d Drag the blue line until you obtain a line that visually fits the data well. What are the slope and intercept of the line that you visually fit to the data? What is the value of SSE for the line that you visually fit to the data? Click the button "Find Best Model" to obtain the least-squares line. How does the value of SSE compare to the SSE associated with the line that you visually fit to the data? How do the slope and intercept of the line that you visually fit to the data compare to slope and intercept of the least-squares line?

11.3 Fit a straight line to the five data points in the accompanying table. Give the estimates of β_0 and β_1. Plot the points and sketch the fitted line as a check on the calculations.

y	3.0	2.0	1.0	1.0	0.5
x	−2.0	−1.0	0.0	1.0	2.0

11.4 Auditors are often required to compare the audited (or current) value of an inventory item with the book (or listed) value. If a company is keeping its inventory and books up to date, there should be a strong linear relationship between the audited and book values. A company sampled ten inventory items and obtained the audited and book values given in the accompanying table. Fit the model $Y = \beta_0 + \beta_1 x + \varepsilon$ to these data.

Item	Audit Value (y_i)	Book Value (x_i)
1	9	10
2	14	12
3	7	9
4	29	27
5	45	47
6	109	112
7	40	36
8	238	241
9	60	59
10	170	167

a What is your estimate for the expected change in audited value for a one-unit change in book value?

b If the book value is $x = 100$, what would you use to estimate the audited value?

11.5 What did housing prices look like in the "good old days"? The median sale prices for new single-family houses are given in the accompanying table for the years 1972 through 1979.[1] Letting Y denote the median sales price and x the year (using integers $1, 2, \dots, 8$), fit the model $Y = \beta_0 + \beta_1 x + \varepsilon$. What can you conclude from the results?

Year	Median Sales Price (×1000)
1972 (1)	\$27.6
1973 (2)	\$32.5
1974 (3)	\$35.9
1975 (4)	\$39.3
1976 (5)	\$44.2
1977 (6)	\$48.8
1978 (7)	\$55.7
1979 (8)	\$62.9

1. *Source:* Adapted from *Time*, 23 July 1979, p. 67.

11.9 Information about eight four-cylinder automobiles judged to be among the most fuel efficient in 2006 is given in the following table. Engine sizes are in total cylinder volume, measured in liters (L).

Car	Cylinder Volume (x)	Horsepower (y)
Honda Civic	1.8	51
Toyota Prius	1.5	51
VW Golf	2.0	115
VW Beetle	2.5	150
Toyota Corolla	1.8	126
VW Jetta	2.5	150
Mini Cooper	1.6	118
Toyota Yaris	1.5	106

a Plot the data points on graph paper.

b Find the least-squares line for the data.

c Graph the least-squares line to see how well it fits the data.

d Use the least-squares line to estimate the mean horsepower rating for a fuel-efficient automobile with cylinder volume 1.9 L.

11.10 Suppose that we have postulated the model

$$Y_i = \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n,$$

where the ε_i's are independent and identically distributed random variables with $E(\varepsilon_i) = 0$. Then $\hat{y}_i = \hat{\beta}_1 x_i$ is the predicted value of y when $x = x_i$ and $\text{SSE} = \sum_{i=1}^{n}[y_i - \hat{\beta}_1 x_i]^2$. Find the least-squares estimator of β_1. (Notice that the equation $y = \beta x$ describes a straight line passing through the origin. The model just described often is called the *no-intercept* model.)

11.11 Some data obtained by C. E. Marcellari[2] on the height x and diameter y of shells appear in the following table. If we consider the model

$$E(Y) = \beta_1 x,$$

then the slope β_1 is the ratio of the mean diameter to the height. Use the following data and the result of Exercise 11.10 to obtain the least-squares estimate of the mean diameter to height ratio.

Specimen	Diameter (y)	Height (x)
OSU 36651	185	78
OSU 36652	194	65
OSU 36653	173	77
OSU 36654	200	76
OSU 36655	179	72
OSU 36656	213	76
OSU 36657	134	75
OSU 36658	191	77
OSU 36659	177	69
OSU 36660	199	65

2. *Source:* Carlos E. Marcellari, "Revision of Serpulids of the Genus Rotularia (Annelida) at Seymour Island (Antarctic Peninsula) and Their Value in Stratigraphy," *Journal of Paleontology* 58(4) (1984).

11.12 Processors usually preserve cucumbers by fermenting them in a low-salt brine (6% to 9% sodium chloride) and then storing them in a high-salt brine until they are used by processors to produce various types of pickles. The high-salt brine is needed to retard softening of the pickles and to prevent freezing when they are stored outside in northern climates. Data showing the reduction in firmness of pickles stored over time in a low-salt brine (2% to 3%) are given in the accompanying table.[3]

	Weeks (x) in Storage at 72°F				
	0	4	14	32	52
Firmness (y) in pounds	19.8	16.5	12.8	8.1	7.5

a Fit a least-squares line to the data.

b As a check on your calculations, plot the five data points and graph the line. Does the line appear to provide a good fit to the data points?

c Use the least-squares line to estimate the mean firmness of pickles stored for 20 weeks.

11.13 The accompanying table gives the catches of Peruvian anchovies (in millions of metric tons) and the prices of fish meal (in current dollars per ton) for 14 consecutive years.[4]

Price of fish meal (y)	190	160	134	129	172	197	167
Anchovy catch (x)	7.23	8.53	9.82	10.26	8.96	12.27	10.28
Price of fish meal (y)	239	542	372	245	376	454	410
Anchovy catch (x)	4.45	1.78	4.0	3.3	4.3	0.8	0.5

a Find the least-squares line appropriate for these data.

b Plot the points and graph the line as a check on your calculations.

11.14 J. H. Matis and T. E. Wehrly[5] report the following table of data on the proportion of green sunfish that survive a fixed level of thermal pollution for varying lengths of time.

Proportion of Survivors (y)	Scaled Time (x)
1.00	.10
.95	.15
.95	.20
.90	.25
.85	.30
.70	.35
.65	.40
.60	.45
.55	.50
.40	.55

a Fit the linear model $Y = \beta_0 + \beta_1 x + \varepsilon$. Give your interpretation.

b Plot the points and graph the result of part (a). Does the line fit through the points?

3. *Source:* R. W. Buescher, J. M. Hudson, J. R. Adams, and D. H. Wallace, "Calcium Makes It Possible to Store Cucumber Pickles in Low-Salt Brine," *Arkansas Farm Research* 30(4) (1981).

4. *Source:* John E. Bardach and Regina M. Santerre, "Climate and the Fish in the Sea," *BioScience* 31(3) (March 1981): 206ff. Copyright ©1981 by the American Institute of Biological Sciences.

5. *Source:* J. H. Matis and T. E. Wehrly, "Stochastic Models of Compartmental Systems," *Biometrics* 35(1) (1979): 199–220.

11.16 An experiment was conducted to observe the effect of an increase in temperature on the potency of an antibiotic. Three 1-ounce portions of the antibiotic were stored for equal lengths of time at each of the following Fahrenheit temperatures: 30°, 50°, 70°, and 90°. The potency readings observed at the end of the experimental period were as shown in the following table.

Potency Readings (y)	38, 43, 29	32, 26, 33	19, 27, 23	14, 19, 21
Temperature (x)	30°	50°	70°	90°

a Find the least-squares line appropriate for this data.
b Plot the points and graph the line as a check on your calculations.
c Calculate S^2.

11.17 **a** Calculate SSE and S^2 for Exercise 11.5.
b It is sometimes convenient, for computational purposes, to have x-values spaced symmetrically and equally about zero. The x-values can be rescaled (or coded) in any convenient manner, with no loss of information in the statistical analysis. Refer to Exercise 11.5. Code the x-values (originally given on a scale of 1 to 8) by using the formula

$$x^* = \frac{x - 4.5}{.5}.$$

Then fit the model $Y = \beta_0^* + \beta_1^* x^* + \varepsilon$. Calculate SSE. (Notice that the x^*-values are integers symmetrically spaced about zero.) Compare the SSE with the value obtained in part (a).

11.18 **a** Calculate SSE and S^2 for Exercise 11.8.
b Refer to Exercise 11.8. Code the x-values in a convenient manner and fit a simple linear model to the LC50 measurements presented there. Compute SSE and compare your answer to the result of part (a).

11.19 A study was conducted to determine the effects of sleep deprivation on subjects' ability to solve simple problems. The amount of sleep deprivation varied over 8, 12, 16, 20, and 24 hours without sleep. A total of ten subjects participated in the study, two at each sleep-deprivation level. After his or her specified sleep-deprivation period, each subject was administered a set of simple addition problems, and the number of errors was recorded. The results shown in the following table were obtained.

Number of Errors (y)	8, 6	6, 10	8, 14	14, 12	16, 12
Number of Hours without Sleep (x)	8	12	16	20	24

a Find the least-squares line appropriate to these data.
b Plot the points and graph the least-squares line as a check on your calculations.
c Calculate S^2.

11.20 Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent normal random variables with $E(Y_i) = \beta_0 + \beta_1 x_i$ and $V(Y_i) = \sigma^2$, for $i = 1, 2, \ldots, n$. Show that the maximum-likelihood estimators (MLEs) of β_0 and β_1 are the same as the least-squares estimators of Section 11.3.

11.21 Under the assumptions of Exercise 11.20, find $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$. Use this answer to show that $\hat{\beta}_0$ and $\hat{\beta}_1$ are independent if $\sum_{i=1}^n x_i = 0$. [*Hint:* $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{Cov}(\overline{Y} - \hat{\beta}_1\overline{x}, \hat{\beta}_1)$. Use Theorem 5.12 and the results of this section.]

11.22 Under the assumptions of Exercise 11.20, find the MLE of σ^2.

c **Applet Exercise** What can be said about the attained significance level associated with the test implemented in part (a) using the appropriate applet?

d Find a 95% confidence interval for β_1.

11.24 Refer to Exercise 11.13. Do the data present sufficient evidence to indicate that the size x of the anchovy catch contributes information for the prediction of the price y of the fish meal?

a Give bounds on the attained significance level.

b **Applet Exercise** What is the exact p-value?

c Based on your answers to parts (a) and/or (b), what would you conclude at the $\alpha = .10$ level of significance?

11.25 Do the data in Exercise 11.19 present sufficient evidence to indicate that the number of errors is linearly related to the number of hours without sleep?

a Give bounds on the attained significance level.

b **Applet Exercise** Determine the exact p-value.

c Based on your answers to parts (a) and/or (b), what would you conclude at the $\alpha = .05$ level of significance?

d Would you expect the relationship between y and x to be linear if x were varied over a wider range, say, from $x = 4$ to $x = 48$?

e Give a 95% confidence interval for the slope. Provide a practical interpretation for this interval estimate.

11.26 Most sophomore physics students are required to conduct an experiment verifying Hooke's law. Hooke's law states that when a force is applied to a body that is long in comparison to its cross-sectional area, the change y in its length is proportional to the force x; that is,

$$y = \beta_1 x,$$

where β_1 is a constant of proportionality. The results of a physics student's laboratory experiment are shown in the following table. Six lengths of steel wire, .34 millimeter (mm) in diameter and 2 meters (m) long, were used to obtain the six force-length change measurements.

Force x (kg)	Change in Length (y) (mm)
29.4	4.25
39.2	5.25
49.0	6.50
58.8	7.85
68.6	8.75
78.4	10.00

a Fit the model, $Y = \beta_0 + \beta_1 x + \varepsilon$, to the data, using the method of least squares.

b Find a 95% confidence interval for the slope of the line.

c According to Hooke's law, the line should pass through the point (0, 0); that is, β_0 should equal 0. Test the hypothesis that $E(Y) = 0$ when $x = 0$. Give bounds for the attained significance level.

d **Applet Exercise** What is the exact p-value?

e What would you conclude at the $\alpha = .05$ level?

11.27 Use the properties of the least-squares estimators given in Section 11.4 to complete the following.

11.38 Refer to Exercise 11.3. Find a 90% confidence interval for $E(Y)$ when $x^* = 0$. Then find 90% confidence intervals for $E(Y)$ when $x^* = -2$ and $x^* = +2$. Compare the lengths of these intervals. Plot these confidence limits on the graph you constructed for Exercise 11.3.

11.39 Refer to Exercise 11.16. Find a 95% confidence interval for the mean potency of a 1-ounce portion of antibiotic stored at 65°F.

11.40 Refer to Exercise 11.14. Find a 90% confidence interval for the expected proportion of survivors at time period .30.

***11.41** Refer to Exercise 11.4. Suppose that the sample given there came from a large but finite population of inventory items. We wish to estimate the population mean of the audited values, using the fact that book values are known for every item on inventory. If the population contains N items and

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i,$$

then the population mean is given by

$$\mu_Y = \frac{1}{N}\sum_{i=1}^{N}\mu_i = \beta_0 + \beta_1\left(\frac{1}{N}\right)\sum_{i=1}^{N}x_i = \beta_0 + \beta_1\mu_x.$$

a Using the least-squares estimators of β_0 and β_1, show that μ_Y can be estimated by

$$\hat{\mu}_Y = \overline{y} + \hat{\beta}_1(\mu_x - \overline{x}).$$

(Notice that $\overline{y}$ is adjusted up or down, depending on whether $\overline{x}$ is larger or smaller than μ_x.)

b Using the data of Exercise 11.4 and the fact that $\mu_x = 74.0$, estimate μ_Y, the mean of the audited values, and place a 2-standard-deviation bound on the error of estimation. (Regard the x_i-values as constants when computing the variance of $\hat{\mu}_Y$.)

11.7 Predicting a Particular Value of Y by Using Simple Linear Regression

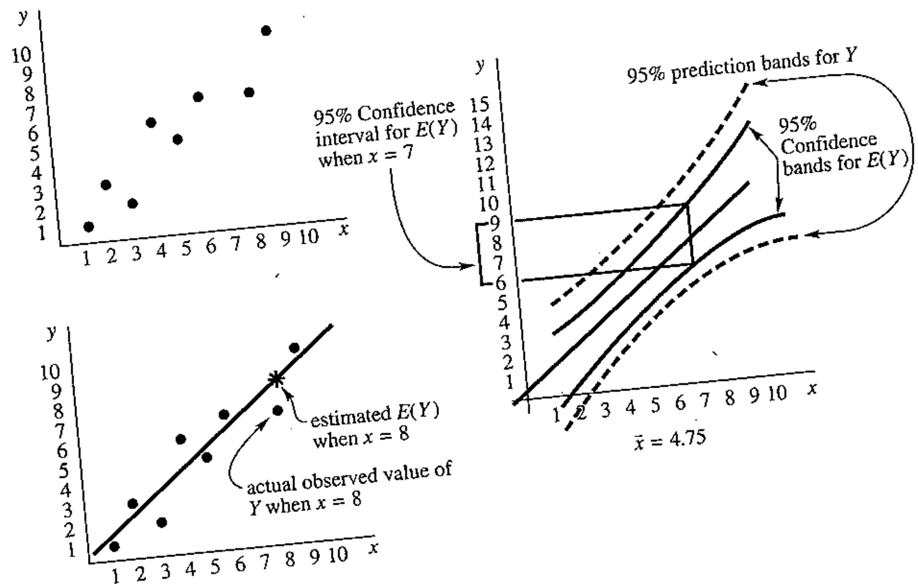
Suppose that for a fixed pressure the yield Y for a chemical experiment is a function of the temperature x at which the experiment is run. Assume that a linear model of the form

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

adequately represents the response function traced by Y over the experimental region of interest. In Section 11.6, we discussed methods for estimating $E(Y)$ for a given temperature, say, x^*. That is, we know how to estimate the mean yield $E(Y)$ of the process at the setting $x = x^*$.

Now consider a different problem. Instead of estimating the mean yield at x^*, we wish to *predict* the particular response Y that we will observe if the experiment is run at some time in the future (such as next Monday). This situation would occur if, for some reason, the response next Monday held a special significance to us. Prediction problems frequently occur in business where we may be interested in next month's profit on a specific investment rather than the average gain per investment in a large portfolio of similar stocks.

FIGURE 11.7 Some hypothetical data and associated confidence and prediction bands



the same approach, we computed prediction bands for the prediction of an actual Y-value for each setting of x. As discussed earlier, for each fixed value of x, the prediction interval is wider than the corresponding confidence interval. The result is that the prediction bands fall uniformly farther from the prediction line than do the confidence bands. The prediction bands are also closest together when $x = \bar{x}$.

Exercises

11.42 Suppose that the model $Y = \beta_0 + \beta_1 x + \varepsilon$ is fit to the n data points $(y_1, x_1), \ldots, (y_n, x_n)$. At what value of x will the length of the prediction interval for Y be minimized?

11.43 Refer to Exercises 11.5 and 11.17. Use the data and model given there to construct a 95% prediction interval for the median sale price in 1980.

11.44 Refer to Exercise 11.43. Find a 95% prediction interval for the median sale price for the year 1981. Repeat for 1982. Would you feel comfortable in using this model and the data of Exercise 11.5 to predict the median sale price for the year 1988?

11.45 Refer to Exercises 11.8 and 11.18. Find a 95% prediction interval for a flow-through LC50 if the static LC50 is observed to be 12 parts per million. Compare the length of this interval to that of the interval found in Exercise 11.37.

11.46 Refer to Exercise 11.16. Find a 95% prediction interval for the potency of a 1-ounce portion of antibiotic stored at 65°F. Compare this interval to that calculated in Exercise 11.39.

11.47 Refer to Exercise 11.14. Find a 95% prediction interval for the proportion of survivors at time $x = .60$.

Exercises

11.48 The accompanying table gives the peak power load for a power plant and the daily high temperature for a random sample of 10 days. Test the hypothesis that the population correlation coefficient ρ between peak power load and high temperature is zero versus the alternative that it is positive. Use $\alpha = .05$. Bound or determine the attained significance level.

Day	High Temperature (°F)	Peak Load
1	95	214
2	82	152
3	90	156
4	81	129
5	99	254
6	100	266
7	93	210
8	95	204
9	93	213
10	87	150

11.49 **Applet Exercise** Refer to Example 11.1 and Exercise 11.2. Access the applet *Fitting a Line Using Least Squares*. The data that appear on the first graph is from Example 11.1.

a Drag the blue line to obtain an equation that visually fits the data well. What do you notice about the values of SSE and r^2 as the fit of the line improves? Why does r^2 increase as SSE decreases?

b Click the button "Find Best Model" to obtain the least-squares line. What is the value of r^2? What is the value of the correlation coefficient?

11.50 **Applet Exercise** Refer to Exercises 11.5 and 11.6. The data from Exercise 11.5 appear in the graph under the heading "Another Example" in the applet *Fitting a Line Using Least Squares*.

a Drag the blue line to obtain an equation that visually fits the data well. What do you notice about the value of r^2 as the fit of the line improves?

b Click the button "Find Best Model" to obtain the least-squares line. What is the value of r^2? What is the value of the correlation coefficient?

c Why is the value of r^2 so much larger than the value of r^2 that you obtained in Exercise 11.49(b) that used the data from Example 11.1?

11.51 In Exercise 11.8 both the flow-through and static LC50 values could be considered random variables. Using the data of Exercise 11.8, test to see whether the correlation between static and flow-through values significantly differs from zero. Use $\alpha = .01$. Bound or determine the associated p-value.

11.52 Is the plant density of a species related to the altitude at which data are collected? Let Y denote the species density and X denote the altitude. A fit of a simple linear regression model using 14 observations yielded $\hat{y} = 21.6 - 7.79x$ and $r^2 = .61$.

a What is the value of the correlation coefficient r?

b What proportion of the variation in densities is explained by the linear model using altitude as the independent variable?

c Is there sufficient evidence at the $\alpha = .05$ to indicate that plant densities decrease with an increase in altitude?