11.5 In each case, graph the line that passes through the given points.

a. $(1, 1)$ and $(5, 5)$

b. $(0, 3)$ and $(3, 0)$

c. $(-1, 1)$ and $(4, 2)$

d. $(-6, -3)$ and $(2, 6)$

11.9 Plot the following lines:

a. $y = 4 + x$

b. $y = 5 - 2x$

c. $y = -4 + 3x$

d. $y = -2x$

e. $y = x$

f. $y = .50 + 1.5x$

11.18 The accompanying table is used to make the preliminary computations for finding the least squares line for the given pairs of x and y values.

a. Complete the table.

b. Find SS_{xy}.

c. Find SS_{xx}.

d. Find $\hat{\beta}_1$.

e. Find $\bar{x}$ and $\bar{y}$.

f. Find $\hat{\beta}_0$.

g. Find the least squares line.

[*Hint*: Use the formulas in the box on p. 592.]

	x_i	y_i	x_i^2	$x_i y_i$
	7	2	—	—
	4	4	—	—
	6	2	—	—
	2	5	—	—
	1	7	—	—
	1	6	—	—
	3	5	—	—
Totals	$\sum x_i =$	$\sum y_i =$	$\sum x_i^2 =$	$\sum x_i y_i =$

11.21 Consider the following pairs of measurements.

D
L11021

x	5	3	−1	2	7	6	4
y	4	3	0	1	8	5	3

a. Construct a scatterplot of these data.

b. What does the scatterplot suggest about the relationship between x and y?

c. Given that $SS_{xx} = 43.4286$, $SS_{xy} = 39.8571$, $\bar{y} = 3.4286$, and $\bar{x} = 3.7143$, calculate the least squares estimates of β_0 and β_1.

d. Plot the least squares line on your scatterplot. Does the line appear to fit the data well? Explain.

e. Interpret the y-intercept and slope of the least squares line. Over what range of x are these interpretations meaningful?

11.26 Mongolian desert ants. Refer to the *Journal of Biogeography* (Dec. 2003) study of ants in Mongolia, presented in Exercise 2.167 (p. 97). Data on annual rainfall, maximum daily temperature, and number of ant species recorded at each of 11 study sites are listed in the table.

D
ANTS

Site	Region	Annual Rainfall (mm)	Max. Daily Temp. (°C)	Number of Ant Species
1	Dry Steppe	196	5.7	3
2	Dry Steppe	196	5.7	3
3	Dry Steppe	179	7.0	52
4	Dry Steppe	197	8.0	7
5	Dry Steppe	149	8.5	5
6	Gobi Desert	112	10.7	49
7	Gobi Desert	125	11.4	5
8	Gobi Desert	99	10.9	4
9	Gobi Desert	125	11.4	4
10	Gobi Desert	84	11.4	5
11	Gobi Desert	115	11.4	4

a. Consider a straight-line model relating annual rainfall (y) and maximum daily temperature (x). A MINITAB printout of the simple linear regression is shown below. Give the least squares prediction equation.

Regression Analysis: Rain versus Temp

```
The regression equation is
Rain = 295 - 16.4 Temp

Predictor      Coef  SE Coef      T      P
Constant     295.25    22.41  13.18  0.000
Temp        -16.364    2.346  -6.97  0.000

S = 17.5111   R-Sq = 84.4%   R-Sq(adj) = 82.7%

Analysis of Variance
Source          DF     SS     MS      F      P
Regression       1  14915  14915  48.64  0.000
Residual Error   9   2760    307
Total           10  17675
```

b. Construct a scatterplot for the analysis you performed in part **a.** Include the least squares line on the plot. Does the line appear to be a good predictor of annual rainfall?

c. Now consider a straight-line model relating number of ant species (y) to annual rainfall (x). On the basis of the MINITAB printout below. repeat parts **a** and **b.**

Regression Analysis: AntSpecies versus Rain

```
The regression equation is
AntSpecies = 10.5 + 0.016 Rain

Predictor     Coef  SE Coef     T      P
Constant     10.52    22.03  0.48  0.644
Rain        0.0160   0.1480  0.11  0.916

S = 19.6726   R-Sq = 0.1%   R-Sq(adj) = 0.0%

Analysis of Variance
Source          DF      SS     MS     F      P
Regression       1     4.5    4.5  0.01  0.916
Residual Error   9  3483.1  387.0
Total           10  3487.6
```

11.29 Ranking driving performance of professional golfers. Refer to *The Sport Journal* (Winter 2007) study of a new method for ranking the total driving performance of golfers on the Professional Golf Association (PGA) tour, presented in Exercise 2.66 (p. 62). Recall that the method

PGA

computes a driving performance index based on a golfer's average driving distance (yards) and driving accuracy (percent of drives that land in the fairway). Data for the top 40 PGA golfers (as ranked by the new method) are saved in the **PGA** file. (The first five and last five observations are listed in the next table.)

a. Write the equation of a straight-line model relating driving accuracy (y) to diving distance (x).

Rank	Player	Driving Distance (yards)	Driving Accuracy (%)	Driving Performance Index
1	Woods	316.1	54.6	3.58
2	Perry	304.7	63.4	3.48
3	Gutschewski	310.5	57.9	3.27
4	Wetterich	311.7	56.6	3.18
5	Hearn	295.2	68.5	2.82
⋮	⋮	⋮	⋮	⋮
36	Senden	291	66	1.31
37	Mickelson	300	58.7	1.30
38	Watney	298.9	59.4	1.26
39	Trahan	295.8	61.8	1.23
40	Pappas	309.4	50.6	1.17

Based on Frederick Wiseman, Ph.D., Mohamed Habibullah, Ph.D., and Mustafa Yilmaz, Ph.D, *Sports Journal*, Vol. 10, No. 1.

b. Use simple linear regression to fit the model you found in part **a** to the data. Give the least squares prediction equation.

c. Interpret the estimated y-intercept of the line.

d. Interpret the estimated slope of the line.

e. A professional golfer, practicing a new swing to increase his average driving distance, is concerned that his driving accuracy will be lower. Which of the two estimates, y-intercept or slope, will help you determine whether the golfer's concern is a valid one? Explain.

11.32 Sweetness of orange juice The quality of the orange juice produced by a manufacturer is constantly monitored. There are numerous sensory and chemical components that combine to make the best-tasting orange juice. For example, one manufacturer has developed a quantitative index of the "sweetness" of orange juice. (The higher the index, the sweeter is the juice.) Is there a relationship between the sweetness index and a chemical measure such as the amount of water-soluble pectin (parts per million) in the orange juice? Data collected on these two variables during 24 production runs at a juice-manufacturing plant are shown in the table. Suppose a manufacturer wants to use simple linear regression to predict the sweetness (y) from the amount of pectin (x).

D OJUICE

a. Find the least squares line for the data.

b. Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ in the words of the problem.

c. Predict the sweetness index if the amount of pectin in the orange juice is 300 ppm. [*Note:* A measure of reliability of such a prediction is discussed in Section 11.6.]

Run	Sweetness Index	Pectin (ppm)
1	5.2	220
2	5.5	227
3	6.0	259
4	5.9	210
5	5.8	224
6	6.0	215
7	5.8	231
8	5.6	268
9	5.6	239
10	5.9	212
11	5.4	410
12	5.6	256
13	5.8	306
14	5.5	259
15	5.3	284
16	5.3	383
17	5.7	271
18	5.5	264
19	5.7	227
20	5.3	263
21	5.9	232
22	5.8	220
23	5.8	246
24	5.9	241

Note: The data in the table are authentic. For reasons of confidentiality, the name of the manufacturer cannot be disclosed.

11.47 Mongolian desert ants. Refer to the *Journal of Biogeography* (Dec. 2003) study of ant sites in Mongolia, presented in Exercise 11.26 (p. 599). The data were used to estimate the straight-line model relating annual rainfall (y) to maximum daily temperature (x).

D ANTS

a. Give the values of SSE, s^2, and s, shown on the MINITAB printout (p. 599).

b. Give a practical interpretation of the value of s.

11.50 Sweetness of orange juice. Refer to the study of the quality of orange juice produced at a juice manufacturing plant, Exercise 11.32 (p. 601). Recall that simple linear regression was used to predict the sweetness index (y) from the amount of pectin (x) in orange juice manufactured during a production run.

D OJUICE

a. Give the values of SSE, s^2, and s for this regression.

b. Explain why it is difficult to give a practical interpretation to s^2.

c. Use the value of s to derive a range within which most (about 95%) of the errors of prediction of sweetness index fall.

11.78 Describe the slope of the least squares line if

a. $r = .7$ **b.** $r = -.7$

c. $r = 0$ **d.** $r^2 = .64$

11.79 Explain what each of the following sample correlation coefficients tells you about the relationship between the x and y values in the sample:

NW

a. $r = 1$ **b.** $r = -1$

c. $r = 0$ **d.** $r = .90$

e. $r = .10$ **f.** $r = -.88$

11.81 Construct a scatterplot for each data set. Then calculate r and r^2 for each data set.

a.

x	−2	−1	0	1	2
y	−2	1	2	5	6

b.

x	−2	−1	0	1	2
y	6	5	3	2	0

c.

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

d.

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

11.93 View of rotated objects. *Perception & Psychophysics* (July 1998) reported on a study of how people view three-dimensional objects projected onto a rotating two-dimensional image. Each in a sample of 25 university students viewed various depth-rotated objects (e.g., a hairbrush, a duck, and a shoe) until they recognized the object. The recognition exposure time—that is, the minimum time (in milliseconds) required for the subject to recognize the object—was recorded

for each object. In addition, each subject rated the "goodness of view" of the object on a numerical scale, with lower scale values corresponding to better views. The following table gives the correlation coefficient r between recognition exposure time and goodness of view for several different rotated objects:

Object	r	t
Piano	.447	2.40
Bench	−.057	.27
Motorbike	.619	3.78
Armchair	.294	1.47
Teapot	.949	14.50

a. Interpret the value of r for each object.
b. Calculate and interpret the value of r^2 for each object.
c. The table also includes the t-value for testing the null hypothesis of no correlation (i.e., for testing H_0: $\beta_1 = 0$). Interpret these results using $x = .05$.

11.97 Effect of massage on boxing. Refer to the *British Journal of Sports Medicine* (Apr. 2000) study of the effect of massage on boxing performance, presented in Exercise 11.70 (p. 615). Find and interpret the values of r and r^2 for the simple

D BOXING2

linear regression relating the blood lactate concentration and the boxer's perceived recovery.

11.57 Construct both a 95% and a 90% confidence interval for β_1 for each of the following cases:
a. $\hat{\beta}_1 = 31, s = 3, SS_{xx} = 35, n = 12$
b. $\hat{\beta}_1 = 64, SSE = 1{,}960, SS_{xx} = 30, n = 18$
c $\hat{\beta}_1 = -8.4, SSE = 146, SS_{xx} = 64, n = 24$

11.59 Consider the following pairs of observations:

L11059

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

a. Construct a scatterplot of the data.
b. Use the method of least squares to fit a straight line to the six data points.
c. Graph the least squares line on the scatterplot of part **a**.
d. Compute the test statistic for determining whether x and y are linearly related.
e. Carry out the test you set up in part **d**, using $\alpha = .01$.
f. Find a 99% confidence interval for β_1.

11.65 Ranking driving performance of professional golfers. Refer to *The Sport Journal* (Winter 2007) study of a new method for ranking the total driving performance of golfers on the Professional Golf Association (PGA) tour, presented in Exercise 11.29 (p. 600). You fit a straight-line model relating driving accuracy (y) to driving distance (x) to the data.

PGA

a. Give the null and alternative hypotheses for testing whether driving accuracy (y) decreases linearly as driving distance (x) increases.
b. Find the test statistic and p-value of the test you set up in part **a**.
c. Make the appropriate conclusion at $\alpha = .01$.

11.72 Pain empathy and brain activity. *Empathy* refers to being able to understand and vicariously feel what others actually feel. Neuroscientists at University College of London investigated the relationship between brain activity and pain-related empathy in persons who watch others in pain (*Science*, Feb. 20, 2004). Sixteen couples participated in the experiment. The female partner watched while painful stimulation was applied to the finger of her male partner. Two variables were measured for each female: y = pain-related brain activity (measured on a scale ranging from -2 to 2) and x = score on the Empathic Concern Scale (0 to 25 points). The data are listed in the accompanying table. The research question of interest was "Do people scoring higher in empathy show higher pain-related brain activity?" Use simple linear regression analysis to answer this question.

D EMPATHY

Couple	Brain Activity (y)	Empathic Concern (x)
1	.05	12
2	$-.03$	13
3	.12	14
4	.20	16
5	.35	16
6	0	17
7	.26	17
8	.50	18
9	.20	18
10	.21	18
11	.45	19
12	.30	20
13	.20	21
14	.22	22
15	.76	23
16	.35	24

11.104 For each of the following, decide whether the proper inference is a prediction interval for y or a confidence interval for $E(y)$:

a. A jeweler wants to predict the selling price of a diamond stone on the basis of its size (number of carats).

b. A psychologist wants to estimate the average IQ of all patients who have a certain income level.

Learning the Mechanics

11.105 In fitting a least squares line to $n = 10$ data points, the following quantities were computed:

$$SS_{xx} = 32, \bar{x} = 3, SS_{yy} = 26, \bar{y} = 4, SS_{xy} = 28$$

a. Find the least squares line.

b. Graph the least squares line.

c. Calculate SSE.

d. Calculate s^2.

e. Find a 95% confidence interval for the mean value of y when $x_p = 2.5$.

f. Find a 95% prediction interval for y when $x_p = 4$.

11.118 **The "name game."** Refer to the *Journal of Experimental Psychology—Applied* (June 2000) name-retrieval study, presented in Exercise 11.34 (p. 602).

D NAME2

a. Find a 99% confidence interval for the mean recall proportion for students in the fifth position during the "name game." Interpret the result.

b. Find a 99% prediction interval for the recall proportion of a particular student in the fifth position during the "name game." Interpret the result.

c. Compare the intervals you found in parts **a** and **b**. Which interval is wider? Will this always be the case? Explain.

11.119 **Spreading rate of spilled liquid.** Refer to the *Chemicial Engineering Progress* (Jan. 2005) study of the rate at which a spilled volatile liquid will spread across a surface, presented in Exercise 11.35 (p. 602). Recall that simple linear regression was used to model y = mass of the spill as a function of y = elapsed time of the spill.

D LSPILL

a. Find a 99% confidence interval for the mean mass of all spills with an elapsed time of 15 minutes. Interpret the result.

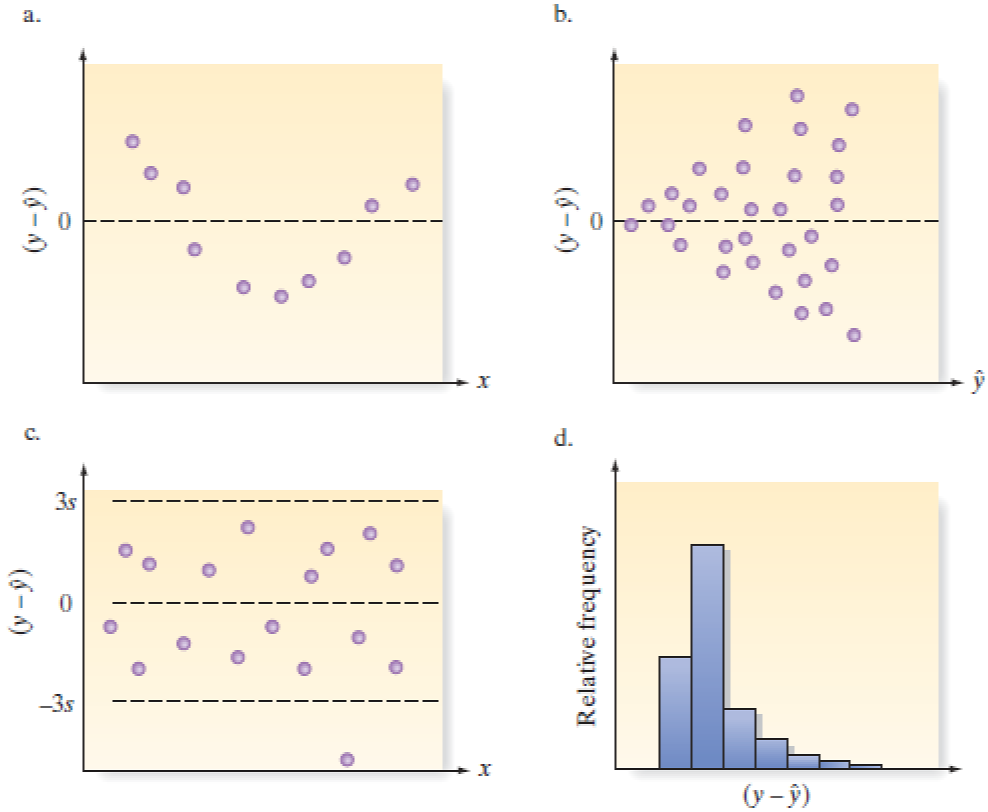
b. Find a 99% prediction interval for the mass of a single spill with an elapsed time of 15 minutes. Interpret the result.

c. Compare the intervals you found in parts **a** and **b**. Which interval is wider? Will this always be the case? Explain.

12.154 Identify the problem(s) in each of the residual plots shown on page 752.

NW

Residual plots for Exercise 12.154



14.101 Effect of massage on boxers. Refer to the *British Journal of Sports Medicine* (Apr. 2000) study of the effect of massaging boxers between rounds, presented in Exercise 11.70 (p. 615). Two variables measured on the boxers were blood lactate level (y) and the boxer's perceived recovery (x). The data for 16 five-round boxing performances are reproduced in the table.

D BOXING2

Blood Lactate Level	Perceived Recovery
3.8	7
4.2	7
4.8	11
4.1	12
5.0	12
5.3	12
4.2	13
2.4	17
3.7	17
5.3	17
5.8	18
6.0	18
5.9	21
6.3	21
5.5	20
6.5	24

Based on Hemmings, B., Smith, M., Graydon, J., and Dyson, R. "Effects of massage on physiological restoration, perceived recovery, and repeated sports performance." *British Journal of Sports Medicine*, Vol. 34, No. 2, Apr. 2000 (data adapted from Figure 3).

a. Rank the values of the 16 blood lactate levels.
b. Rank the values of the 16 perceived recovery values.
c. Use the ranks from parts **a** and **b** to compute Spearman's rank correlation coefficient. Give a practical interpretation of the result.
d. Find the rejection region for a test to determine whether y and x are rank correlated. Use $\alpha = .10$.
e. What is the conclusion of the test you conducted in part **d**? State your answer in the words of the problem.

14.134 Ranking wines. Two expert wine tasters were asked to rank six brands of wine. Their rankings are shown in the following table. Do the data indicate a positive correlation in the rankings of the two experts? Test, using $\alpha = .10$.

D WINE

Brand	Expert 1	Expert 2
A	6	5
B	5	6
C	1	2
D	3	1
E	2	4
F	4	3

12.1 Write a first-order model relating $E(y)$ to

a. two quantitative independent variables
b. four quantitative independent variables
c. five quantitative independent variables

12.6 MINITAB was used to fit the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ to $n = 20$ data points, and the printout on p. 669 was obtained.

NW

a. What are the sample estimates of β_0, β_1, and β_2?
b. What is the least squares prediction equation?
c. Find SSE, MSE, and s. Interpret the standard deviation in the context of the problem.
d. Test H_0: $\beta_1 = 0$ against H_a: $\beta_1 \neq 0$. Use $\alpha = .05$.
e. Use a 95% confidence interval to estimate β_2.
f. Find R^2 and R_a^2 and interpret these values.
g. Use the two formulas given in this section to calculate the test statistic for the null hypothesis H_0: $\beta_1 = \beta_2 = 0$.

12.19 Predicting runs scored in baseball. Consider a multiple-regression model for predicting the total number of runs scored by a Major League Baseball (MLB) team during a season. Using data on number of walks (x_1), singles (x_2), doubles (x_3), triples (x_4), home runs (x_5), stolen bases (x_6), times caught stealing (x_7), strike outs (x_8), and ground outs (x_9) for each of the 30 teams during the 2014 MLB season, a 1st-order model for total number of runs scored (y) was fit. The results are shown in the accompanying Minitab printout.

NW

```
The regression equation is
Runs = 40 + 0.254 Walks + 0.407 Singles + 0.841 Doubles + 1.63 Triples
       + 1.09 HomeRuns + 0.050 StolenBases - 0.044 CaughtStealing
       - 0.0911 StrikeOuts - 0.122 GroundOuts
```

Predictor	Coef	SE Coef	T	P
Constant	39.7	155.7	0.25	0.801
Walks	0.25380	0.08349	3.04	0.006
Singles	0.4072	0.1052	3.87	0.001
Doubles	0.8411	0.2072	4.06	0.001
Triples	1.6327	0.5460	2.99	0.007
HomeRuns	1.0925	0.1945	5.62	0.000
StolenBases	0.0499	0.1786	0.28	0.783
CaughtStealing	-0.0436	0.4842	-0.09	0.929
StrikeOuts	-0.09114	0.04120	-2.21	0.039
GroundOuts	-0.12185	0.06116	-1.99	0.060

S = 19.4318   R-Sq = 91.8%   R-Sq(adj) = 88.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	84174.4	9352.7	24.77	0.000
Residual Error	20	7551.9	377.6		
Total	29	91726.3			

a. Write the least squares prediction equation for y = total number of runs scored by a team during the 2014 season.

b. Give practical interpretations of the beta estimates.

c. Conduct a test of $H_0: \beta_7 = 0$ against $H_a: \beta_7 < 0$ at $\alpha = .05$. Interpret the results.

d. Form a 95% confidence interval for β_5. Interpret the results.

e. Predict the number of runs scored in 2014 by your favorite Major League Baseball team. How close is the predicted value to the actual number of runs scored by your team? (*Note*: You can find data on your favorite team on the Internet at www.majorleaguebaseball.com.)

12.23 Arsenic in groundwater. *Environmental Science & Technology* (Jan. 2005) reported on a study of the reliability of a commercial kit designed to test for arsenic in groundwater. The field kit was used to test a sample of 328 groundwater wells in Bangladesh. In addition to the arsenic level (in micrograms per liter), the latitude (degrees), longitude (degrees), and depth (feet) of each well were measured. The first and last five observations of the data are listed in the following table.

D ASWELLS

Wellid	Latitude	Longitude	Depth	Arsenic
10	23.7887	90.6522	60	331
14	23.7886	90.6523	45	302
30	23.7880	90.6517	45	193
59	23.7893	90.6525	125	232
85	23.7920	90.6140	150	19
⋮	⋮	⋮	⋮	⋮
7353	23.7949	90.6515	40	48
7357	23.7955	90.6515	30	172
7890	23.7658	90.6312	60	175
7893	23.7656	90.6315	45	624
7970	23.7644	90.6303	30	254

a. Write a first-order model for arsenic level (y) as a function of latitude, longitude, and depth.
b. Use the method of least squares to fit the model to the data.
c. Give practical interpretations of the β estimates.
d. Find the standard deviation s of the model, and interpret its value.
e. Find and interpret the values of R^2 and R_a^2.
f. Conduct a test of overall model utility at $\alpha = .05$.
g. On the basis of the results you obtained in parts **d–f**, would you recommend using the model to predict arsenic level (y)? Explain.

12.27 Cooling method for gas turbines. Refer to the *Journal of Engineering for Gas Turbines and Power* (Jan. 2005) study of a high-pressure inlet fogging method for a gas turbine engine, presented in Exercise 8.46 (p. 392). Recall that the heat rate (kilojoules per kilowatt per hour) was measured for each in a sample of 67 gas turbines augmented with high-pressure inlet fogging. In addition, several other variables were measured, including cycle speed (revolutions per minute), inlet temperature (°C), exhaust gas temperature (°C), cycle pressure ratio, and air mass flow rate (kilograms per second). The first and last five observations of the data are listed in the following table.

D TURBINE

Rpm	Cpratio	Inlet-Temp	Exh-Temp	Airflow	Heatrate
27245	9.2	1134	602	7	14622
14000	12.2	950	446	15	13196
17384	14.8	1149	537	20	11948
11085	11.8	1024	478	27	11289
14045	13.2	1149	553	29	11964
⋮	⋮	⋮	⋮	⋮	⋮
18910	14.0	1066	532	8	12766
3600	35.0	1288	448	152	8714
3600	20.0	1160	456	84	9469
16000	10.6	1232	560	14	11948
14600	13.4	1077	536	20	12414

Based on Bhargava, R., and Meher-Homji, C. B. "Parametric analysis of existing gas turbines with inlet evaporative and overspray fogging." *Journal of Engineering for Gas Turbines and Power*, Vol. 127, No. 1, Jan. 2005. Table from pp. 156–157.

a. Write a first-order model for heat rate (y) as a function of speed, inlet temperature, exhaust temperature, cycle pressure ratio, and air mass flow rate.

b. Use the method of least squares to fit the model to the data.

c. Give practical interpretations of the β estimates.

d. Find the standard deviation s of the model, and interpret its value.

e. Find R_a^2 and interpret its value.

f. Is the overall model statistically useful in predicting heat rate (y)? Test, using $\alpha = .01$.

a. Interpret the confidence interval for $E(y)$ for student 1.
b. Interpret the confidence interval for $E(y)$ for student 4.

12.36 Usability Professionals, salary survey. The Usability Professionals' Association (UPA) supports people who research, design, and evaluate the user experience of products and services. The UPA conducted a salary survey of its members (*UPA Salary Survey*, Aug. 18, 2009). One of the report's authors, Jeff Sauro, investigated how much having a PhD affects salaries in this profession and discussed his analysis on the blog www.measuringusability.com. Sauro fit a first-order multiple regression model for salary (y, in dollars) as a function of years of experience (x_1), PhD status ($x_2 = 1$ if PhD, 0 if not), and manager status ($x_3 = 1$ if manager, 0 if not). The following prediction equation was obtained:

$$\hat{y} = 52{,}484 + 2{,}941x_1 + 16{,}880x_2 + 11{,}108x_3$$

a. Predict the salary of a UPA member with 10 years of experience who does not have a PhD but is a manager.
b. Predict the salary of a UPA member with 10 years of experience who does have a PhD but is not a manager.
c. Why is a 95% prediction interval preferred over the predicted values given in parts **a** and **b**?

12.37 Cooling method for gas turbines. Refer to the *Journal of Engineering for Gas Turbines and Power* (Jan. 2005) study of a high-pressure inlet fogging method for a gas turbine engine, presented in Exercise 12.27 (p. 673). Recall that you fitted a first-order model for heat rate (y) as a function of speed (x_1), inlet temperature (x_2), exhaust temperature (x_3), cycle pressure ratio (x_4), and air mass flow rate (x_5). A MINITAB printout with both a 95% confidence interval for $E(y)$ and a prediction interval for y, for selected values of the x's, is shown above.

NW

a. Interpret the 95% prediction interval for y in the words of the problem.
b. Interpret the 95% confidence interval for $E(y)$ in the words of the problem.
c. Will the confidence interval for $E(y)$ always be narrower than the prediction interval for y? Explain.

Applying the Concepts—Intermediate

12.38 California rain levels. An article published in *Geography* (July 1980) used multiple regression to predict annual rainfall levels in California. Data on the average annual precipitation (y), altitude (x_1), latitude (x_2), and distance from the Pacific coast (x_3) for 30 meteorological stations scattered throughout California are saved in the **CALRAIN** file. (Selected observations are listed in the table below.) Consider the first-order model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$.

D CALRAIN

a. Fit the model to the data and give the least squares prediction equation.
b. Is there evidence that the model is useful in predicting annual precipitation y? Test, using $\alpha = .05$.
c. Find a 95% prediction interval for y for the Giant Forest meteorological station (station 9). Interpret the interval.

Data for Exercise 12.38

Station	Avg. Annual Precipitation y (inches)	Altitude x_1 (feet)	Latitude x_2 (degrees)	Distance from Coast x_3 (miles)
1. Eureka	39.57	43	40.8	1
2. Red Bluff	23.27	341	40.2	97
3. Thermal	18.20	4152	33.8	70
4. Fort Bragg	37.48	74	39.4	1
5. Soda Springs	49.26	6752	39.3	150
⋮	⋮	⋮	⋮	⋮
26. San Diego	9.94	19	32.7	5
27. Daggett	4.25	2105	34.1	85
28. Death Valley	1.66	−178	36.5	194
29. Crescent City	74.87	35	41.7	1
30. Colusa	15.95	60	39.2	91

12.40 Boiler drum production. In a production facility, an accurate estimate of hours needed to complete a task is crucial to management in making such decisions as hiring the proper number of workers, quoting an accurate deadline for a client, or performing cost analyses regarding budgets. A manufacturer of boiler drums wants to use regression to predict the number of hours needed to erect the drums in future projects. To accomplish this task, data on 36 boilers were collected. In addition to hours (y), the variables measured were boiler capacity (x_1 = lb/hr), boiler design pressure (x_2 = pounds per square inch, or psi), boiler type ($x_3 = 1$ if industry field erected, 0 if utility field erected), and drum type ($x_4 = 1$ if steam, 0 if mud). The data are saved in the **BOILERS** file. (Selected observations are shown in the table below.)

D BOILERS

Hours y	Boiler Capacity x_1	Design Pressure x_2	Boiler Type x_3	Drum Type x_4
3,137	120,000	375	1	1
3,590	65,000	750	1	1
4,526	150,000	500	1	1
10,825	1,073,877	2,170	0	1
4,023	150,000	325	1	1
⋮	⋮	⋮	⋮	⋮
4,206	441,000	410	1	0
4,006	441,000	410	1	0
3,728	627,000	1,525	0	0
3,211	610,000	1,500	0	0
1,200	30,000	325	1	0

Based on data provided by Dr. Kelly Uscategui, University of Connecticut.

a. Fit the model $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$ to the data and give the prediction equation.

b. Conduct a test for the global utility of the model. Use $\alpha = .01$.

c. Find a 95% confidence interval for $E(y)$ when $x_1 = 150{,}000$, $x_2 = 500$, $x_3 = 1$, and $x_4 = 0$. Interpret the result.

d. What type of interval would you use if you want to estimate the average number of hours required to erect all industrial mud boilers with a capacity of 150,000 lb/hr and a design pressure of 500 psi?