502$;
$\hat{\sigma}_{B_1,B_{11}} = -1103.5$

12.55 (a) Intercept model is the best.

12.57 (a) $\hat{y} = 3.1368 + 0.6444x_1 - 0.0104x_2 + 0.5046x_3 - 0.1197x_4 - 2.4618x_5 + 1.5044x_6$;
(b) $\hat{y} = 4.6563 + 0.5133x_3 - 0.1242x_4$;
(c) C_p criterion: variables x_1 and x_2 with $s^2 = 0.7317$ and $R^2 = 0.6476$; s^2 criterion: variables x_1, x_3 and x_4 with $s^2 = 0.7251$ and $R^2 = 0.6726$;
(d) $\hat{y} = 4.6563 + 0.5133x_3 - 0.1242x_4$; This one does not lose much in s^2 and R^2;
(e) two observations have large R-Student values and should be checked.

12.59 (a) $\hat{y} = 125.8655 + 7.7586x_1 + 0.0943x_2 - 0.0092x_1x_2$;
(b) the model with x_2 alone is the best.

12.61 (a) $\hat{p} = (1 + e^{2.9949 - 0.0308x})^{-1}$;
(b) 1.8515

Chapter 13

13.1 $f = 0.31$; not sufficient evidence to support the hypothesis that there are differences among the 6 machines.

13.3 $f = 14.52$; yes, the difference is significant.

13.5 $f = 8.38$; the average specific activities differ significantly.

13.7 $f = 2.25$; not sufficient evidence to support the hypothesis that the different concentrations of $MgNH_4PO_4$ significantly affect the attained height of chrysanthemums.

13.9 $b = 0.79 > b_4(0.01, 4, 4, 4, 9) = 0.4939$. Do not reject H_0. There is not sufficent evidence to claim that variances are different.

13.11 $b = 0.7822 < b_4(0.05, 9, 8, 15) = 0.8055$. The variances are significantly different.

13.13 (a) P-value < 0.0001, significant,

(b) for contrast 1 vs. 2, P-value $< 0.0001Z$, significantly different; for contrast 3 vs. 4, P-value $= 0.0648$, not significantly different

13.15 Results of Tukey's tests are given below.

$\bar{y}_{4.}$	$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{5.}$	$\bar{y}_{2.}$
2.98	4.30	5.44	6.96	7.90

13.17 (a) P-value $= 0.0121$; yes, there is a significant difference.

(b)

Depletion	Modified Hess	Substrate Removal Kicknet	Surber	Kicknet

13.19 $f = 70.27$ with P-value < 0.0001; reject H_0.

$\bar{x}_0$	$\bar{x}_{25}$	$\bar{x}_{100}$	$\bar{x}_{75}$	$\bar{x}_{50}$
55.167	60.167	64.167	70.500	72.833

Temperature is important; both 75° and 50°(C) yielded batteries with significantly longer activated life.

13.21 The mean absorption is significantly lower for aggregate 4 than for the other aggregates.

13.23 Comparing the control to 1 and 2: significant; comparing the control to 3 and 4: insignificant

13.25 $f(\text{fertilizer}) = 6.11$; there is significant difference among the fertilizers.

13.27 $f = 5.99$; percent of foreign additives is not the same for all three brands of jam; brand A

13.29 P-value < 0.0001; significant

13.31 P-value $= 0.0023$; significant

13.33 P-value $= 0.1250$; not significant

13.35 P-value < 0.0001;
$f = 122.37$; the amount of dye has an effect on the color of the fabric.

13.37 (a) $y_{ij} = \mu + A_i + \epsilon_{ij}$, $A_i \sim n(x; 0, \sigma_\alpha)$, $\epsilon_{ij} \sim n(x; 0, \sigma)$;

(b) $\hat{\sigma}_\alpha^2 = 0$ (the estimated variance component is -0.00027); $\hat{\sigma}^2 = 0.0206$.

13.39 (a) $f = 14.9$; operators differ significantly;

(b) $\hat{\sigma}_\alpha^2 = 28.91$; $s^2 = 8.32$.

13.41 (a) $y_{ij} = \mu + A_i + \epsilon_{ij}$, $A_i \sim n(x; 0, \sigma_\alpha)$;

(b) yes; $f = 5.63$ with P-value $= 0.0121$;

(c) there is a significant loom variance component.

Chapter 14

14.1 (a) $f = 8.13$; significant;

(b) $f = 5.18$; significant;

(c) $f = 1.63$; insignificant

14.3 (a) $f = 14.81$; significant;

(b) $f = 9.04$; significant;

(c) $f = 0.61$; insignificant

14.5 (a) $f = 34.40$; significant;

(b) $f = 26.95$; significant;

(c) $f = 20.30$; significant

14.7 Test for effect of temperature: $f_1 = 10.85$ with P-value $= 0.0002$;
Test for effect of amount of catalyst: $f_2 = 46.63$ with P-value < 0.0001;
Test for effect of interaction: $f = 2.06$ with P-value $= 0.074$.

14.9 (a)

Source of Variation	**df**	**Sum of Squares**	**Mean Squares**	***f***	***P***
Cutting speed	1	12.000	12.000	1.32	0.2836
Tool geometry	1	675.000	675.000	74.31	< 0.0001
Interaction	1	192.000	192.000	21.14	0.0018
Error	8	72.667	9.083		
Total	11	951.667			

(b) The interaction effect masks the effect of cutting speed;

(c) $f_{\text{tool geometry}=1} = 16.51$ and P-value $= 0.0036$;

$f_{\text{tool geometry}=2} = 5.94$ and P-value $= 0.0407$.

14.11 (a)

Source of Variation	df	Sum of Squares	Mean Squares	f	P
Method	1	0.000104	0.000104	6.57	0.0226
Laboratory	6	0.008058	0.001343	84.70	< 0.0001
Interaction	6	0.000198	0.000033	2.08	0.1215
Error	14	0.000222	0.000016		
Total	27	0.008582			

(b) The interaction is not significant;

(c) Both main effects are significant;

(e) $f_{\text{laboratory}=1} = 0.01576$ and P-value $= 0.9019$; no significant difference between the methods in laboratory 1; $f_{\text{tool geometry}=2} = 9.081$ and P-value $= 0.0093$.

14.13 (b)

Source of Variation	df	Sum of Squares	Mean Squares	f	P
Time	1	0.060208	0.060208	157.07	< 0.0001
Treatment	1	0.060208	0.060208	157.07	< 0.0001
Interaction	1	0.000008	0.000008	.02	0.8864
Error	8	0.003067	0.000383		
Total	11	0.123492			

(c) Both time and treatment influence the magnesium uptake significantly, although there is no significant interaction between them.

(d) $Y = \mu + \beta_T \text{Time} + \beta_Z Z + \beta_{TZ} \text{Time } Z + \epsilon$, where $Z = 1$ when treatment $= 1$ and $Z = 0$ when treatment $= 2$;

(e) $f = 0.02$ with P-value $= 0.8864$; the interaction in the model is insignificant.

14.15 (a) Interaction is significant at a level of 0.05, with P-value of 0.0166.

(b) Both main effects are significant.

14.17 (a) AB: $f = 3.83$; significant;
AC: $f = 3.79$; significant;
BC: $f = 1.31$; not significant;
ABC: $f = 1.63$; not significant;

(b) A: $f = 0.54$; not significant;
B: $f = 6.85$; significant;
C: $f = 2.15$; not significant;

(c) The presence of AC interaction masks the main effect C.

14.19 (a) Stress: $f = 45.96$ with P-value < 0.0001;
coating: $f = 0.05$ with P-value $= 0.8299$;
humidity: $f = 2.13$ with P-value $= 0.1257$;
coating × humidity: $f = 3.41$ with P-value $= 0.0385$;
coating × stress: $f = 0.08$ with P-value $= 0.9277$;
humidity × stress: $f = 3.15$ with P-value $= 0.0192$;
coating × humidity × stress: $f = 1.93$ with P-value $= 0.1138$.

(b) The best combination appears to be uncoated, medium humidity, and a stress level of 20.

14.21

Effect	f	P
Temperature	14.22	< 0.0001
Surface	6.70	0.0020
HRC	1.67	0.1954
T × S	5.50	0.0006
T × HRC	2.69	0.0369
S × HRC	5.41	0.0007
T × S × HRC	3.02	0.0051

14.23 (a) Yes; brand × type; brand × temperature;

(b) yes;

(c) brand Y, powdered detergent, hot temperature.

14.25 (a)

Effect	f	P
Time	543.53	< 0.0001
Temp	209.79	< 0.0001
Solvent	4.97	0.0457
Time × Temp	2.66	0.1103
Time × Solvent	2.04	0.1723
Temp × Solvent	0.03	0.8558
Time × Temp × Solvent	6.22	0.0140

Although three two-way interactions are shown to be insignificant, they may be masked by the significant three-way interaction.

14.27 (a) $f = 1.49$; no significant interaction;

(b) $f(\text{operators}) = 12.45$; significant;
$f(\text{filters}) = 8.39$; significant;

(c) $\hat{\sigma}_\alpha^2 = 0.1777$ (filters);
$\hat{\sigma}_\beta^2 = 0.3516$ (operators);
$s^2 = 0.185$

14.29 (a) $\hat{\sigma}_\beta^2$, $\hat{\sigma}_\gamma^2$, $\hat{\sigma}_{\alpha\gamma}^2$ are significant;

(b) $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_{\alpha\gamma}^2$ are significant.

14.31 (a) Mixed model;

(b) Material: $f = 47.42$ with P-value < 0.0001;
brand: $f = 1.73$ with P-value $= 0.2875$;
material × brand: $f = 16.06$ with P-value $= 0.0004$;

(c) no

Chapter 15

15.1 B and C are significant at level 0.05.

15.3 Factors A, B, and C have negative effects on the phosphorus compound, and factor D has a positive effect. However, the interpretation of the effect of individual factors should involve the use of interaction plots.

15.5 Significant effects:
A: $f = 9.98$; BC: $f = 19.03$.
Insignificant effects:
B: $f = 0.20$; C: $f = 6.54$; D: $f = 0.02$; AB: $f = 1.83$;
AC: $f = 0.20$; AD: $f = 0.57$; BD: $f = 1.83$;
CD: $f = 0.02$. Since the BC interaction is significant, both B and C would be investigated further.

15.9 (a) $b_A = 5.5$, $b_B = -3.25$, and $b_{AB} = 2.5$;

(b) the values of the coefficients are one-half those of the effects;

(c) $t_A = 5.99$ with P-value $= 0.0039$;
$t_B = -3.54$ with P-value $= 0.0241$;
$t_{AB} = 2.72$ with P-value $= 0.0529$;
$t^2 = F$.

15.11 (a) $A = -0.8750$, $B = 5.8750$, $C = 9.6250$, $AB = -3.3750$, $AC = -9.6250$, $BC = 0.1250$, and $ABC = -1.1250$;
B, C, AB, and AC appear important based on their magnitude.

(b)

Effects	***P*-Value**
A	0.7528
B	0.0600
C	0.0071
AB	0.2440
AC	0.0071
BC	0.9640
ABC	0.6861

(c) Yes;

(d) At a high level of A, C essentially has no effect. At a low level of A, C has a positive effect.

15.13 (a)

Machine			
1	**2**	**3**	**4**
(1)	*c*	*a*	*ac*
ab	*d*	*b*	*ad*
cd	*e*	*acd*	*ae*
ce	*abc*	*ace*	*bc*
de	*abd*	*ade*	*bd*
abcd	*abe*	*bcd*	*be*
abce	*cde*	*bce*	*acde*
abde	*abcde*	*bde*	*bcde*

(b) ABD, CDE, $ABCDE$ (one possible design)

15.15 (a) x_2, x_3, x_1x_2, and x_1x_3;

(b) Curvature: P-value $= 0.0038$;

(c) One additional design point different from the original ones

15.17 $(0, -1)$, $(0, 1)$, $(-1, 0)$, $(1, 0)$ might be used.

15.19 (a) With BCD as the defining contrast, the principal block contains (1), *a*, *bc*, *abc*, *bd*, *abd*, *cd*, *acd*;

(b)

Block 1	Block 2
(1)	*a*
bc	*abc*
abd	*bd*
acd	*cd*

confounded by ABC;

(c) Defining contrast BCD produces the following aliases: $A \equiv ABCD$, $B \equiv CD$, $C \equiv BD$, $D \equiv BC$, $AB \equiv ACD$, $AC \equiv ABD$, and $AD \equiv ABC$. Since AD and ABC are confounded with blocks, there are only 2 degrees of freedom for error from the interactions not confounded.

Source of Variation	**Degrees of Freedom**
A	1
B	1
C	1
D	1
Blocks	1
Error	2
Total	7

15.21 (a) With the defining contrasts $ABCE$ and $ABDF$, the principal block contains (1), *ab*, *acd*, *bcd*, *ce*, *abce*, *ade*, *bde*, *acf*, *bcf*, *df*, *abdf*, *aef*, *bef*, *cdef*, *abcdef*;

(b) $A \equiv BCE \equiv BDF \equiv ACDEF$,
$AD \equiv BCDE \equiv BF \equiv ACEF$,
$B \equiv ACE \equiv ADF \equiv BCDEF$,
$AE \equiv BC \equiv BDEF \equiv ACDF$,
$C \equiv ABE \equiv ABCDF \equiv DEF$,
$AF \equiv BCEF \equiv BD \equiv ACDE$,
$D \equiv ABCDE \equiv ABF \equiv CEF$,
$CE \equiv AB \equiv ABCDEF \equiv DF$,
$E \equiv ABC \equiv ABDEF \equiv CDF$,
$DE \equiv ABCD \equiv ABEF \equiv CF$,
$F \equiv ABCEF \equiv ABD \equiv CDE$,
$BCD \equiv ADE \equiv ACF \equiv BEF$,
$AB \equiv CE \equiv DF \equiv ABCDEF$,
$BCF \equiv AEF \equiv ACD \equiv BDE$,
$AC \equiv BE \equiv BCDF \equiv ADEF$;

Source of Variation	Degrees of Freedom
A	1
B	1
C	1
D	1
E	1
F	1
AB	1
AC	1
AD	1
BC	1
BD	1
CD	1
CF	1
Error	2
Total	15

15.23

Source	df	SS	MS	f	P
A	1	6.1250	6.1250	5.81	0.0949
B	1	0.6050	0.6050	0.57	0.5036
C	1	4.8050	4.8050	4.56	0.1223
D	1	0.2450	0.2450	0.23	0.6626
Error	3	3.1600	1.0533		
Total	7	14.9400			

15.25

Source	df	SS	MS	f	P
A	1	388,129.00	388,129.00	3585.49	0.0001
B	1	277,202.25	277,202.25	2560.76	0.0001
C	1	4692.25	4692.25	43.35	0.0006
D	1	9702.25	9702.25	89.63	0.0001
E	1	1806.25	1806.25	16.69	0.0065
AD	1	1406.25	1406.25	12.99	0.0113
AE	1	462.25	462.25	4.27	0.0843
BD	1	1156.00	1156.00	10.68	0.0171
BE	1	961.00	961.00	8.88	0.0247
Error	6	649.50	108.25		
Total	15	686,167.00			

All main effects are significant at the 0.05 level; AD, BD, and BE are also significant at the 0.05 level.

15.27 The principal block contains af, be, cd, abd, ace, bcf, def, $abcdef$.

15.29 $A \equiv BD \equiv CE \equiv CDF \equiv BEF \equiv ABCF \equiv ADEF \equiv ABCDE$;
$B \equiv AD \equiv CF \equiv CDE \equiv AEF \equiv ABCE \equiv BDEF \equiv ABCDF$;
$C \equiv AE \equiv BF \equiv BDE \equiv ADF \equiv CDEF \equiv ABCD \equiv ABCEF$;
$D \equiv AB \equiv EF \equiv BCE \equiv ACF \equiv BCDF \equiv ACDE \equiv ABDEF$;
$E \equiv AC \equiv DF \equiv ABF \equiv BCD \equiv ABDE \equiv BCEF \equiv ACDEF$;
$F \equiv BC \equiv DE \equiv ACD \equiv ABE \equiv ACEF \equiv ABDF \equiv BCDEF$.

15.31 $x_1 = 1$ and $x_2 = 1$

15.33 (a) Yes;
(b) (i) $E(\hat{y}) = 79.00 + 5.281A$;
(ii) $\text{Var}(\hat{y}) = 6.22^2\sigma_Z^2 + 5.70^2A^2\sigma_Z^2 + 2(6.22)(5.70)A\sigma_Z^2$;
(c) velocity at low level;
(d) velocity at low level;
(e) yes

15.35 $\hat{y} = 12.7519 + 4.7194x_1 + 0.8656x_2 - 1.4156x_3$; units are centered and scaled; test for lack of fit, $F = 81.58$, with P-value < 0.0001.

15.37 AFG, BEG, CDG, DEF, $CEFG$, $BDFG$, $BCDE$, $ADEG$, $ACDF$, $ABEF$, and $ABCDEFG$

Chapter 16

16.1 $x = 7$ with P-value $= 0.1719$; fail to reject H_0.

16.3 $x = 3$ with P-value $= 0.0244$; reject H_0.

16.5 $x = 4$ with P-value $= 0.3770$; fail to reject H_0.

16.7 $x = 4$ with P-value $= 0.1335$; fail to reject H_0.

16.9 $w = 43$; fail to reject H_0.

16.11 $w_+ = 17.5$; fail to reject H_0.

16.13 $w_+ = 15$ with $n = 13$; reject H_0 in favor of $\tilde{\mu}_1 - \tilde{\mu}_2 < 8$.

16.15 $u_1 = 4$; claim is not valid.

16.17 $u_2 = 5$; A operates longer.

16.19 $u = 15$; fail to reject H_0.

16.21 $h = 10.58$; operating times are different.

16.23 $v = 7$ with P-value $= 0.910$; random sample.

16.25 $v = 6$ with P-value $= 0.044$; fail to reject H_0.

16.27 $v = 4$; random sample.

16.29 0.70

16.31 0.995

16.33 (a) $r_s = 0.39$; (b) fail to reject H_0.

16.35 (a) $r_s = 0.72$; (b) reject H_0, so $\rho > 0$.

16.37 (a) $r_s = 0.71$; (b) reject H_0, so $\rho > 0$.

Chapter 18

18.1 $p^* = 0.173$

18.3 (a) $\pi(p \mid x = 1) = 40p(1-p)^3/0.2844$; $0.05 < p < 0.15$;
(b) $p^* = 0.106$

18.5 (a) $beta(95, 45)$; (b) 1

18.7 $8.077 < \mu < 8.692$

18.9 (a) 0.2509; (b) $68.71 < \mu < 71.69$; (c) 0.0174

18.13 $p^* = \frac{6}{x+2}$

18.15 2.21

Index