Chapter 14
Simple Linear Regression

Learning Objectives

1. Understand how regression analysis can be used to develop an equation that estimates mathematically how two variables are related.

2. Understand the differences between the regression model, the regression equation, and the estimated regression equation.

3. Know how to fit an estimated regression equation to a set of sample data based upon the least-squares method.

4. Be able to determine how good a fit is provided by the estimated regression equation and compute the sample correlation coefficient from the regression analysis output.

5. Understand the assumptions necessary for statistical inference and be able to test for a significant relationship.

6. Learn how to use a residual plot to make a judgement as to the validity of the regression assumptions, recognize outliers, and identify influential observations.

7. Know how to develop confidence interval estimates of $y$ given a specific value of $x$ in both the case of a mean value of $y$ and an individual value of $y$.

8. Be able to compute the sample correlation coefficient from the regression analysis output.

9. Know the definition of the following terms:
   - independent and dependent variable
   - simple linear regression
   - regression model
   - regression equation and estimated regression equation
   - scatter diagram
   - coefficient of determination
   - standard error of the estimate
   - confidence interval
   - prediction interval
   - residual plot
   - standardized residual plot
   - outlier
   - influential observation
   - leverage
Chapter 14

Solutions:

1. a.

b. There appears to be a linear relationship between $x$ and $y$.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between $x$ and $y$; in part d we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

d. Summations needed to compute the slope and y-intercept are:

\[
\begin{align*}
\Sigma x_i &= 15 \\
\Sigma y_i &= 40 \\
\Sigma (x_i - \bar{x})(y_i - \bar{y}) &= 26 \\
\Sigma (x_i - \bar{x})^2 &= 10
\end{align*}
\]

\[
b_y = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{26}{10} = 2.6
\]

\[
b_x = \bar{y} - b_y \bar{x} = 8 - (2.6)(3) = 0.2
\]

\[
\hat{y} = 0.2 - 2.6x
\]

e. \hat{y} = 0.2 - 2.6(4) = 10.6
2. a. There appears to be a linear relationship between $x$ and $y$.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between $x$ and $y$; in part d we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

d. Summations needed to compute the slope and $y$-intercept are:

$$
\begin{align*}
\sum x_i &= 19 \quad \sum y_i = 116 \\
\sum (x_i - \bar{x})(y_i - \bar{y}) &= -57.8 \\
\sum (x_i - \bar{x})^2 &= 30.8
\end{align*}
$$

$$
\begin{align*}
b_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{-57.8}{30.8} = -1.8766 \\
b_0 &= \bar{y} - b_1 \bar{x} = 23.2 - (-18766)(3.8) = 303311 \\
\hat{y} &= 30.33 - 1.88x
\end{align*}
$$

e. $\hat{y} = 30.33 - 1.88(6) = 19.05$
3. a.

b. Summations needed to compute the slope and y-intercept are:

\[ \Sigma x_i = 26 \quad \Sigma y_i = 17 \quad \Sigma (x_i - \bar{x})(\bar{y} - y) = 11.6 \quad \Sigma (x_i - \bar{x})^2 = 22.8 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(\bar{y} - y)}{\Sigma (x_i - \bar{x})^2} = \frac{11.6}{22.8} = 0.5088 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 3.4 - (0.5088)(5.2) = 0.7542 \]

\[ \hat{y} = 0.75 + 0.51x \]

c. \[ \hat{y} = 0.75 + 0.51(4) = 2.79 \]
Simple Linear Regression

4. a.

b. There:

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between \( x \) and \( y \); in part d we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

d. Summations needed to compute the slope and y-intercept are:

\[
\Sigma x_i = 325 \quad \Sigma y_i = 585 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 110 \quad \Sigma (x_i - \bar{x})^2 = 20
\]

\[
b_y = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{110}{20} = 5.5
\]

\[
b_y = \bar{y} - b_y \bar{x} = 117 - (5.5)(65) = -240.5
\]

\[
\hat{y} = -240.5 + 5.5x
\]

e. \(\hat{y} = -240.5 + 5.5x = -240.5 + 5.5(63) = 106 \text{ pounds}\)
5. a. 

There appears to be a linear relationship between $x$ and $y$.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between $x$ and $y$; in part d we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

Summations needed to compute the slope and y-intercept are:

\[ \Sigma x_i = 420.6 \quad \Sigma y_i = 5958.7 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 142,040.3443 \quad \Sigma (x_i - \bar{x})^2 = 9847.6486 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{142,040.3443}{9847.6486} = 14.4238 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 851.2429 - (14.4238)(60.0857) = -15.42 \]

\[ \hat{y} = -15.42 + 14.42x \]

d. A one million dollar increase in media expenditures will increase case sales by approximately 14.42 million.

e. \[ \hat{y} = -15.42 + 14.42x = -15.42 + 14.42(70) = 993.98 \]
6. a. 

b. There appears to be a linear relationship between $x$ and $y$.

c. Summations needed to compute the slope and $y$-intercept are:

\[ \Sigma x_i = 667.2 \quad \Sigma y_i = 7.18 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = -9.0623 \quad \Sigma (x_i - \bar{x})^2 = 128.7 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{-9.0623}{128.7} = -0.0704 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 0.7978 - (-0.0704)(74.1333) = 6.02 \]

\[ \hat{y} = 6.02 - 0.07x \]

d. A one percent increase in the percentage of flights arriving on time will decrease the number of complaints per 100,000 passengers by 0.07.

e. \[ \hat{y} = 6.02 - 0.07x = 6.02 - 0.07(80) = 0.42 \]
7. a.

b. Let \( x = \text{DJIA} \) and \( y = \text{S&P} \). Summations needed to compute the slope and \( y \)-intercept are:

\[
\Sigma x_i = 104,850 \quad \Sigma y_i = 14,233 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 268,921 \quad \Sigma (x_i - \bar{x})^2 = 1,806,384
\]

\[
b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{268,921}{1,806,384} = 0.14887
\]

\[
b_0 = \bar{y} - b \bar{x} = 1423.3 - (0.14887)(10,485) = -137.629
\]

\[\hat{y} = -137.63 + 0.1489x\]

c. \[\hat{y} = -137.63 + 0.1489(11,000) = 1500.27 \text{ or approximately } 1500\]

8. a. Summations needed to compute the slope and \( y \)-intercept are:

\[
\Sigma x_i = 121 \quad \Sigma y_i = 1120.9 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 544.0429 \quad \Sigma (x_i - \bar{x})^2 = 177.4286
\]

\[
b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{544.0429}{177.4286} = 3.0663
\]

\[
b_0 = \bar{y} - b \bar{x} = 160.1286 - (3.0663)(17.2857) = 107.13
\]

\[\hat{y} = 107.13 + 3.07x\]

b. Increasing the number of times an ad is aired by one will increase the number of household exposures by approximately 3.07 million.
c. \( \hat{y} = 107.13 + 3.07x = 107.13 + 307(15) = 153.2 \)

9. a.

b. Summations needed to compute the slope and y-intercept are:

\[
\Sigma x_i = 70 \quad \Sigma y_i = 1080 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 568 \quad \Sigma (x_i - \bar{x})^2 = 142
\]

\[
b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{568}{142} = 4
\]

\[
b_0 = \bar{y} - b_1 \bar{x} = 108 - (4)(7) = 80
\]

\[
\hat{y} = 80 + 4x
\]

c. \( \hat{y} = 80 + 4x = 80 + 4(9) = 116 \)
10. a. Let \( x \) = performance score and \( y \) = overall rating. Summations needed to compute the slope and \( y \)-intercept are:

\[
\Sigma x_i = 2752 \quad \Sigma y_i = 1177 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 1723.73 \quad \Sigma (x_i - \bar{x})^2 = 11,867.73
\]

\[
b_i = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{1723.73}{11,867.73} = 0.1452
\]

\[
b_0 = \bar{y} - b_i \bar{x} = 78.4667 - (0.1452)(183.4667) = 51.82
\]

\[\hat{y} = 51.82 + 0.145x\]

c. \[\hat{y} = 51.82 + 0.145(225) = 84.4\] or approximately 84
11. a. 

b. There appears to be a linear relationship between the variables.

c. The summations needed to compute the slope and the y-intercept are:

\[ \Sigma x_i = 2973.3 \quad \Sigma y_i = 3925.6 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 453,345.042 \quad \Sigma (x_i - \bar{x})^2 = 483,507.581 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{453,345.042}{483,507.581} = 0.9385 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 392.56 - (0.9385)(297.33) = 113.52 \]

\[ \hat{y} = 113.52 + 0.94x \]

d. \[ \hat{y} = 113.52 + 0.94x = 113.52 + 094(500) = 5835 \]
12. a.

b. There appears to be a positive linear relationship between the number of employees and the revenue.

c. Let $x =$ number of employees and $y =$ revenue. Summations needed to compute the slope and y-intercept are:

\[ \Sigma x_i = 4200 \quad \Sigma y_i = 1669 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 4,658,594.168 \quad \Sigma (x_i - \bar{x})^2 = 14,718,343.803 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{4,658,594.168}{14,718,343.803} = 0.316516 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 14,048 - (0.316516)(40,299) = 1293 \]

\[ \hat{y} = 1293 + 0.3165x \]

d. \[ \hat{y} = 1293 + 0.3165(75,000) = 25,031 \]
13. a. 

b.  The summations needed to compute the slope and the y-intercept are:

\[ \Sigma x_i = 399 \quad \Sigma y_i = 97.1 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 1233.7 \quad \Sigma (x_i - \bar{x})^2 = 7648 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{1233.7}{7648} = 0.16131 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 13.87143 - (0.16131)(57) = 4.67675 \]

\[ \hat{y} = 4.68 + 0.16x \]

c. \[ \hat{y} = 4.68 + 0.16x = 4.68 + 0.16(52.5) = 13.08 \text{ or approximately$13,080.} \]

The agent's request for an audit appears to be justified.
14. a.

The summations needed to compute the slope and the y-intercept are:

\[ \Sigma x_i = 1677.25 \quad \Sigma y_i = 1404.3 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 897.9493 \quad \Sigma (x_i - \bar{x})^2 = 3657.4568 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{897.9493}{3657.4568} = 0.2455 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 70.215 - (0.2455)(838625) = 49.63 \]

\[ \hat{y} = 49.63 + 0.2455x \]

c. \( \hat{y} = 49.63 + 0.2455x = 49.63 + 0.2455(80) = 69.3\% \)

15. a. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y}_i = 0.2 + 2.6x_i \quad \bar{y} = 8 \]

The sum of squares due to error and the total sum of squares are

\[ SSE = \sum (y_i - \hat{y}_i)^2 = 12.40 \quad SST = \sum (y_i - \bar{y})^2 = 80 \]

Thus, \( SSR = SST - SSE = 80 - 12.4 = 67.6 \)

b. \( r^2 = SSR/SST = 67.6/80 = .845 \)

The least squares line provided a very good fit; 84.5\% of the variability in \( y \) has been explained by the least squares line.

c. \( r = \sqrt{.845} = +.9192 \)
16. a. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y}_i = 30.33 - 1.88x \quad \bar{y} = 23.2 \]

The sum of squares due to error and the total sum of squares are

\[ \text{SSE} = \sum (y_i - \hat{y}_i)^2 = 6.33 \quad \text{SST} = \sum (y_i - \bar{y})^2 = 114.80 \]

Thus, \( \text{SSR} = \text{SST} - \text{SSE} = 114.80 - 6.33 = 108.47 \)

b. \( r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{108.47}{114.80} = .945 \)

The least squares line provided an excellent fit; 94.5% of the variability in \( y \) has been explained by the estimated regression equation.

c. \( r = \sqrt{.945} = -.9721 \)

Note: the sign for \( r \) is negative because the slope of the estimated regression equation is negative. \((b_1 = -1.88)\)

17. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y}_i = .75 + .51x \quad \bar{y} = 3.4 \]

The sum of squares due to error and the total sum of squares are

\[ \text{SSE} = \sum (y_i - \hat{y}_i)^2 = 5.3 \quad \text{SST} = \sum (y_i - \bar{y})^2 = 11.2 \]

Thus, \( \text{SSR} = \text{SST} - \text{SSE} = 11.2 - 5.3 = 5.9 \)

\( r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{5.9}{11.2} = .527 \)

We see that 52.7% of the variability in \( y \) has been explained by the least squares line.

\( r = \sqrt{.527} = .7259 \)

18. a. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y} = 1790.5 + 581.1x \quad \bar{y} = 3650 \]

The sum of squares due to error and the total sum of squares are

\[ \text{SSE} = \sum (y_i - \hat{y}_i)^2 = 85,135.14 \quad \text{SST} = \sum (y_i - \bar{y})^2 = 335,000 \]

Thus, \( \text{SSR} = \text{SST} - \text{SSE} = 335,000 - 85,135.14 = 249,864.86 \)

b. \( r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{249,864.86}{335,000} = .746 \)

We see that 74.6% of the variability in \( y \) has been explained by the least squares line.
c. \( r = \sqrt{0.746} = +.8637 \)

19. a. The estimated regression equation and the mean for the dependent variable are:
\[
\hat{y} = -137.63 + 148.91x \\
\bar{y} = 1423.3
\]
The sum of squares due to error and the total sum of squares are
\[
SSE = \sum(y_i - \hat{y}_i)^2 = 7547.14 \\
SST = \sum(y_i - \bar{y})^2 = 47,582.10
\]
Thus, \( SSR = SST - SSE = 47,582.10 - 7547.14 = 40,034.96 \)
\[r^2 = \frac{SSR}{SST} = \frac{40,034.96}{47,582.10} = .84\]
We see that 84% of the variability in \( y \) has been explained by the least squares line.

c. \( r = \sqrt{0.84} = +.92 \)

20. a. Let \( x = \text{income} \) and \( y = \text{home price} \). Summations needed to compute the slope and \( y \)-intercept are:
\[
\Sigma x_i = 1424 \\
\Sigma y_i = 2455.5 \\
\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 4011 \\
\Sigma (x_i - \bar{x})^2 = 1719.618
\]
\[
b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{4011}{1719.618} = 2.3325
\]
\[
b_0 = \bar{y} - b_1\bar{x} = 136.4167 - (2.3325)(79.1111) = -48.11
\]
\[
\hat{y} = -48.11 + 2.3325x
\]

b. The sum of squares due to error and the total sum of squares are
\[
SSE = \sum(y_i - \hat{y}_i)^2 = 2017.37 \\
SST = \sum(y_i - \bar{y})^2 = 11,373.09
\]
Thus, \( SSR = SST - SSE = 11,373.09 - 2017.37 = 9355.72 \)
\[r^2 = \frac{SSR}{SST} = \frac{9355.72}{11,373.09} = .82\]
We see that 82% of the variability in \( y \) has been explained by the least squares line.
\[r = \sqrt{0.82} = +.91\]

c. \( \hat{y} = -48.11 + 2.3325(95) = 173.5 \) or approximately $173,500

21. a. The summations needed in this problem are:
\[
\Sigma x_i = 3450 \\
\Sigma y_i = 33,700 \\
\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 712,500 \\
\Sigma (x_i - \bar{x})^2 = 93,750
\]
\[
b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{712,500}{93,750} = 7.6
\]
Simple Linear Regression

\[ b_0 = \overline{y} - b_1 \overline{x} = 5616.67 - (7.6)(575) = 1246.67 \]

\[ \hat{y} = 1246.67 + 7.6x \]

b. \$7.60

c. The sum of squares due to error and the total sum of squares are:

\[ \text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 233,333.33 \quad \text{SST} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 5,648,333.33 \]

Thus,
\[ \text{SSR} = \text{SST} - \text{SSE} = 5,648,333.33 - 233,333.33 = 5,415,000 \]

\[ r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{5,415,000}{5,648,333.33} = .9587 \]

We see that 95.87% of the variability in \( y \) has been explained by the estimated regression equation.

d. \( \hat{y} = 1246.67 + 7.6x = 1246.67 + 7.6(500) = \$5046.67 \)

22. a. The summations needed in this problem are:

\[ \Sigma x_i = 613.1 \quad \Sigma y_i = 70 \quad \Sigma (x_i - \overline{x})(y_i - \overline{y}) = 5766.7 \quad \Sigma (x_i - \overline{x})^2 = 45,833.9286 \]

\[ b_1 = \frac{\Sigma (x_i - \overline{x})(y_i - \overline{y})}{\Sigma (x_i - \overline{x})^2} = \frac{5766.7}{45,833.9286} = 0.1258 \]

\[ b_0 = \overline{y} - b_1 \overline{x} = 10 - (0.1258)(87.5857) = -1.0183 \]

\[ \hat{y} = -1.0183 + 0.1258x \]

b. The sum of squares due to error and the total sum of squares are:

\[ \text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 1272.4495 \quad \text{SST} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 1998 \]

Thus,
\[ \text{SSR} = \text{SST} - \text{SSE} = 1998 - 1272.4495 = 725.5505 \]

\[ r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{725.5505}{1998} = 0.3631 \]

Approximately 37% of the variability in change in executive compensation is explained by the two-year change in the return on equity.

c. \[ r = \sqrt{0.3631} = +0.6026 \]

It reflects a linear relationship that is between weak and strong.

23. a. \[ s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{12.4}{3} = 4.133 \]

b. \[ s = \sqrt{\text{MSE}} = \sqrt{4.133} = 2.033 \]

c. \[ \Sigma (x_i - \overline{x})^2 = 10 \]


\[
s_h = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{2.033}{\sqrt{10}} = 0.643
\]

d. \[ t = \frac{b_i}{s_h} = \frac{2.6}{.643} = 4.04 \]

\[ t_{0.05} = 3.182 \] (3 degrees of freedom)

Since \( t = 4.04 \) > \( t_{0.05} = 3.182 \), we reject \( H_0: \beta_1 = 0 \)

e. MSR = SSR / 1 = 67.6

\[ F = MSR / MSE = 67.6 / 4.133 = 16.36 \]

\( F_{0.05} = 10.13 \) (1 degree of freedom numerator and 3 denominator)

Since \( F = 16.36 > F_{0.05} = 10.13 \), we reject \( H_0: \beta_1 = 0 \)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>67.6</td>
<td>1</td>
<td>67.6</td>
<td>16.36</td>
</tr>
<tr>
<td>Error</td>
<td>12.4</td>
<td>3</td>
<td>4.133</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>80.0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24. a. \[ s^2 = MSE = SSE / (n - 2) = 6.33 / 3 = 2.11 \]

b. \[ s = \sqrt{MSE} = \sqrt{2.11} = 1.453 \]

c. \[ \sum(x_i - \bar{x})^2 = 30.8 \]

\[ s_h = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{1.453}{\sqrt{30.8}} = 0.262 \]

d. \[ t = \frac{b_i}{s_h} = \frac{-1.88}{.262} = -7.18 \]

\[ t_{0.05} = 3.182 \] (3 degrees of freedom)

Since \( t = -7.18 < -t_{0.05} = -3.182 \), we reject \( H_0: \beta_1 = 0 \)

e. MSR = SSR / 1 = 8.47

\[ F = MSR / MSE = 108.47 / 2.11 = 51.41 \]

\( F_{0.05} = 10.13 \) (1 degree of freedom numerator and 3 denominator)

Since \( F = 51.41 > F_{0.05} = 10.13 \), we reject \( H_0: \beta_1 = 0 \)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>108.47</td>
<td>1</td>
<td>108.47</td>
<td>51.41</td>
</tr>
<tr>
<td>Error</td>
<td>6.33</td>
<td>3</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>114.80</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
25. a. \[ s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{5.30}{3} = 1.77 \]
\[ s = \sqrt{\text{MSE}} = \sqrt{1.77} = 1.33 \]

b. \[ \Sigma(x_i - \bar{x})^2 = 22.8 \]
\[ s_h = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{1.33}{\sqrt{22.8}} = 0.28 \]
\[ t = \frac{b_1}{s_h} = \frac{51}{28} = 1.82 \]
\[ t_{0.025} = 3.182 \quad \text{(3 degrees of freedom)} \]
Since \( t = 1.82 < t_{0.025} = 3.182 \) we cannot reject \( H_0: \beta_1 = 0; x \) and \( y \) do not appear to be related.

c. MSR = \( \frac{\text{SSR}}{1} = \frac{5.90}{1} = 5.90 \)
\[ F = \frac{\text{MSR}}{\text{MSE}} = \frac{5.90}{1.77} = 3.33 \]
\[ F_{0.05} = 10.13 \quad \text{(1 degree of freedom numerator and 3 denominator)} \]
Since \( F = 3.33 < F_{0.05} = 10.13 \) we cannot reject \( H_0: \beta_1 = 0; x \) and \( y \) do not appear to be related.

26. a. \[ s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{85,135.14}{4} = 21,283.79 \]
\[ s = \sqrt{\text{MSE}} = \sqrt{21,283.79} = 145.89 \]
\[ \Sigma(x_i - \bar{x})^2 = 0.74 \]
\[ s_h = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{145.89}{\sqrt{0.74}} = 169.59 \]
\[ t = \frac{b_1}{s_h} = \frac{58.108}{169.59} = 3.43 \]
\[ t_{0.025} = 2.776 \quad \text{(4 degrees of freedom)} \]
Since \( t = 3.43 > t_{0.025} = 2.776 \) we reject \( H_0: \beta_1 = 0 \)

b. MSR = \( \frac{\text{SSR}}{1} = \frac{249,864.86}{1} = 249,864.86 \)
\[ F = \frac{\text{MSR}}{\text{MSE}} = \frac{249,864.86}{21,283.79} = 11.74 \]
\[ F_{0.05} = 7.71 \quad \text{(1 degree of freedom numerator and 4 denominator)} \]
Since \( F = 11.74 > F_{0.05} = 7.71 \) we reject \( H_0: \beta_1 = 0 \)

c.
Chapter 14

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>249864.86</td>
<td>1</td>
<td>249864.86</td>
<td>11.74</td>
</tr>
<tr>
<td>Error</td>
<td>85135.14</td>
<td>4</td>
<td>21283.79</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>335000</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27. The sum of squares due to error and the total sum of squares are:

\[
SSE = \Sigma (y_i - \hat{y}_i)^2 = 170 \\
SST = 2442
\]

Thus, \( SSR = SST - SSE = 2442 - 170 = 2272 \)

\[
MSR = SSR / 1 = 2272
\]

\[
SSE = SST - SSR = 2442 - 2272 = 170
\]

\[
MSE = SSE / (n - 2) = 170 / 8 = 21.25
\]

\[
F = MSR / MSE = 2272 / 21.25 = 106.92
\]

\( F_{.05} = 5.32 \) (1 degree of freedom numerator and 8 denominator)

Since \( F = 106.92 > F_{.05} = 5.32 \) we reject \( H_0: \beta_1 = 0. \)

Years of experience and sales are related.

28. SST = 411.73 \quad SSE = 161.37 \quad SSR = 250.36

\[
MSR = SSR / 1 = 250.36
\]

\[
MSE = SSE / (n - 2) = 161.37 / 13 = 12.413
\]

\[
F = MSR / MSE = 250.36 / 12.413 = 20.17
\]

\( F_{.05} = 4.67 \) (1 degree of freedom numerator and 13 denominator)

Since \( F = 20.17 > F_{.05} = 4.67 \) we reject \( H_0: \beta_1 = 0. \)

29. SSE = 233,333.33 \quad SST = 5,648,333.33 \quad SSR = 5,415,000

\[
MSE = SSE / (n - 2) = 233,333.33 / 6 = 58,333.33
\]

\[
MSR = SSR / 1 = 5,415,000
\]

\[
F = MSR / MSE = 5,415,000 / 58,333.25 = 92.83
\]

\( F_{.05} = 7.71 \) (1 degree of freedom numerator and 4 denominator)

Since \( F = 92.83 > 7.71 \) we reject \( H_0: \beta_1 = 0. \) Production volume and total cost are related.
30. Using the computations from Exercise 22,

\[ \text{SSE} = 1272.4495 \quad \text{SST} = 1998 \quad \text{SSR} = 725.5505 \]

\[ s = \sqrt{254.4899} = 15.95 \]

\[ \sum (x_i - \bar{x})^2 = 45,833.9286 \]

\[ s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{15.95}{\sqrt{45,833.9286}} = 0.0745 \]

\[ t = \frac{b_1}{s_{b_1}} = \frac{0.1258}{0.0745} = 1.69 \]

\[ t_{0.025} = 2.571 \]

Since \( t = 1.69 < 2.571 \), we cannot reject \( H_0: \beta_1 = 0 \)

There is no evidence of a significant relationship between \( x \) and \( y \).

31. \( \text{SST} = 11,373.09 \quad \text{SSE} = 2017.37 \quad \text{SSR} = 9355.72 \)

\[ \text{MSR} = \frac{\text{SSR}}{1} = 9355.72 \]

\[ \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{2017.37}{16} = 126.0856 \]

\[ F = \frac{\text{MSR}}{\text{MSE}} = \frac{9355.72}{126.0856} = 74.20 \]

\[ F_{0.01} = 8.53 \quad (1 \text{ degree of freedom numerator and } 16 \text{ denominator}) \]

Since \( F = 74.20 > F_{0.01} = 8.53 \) we reject \( H_0: \beta_1 = 0 \).

32. a. \( s = 2.033 \)

\[ \bar{x} = 3 \quad \sum (x_i - \bar{x})^2 = 10 \]

\[ s_{\hat{y}_p} = s \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 2.033 \sqrt{\frac{1}{5} + \frac{(4 - 3)^2}{10}} = 1.11 \]

b. \( \hat{y} = 0.2 + 2.6x = 0.2 + 2.6(4) = 10.6 \)

\[ \hat{y}_p \pm t_{0.025}s_{\hat{y}_p} \]

\[ 10.6 \pm 3.182 (1.11) = 10.6 \pm 3.53 \]

or 7.07 to 14.13

c. \( s_{\text{tol}} = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 2.033 \sqrt{1 + \frac{1}{5} + \frac{(4 - 3)^2}{10}} = 2.32 \)
d. \( \hat{y}_p \pm t_{a/2}s_{ind} \)

\[
10.6 \pm 3.182 (2.32) = 10.6 \pm 7.38
\]

or 3.22 to 17.98

33. a. \( s = 1.453 \)

b. \( \bar{x} = 3.8 \quad \Sigma(x_i - \bar{x})^2 = 30.8 \)

\[
s_{yp} = s \sqrt{1 + \frac{(x_p - \bar{x})^2}{\frac{1}{n} \Sigma(x_i - \bar{x})^2}} = 1.453 \sqrt{1 + \frac{(3 - 3.8)^2}{30.8}} = .068
\]

\[
\hat{y} = 30.33 - 1.88x = 30.33 - 188(3) = 24.69
\]

\[
\hat{y}_p \pm t_{a/2}s_{yp}
\]

24.69 \( \pm 3.182(.68) = 24.69 \pm 2.16 \)

or 22.53 to 26.85

c. \( s_{ind} = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\frac{1}{n} \Sigma(x_i - \bar{x})^2}} = 1.453 \sqrt{1 + \frac{1}{5} + \frac{(3 - 3.8)^2}{30.8}} = 1.61
\]

d. \( \hat{y}_p \pm t_{a/2}s_{ind} \)

24.69 \( \pm 3.182(1.61) = 24.69 \pm 5.12 \)

or 19.57 to 29.81

34. \( s = 1.33 \)

\( \bar{x} = 5.2 \quad \Sigma(x_i - \bar{x})^2 = 22.8 \)

\[
s_{yp} = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\frac{1}{n} \Sigma(x_i - \bar{x})^2}} = 1.33 \sqrt{1 + \frac{1}{5} + \frac{(3 - 5.2)^2}{22.8}} = 0.85
\]

\[
\hat{y} = 0.75 + 0.51x = 0.75 + 0.51(3) = 2.28
\]

\[
\hat{y}_p \pm t_{a/2}s_{yp}
\]

2.28 \( \pm 3.182(.85) = 2.28 \pm 2.70 \)

or -.40 to 4.98
Simple Linear Regression

\[ s_{\text{ind}} = \sqrt{\frac{1}{n} \left( \frac{1}{n} \sum (x_p - \bar{x})^2 \right)} = 1.33 \sqrt{\frac{1}{5} + \frac{(3-3.2)^2}{22.8}} = 1.58 \]

\[ \hat{y}_p \pm t_{a/2}s_{\text{ind}} \]

\[ 2.28 \pm 3.182 (1.58) = 2.28 \pm 5.03 \]

or -2.27 to 7.31

35. a. \( s = 145.89 \)

\[ \bar{x} = 3.2 \quad \Sigma (x_i - \bar{x})^2 = 0.74 \]

\[ s_{\bar{x}} = s \sqrt{\frac{1}{n} \left( \frac{1}{n} \sum (x_p - \bar{x})^2 \right)} = 145.89 \sqrt{\frac{1}{6} + \frac{(3-3.2)^2}{0.74}} = 68.54 \]

\[ \hat{y} = 290.54 + 581.08 \bar{x} = 290.54 + 581.08(3) = 2033.78 \]

\[ \hat{y}_p \pm t_{a/2}s_{\bar{x}} \]

\[ 2,033.78 \pm 2.776 (68.54) = 2,033.78 \pm 190.27 \]

or $1,843.51 to $2,224.05

b. \( s_{\text{ind}} = \sqrt{\frac{1}{n} \left( \frac{1}{n} \sum (x_p - \bar{x})^2 \right)} = 145.89 \sqrt{\frac{1}{15} + \frac{(200-183.4667)^2}{11,867.73}} = 161.19 \]

\[ \hat{y}_p \pm t_{a/2}s_{\text{ind}} \]

\[ 2,033.78 \pm 2.776 (161.19) = 2,033.78 \pm 447.46 \]

or $1,586.32 to $2,481.24

36. a. \( \hat{y} = 51.819 + .1452x = 51.819 + .1452(200) = 80.859 \)

b. \( s = 3.5232 \)

\[ \bar{x} = 183.4667 \quad \Sigma (x_i - \bar{x})^2 = 11,867.73 \]

\[ s_{\bar{x}} = s \sqrt{\frac{1}{n} \left( \frac{1}{n} \sum (x_p - \bar{x})^2 \right)} = 3.5232 \sqrt{\frac{1}{15} + \frac{(200-183.4667)^2}{11,867.73}} = 1.055 \]

\[ \hat{y}_p \pm t_{a/2}s_{\bar{x}} \]

\[ 80.859 \pm 2.160 (1.055) = 80.859 \pm 2.279 \]
c. 

\[ s_{\text{ind}} = s \sqrt{1 + \frac{1}{n} \cdot \frac{(x_p - \bar{x})}{\frac{\Sigma (x - \bar{x})^2}{n}}} = 3.5232 \sqrt{1 + \frac{1}{15} \cdot \frac{(200 - 183.4667)^2}{11.867.73}} = 3.678 \]

\[ \hat{y}_p \pm t_{a/2} s_{\text{ind}} \]

\[ 80.859 \pm 2.160 (3.678) = 80.859 \pm 7.944 \]

or 72.92 to 88.80

37. a. \[ \bar{x} = 57 \quad \Sigma (x_i - \bar{x})^2 = 7648 \]

\[ s^2 = 1.88 \quad s = 1.37 \]

\[ s_{y_0} = s \sqrt{1 + \frac{1}{n} \cdot \frac{(x_{y_0} - \bar{x})}{\frac{\Sigma (x - \bar{x})^2}{n}}} = 1.37 \sqrt{1 + \frac{(52.5 - 57)^2}{7648}} = 0.52 \]

\[ \hat{y}_p \pm t_{a/2} s_{y_0} \]

\[ 13.08 \pm 2.571 (0.52) = 13.08 \pm 1.34 \]

or 11.74 to 14.42 or $11,740 to $14,420

b. \[ s_{\text{ind}} = 1.47 \]

\[ 13.08 \pm 2.571 (1.47) = 13.08 \pm 3.78 \]

or 9.30 to 16.86 or $9,300 to $16,860

c. Yes, $20,400 is much larger than anticipated.

d. Any deductions exceeding the $16,860 upper limit could suggest an audit.

38. a. \[ \hat{y} = 1246.67 + 7.6(500) = 5046.67 \]

b. \[ \bar{x} = 575 \quad \Sigma (x_i - \bar{x})^2 = 93,750 \]

\[ s^2 = \text{MSE} = 58,333.33 \quad s = 241.52 \]

\[ s_{\text{ind}} = s \sqrt{1 + \frac{1}{n} \cdot \frac{(x_p - \bar{x})}{\frac{\Sigma (x - \bar{x})^2}{n}}} = 241.52 \sqrt{1 + \frac{1}{6} \cdot \frac{(500 - 575)^2}{93,750}} = 267.50 \]

\[ \hat{y}_p \pm t_{a/2} s_{\text{ind}} \]

\[ 5046.67 \pm 4.604 (267.50) = 5046.67 \pm 1231.57 \]
Simple Linear Regression

or $3815.10 to $6278.24

c. Based on one month, $6000 is not out of line since $3815.10 to $6278.24 is the prediction interval. However, a sequence of five to seven months with consistently high costs should cause concern.

39. a. Summations needed to compute the slope and \( y \)-intercept are:

\[
\begin{align*}
\Sigma x_i &= 227 \\
\Sigma y_i &= 2281.7 \\
\Sigma(x_i - \bar{x})(y_i - \bar{y}) &= 6003.41 \\
\Sigma(x_i - \bar{x})^2 &= 1032.1
\end{align*}
\]

\[
b = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{6003.41}{1032.1} = 5.816694
\]

\[
b_0 = \bar{y} - b\bar{x} = 228.17 - (5.816694)(27.7) = 67.047576
\]

\[
\hat{y} = 67.0476 + 5.8167x
\]

b. \( SST = 39,065.14 \quad SSE = 4145.141 \quad SSR = 34,920.000 \)

\[
r^2 = \frac{SSR}{SST} = \frac{34,920.000}{39,065.141} = 0.894
\]

The estimated regression equation explained 89.4% of the variability in \( y \); a very good fit.

c. \( s^2 = MSE = 4145.141/8 = 518.143 \)

\[
s = \sqrt{518.143} = 22.76
\]

\[
s_{\hat{y}} = s\sqrt{\frac{1 + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 22.76\sqrt{\frac{1 + \frac{(35 - 27.7)^2}{1032.1}} = 8.86}}
\]

\[
\hat{y} = 67.0476 + 5.8167x = 67.0476 + 5.8167(35) = 270.63
\]

\[
\hat{y}_p \pm t_{\alpha/2}s_{\hat{y}} = 270.63 \pm 2.262 (8.86) = 270.63 \pm 20.04
\]

or 250.59 to 290.67

d. \( s_{\hat{y}_{ind}} = s\sqrt{\frac{1 + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 22.76\sqrt{\frac{1 + \frac{(35 - 27.7)^2}{1032.1}} = 24.42}}
\]

\[
\hat{y}_p \pm t_{\alpha/2}s_{\hat{y}_{ind}} = 270.63 \pm 2.262 (24.42) = 270.63 \pm 55.24
\]

or 215.39 to 325.87

40. a. 9

b. \( \hat{y} = 20.0 + 7.21x \)
c. 1.3626

d. \[ \text{SSE} = \text{SST} - \text{SSR} = 51,984.1 - 41,587.3 = 10,396.8 \]
\[ \text{MSE} = \frac{10,396.8}{7} = 1,485.3 \]
\[ F = \frac{\text{MSR}}{\text{MSE}} = \frac{41,587.3}{1,485.3} = 28.00 \]
\[ F_{.05} = 5.59 \text{ (1 degree of freedom numerator and 7 denominator)} \]
Since \( F = 28 > F_{.05} = 5.59 \) we reject \( H_0: B_1 = 0 \).

e. \[ \hat{y} = 20.0 + 7.21(50) = 380.5 \text{ or } \$380,500 \]

41. a. \[ \hat{y} = 6.1092 + .8951x \]

b. \[ t = \frac{b_1 - B_1}{s_{b_1}} = \frac{.8951 - 0}{.149} = 6.01 \]
\[ t_{.025} = 2.306 \text{ (1 degree of freedom numerator and 8 denominator)} \]
Since \( t = 6.01 > t_{.025} = 2.306 \) we reject \( H_0: B_1 = 0 \).

c. \[ \hat{y} = 6.1092 + .8951(25) = 28.49 \text{ or } \$28.49 \text{ per month} \]

42. a. \[ \hat{y} = 80.0 + 50.0x \]

b. 30

c. \[ F = \frac{\text{MSR}}{\text{MSE}} = \frac{6828.6}{82.1} = 83.17 \]
\[ F_{.05} = 4.20 \text{ (1 degree of freedom numerator and 28 denominator)} \]
Since \( F = 83.17 > F_{.05} = 4.20 \) we reject \( H_0: B_1 = 0 \).
Branch office sales are related to the salespersons.

d. \[ \hat{y} = 80 + 50(12) = 680 \text{ or } \$680,000 \]

43. a. The Minitab output is shown below:

The regression equation is
\[ \text{Price} = -11.8 + 2.18 \text{ Income} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-11.8</td>
<td>12.84</td>
<td>-0.92</td>
<td>0.380</td>
</tr>
<tr>
<td>Income</td>
<td>2.1843</td>
<td>0.2780</td>
<td>7.86</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 6.634 \quad \text{R-Sq} = 86.1\% \quad \text{R-Sq(adj)} = 84.7\% \]
### Simple Linear Regression

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>2717.9</td>
<td>2717.9</td>
<td>61.75</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>10</td>
<td>440.1</td>
<td>44.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>3158.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95.0% CI</th>
<th>95.0% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75.79</td>
<td>2.47</td>
<td>( 70.29, 81.28)</td>
<td>( 60.02, 91.56)</td>
</tr>
</tbody>
</table>

#### b.
\( r^2 = .861 \). The least squares line provided a very good fit.

#### c.
The 95% confidence interval is 70.29 to 81.28 or $70,290 to $81,280.

#### d.
The 95% prediction interval is 60.02 to 91.56 or $60,020 to $91,560.

### 44. a/b.

The scatter diagram shows a linear relationship between the two variables.

#### c.
The Minitab output is shown below:

The regression equation is

\[
\text{Rental} = 37.1 - 0.779 \times \text{Vacancy}\%
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>37.066</td>
<td>3.530</td>
<td>10.50</td>
<td>0.000</td>
</tr>
<tr>
<td>Vacancy%</td>
<td>-0.7791</td>
<td>0.2226</td>
<td>-3.50</td>
<td>0.003</td>
</tr>
</tbody>
</table>

\( S = 4.889 \) \hspace{1cm} R-Sq = 43.4% \hspace{1cm} R-Sq(adj) = 39.8%

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>292.89</td>
<td>292.89</td>
<td>12.26</td>
<td>0.003</td>
</tr>
<tr>
<td>Residual Error</td>
<td>16</td>
<td>382.37</td>
<td>23.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>675.26</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95.0% CI</th>
<th>95.0% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.59</td>
<td>2.51</td>
<td>( 12.27, 22.90)</td>
<td>(  5.94, 29.23)</td>
</tr>
<tr>
<td>2</td>
<td>28.26</td>
<td>1.42</td>
<td>( 25.26, 31.26)</td>
<td>( 17.47, 39.05)</td>
</tr>
</tbody>
</table>

#### Values of Predictors for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Vacancy%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.0</td>
</tr>
<tr>
<td>2</td>
<td>11.3</td>
</tr>
</tbody>
</table>

#### d.
Since the \( p \)-value = 0.003 is less than \( \alpha = .05 \), the relationship is significant.

#### e.
\( r^2 = .434 \). The least squares line does not provide a very good fit.
f. The 95% confidence interval is 12.27 to 22.90 or $12.27 to $22.90.
g. The 95% prediction interval is 17.47 to 39.05 or $17.47 to $39.05.

45. a. \[ \Sigma x_i = 14 \quad \Sigma y_i = 76 \quad \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 200 \quad \Sigma (x_i - \bar{x})^2 = 126 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{200}{126} = 1.5873 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 15.2 - (1.5873)(14) = -7.0222 \]

\[ \hat{y} = -7.02 + 1.59x \]

b. The residuals are 3.48, -2.47, -4.83, -1.6, and 5.22

c. 

With only 5 observations it is difficult to determine if the assumptions are satisfied. However, the plot does suggest curvature in the residuals that would indicate that the error term assumptions are not satisfied. The scatter diagram for these data also indicates that the underlying relationship between x and y may be curvilinear.

d. \[ s^2 = 23.78 \]

\[ h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma (x_i - \bar{x})^2} = \frac{1}{5} + \frac{(x_i - 14)^2}{126} \]

The standardized residuals are 1.32, -0.59, -1.11, -0.40, 1.49.
e. The standardized residual plot has the same shape as the original residual plot. The curvature observed indicates that the assumptions regarding the error term may not be satisfied.

46. a. \( \hat{y} = 2.32 + 0.64x \)

b.

![Residual plot](image)

The assumption that the variance is the same for all values of \( x \) is questionable. The variance appears to increase for larger values of \( x \).

47. a. Let \( x = \) advertising expenditures and \( y = \) revenue

\[
\hat{y} = 29.4 + 1.55x
\]

b. \( \text{SST} = 1002 \quad \text{SSE} = 310.28 \quad \text{SSR} = 691.72 \)

\[\text{MSR} = \frac{\text{SSR}}{1} = 691.72\]

\[\text{MSE} = \frac{\text{SSE}}{n - 2} = \frac{310.28}{5} = 62.0554\]

\[F = \frac{\text{MSR}}{\text{MSE}} = \frac{691.72}{62.0554} = 11.15\]

\( F_{0.05} = 6.61 \) (1 degree of freedom numerator and 5 denominator)

Since \( F = 11.15 > F_{0.05} = 6.61 \) we conclude that the two variables are related.

c.
The residual plot leads us to question the assumption of a linear relationship between $x$ and $y$. Even though the relationship is significant at the .05 level of significance, it would be extremely dangerous to extrapolate beyond the range of the data.

48. a. $\hat{y} = 80 + 4x$
b. The assumptions concerning the error term appear reasonable.

49. a. Let $x = \text{return on investment (ROE)}$ and $y = \text{price/earnings (P/E) ratio}$.

$\hat{y} = -32.13 + 3.22x$

b.

c. There is an unusual trend in the residuals. The assumptions concerning the error term appear questionable.
50. a. The MINITAB output is shown below:

The regression equation is
\[ Y = 66.1 + 0.402 \, X \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>66.10</td>
<td>32.06</td>
<td>2.06</td>
<td>0.094</td>
</tr>
<tr>
<td>X</td>
<td>0.4023</td>
<td>0.2276</td>
<td>1.77</td>
<td>0.137</td>
</tr>
</tbody>
</table>

\[ s = 12.62 \quad R\text{-sq} = 38.5\% \quad R\text{-sq(adj)} = 26.1\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>497.2</td>
<td>497.2</td>
<td>3.12</td>
<td>0.137</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>795.7</td>
<td>159.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>1292.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unusual Observations

<table>
<thead>
<tr>
<th>Obs.</th>
<th>X</th>
<th>Y</th>
<th>Fit</th>
<th>Stdev.Fit</th>
<th>Residual</th>
<th>St.Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>135</td>
<td>145</td>
<td>120.42</td>
<td>4.87</td>
<td>24.58</td>
<td>2.11R</td>
</tr>
</tbody>
</table>

R denotes an obs. with a large st. resid.

The standardized residuals are: 2.11, -1.08, -0.38, -0.78, -0.04, -0.41

The first observation appears to be an outlier since it has a large standardized residual.

b.

The standardized residual plot indicates that the observation \( x = 135, y = 145 \) may be an outlier; note that this observation has a standardized residual of 2.11.
c. The scatter diagram is shown below

The scatter diagram also indicates that the observation \( x = 135, y = 145 \) may be an outlier; the implication is that for simple linear regression an outlier can be identified by looking at the scatter diagram.

51. a. The Minitab output is shown below:

The regression equation is
\[ Y = 13.0 + 0.425 X \]

The regression equation is
\[ Y = 13.0 + 0.425 X \]

\[
\begin{array}{lrcrr}
\text{Predictor} & \text{Coef} & \text{Stdev} & \text{t-ratio} & \text{p} \\
\text{Constant} & 13.002 & 2.396 & 5.43 & 0.002 \\
X & 0.4248 & 0.2116 & 2.01 & 0.091 \\
\hline
s & 3.181 & R-sq & 40.2\% & R-sq(adj) & 30.2\%
\end{array}
\]

Analysis of Variance

\[
\begin{array}{lrrrrr}
\text{SOURCE} & \text{DF} & \text{SS} & \text{MS} & F & \text{p} \\
\text{Regression} & 1 & 40.78 & 40.78 & 4.03 & 0.091 \\
\text{Error} & 6 & 60.72 & 10.12 & & \\
\text{Total} & 7 & 101.50 & & & \\
\end{array}
\]

Unusual Observations

\[
\begin{array}{llllllll}
\text{Obs.} & \text{X} & \text{Y} & \text{Fit} & \text{Stdev.Fit} & \text{Residual} & \text{St.Resid} \\
7 & 12.0 & 24.00 & 18.10 & 1.20 & 5.90 & 2.00R \\
8 & 22.0 & 19.00 & 22.35 & 2.78 & -3.35 & -2.16RX \\
\end{array}
\]

R denotes an obs. with a large st. resid.
X denotes an obs. whose X value gives it large influence.
The standardized residuals are: -1.00, -.41, .01, -.48, .25, .65, -2.00, -2.16

The last two observations in the data set appear to be outliers since the standardized residuals for these observations are 2.00 and -2.16, respectively.

b. Using MINITAB, we obtained the following leverage values:

.28, .24, .16, .14, .13, .14, .14, .76

MINITAB identifies an observation as having high leverage if $h_i > 6/n$; for these data, $6/n = 6/8 = .75$. Since the leverage for the observation $x = 22, y = 19$ is .76, MINITAB would identify observation 8 as a high leverage point. Thus, we conclude that observation 8 is an influential observation.

c. The scatter diagram indicates that the observation $x = 22, y = 19$ is an influential observation.

52. a. The Minitab output is shown below:

The regression equation is

$\text{Amount} = 4.09 + 0.196 \text{MediaExp}$

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4.089</td>
<td>2.168</td>
<td>1.89</td>
<td>0.096</td>
</tr>
<tr>
<td>MediaExp</td>
<td>0.19552</td>
<td>0.03635</td>
<td>5.38</td>
<td>0.001</td>
</tr>
</tbody>
</table>

$S = 5.044 \hspace{1cm} \text{R-Sq} = 78.3\% \hspace{1cm} \text{R-Sq(adj)} = 75.6\%$

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>735.84</td>
<td>735.84</td>
<td>28.93</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Simple Linear Regression

Residual Error 8 203.51 25.44
Total 9 939.35

Unusual Observations
Obs MediaExp Amount Fit SE Fit Residual St Resid
1 120 36.30 27.55 3.30 8.75 2.30R

R denotes an observation with a large standardized residual

53. a. The Minitab output is shown below:

The regression equation is
Exposure = - 8.6 + 7.71 Aired

Predictor Coef SE Coef T P
Constant -8.55 21.65 -0.39 0.703
Aired 7.7149 0.5119 15.07 0.000

S = 34.88 R-Sq = 96.6% R-Sq(adj) = 96.2%

Analysis of Variance
Source DF SS MS F P
Regression 1 276434 276434 227.17 0.000
Residual Error 8 9735 1217
Total 9 286169

Unusual Observations
Obs Aired Exposure Fit SE Fit Residual St Resid
1 95.0 758.8 724.4 32.0 34.4 2.46RX

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

b. Minitab identifies observation 1 as having a large standardized residual; thus, we would consider observation 1 to be an outlier.

54. a. The Minitab output is shown below:

The regression equation is
Salary = 707 + 0.00482 MktCap

Predictor Coef SE Coef T P
Constant 707.0 118.0 5.99 0.000
MktCap 0.0048154 0.0008076 5.96 0.000

S = 379.8 R-Sq = 66.4% R-Sq(adj) = 64.5%

Analysis of Variance
### Chapter 14

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>5129071</td>
<td>5129071</td>
<td>35.55</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>18</td>
<td>2596647</td>
<td>144258</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>7725718</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>MktCap</th>
<th>Salary</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>507217</td>
<td>3325.0</td>
<td>3149.5</td>
<td>338.6</td>
<td>175.5</td>
<td>1.02 X</td>
</tr>
<tr>
<td>17</td>
<td>120967</td>
<td>116.2</td>
<td>1289.5</td>
<td>86.4</td>
<td>-1173.3</td>
<td>-3.17R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

b. Minitab identifies observation 6 as having a large standardized residual and observation 17 as an observation whose x value gives it large influence. A standardized residual plot against the predicted values is shown below:

---

55. No. Regression or correlation analysis can never prove that two variables are casually related.

56. The estimate of a mean value is an estimate of the average of all y values associated with the same x. The estimate of an individual y value is an estimate of only one of the y values associated with a particular x.

57. To determine whether or not there is a significant relationship between x and y. However, if we reject $B_1 = 0$, it does not imply a good fit.

58. a. The Minitab output is shown below:

The regression equation is

\[
\text{Price} = 9.26 + 0.711 \text{ Shares}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.265</td>
<td>1.099</td>
<td>8.43</td>
<td>0.000</td>
</tr>
<tr>
<td>Shares</td>
<td>0.7105</td>
<td>0.1474</td>
<td>4.82</td>
<td>0.001</td>
</tr>
</tbody>
</table>

\[
S = 1.419 \quad \text{R-Sq} = 74.4\% \quad \text{R-Sq(adj)} = 71.2\%
\]

Analysis of Variance
Simple Linear Regression

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>46.784</td>
<td>46.784</td>
<td>23.22</td>
<td>0.001</td>
</tr>
<tr>
<td>Residual Error</td>
<td>8</td>
<td>16.116</td>
<td>2.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>62.900</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Since the $p$-value corresponding to $F = 23.22 = .001 < \alpha = .05$, the relationship is significant.

c. $r^2 = .744$; a good fit. The least squares line explained 74.4% of the variability in Price.

d. $\hat{y} = 9.26 + .711(6) = 13.53$

59. a. The Minitab output is shown below:

The regression equation is
Options = - 3.83 + 0.296 Common

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.834</td>
<td>5.903</td>
<td>-0.65</td>
<td>0.529</td>
</tr>
<tr>
<td>Common</td>
<td>0.29567</td>
<td>0.02648</td>
<td>11.17</td>
<td>0.000</td>
</tr>
</tbody>
</table>

$S = 11.04$ \hspace{1cm} R-Sq = 91.9\% \hspace{1cm} R-Sq(adj) = 91.2\%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>15208</td>
<td>15208</td>
<td>124.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>11</td>
<td>1341</td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>16550</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. $\hat{y} = -3.83 + .296(150) = 40.57$; approximately 40.6 million shares of options grants outstanding.

c. $r^2 = .919$; a very good fit. The least squares line explained 91.9\% of the variability in Options.

60. a. The Minitab output is shown below:

The regression equation is
IBM = 0.275 + 0.950 S&P 500

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.2747</td>
<td>0.9004</td>
<td>0.31</td>
<td>0.768</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.9498</td>
<td>0.3569</td>
<td>2.66</td>
<td>0.029</td>
</tr>
</tbody>
</table>

$S = 2.664$ \hspace{1cm} R-Sq = 47.0\% \hspace{1cm} R-Sq(adj) = 40.3\%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>50.255</td>
<td>50.255</td>
<td>7.08</td>
<td>0.029</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>56.781</td>
<td>7.098</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
b. Since the $p$-value = 0.029 is less than $\alpha = .05$, the relationship is significant.

c. $r^2 = .470$. The least squares line does not provide a very good fit.

d. Woolworth has higher risk with a market beta of 1.25.

61. a. 

![Temperature Chart]

b. It appears that there is a positive linear relationship between the two variables.

c. The Minitab output is shown below:

The regression equation is

High = 23.9 + 0.898 Low

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>23.899</td>
<td>6.481</td>
<td>3.69</td>
<td>0.002</td>
</tr>
<tr>
<td>Low</td>
<td>0.8980</td>
<td>0.1121</td>
<td>8.01</td>
<td>0.000</td>
</tr>
</tbody>
</table>

$S = 5.285$  R-Sq = 78.1%  R-Sq(adj) = 76.9%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>1792.3</td>
<td>1792.3</td>
<td>64.18</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Simple Linear Regression

<table>
<thead>
<tr>
<th>Residual Error</th>
<th>18</th>
<th>502.7</th>
<th>27.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>19</td>
<td>2294.9</td>
<td></td>
</tr>
</tbody>
</table>

d. Since the \( p \)-value corresponding to \( F = 64.18 \) = .000 < \( \alpha = .05 \), the relationship is significant.

e. \( r^2 = .781 \); a good fit. The least squares line explained 78.1% of the variability in high temperature.

f. \( r = \sqrt{.781} = .88 \)

62. The MINITAB output is shown below:

The regression equation is

\[ Y = 10.5 + 0.953 \times \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.528</td>
<td>3.745</td>
<td>2.81</td>
<td>0.023</td>
</tr>
<tr>
<td>X</td>
<td>0.9534</td>
<td>0.1382</td>
<td>6.90</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ s = 4.250 \quad R\text{-sq} = 85.6\% \quad R\text{-sq(adj)} = 83.8\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>860.05</td>
<td>860.05</td>
<td>47.62</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>144.47</td>
<td>18.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>1004.53</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{y} = 10.5 + 0.953 x \]

b. Since the \( p \)-value corresponding to \( F = 47.62 \) = .000 < \( \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \).

c. The 95% prediction interval is 28.74 to 49.52 or $2874 to $4952

d. Yes, since the expected expense is $3913.

63. a. The Minitab output is shown below:

The regression equation is

\[ \text{Defects} = 22.2 - 0.148 \times \text{Speed} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>22.174</td>
<td>1.653</td>
<td>13.42</td>
<td>0.000</td>
</tr>
<tr>
<td>Speed</td>
<td>-0.14783</td>
<td>0.04391</td>
<td>-3.37</td>
<td>0.028</td>
</tr>
</tbody>
</table>

\[ S = 1.489 \quad R\text{-Sq} = 73.9\% \quad R\text{-Sq(adj)} = 67.4\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>25.130</td>
<td>25.130</td>
<td>11.33</td>
<td>0.028</td>
</tr>
<tr>
<td>Residual Error</td>
<td>4</td>
<td>8.870</td>
<td>2.217</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 14

| Total   | 5   | 34.000 |

Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95.0% CI</th>
<th>95.0% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.783</td>
<td>0.896</td>
<td>( 12.294, 17.271)</td>
<td>(  9.957, 19.608)</td>
</tr>
</tbody>
</table>

b. Since the $p$-value corresponding to $F = 11.33 = .028 < \alpha = .05$, the relationship is significant.

c. $r^2 = .739$; a good fit. The least squares line explained 73.9% of the variability in the number of defects.

d. Using the Minitab output in part (a), the 95% confidence interval is 12.294 to 17.271.

64. a. There appears to be a negative linear relationship between distance to work and number of days absent.

b. The MINITAB output is shown below:

The regression equation is

\[ Y = 8.10 - 0.344 \times X \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>8.0978</td>
<td>0.8088</td>
<td>10.01</td>
<td>0.000</td>
</tr>
<tr>
<td>X</td>
<td>-0.34420</td>
<td>0.07761</td>
<td>-4.43</td>
<td>0.002</td>
</tr>
</tbody>
</table>

\[ s = 1.289 \quad R\text{-sq} = 71.1\% \quad R\text{-sq(adj)} = 67.5\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>32.699</td>
<td>32.699</td>
<td>19.67</td>
<td>0.002</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>13.301</td>
<td>1.663</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>46.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit</th>
<th>Stdev.Fit</th>
<th>95% C.I.</th>
<th>95% P.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.377</td>
<td>0.512</td>
<td>( 5.195, 7.559)</td>
<td>( 3.176, 9.577)</td>
</tr>
</tbody>
</table>

c. Since the $p$-value corresponding to $F = 419.67$ is $.002 < \alpha = .05$. We reject $H_0: \beta_1 = 0$.

d. $r^2 = .711$. The estimated regression equation explained 71.1% of the variability in $y$; this is a reasonably good fit.

e. The 95% confidence interval is 5.195 to 7.559 or approximately 5.2 to 7.6 days.

65. a. Let $X =$ the age of a bus and $Y =$ the annual maintenance cost.

The MINITAB output is shown below:
The regression equation is
\[ Y = 220 + 132 \, X \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>220.00</td>
<td>58.48</td>
<td>3.76</td>
<td>0.006</td>
</tr>
<tr>
<td>X</td>
<td>131.67</td>
<td>17.80</td>
<td>7.40</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ s = 75.50 \quad R\text{-sq} = 87.3\% \quad R\text{-sq(adj)} = 85.7\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>312050</td>
<td>312050</td>
<td>54.75</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>45600</td>
<td>5700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>357650</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fit  Stdev.Fit 95% C.I. 95% P.I.
746.7 29.8 ( 678.0, 815.4) ( 559.5, 933.9)

b. Since the \( p \)-value corresponding to \( F = 54.75 \) is \( .000 < \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \).

c. \( r^2 = .873 \). The least squares line provided a very good fit.

d. The 95% prediction interval is 559.5 to 933.9 or $559.50 to $933.90

66. a. Let \( X = \) hours spent studying and \( Y = \) total points earned

The MINITAB output is shown below:

The regression equation is
\[ Y = 5.85 + 0.830 \, X \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.847</td>
<td>7.972</td>
<td>0.73</td>
<td>0.484</td>
</tr>
<tr>
<td>X</td>
<td>0.8295</td>
<td>0.1095</td>
<td>7.58</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ s = 7.523 \quad R\text{-sq} = 87.8\% \quad R\text{-sq(adj)} = 86.2\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>3249.7</td>
<td>3249.7</td>
<td>57.42</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>452.8</td>
<td>56.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>3702.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fit  Stdev.Fit 95% C.I. 95% P.I.
84.65 3.67 ( 76.19, 93.11) ( 65.35, 103.96)

b. Since the \( p \)-value corresponding to \( F = 57.42 \) is \( .000 < \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \).

c. 84.65 points

d. The 95% prediction interval is 65.35 to 103.96
67. a. The Minitab output is shown below:

The regression equation is
Audit% = -0.471 + 0.000039 Income

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.4710</td>
<td>0.5842</td>
<td>-0.81</td>
<td>0.431</td>
</tr>
<tr>
<td>Income</td>
<td>0.00003868</td>
<td>0.00001731</td>
<td>2.23</td>
<td>0.038</td>
</tr>
</tbody>
</table>

S = 0.2088    R-Sq = 21.7%    R-Sq(adj) = 17.4%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.21749</td>
<td>0.21749</td>
<td>4.99</td>
<td>0.038</td>
</tr>
<tr>
<td>Residual Error</td>
<td>18</td>
<td>0.78451</td>
<td>0.04358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>1.00200</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95.0% CI</th>
<th>95.0% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8828</td>
<td>0.0523</td>
<td>(0.7729, 0.9927)</td>
<td>(0.4306, 1.3349)</td>
</tr>
</tbody>
</table>

b. Since the p-value = 0.038 is less than α = 0.05, the relationship is significant.

c. $r^2 = .217$. The least squares line does not provide a very good fit.

d. The 95% confidence interval is .7729 to .9927.