x	0.1	0.2	0.3	0.4	0.5
y	0.06	0.12	0.36	0.65	0.95

Compute D and $d_{\max}$ to bound $c_{\max}$. Compare the results with your solution to Problem 3 in Section 3.2.

4. Make an appropriate transformation to fit the model $P = ae^{bt}$ using Equation (3.4). Estimate a and b.

t	7	14	21	28	35	42
P	8	41	133	250	280	297

5. Examine closely the system of equations that result when you fit the quadratic in Problem 3. Suppose $c_2 = 0$. What would be the corresponding system of equations? Repeat for the cases $c_1 = 0$ and $c_3 = 0$. Suggest a system of equations for a cubic. Check your result. Explain how you would generalize the system of Equation (3.4) to fit any polynomial. Explain what you would do if one or more of the coefficients in the polynomial were zero.

6. A general rule for computing a person's weight is as follows: For a female, multiply the height in inches by 3.5 and subtract 108; for a male, multiply the height in inches by 4.0 and subtract 128. If the person is small bone-structured, adjust this computation by subtracting 10%; for a large bone-structured person, add 10%. No adjustment is made for an average-size person. Gather data on the weight versus height of people of differing age, size, and gender. Using Equation (3.4), fit a straight line to your data for males and another straight line to your data for females. What are the slopes and intercepts of those lines? How do the results compare with the general rule?

In Problems 7–10, fit the data with the models given, using least squares.

7.

x	1	2	3	4	5
y	1	1	2	2	4

a. $y = b + ax$

b. $y = ax^2$

8. Data for stretch of a spring

$x(\times 10^{-3})$	5	10	20	30	40	50	60	70	80	90	100
$y(\times 10^{-5})$	0	19	57	94	134	173	216	256	297	343	390

a. $y = ax$

b. $y = b + ax$

c. $y = ax^2$

9. Data for the ponderosa pine

x	17	19	20	22	23	25	28	31	32	33	36	37	39	42
y	19	25	32	51	57	71	113	140	153	187	192	205	250	260

a. $y = ax + b$

b. $y = ax^2$

c. $y = ax^3$

d. $y = ax^3 + bx^2 + c$

10. Data for planets

Body	Period (sec)	Distance from sun (m)
Mercury	7.60×10^6	5.79×10^{10}
Venus	1.94×10^7	1.08×10^{11}
Earth	3.16×10^7	1.5×10^{11}
Mars	5.94×10^7	2.28×10^{11}
Jupiter	3.74×10^8	7.79×10^{11}
Saturn	9.35×10^8	1.43×10^{12}
Uranus	2.64×10^9	2.87×10^{12}
Neptune	5.22×10^9	4.5×10^{12}

Fit the model $y = ax^{3/2}$.

3.3 PROJECTS

1. Compete the requirements of the module "Curve Fitting via the Criterion of Least Squares," by John W. Alexander, Jr., UMAP 321. (See enclosed CD for UMAP module.) This unit provides an easy introduction to correlations, scatter diagrams (polynomial, logarithmic, and exponential scatters), and lines and curves of regression. Students construct scatter diagrams, choose appropriate functions to fit specific data, and use a computer program to fit curves. Recommended for students who wish an introduction to statistical measures of correlation.

2. Select a project from Projects 1–7 in Section 2.3 and use least squares to fit your proposed proportionality model. Compare your least-squares results with the model used from Section 2.3. Find the bounds on the Chebyshev criterion and interpret the results.

Further Reading

Burden, Richard L., & J. Douglas Faires. *Numerical Analysis*, 7th ed. Pacific Grove, CA: Brooks/Cole, 2001.

Cheney, E. Ward, & David Kincaid. *Numerical Mathematics and Computing.* Monterey, CA: Brooks/Cole, 1984.

Cheney, E. Ward, & David Kincaid. *Numerical Analysis*, 4th ed. Pacific Grove, CA: Brooks/Cole, 1999.

Hamming, R. W. *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill, 1973.

Stiefel, Edward L. *An Introduction to Numerical Mathematics*. New York: Academic Press, 1963.

3.4 Choosing a Best Model

Let's consider the adequacy of the various models of the form $y = Ax^2$ that we fit using the least-squares and transformed least-squares criteria in the previous section. Using the least-squares criterion, we obtained the model $y = 3.1869x^2$. One way of evaluating how well the model fits the data is to compute the deviations between the model and the actual data. If we compute the sum of the squares of the deviations, we can bound $c_{\max}$ as well. For the model $y = 3.1869x^2$ and the data given in Table 3.3, we compute the deviations shown in Table 3.4.

Table 3.4 Deviations between the data in Table 3.3 and the fitted model $y = 3.1869x^2$

x_i	0.5	1.0	1.5	2.0	2.5
y_i	0.7	3.4	7.2	12.4	20.1
$y_i - y(x_i)$	−0.0967	0.2131	0.02998	−0.3476	0.181875

From Table 3.4 we compute the sum of the squares of the deviations as 0.20954, so $D = (0.20954/5)^{1/2} = 0.204714$. Because the largest absolute deviation is 0.3476 when $x = 2.0$, $c_{\max}$ can be bounded as follows:

$$D = 0.204714 \le c_{\max} \le 0.3476 = d_{\max}$$

Let's find $c_{\max}$. Because there are five data points, the mathematical problem is to minimize the largest of the five numbers $|r_i| = |y_i - y(x_i)|$. Calling that largest number r, we want to minimize r subject to $r \ge r_i$ and $r \ge -r_i$ for each $i = 1, 2, 3, 4, 5$. Denote our model by $y(x) = a_2 x^2$. Then, substitution of the observed data points in Table 3.3 into the inequalities $r \ge r_i$ and $r \ge -r_i$ for each $i = 1, 2, 3, 4, 5$ yields the following linear program:

$$\text{Minimize } r$$

subject to

$$\begin{aligned}
r - r_1 &= r - (0.7 - 0.25a_2) \ge 0 \\
r + r_1 &= r + (0.7 - 0.25a_2) \ge 0 \\
r - r_2 &= r - (3.4 - a_2) \ge 0 \\
r + r_2 &= r + (3.4 - a_2) \ge 0
\end{aligned}$$

$$
\begin{aligned}
r - r_3 &= r - (7.2 - 2.25a_2) \geq 0 \\
r + r_3 &= r + (7.2 - 2.25a_2) \geq 0 \\
r - r_4 &= r - (12.4 - 4a_2) \geq 0 \\
r + r_4 &= r + (12.4 - 4a_2) \geq 0 \\
r - r_5 &= r - (20.1 - 6.25a_2) \geq 0 \\
r + r_5 &= r + (20.1 - 6.25a_2) \geq 0
\end{aligned}
$$

In Chapter 7 we show that the solution of the preceding linear program yields $r = 0.28293$ and $a_2 = 3.17073$. Thus, we have reduced our largest deviation from $d_{\max} = 0.3476$ to $c_{\max} = 0.28293$. Note that we can reduce the largest deviation no further than 0.28293 for the model type $y = Ax^2$.

We have now determined three estimates of the parameter A for the model type $y = Ax^2$. Which estimate is best? For each model we can readily compute the deviations from each data point as recorded in Table 3.5.

Table 3.5 Summary of the deviations for each model $y = Ax^2$

x_i	y_i	$y_i - 3.1869x_i^2$	$y_i - 3.1368x_i^2$	$y_i - 3.17073x_i^2$
0.5	0.7	−0.0967	−0.0842	−0.0927
1.0	3.4	0.2131	0.2632	0.2293
1.5	7.2	0.029475	0.1422	0.0659
2.0	12.4	−0.3476	−0.1472	−0.2829
2.5	20.1	0.181875	0.4950	0.28293

For each of the three models we can compute the sum of the squares of the deviations and the maximum absolute deviation. The results are shown in Table 3.6.

As we would expect, each model has something to commend it. However, notice the increase in the sum of the squares of the deviations in the transformed least-squares model. It is tempting to apply a simple rule, such as choose the model with the smallest absolute deviation. (Other statistical indicators of goodness of fit exist as well. For example, see *Probability and Statistics in Engineering and Management Science*, by William W. Hines and Douglas C. Montgomery, New York: Wiley, 1972.) These indicators are useful for eliminating obviously poor models, but there is no easy answer to the question, Which model is best? The model with the smallest absolute deviation or the smallest sum of squares may fit very poorly over the range where you intend to use it most. Furthermore, as you will see in Chapter 4, models can easily be constructed that pass through each data point, thereby yielding a zero sum of squares and zero maximum deviation. So we need to

Table 3.6 Summary of the results for the three models

Criterion	Model	$\sum [y_i - y(x_i)]^2$	Max $\lvert y_i - y(x_i)\rvert$
Least-squares	$y = 3.1869x^2$	0.2095	0.3476
Transformed least-squares	$y = 3.1368x^2$	0.3633	0.4950
Chebyshev	$y = 3.17073x^2$	0.2256	0.28293

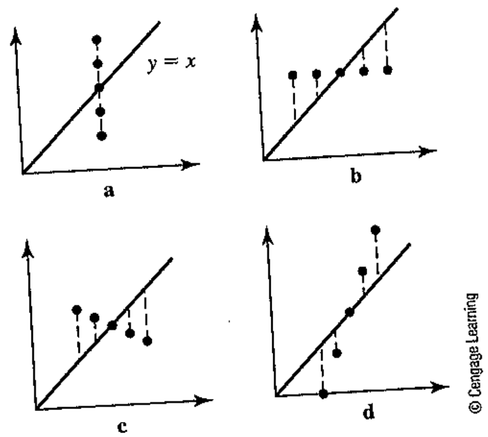

Figure 3.15
In all of these graphs, the model $y = x$ has the same sum of squared deviations.

answer the question of which model is best on a case-by-case basis, taking into account such things as the purpose of the model, the precision demanded by the scenario, the accuracy of the data, and the range of values for the independent variable over which the model will be used.

When choosing among models or judging the adequacy of a model, we may find it tempting to rely on the value of the best-fit criterion being used. For example, it is tempting to choose the model that has the smallest sum of squared deviations for the given data set or to conclude that a sum of squared deviations less than a predetermined value indicates a good fit. However, in isolation these indicators may be very misleading. For example, consider the data displayed in Figure 3.15. In all of the four cases, the model $y = x$ results in exactly the same sum of squared deviations. Without the benefit of the graphs, therefore, we might conclude that in each case the model fits the data about the same. However, as the graphs show, there is a significant variation in each model's ability to capture the trend of the data. The following examples illustrate how the various indicators may be used to help in reaching a decision on the adequacy of a particular model. Normally, a graphical plot is of great benefit.

EXAMPLE 1 *Vehicular Stopping Distance*

Let's reconsider the problem of predicting a motor vehicle's stopping distance as a function of its speed. (This problem was addressed in Sections 2.2 and 3.3.) In Section 3.3 the submodel in which reaction distance d_r was proportional to the velocity v was tested graphically, and the constant of proportionality was estimated to be 1.1. Similarly, the submodel predicting a proportionality between braking distance d_b and the square of the velocity was tested. We found reasonable agreement with the submodel and estimated the proportionality constant to be 0.054. Hence, the model for stopping distance was given by

$$d = 1.1v + 0.054v^2$$

We now fit these submodels analytically and compare the various fits.

To fit the model using the least-squares criterion, we use the formula from Equation (3.7):

$$A = \frac{\sum x_i y_i}{\sum x_i^2}$$

where y_i denotes the driver reaction distance and x_i denotes the speed at each data point. For the 13 data points given in Table 2.4, we compute $\sum x_i y_i = 40905$ and $\sum x_i^2 = 37050$, giving $A = 1.104049$.

For the model type $d_b = Bv^2$, we use the formula

$$B = \frac{\sum x_i^2 y_i}{\sum x_i^4}$$

where y_i denotes the average braking distance and x_i denotes the speed at each data point. For the 13 data points given in Table 2.4, we compute $\sum x_i^2 y_i = 8258350$ and $\sum x_i^4 = 152343750$, giving $B = 0.054209$. Because the data are relatively imprecise and the modeling is done qualitatively, we round the coefficients to obtain the model

$$d = 1.104v + 0.0542v^2 \tag{3.10}$$

Model (3.10) does not differ significantly from that obtained graphically in Chapter 3. Next, let's analyze how well the model fits. We can readily compute the deviations between the observed data points in Table 2.4 and the values predicted by Models (3.9) and (3.10). The deviations are summarized in Table 3.7. The fits of both models are very similar. The largest absolute deviation for Model (3.9) is 30.4 and for Model (3.10) it is 28.8. Note that both models overestimate the stopping distance up to 70 mph, and then they begin to underestimate the stopping distance. A better-fitting model would be obtained by directly fitting the data for total stopping distance to

Table 3.7 Deviations from the observed data points and Models (3.9) and (3.10)

Speed	Graphical model (3.9)	Least-squares model (3.10)
20	1.6	1.76
25	5.25	5.475
30	8.1	8.4
35	13.15	13.535
40	14.4	14.88
45	16.35	16.935
50	17	17.7
55	14.35	15.175
60	12.4	13.36
65	7.15	8.255
70	−1.4	−0.14
75	−14.75	−13.325
80	−30.4	−28.8

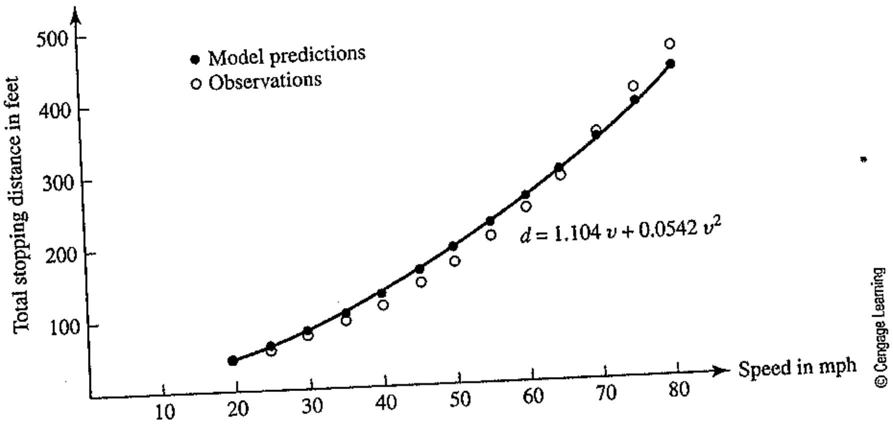

Figure 3.16
A plot of the proposed model and the observed data points provides a visual check on the adequacy of the model.

$$d = k_1 v + k_2 v^2$$

instead of fitting the submodels individually as we did. The advantage of fitting the submodels individually and then testing each submodel is that we can measure how well they explain the behavior.

A plot of the proposed model(s) and the observed data points is useful to determine how well the model fits the data. Model (3.10) andthe observations are plotted in Figure 3.16. It is evident from the figure that a definite trend exists in the data and that Model (3.10) does a reasonable job of capturing that trend, especially at the lower speeds.

A powerful technique for quickly determining where the model is breaking down is to plot the deviations (residuals) as a function of the independent variable(s). For Model (3.10), a plot of the deviations is given in Figure 3.17 showing that the model is indeed reasonable up to 70 mph. Beyond 70 mph there is a breakdown in the model's ability to predict the observed behavior.

Let's examine Figure 3.17 more closely. Note that although the deviations up to 70 mph are relatively small, they are all positive. If the model fully explains the behavior, not only should the deviations be small, but some should be positive and some negative. Why? In Figure 3.17 we note a definite pattern in the nature of the deviations, which might cause us to reexamine the model and/or the data. The nature of the pattern in the deviations can give us clues on how to refine the model further. In this case, the imprecision in the data collection process probably does not warrant further model refinement. ■

EXAMPLE 2 *Comparing the Criterion*

We consider the following data for diameter, height, volume, and diameter^3. We see the trend in Figure 3.18.

We want to fit the model, $V = kD^3$. We compare all three criterion: (a) least squares, (b) sum of absolute deviations, and (c) minimize the largest error (Chebyshev's criterion). Although the solutions to some of these methods, (b) & (c), have yet to be covered in the text, we illustrate the results from those models here.

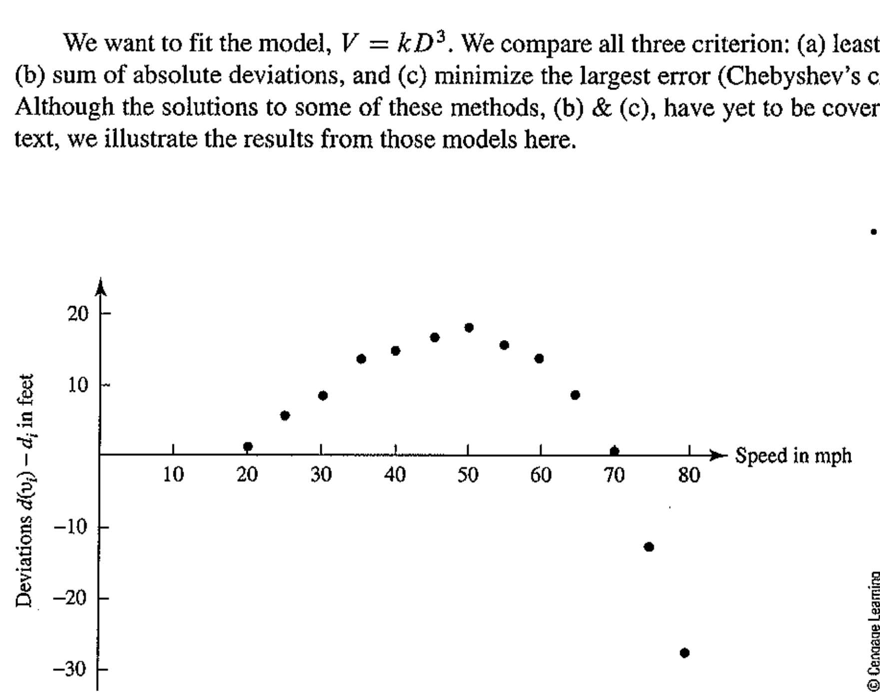

Figure 3.17
A plot of the deviations (residuals) reveals those regions where the model does not fit well.

Diameter	Volume
8.3	10.3
8.8	10.2
10.5	16.4
11.3	24.2
11.4	21.4
12.0	19.1
12.9	22.2
13.7	25.7
14.0	34.5
14.5	36.3
16.0	38.3
17.3	55.4
18.0	51.0
20.6	77.0

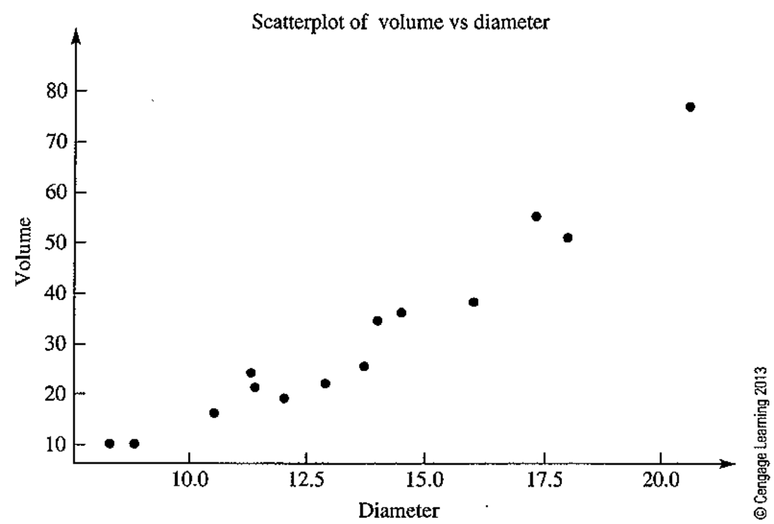

Figure 3.18
Scatterplot of diameter versus volume.

(a) Method of Least Squares Using our formula to find the least squares estimate k,

$$k = \frac{\sum D_i^3 V}{\sum D_i^6}$$

$$k = 1864801/19145566 = 0.00974$$

The regression equation is

$$\text{Volume} = 0.00974\ \text{Diameter}^3$$

The total SSE is 451.

(b) Sum of Absolute Deviations We use the numerical optimization methods described in Chapter 7 to solve this problem.

We solve minimize $S = \sum |y_i - ax_i^3|$ for $i = 1, 2, \ldots, 14$. A summary of the model and absolute error are provided. Using numerical methods, we find the best coefficient value for a is 0.009995, which yields a sum of absolute error of 68.60255.

Diameter	Volume	Model	ABS_Error	Coefficient	0.009995
8.3	10.3	5.714861	4.585139		
8.8	10.2	6.811134	3.388866		
10.5	16.4	11.57016	4.829842		
11.3	24.2	14.42138	9.778624		
11.4	21.4	14.80764	6.592357		
12	19.1	17.27091	1.829094		
12.9	22.2	21.45559	0.744408		
13.7	25.7	25.7	2.45E-06		
14	34.5	27.42556	7.074441		
14.5	36.3	30.47021	5.829794		
16	38.3	40.93844	2.638444		
17.3	55.4	51.74992	3.650079		
18	51	58.28931	7.289307		
20.6	77	87.37215	10.37215		
	Sum of abs_error		68.60255		

The sum of the absolute errors is 68.6055. The model is

$$\text{Volume} = 0.009995\ \text{Diameter}^3$$

Note: We point out that if you used this model and calculated its SSE, it would be greater than 451, the value from the least-squares model. Additionally, if you computed the sum of absolute errors for the least-squares model, it would be greater than 68.60255.

(c) The Chebyshev Method We use the linear programming formulation process discussed in Chapter 7. We will utilize appropriate technology to solve the formulation. The model formulation for the Chebyshevs method is

$$\text{Minimize } R$$

Subject to:

$$
\begin{aligned}
&R + 10.3 - 571.787k \geq 0 \\
&R - (10.3 - 571.787k) \geq 0 \\
&R + 10.2 - 681.472k \geq 0 \\
&R - (10.2 - 681.472k) \geq 0 \\
&\dots \\
&R + 77 - 8741.816k \geq 0 \\
&R - (77 - 8741.816k) \geq 0 \\
&R, k \geq 0
\end{aligned}
$$

The optimal solution is found as $k = 0.009936453825$ and $R = 9.862690776$. The model is

$$\text{Volume} = 0.009936453825 * \text{Diameter}^3$$

The objective function value, Minimize R, the largest error is 9.862690776.

We note that it is the criterion that we choose that determines which model form we pursue: minimize the sum of squared error, minimize the sum of the absolute errors, or minimize the largest absolute error.

3.4 PROBLEMS

For Problems 1–6, find a model using the least-squares criterion either on the data or on the transformed data (as appropriate). Compare your results with the graphical fits obtained in the problem set 3.1 by computing the deviations, the maximum absolute deviation, and the sum of the squared deviations for each model. Find a bound on $c_{\max}$ if the model was fit using the least-squares criterion.

1. Problem 3 in Section 3.1
2. Problem 4a in Section 3.1
3. Problem 4b in Section 3.1
4. Problem 5a in Section 3.1
5. Problem 2 in Section 3.1
6. Problem 6 in Section 3.1
7. **a.** In the following data, W represents the weight of a fish (bass) and l represents its length. Fit the model $W = kl^3$ to the data using the least-squares criterion.

Length, l (in.)	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
Weight, W (oz)	27	17	41	26	17	49	23	16

b. In the following data, g represents the girth of a fish. Fit the model $W = klg^2$ to the data using the least-squares criterion.

Length, l (in.)	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
Girth, g (in.)	9.75	8.375	11.0	9.75	8.5	12.5	9.0	8.5
Weight, W (oz)	27	17	41	26	17	49	23	16

c. Which of the two models fits the data better? Justify fully. Which model do you prefer? Why?

8. Use the data presented in Problem 7b to fit the models $W = cg^3$ and $W = kgl^2$. Interpret these models. Compute appropriate indicators and determine which model is best. Explain.

3.4 PROJECTS

1. Write a computer program that finds the least-squares estimates of the coefficients in the following models.

a. $y = ax^2 + bx + c$

b. $y = ax^n$

2. Write a computer program that computes the deviation from the data points and any model that the user enters. Assuming that the model was fitted using the least-squares criterion, compute D and d_{max}. Output each data point, the deviation from each data point, D, d_{max}, and the sum of the squared deviations.

3. Write a computer program that uses Equations (3.4) and the appropriate transformed data to estimate the parameters of the following models.

a. $y = bx^n$

b. $y = be^{ax}$

c. $y = a \ln x + b$

d. $y = ax^2$

e. $y = ax^3$