$$P(Y_1 - Y_2 > 0) = P\left(Z > \frac{0 - (-1.2)}{11.314}\right) = P(Z > .11) = .4562$$

That is, even though we expected Y to decrease when x increases by 1 unit, it is not unlikely that the observed Y at $x + 1$ will be larger than the observed Y at x. ■

EXERCISES Section 12.1 (1–11)

1. The efficiency ratio for a steel specimen immersed in a phosphating tank is the weight of the phosphate coating divided by the metal loss (both in mg/ft^2). The article "Statistical Process Control of a Phosphate Coating Line" (*Wire J. Intl.,* May 1997: 78–81) gave the accompanying data on tank temperature (x) and efficiency ratio (y).

Temp.	170	172	173	174	174	175	176
Ratio	.84	1.31	1.42	1.03	1.07	1.08	1.04
Temp.	177	180	180	180	180	180	181
Ratio	1.80	1.45	1.60	1.61	2.13	2.15	.84
Temp.	181	182	182	182	182	184	184
Ratio	1.43	.90	1.81	1.94	2.68	1.49	2.52
Temp.	185	186	188				
Ratio	3.00	1.87	3.08				

a. Construct stem-and-leaf displays of both temperature and efficiency ratio, and comment on interesting features.

b. Is the value of efficiency ratio completely and uniquely determined by tank temperature? Explain your reasoning.

c. Construct a scatter plot of the data. Does it appear that efficiency ratio could be very well predicted by the value of temperature? Explain your reasoning.

2. The article "Exhaust Emissions from Four-Stroke Lawn Mower Engines" (*J. of the Air and Water Mgmnt. Assoc.,* 1997: 945–952) reported data from a study in which both a baseline gasoline mixture and a reformulated gasoline were used. Consider the following observations on age (yr) and NO_x emissions (g/kWh):

Engine	1	2	3	4	5
Age	0	0	2	11	7
Baseline	1.72	4.38	4.06	1.26	5.31
Reformulated	1.88	5.93	5.54	2.67	6.53
Engine	6	7	8	9	10
Age	16	9	0	12	4
Baseline	.57	3.37	3.44	.74	1.24
Reformulated	.74	4.94	4.89	.69	1.42

Construct scatter plots of NO_x emissions versus age. What appears to be the nature of the relationship between these two variables? [*Note:* The authors of the cited article commented on the relationship.]

3. Bivariate data often arises from the use of two different techniques to measure the same quantity. As an example, the accompanying observations on x = hydrogen concentration (ppm) using a gas chromatography method and y = concentration using a new sensor method were read from a graph in the article "A New Method to Measure the Diffusible Hydrogen Content in Steel Weldments Using a Polymer Electrolyte-Based Hydrogen Sensor" (*Welding Res.,* July 1997: 251s–256s).

x	47	62	65	70	70	78	95	100	114	118
y	38	62	53	67	84	79	93	106	117	116
x	124	127	140	140	140	150	152	164	198	221
y	127	114	134	139	142	170	149	154	200	215

Construct a scatter plot. Does there appear to be a very strong relationship between the two types of concentration measurements? Do the two methods appear to be measuring roughly the same quantity? Explain your reasoning.

4. A study to assess the capability of subsurface flow wetland systems to remove biochemical oxygen demand (BOD) and various other chemical constituents resulted in the accompanying data on x = BOD mass loading (kg/ha/d) and y = BOD mass removal (kg/ha/d) ("Subsurface Flow Wetlands—A Performance Evaluation," *Water Envir. Res.,* 1995: 244–247).

x	3	8	10	11	13	16	27	30	35	37	38	44	103	142
y	4	7	8	8	10	11	16	26	21	9	31	30	75	90

a. Construct boxplots of both mass loading and mass removal, and comment on any interesting features.

b. Construct a scatter plot of the data, and comment on any interesting features.

5. The article "Objective Measurement of the Stretchability of Mozzarella Cheese" (*J. of Texture Studies,* 1992: 185–194) reported on an experiment to investigate how the behavior of mozzarella cheese varied with temperature. Consider the accompanying data on x = temperature and y = elongation (%) at failure of the cheese. [*Note:* The researchers were Italian and used *real* mozzarella cheese, not the poor cousin widely available in the United States.]

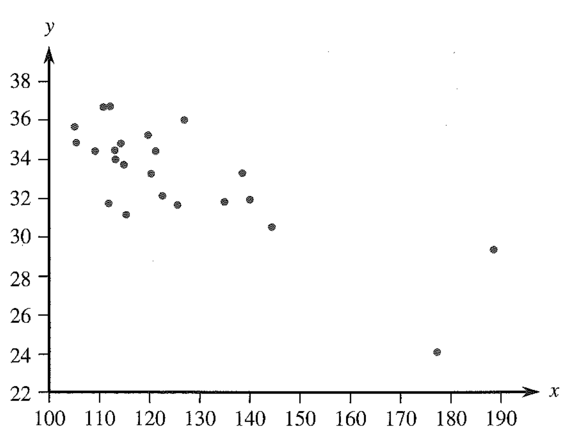
x	59	63	68	72	74	78	83
y	118	182	247	208	197	135	132

a. Construct a scatter plot in which the axes intersect at (0, 0). Mark 0, 20, 40, 60, 80, and 100 on the horizontal axis and 0, 50, 100, 150, 200, and 250 on the vertical axis.

b. Construct a scatter plot in which the axes intersect at (55, 100), as was done in the cited article. Does this plot seem preferable to the one in part (a)? Explain your reasoning.

c. What do the plots of parts (a) and (b) suggest about the nature of the relationship between the two variables?

6. One factor in the development of tennis elbow, a malady that strikes fear in the hearts of all serious tennis players, is the impact-induced vibration of the racket-and-arm system at ball contact. It is well known that the likelihood of getting tennis elbow depends on various properties of the racket used. Consider the scatter plot of x = racket resonance frequency (Hz) and y = sum of peak-to-peak acceleration (a characteristic of arm vibration, in m/sec/sec) for $n = 23$ different rackets ("Transfer of Tennis Racket Vibrations into the Human Forearm," *Medicine and Science in Sports and Exercise,* 1992: 1134–1140). Discuss interesting features of the data and scatter plot.



7. The article "Some Field Experience in the Use of an Accelerated Method in Estimating 28-Day Strength of Concrete" (*J. of Amer. Concrete Institute,* 1969: 895) considered regressing y = 28-day standard-cured strength (psi) against x = accelerated strength (psi). Suppose the equation of the true regression line is $y = 1800 + 1.3x$.

a. What is the expected value of 28-day strength when accelerated strength = 2500?

b. By how much can we expect 28-day strength to change when accelerated strength increases by 1 psi?

c. Answer part (b) for an increase of 100 psi.

d. Answer part (b) for a decrease of 100 psi.

8. Referring to Exercise 7, suppose that the standard deviation of the random deviation ϵ is 350 psi.

a. What is the probability that the observed value of 28-day strength will exceed 5000 psi when the value of accelerated strength is 2000?

b. Repeat part (a) with 2500 in place of 2000.

c. Consider making two independent observations on 28-day strength, the first for an accelerated strength of 2000 and the second for $x = 2500$. What is the probability that the second observation will exceed the first by more than 1000 psi?

d. Let Y_1 and Y_2 denote observations on 28-day strength when $x = x_1$ and $x = x_2$ respectively. By how much would x_2 have to exceed x_1 in order that $P(Y_2 > Y_1) = .95$?

9. The flow rate y (m^3/min) in a device used for air-quality measurement depends on the pressure drop x (in. of water) across the device's filter. Suppose that for x values between 5 and 20, the two variables are related according to the simple linear regression model with true regression line $y = -.12 + .095x$.

a. What is the expected change in flow rate associated with a 1-in. increase in pressure drop? Explain.

b. What change in flow rate can be expected when pressure drop decreases by 5 in.?

c. What is the expected flow rate for a pressure drop of 10 in.? A drop of 15 in.?

d. Suppose $\sigma = .025$ and consider a pressure drop of 10 in. What is the probability that the observed value of flow rate will exceed .835? That observed flow rate will exceed .840?

e. What is the probability that an observation on flow rate when pressure drop is 10 in. will exceed an observation on flow rate made when pressure drop is 11 in.?

10. Suppose the expected cost of a production run is related to the size of the run by the equation $y = 4000 + 10x$. Let Y denote an observation on the cost of a run. If the variables' *size* and *cost* are related according to the simple linear regression model, could it be the case that $P(Y > 5500$ when $x = 100) = .05$ and $P(Y > 6500$ when $x = 200) = .10$? Explain.

11. Suppose that in a certain chemical process the reaction time y (hr) is related to the temperature (°F) in the chamber in which the reaction takes place according to the simple linear regression model with equation $y = 5.00 - .01x$ and $\sigma = .075$.

a. What is the expected change in reaction time for a 1°F increase in temperature? For a 10°F increase in temperature?

and son's height y. After collecting a number of pairs (x_i, y_i), Galton used the principle of least squares to obtain the equation of the estimated regression line, with the objective of using it to predict son's height from father's height. In using the derived line, Galton found that if a father was above average in height, the son would also be expected to be above average in height, *but not by as much as the father was.* Similarly, the son of a shorter-than-average father would also be expected to be shorter than average, but not by as much as the father. Thus the predicted height of a son was "pulled back in" toward the mean; because *regression* means a coming or going back, Galton adopted the terminology *regression line.* This phenomenon of being pulled back in toward the mean has been observed in many other situations (e.g., batting averages from year to year in baseball) and is called the **regression effect.**

Our discussion thus far has presumed that the independent variable is under the control of the investigator, so that only the dependent variable Y is random. This was not, however, the case with Galton's experiment; fathers' heights were not preselected, but instead both X and Y were random. Methods and conclusions of regression analysis can be applied both when the values of the independent variable are fixed in advance and when they are random, but because the derivations and interpretations are more straightforward in the former case, we will continue to work explicitly with it. For more commentary, see the excellent book by John Neter et al. listed in the chapter bibliography.

EXERCISES Section 12.2 (12–29)

12. Exercise 4 gave data on x = BOD mass loading and y = BOD mass removal. Values of relevant summary quantities are

$$n = 14 \quad \sum x_i = 517$$
$$\sum y_i = 346 \quad \sum x_i^2 = 39{,}095$$
$$\sum y_i^2 = 17{,}454 \quad \sum x_i y_i = 25{,}825$$

a. Obtain the equation of the least squares line.
b. Predict the value of BOD mass removal for a single observation made when BOD mass loading is 35, and calculate the value of the corresponding residual.
c. Calculate SSE and then a point estimate of σ.
d. What proportion of observed variation in removal can be explained by the approximate linear relationship between the two variables?
e. The last two x values, 103 and 142, are much larger than the others. How are the equation of the least squares line and the value of r^2 affected by deletion of the two corresponding observations from the sample? Adjust the given values of the summary quantities, and use the fact that the new value of SSE is 311.79.

13. The accompanying data on x = current density (mA/cm^2) and y = rate of deposition (μm/min) appeared in the article "Plating of 60/40 Tin/Lead Solder for Head Termination Metallurgy" (*Plating and Surface Finishing,* Jan. 1997: 38–40). Do you agree with the claim by the article's author that "a linear relationship was obtained from the tin-lead rate of deposition as a function of current density"? Explain your reasoning.

x	20	40	60	80
y	.24	1.20	1.71	2.22

14. Refer to the tank temperature–efficiency ratio data given in Exercise 1.
a. Determine the equation of the estimated regression line.
b. Calculate a point estimate for true average efficiency ratio when tank temperature is 182.
c. Calculate the values of the residuals from the least squares line for the four observations for which temperature is 182. Why do they not all have the same sign?
d. What proportion of the observed variation in efficiency ratio can be attributed to the simple linear regression relationship between the two variables?

15. Values of modulus of elasticity (MOE, the ratio of stress, i.e., force per unit area, to strain, i.e., deformation per unit length, in GPa) and flexural strength (a measure of the ability to resist failure in bending, in MPa) were determined for a sample of concrete beams of a certain type, resulting in the following data (read from a graph in the article "Effects of Aggregates and Microfillers on the Flexural Properties of Concrete," *Magazine of Concrete Research,* 1997: 81–98):

MOE	29.8	33.2	33.7	35.3	35.5	36.1	36.2
Strength	5.9	7.2	7.3	6.3	8.1	6.8	7.0
MOE	36.3	37.5	37.7	38.7	38.8	39.6	41.0
Strength	7.6	6.8	6.5	7.0	6.3	7.9	9.0
MOE	42.8	42.8	43.5	45.6	46.0	46.9	48.0
Strength	8.2	8.7	7.8	9.7	7.4	7.7	9.7
MOE	49.3	51.7	62.6	69.8	79.5	80.0	
Strength	7.8	7.7	11.6	11.3	11.8	10.7	

a. Construct a stem-and-leaf display of the MOE values, and comment on any interesting features.
b. Is the value of strength completely and uniquely determined by the value of MOE? Explain.
c. Use the accompanying Minitab output to obtain the equation of the least squares line for predicting strength from modulus of elasticity, and then predict strength for a beam whose modulus of elasticity is 40. Would you feel comfortable using the least squares line to predict strength when modulus of elasticity is 100? Explain.

```
Predictor        Coef      Stdev   t-ratio       P
Constant       3.2925     0.6008      5.48   0.000
mod elas      0.10748    0.01280      8.40   0.000

s = 0.8657   R-sq = 73.8%   R-sq(adj) = 72.8%

Analysis of Variance

SOURCE        DF       SS       MS       F       P
Regression     1   52.870   52.870   70.55   0.000
Error         25   18.736    0.749
Total         26   71.605
```

d. What are the values of SSE, SST, and the coefficient of determination? Do these values suggest that the simple linear regression model effectively describes the relationship between the two variables? Explain.

16. The article "Characterization of Highway Runoff in Austin, Texas, Area" (*J. of Envir. Engr.*, 1998: 131–137) gave a scatter plot, along with the least squares line, of x = rainfull volume (m^3) and y = runoff volume (m^3) for a particular location. The accompanying values were read from the plot.

x	5	12	14	17	23	30	40	47
y	4	10	13	15	15	25	27	46

x	55	67	72	81	96	112	127
y	38	46	53	70	82	99	100

a. Does a scatter plot of the data support the use of the simple linear regression model?
b. Calculate point estimates of the slope and intercept of the population regression line.
c. Calculate a point estimate of the true average runoff volume when rainfall volume is 50.
d. Calculate a point estimate of the standard deviation σ.
e. What proportion of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall?

17. No-fines concrete, made from a uniformly graded coarse aggregate and a cement-water paste, is beneficial in areas prone to excessive rainfall because of its excellent drainage properties. The article "Pavement Thickness Design for No-Fines Concrete Parking Lots," *J. of Trans. Engr.*, 1995: 476–484) employed a least squares analysis in studying how y = porosity (%) is related to x = unit weight (pcf) in concrete specimens. Consider the following representative data:

x	99.0	101.1	102.7	103.0	105.4	107.0	108.7	110.8
y	28.8	27.9	27.0	25.2	22.8	21.5	20.9	19.6

x	112.1	112.4	113.6	113.8	115.1	115.4	120.0
y	17.1	18.9	16.0	16.7	13.0	13.6	10.8

Relevant summary quantities are $\sum x_i = 1640.1$, $\sum y_i = 299.8$, $\sum x_i^2 = 179{,}849.73$, $\sum x_i y_i = 32{,}308.59$, $\sum y_i^2 = 6430.06$.
a. Obtain the equation of the estimated regression line. Then create a scatter plot of the data and graph the estimated line. Does it appear that the model relationship will explain a great deal of the observed variation in y?
b. Interpret the slope of the least squares line.
c. What happens if the estimated line is used to predict porosity when unit weight is 135? Why is this not a good idea?
d. Calculate the residuals corresponding to the first two observations.
e. Calculate and interpret a point estimate of σ.
f. What proportion of observed variation in porosity can be attributed to the approximate linear relationship between unit weight and porosity?

18. For the past decade, rubber powder has been used in asphalt cement to improve performance. The article "Experimental Study of Recycled Rubber-Filled High-Strength Concrete" (*Magazine of Concrete Res.*, 2009: 549–556) includes a regression of y = *axial strength* (MPa) on x = *cube strength* (MPa) based on the following sample data:

x	112.3	97.0	92.7	86.0	102.0	99.2	95.8	103.5	89.0	86.7
y	75.0	71.0	57.7	48.7	74.3	73.3	68.0	59.3	57.8	48.5

a. Obtain the equation of the least squares line, and interpret its slope.
b. Calculate and interpret the coefficient of determination.
c. Calculate and interpret an estimate of the error standard deviation σ in the simple linear regression model.

19. The following data is representative of that reported in the article "An Experimental Correlation of Oxides of Nitrogen Emissions from Power Boilers Based on Field Data" (*J. of Engr. for Power*, July 1973: 165–170), with x = burner-area liberation rate (MBtu/hr-ft^2) and y = NO_x emission rate (ppm):

x	100	125	125	150	150	200	200
y	150	140	180	210	190	320	280
x	250	250	300	300	350	400	400
y	400	430	440	390	600	610	670

a. Assuming that the simple linear regression model is valid, obtain the least squares estimate of the true regression line.
b. What is the estimate of expected NO_x emission rate when burner area liberation rate equals 225?
c. Estimate the amount by which you expect NO_x emission rate to change when burner area liberation rate is decreased by 50.
d. Would you use the estimated regression line to predict emission rate for a liberation rate of 500? Why or why not?

20. A number of studies have shown lichens (certain plants composed of an alga and a fungus) to be excellent bioindicators of air pollution. The article "The Epiphytic Lichen Hypogymnia Physodes as a Biomonitor of Atmospheric Nitrogen and Sulphur Deposition in Norway" (*Envir. Monitoring and Assessment,* 1993: 27–47) gives the following data (read from a graph) on $x = NO_3^-$ wet deposition (g N/m^2) and y = lichen N (% dry weight):

x	.05	.10	.11	.12	.31	.37	.42
y	.48	.55	.48	.50	.58	.52	1.02
x	.58	.68	.68	.73	.85	.92	
y	.86	.86	1.00	.88	1.04	1.70	

The author used simple linear regression to analyze the data. Use the accompanying Minitab output to answer the following questions:
a. What are the least squares estimates of β_0 and β_1?
b. Predict lichen N for an NO_3^- deposition value of .5.
c. What is the estimate of σ?
d. What is the value of total variation, and how much of it can be explained by the model relationship?

```
The regression equation is
lichen N = 0.365 + 0.967 no3 depo

Predictor      Coef      Stdev    t-ratio       P
Constant    0.36510    0.09904       3.69   0.004
no3 depo     0.9668     0.1829       5.29   0.000

s = 0.1932     R-sq = 71.7%     R-sq(adj) = 69.2%

Analysis of Variance
SOURCE       DF       SS       MS        F       P
Regression    1   1.0427   1.0427    27.94   0.000
Error        11   0.4106   0.0373
Total        12   1.4533
```

21. Wrinkle recovery angle and tensile strength are the two most important characteristics for evaluating the performance of crosslinked cotton fabric. An increase in the degree of crosslinking, as determined by ester carboxyl band absorbence, improves the wrinkle resistance of the fabric (at the expense of reducing mechanical strength). The accompanying data on x = absorbance and y = wrinke resistance angle was read from a graph in the paper "Predicting the Performance of Durable Press Finished Cotton Fabric with Infrared Spectroscopy" (*Textile Res. J.,* 1999: 145–151).

x	.115	.126	.183	.246	.282	.344	.355	.452	.491	.554	.651
y	334	342	355	363	365	372	381	392	400	412	420

Here is regression output from Minitab:

```
Predictor         Coef    SE Coef        T       P
Constant       321.878      2.483   129.64   0.000
absorb         156.711      6.464    24.24   0.000

S = 3.60498    R-Sq = 98.5%    R-Sq(adj) = 98.3%

Source            DF      SS       MS        F       P
Regression         1  7639.0   7639.0   587.81   0.000
Residual Error     9   117.0     13.0
Total             10  7756.0
```

a. Does the simple linear regression model appear to be appropriate? Explain.
b. What wrinkle resistance angle would you predict for a fabric specimen having an absorbance of .300?
c. What would be the estimate of expected wrinkle resistence angle when absorbance is .300?

22. Calcium phosphate cement is gaining increasing attention for use in bone repair applications. The article "Short-Fibre Reinforcement of Calcium Phosphate Bone Cement" (*J. of Engr. in Med.,* 2007: 203–211) reported on a study in which polypropylene fibers were used in an attempt to improve fracture behavior. The following data on x = fiber weight (%) and y = compressive strength (MPa) was provided by the article's authors.

x	0.00	0.00	0.00	0.00	0.00	1.25	1.25	1.25	1.25
y	9.94	11.67	11.00	13.44	9.20	9.92	9.79	10.99	11.32
x	2.50	2.50	2.50	2.50	2.50	5.00	5.00	5.00	5.00
y	12.29	8.69	9.91	10.45	10.25	7.89	7.61	8.07	9.04
x	7.50	7.50	7.50	7.50	10.00	10.00	10.00	10.00	
y	6.63	6.43	7.03	7.63	7.35	6.94	7.02	7.67	

a. Fit the simple linear regression model to this data. Then determine the proportion of observed variation in strength that can be attributed to the model relationship between strength and fiber weight. Finally, obtain a point estimate of the standard deviation of ϵ, the random deviation in the model equation.
b. The average strength values for the six different levels of fiber weight are 11.05, 10.51, 10.32, 8.15, 6.93, and 7.24, respectively. The cited paper included a figure in which the average strength was regressed against fiber weight. Obtain the equation of this regression line and calculate the corresponding coefficient of determination. Explain

oil (cm/sec) and y = the extent of mist droplets having diameters smaller than 10 μm (mg/m^3):

x	89	177	189	354	362	442	965
y	.40	.60	.48	.66	.61	.69	.99

a. The investigators performed a simple linear regression analysis to relate the two variables. Does a scatter plot of the data support this strategy?

b. What proportion of observed variation in mist can be attributed to the simple linear regression relationship between velocity and mist?

c. The investigators were particularly interested in the impact on mist of increasing velocity from 100 to 1000 (a factor of 10 corresponding to the difference between the smallest and largest x values in the sample). When x increases in this way, is there substantial evidence that the true average increase in y is less than .6?

d. Estimate the true average change in mist associated with a 1 cm/sec increase in velocity, and do so in a way that conveys information about precision and reliability.

37. Magnetic resonance imaging (MRI) is well established as a tool for measuring blood velocities and volume flows. The article "Correlation Analysis of Stenotic Aortic Valve Flow Patterns Using Phase Contrast MRI," referenced in Exercise 1.67, proposed using this methodology for determination of valve area in patients with aortic stenosis. The accompanying data on peak velocity (m/s) from scans of 23 patients in two different planes was read from a graph in the cited paper.

Level-:	.60	.82	.85	.89	.95	1.01	1.01	1.05
Level--:	.50	.68	.76	.64	.68	.86	.79	1.03
Level-:	1.08	1.11	1.18	1.17	1.22	1.29	1.28	1.32
Level--:	.75	.90	.79	.86	.99	.80	1.10	1.15
Level-:	1.37	1.53	1.55	1.85	1.93	1.93	2.14	
Level--:	1.04	1.16	1.28	1.39	1.57	1.39	1.32	

a. Does there appear to be a difference between true average velocity in the two different planes? Carry out an appropriate test of hypotheses (as did the authors of the article).

b. The authors of the article also regressed level--velocity against level- velocity. The resulting estimated intercept and slope are .14701 and .65393, with corresponding estimated standard errors .07877 and .05947, coefficient of determination .852, and s = .110673. The article included a comment that this regression showed evidence of a strong linear relationship but a regression slope well below 1. Do you agree?

38. Refer to the data on x = liberation rate and y = NO_x emission rate given in Exercise 19.

a. Does the simple linear regression model specify a useful relationship between the two rates? Use the appropriate test procedure to obtain information about the P-value, and then reach a conclusion at significance level .01.

b. Compute a 95% CI for the expected change in emission rate associated with a 10 MBtu/hr-ft^2 increase in liberation rate.

39. Carry out the model utility test using the ANOVA approach for the filtration rate–moisture content data of Example 12.6. Verify that it gives a result equivalent to that of the t test.

40. Use the rules of expected value to show that $\hat{\beta}_0$ is an unbiased estimator for β_0 (assuming that $\hat{\beta}_1$ is unbiased for β_1).

41. a. Verify that $E(\hat{\beta}_1) = \beta_1$ by using the rules of expected value from Chapter 5.

b. Use the rules of variance from Chapter 5 to verify the expression for $V(\hat{\beta}_1)$ given in this section.

42. Verify that if each x_i is multiplied by a positive constant c and each y_i is multiplied by another positive constant d, the t statistic for testing H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ is unchanged in value (the value of $\hat{\beta}_1$ will change, which shows that the magnitude of $\hat{\beta}_1$ is not by itself indicative of model utility).

43. The probability of a type II error for the t test for H_0: $\beta_1 = \beta_{10}$ can be computed in the same manner as it was computed for the t tests of Chapter 8. If the alternative value of β_1 is denoted by β_1', the value of

$$d = \frac{|\beta_{10} - \beta_1'|}{\sigma\sqrt{\dfrac{n-1}{\Sigma x_i^2 - (\Sigma x_i)^2/n}}}$$

is first calculated, then the appropriate set of curves in Appendix Table A.17 is entered on the horizontal axis at the value of d, and β is read from the curve for $n - 2$ df. An article in the *Journal of Public Health Engineering* reports the results of a regression analysis based on $n = 15$ observations in which x = filter application temperature (°C) and y = % efficiency of BOD removal. Calculated quantities include $\Sigma x_i = 402$, $\Sigma x_i^2 = 11{,}098$, $s = 3.725$, and $\hat{\beta}_1 = 1.7035$. Consider testing at level .01 H_0: $\beta_1 = 1$, which states that the expected increase in % BOD removal is 1 when filter application temperature increases by 1°C, against the alternative H_a: $\beta_1 > 1$. Determine P(type II error) when $\beta_1' = 2$, $\sigma = 4$.

12.4 Inferences Concerning $\mu_{Y \cdot x^*}$ and the Prediction of Future Y Values

Let x^* denote a specified value of the independent variable x. Once the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ have been calculated, $\hat{\beta}_0 + \hat{\beta}_1 x^*$ can be regarded either as a point estimate of $\mu_{Y \cdot x^*}$ (the expected or true average value of Y when $x = x^*$) or as a prediction of the Y value that will result from a single observation made when associated with a 10,000 liter increase in the amount filtered? Test appropriate hypotheses using $\alpha = .05$.

f. Calculate and interpret a 95% CI for true average % removed when amount filtered is 100,000 liters. How does this interval compare in width to a CI when amount filtered is 200,000 liters?

g. Calculate and interpret a 95% PI for % removed when amount filtered is 100,000 liters. How does this interval compare in width to the CI calculated in (f) and to a PI when amount filtered is 200,000 liters?

49. You are told that a 95% CI for expected lead content when traffic flow is 15, based on a sample of $n = 10$ observations, is (462.1, 597.7). Calculate a CI with confidence level 99% for expected lead content when traffic flow is 15.

50. Silicon-germanium alloys have been used in certain types of solar cells. The paper "Silicon-Germanium Films Deposited by Low-Frequency Plasma-Enhanced Chemical Vapor Deposition" (*J. of Material Res.*, 2006: 88–104) reported on a study of various structural and electrical properties. Consider the accompanying data on x = Ge concentration in solid phase (ranging from 0 to 1) and y = Fermi level position (eV):

x	0	.42	.23	.33	.62	.60	.45	.87	.90	.79	1	1	1
y	.62	.53	.61	.59	.50	.55	.59	.31	.43	.46	.23	.22	.19

A scatter plot shows a substantial linear relationship. Here is Minitab output from a least squares fit. [*Note:* There are several inconsistencies between the data given in the paper, the plot that appears there, and the summary information about a regression analysis.]

```
The regression equation is
Fermi pos = 0.7217 - 0.4327 Ge conc

S = 0.0737573   R-Sq = 80.2%   R-Sq(adj) = 78.4%

Analysis of Variance

Source      DF        SS        MS       F      P
Regression   1  0.241728  0.241728   44.43  0.000
Error       11  0.059842  0.005440
Total       12  0.301569
```

a. Obtain an interval estimate of the expected change in Fermi-level position associated with an increase of .1 in Ge concentration, and interpret your estimate.

b. Obtain an interval estimate for mean Fermi-level position when concentration is .50, and interpret your estimate.

c. Obtain an interval of plausible values for position resulting from a single observation to be made when concentration is .50, interpret your interval, and compare to the interval of (b).

d. Obtain simultaneous CIs for expected position when concentration is .3, .5, and .7; the joint confidence level should be at least 97%.

51. Refer to Example 12.12 in which x = test track speed and y = rolling test speed.

a. Minitab gave $s_{\hat{\beta}_0+\hat{\beta}_1(45)} = .120$ and $s_{\hat{\beta}_0+\hat{\beta}_1(47)} = .186$. Why is the former estimated standard deviation smaller than the latter one?

b. Use the Minitab output from the example to calculate a 95% CI for expected rolling speed when test speed = 45.

c. Use the Minitab output to calculate a 95% PI for a single value of rolling speed when test speed = 47.

52. Plasma etching is essential to the fine-line pattern transfer in current semiconductor processes. The article "Ion Beam-Assisted Etching of Aluminum with Chlorine" (*J. of the Electrochem. Soc.*, 1985: 2010– 2012) gives the accompanying data (read from a graph) on chlorine flow (x, in SCCM) through a nozzle used in the etching mechanism and etch rate (y, in 100 A/min).

x	1.5	1.5	2.0	2.5	2.5	3.0	3.5	3.5	4.0
y	23.0	24.5	25.0	30.0	33.5	40.0	40.5	47.0	49.0

The summary statistics are $\sum x_i = 24.0$, $\sum y_i = 312.5$, $\sum x_i^2 = 70.50$, $\sum x_i y_i = 902.25$, $\sum y_i^2 = 11{,}626.75$, $\hat{\beta}_0 = 6.448718$, $\hat{\beta}_1 = 10.602564$.

a. Does the simple linear regression model specify a useful relationship between chlorine flow and etch rate?

b. Estimate the true average change in etch rate associated with a 1-SCCM increase in flow rate using a 95% confidence interval, and interpret the interval.

c. Calculate a 95% CI for $\mu_{Y \cdot 3.0}$, the true average etch rate when flow = 3.0. Has this average been precisely estimated?

d. Calculate a 95% PI for a single future observation on etch rate to be made when flow = 3.0. Is the prediction likely to be accurate?

e. Would the 95% CI and PI when flow = 2.5 be wider or narrower than the corresponding intervals of parts (c) and (d)? Answer without actually computing the intervals.

f. Would you recommend calculating a 95% PI for a flow of 6.0? Explain.

53. Consider the following four intervals based on the data of Exercise 12.17 (Section 12.2):

a. A 95% CI for mean porosity when unit weight is 110

b. A 95% PI for porosity when unit weight is 110

c. A 95% CI for mean porosity when unit weight is 115

d. A 95% PI for porosity when unit weight is 115

Without computing any of these intervals, what can be said about their widths relative to one another?

54. The decline of water supplies in certain areas of the United States has created the need for increased understanding of relationships between economic factors such as crop yield and hydrologic and soil factors. The article "Variability of Soil Water Properties and Crop Yield in a Sloped Watershed" (*Water Resources Bull.*, 1988: 281–288) gives data on grain sorghum yield (y, in g/m-row) and distance upslope (x, in m) on a sloping watershed. Selected observations are given in the accompanying table.

Example 12.19 The article "A Study of a Partial Nutrient Removal System for Wastewater Treatment Plants" (*Water Research,* 1972: 1389–1397) reports on a method of nitrogen removal that involves the treatment of the supernatant from an aerobic digester. Both the influent total nitrogen x (mg/L) and the percentage y of nitrogen removed were recorded for 20 days, with resulting summary statistics $\Sigma x_i = 285.90$, $\Sigma x_i^2 = 4409.55$, $\Sigma y_i = 690.30$, $\Sigma y_i^2 = 29{,}040.29$, and $\Sigma x_i y_i = 10{,}818.56$. The sample correlation coefficient between influent nitrogen and percentage nitrogen removed is $r = .733$, giving $v = .935$. With $n = 20$, a 95% confidence interval for μ_V is $(.935 - 1.96/\sqrt{17}, .935 + 1.96/\sqrt{17}) = (.460, 1.410) = (c_1, c_2)$. The 95% interval for ρ is

$$\left[\frac{e^{2(.46)} - 1}{e^{2(.46)} + 1}, \frac{e^{2(1.41)} - 1}{e^{2(1.41)} + 1}\right] = (.43, .89)$$ ■

In Chapter 5, we cautioned that a large value of the correlation coefficient (near 1 or -1) implies only association and not causation. This applies to both ρ and r.

EXERCISES Section 12.5 (57–67)

57. The article "Behavioural Effects of Mobile Telephone Use During Simulated Driving" (*Ergonomics,* 1995: 2536–2562) reported that for a sample of 20 experimental subjects, the sample correlation coefficient for x = age and y = time since the subject had acquired a driving license (yr) was .97. Why do you think the value of r is so close to 1? (The article's authors give an explanation.)

58. The Turbine Oil Oxidation Test (TOST) and the Rotating Bomb Oxidation Test (RBOT) are two different procedures for evaluating the oxidation stability of steam turbine oils. The article "Dependence of Oxidation Stability of Steam Turbine Oil on Base Oil Composition" (*J. of the Society of Tribologists and Lubrication Engrs.,* Oct. 1997: 19–24) reported the accompanying observations on x = TOST time (hr) and y = RBOT time (min) for 12 oil specimens.

TOST	4200	3600	3750	3675	4050	2770
RBOT	370	340	375	310	350	200
TOST	4870	4500	3450	2700	3750	3300
RBOT	400	375	285	225	345	285

a. Calculate and interpret the value of the sample correlation coefficient (as do the article's authors).
b. How would the value of r be affected if we had let x = RBOT time and y = TOST time?
c. How would the value of r be affected if RBOT time were expressed in hours?
d. Construct normal probability plots and comment.
e. Carry out a test of hypotheses to decide whether RBOT time and TOST time are linearly related.

59. Toughness and fibrousness of asparagus are major determinants of quality. This was the focus of a study reported in "Post-Harvest Glyphosphate Application Reduces Toughening, Fiber Content, and Lignification of Stored Asparagus Spears" (*J. of the Amer. Soc. of Hort. Science,* 1988: 569–572). The article reported the accompanying data (read from a graph) on x = shear force (kg) and y = percent fiber dry weight.

x	46	48	55	57	60	72	81	85	94
y	2.18	2.10	2.13	2.28	2.34	2.53	2.28	2.62	2.63
x	109	121	132	137	148	149	184	185	187
y	2.50	2.66	2.79	2.80	3.01	2.98	3.34	3.49	3.26

$n = 18$, $\Sigma x_i = 1950$, $\Sigma x_i^2 = 251{,}970$,
$\Sigma y_i = 47.92$, $\Sigma y_i^2 = 130.6074$, $\Sigma x_i y_i = 5530.92$

a. Calculate the value of the sample correlation coefficient. Based on this value, how would you describe the nature of the relationship between the two variables?
b. If a first specimen has a larger value of shear force than does a second specimen, what tends to be true of percent dry fiber weight for the two specimens?
c. If shear force is expressed in pounds, what happens to the value of r? Why?
d. If the simple linear regression model were fit to this data, what proportion of observed variation in percent fiber dry weight could be explained by the model relationship?
e. Carry out a test at significance level .01 to decide whether there is a positive linear association between the two variables.

60. Head movement evaluations are important because individuals, especially those who are disabled, may be able to operate communications aids in this manner. The article "Constancy