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. 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.

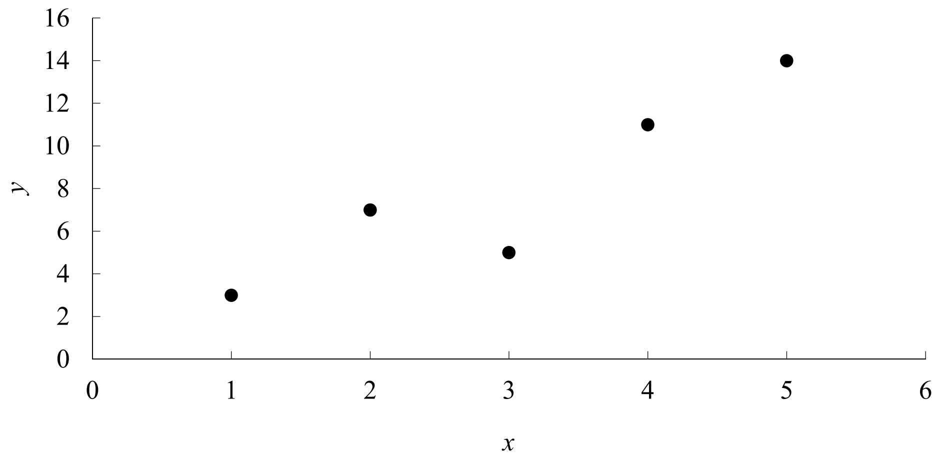
7. Learn how to use a residual plot to make a judgement as to the validity of the regression assumptions.

8. 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

Solutions:

1 a.



b. There appears to be a positive 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. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{15}{5} = 3 \quad \bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{40}{5} = 8$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 26 \quad \Sigma(x_i - \bar{x})^2 = 10$

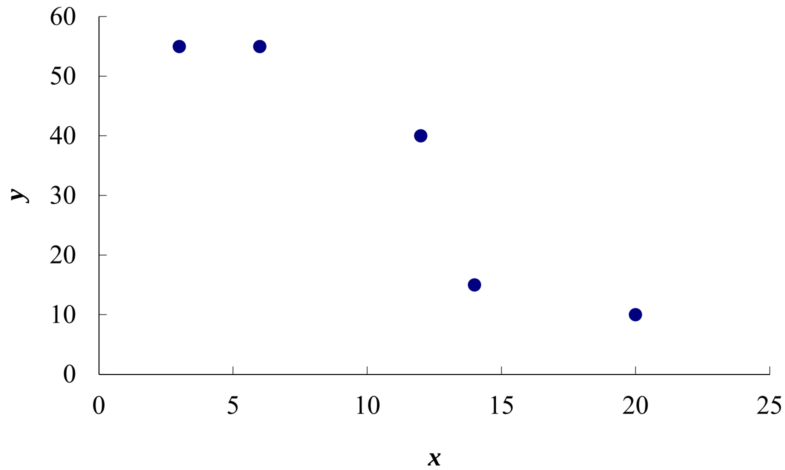
$b_1 = \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \dfrac{26}{10} = 2.6$

$b_0 = \bar{y} - b_1\bar{x} = 8 - (2.6)(3) = 0.2$

$\hat{y} = 0.2 + 2.6x$

e. $\hat{y} = 0.2 + 2.6(4) = 10.6$

2. a.



b. There appears to be a negative 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. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{55}{5} = 11 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{175}{5} = 35$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -540 \qquad \Sigma(x_i - \bar{x})^2 = 180$

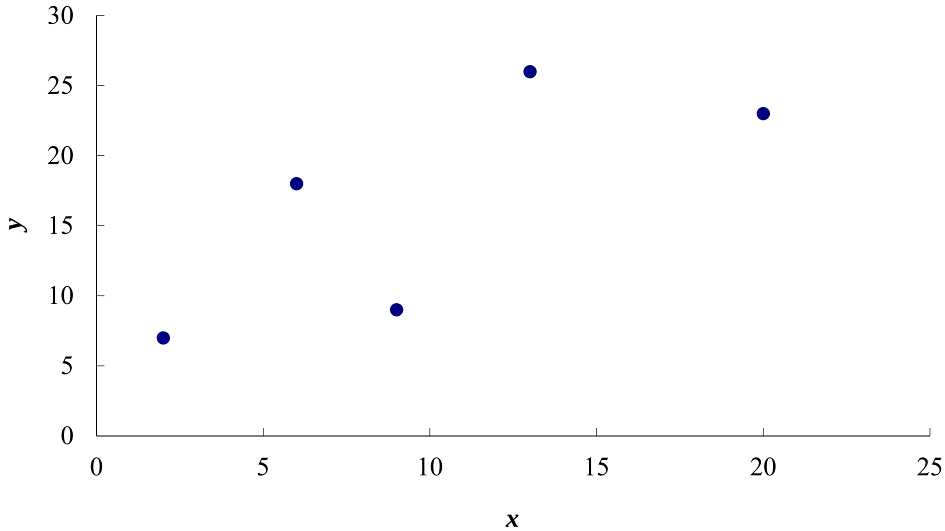
$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-540}{180} = -3$

$b_0 = \bar{y} - b_1\bar{x} = 35 - (-3)(11) = 68$

$\hat{y} = 68 - 3x$

e. $\hat{y} = 68 - 3(10) = 38$

3. a.



b. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{50}{5} = 10 \qquad \bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{83}{5} = 16.6$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 171 \qquad \Sigma(x_i - \bar{x})^2 = 190$

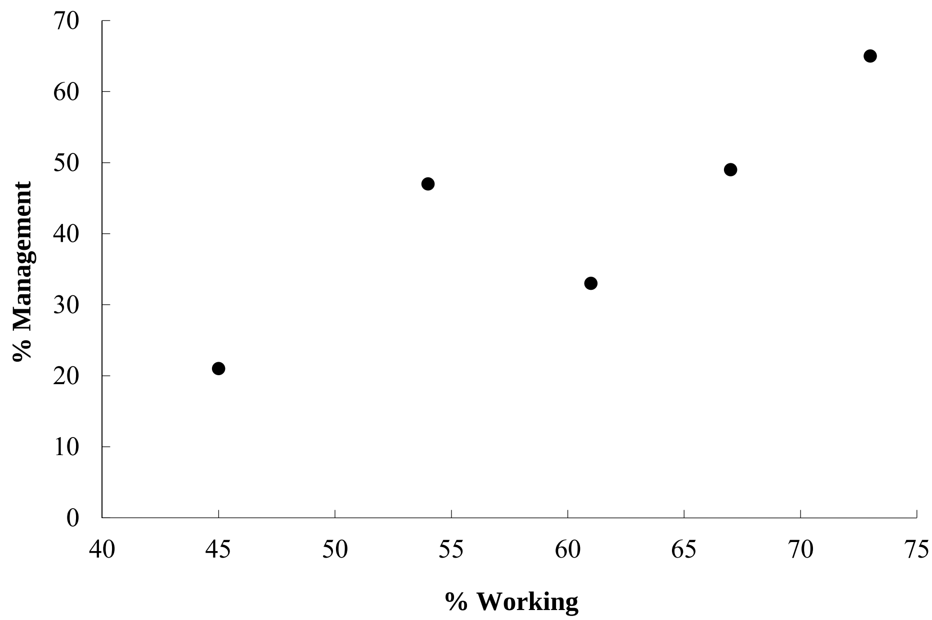
$b_1 = \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \dfrac{171}{190} = 0.9$

$b_0 = \bar{y} - b_1\bar{x} = 16.6 - (0.9)(10) = 7.6$

$\hat{y} = 7.6 + 0.9x$

c. $\hat{y} = 7.6 + 0.9(6) = 13$

4. a.



b. There appears to be a positive linear relationship between the percentage of women working in the five companies (x) and the percentage of management jobs held by women in that company (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. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{300}{5} = 60 \quad \bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{215}{5} = 43$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 624 \quad \Sigma(x_i - \bar{x})^2 = 480$

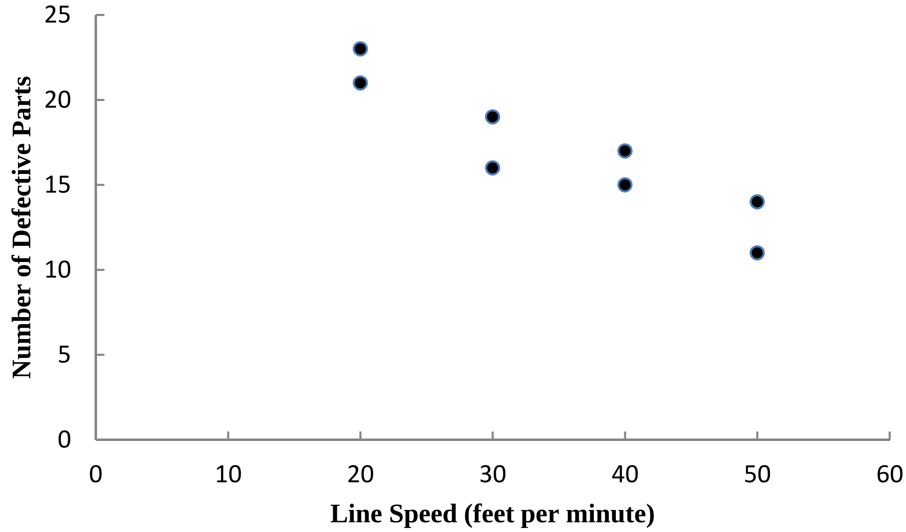
$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{624}{480} = 1.3$$

$b_0 = \bar{y} - b_1\bar{x} = 43 - 1.3(60) = -35$

$\hat{y} = -35 + 1.3x$

e. $\hat{y} = -35 + 1.3x = -35 + 1.3(60) = 43\%$

5. a.



b. There appears to be a negative relationship between line speed (feet per minute) and the number of defective parts.

c. Let x = line speed (feet per minute) and y = number of defective parts.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{280}{8} = 35 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{136}{8} = 17$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -300 \qquad \Sigma(x_i - \bar{x})^2 = 1000$$

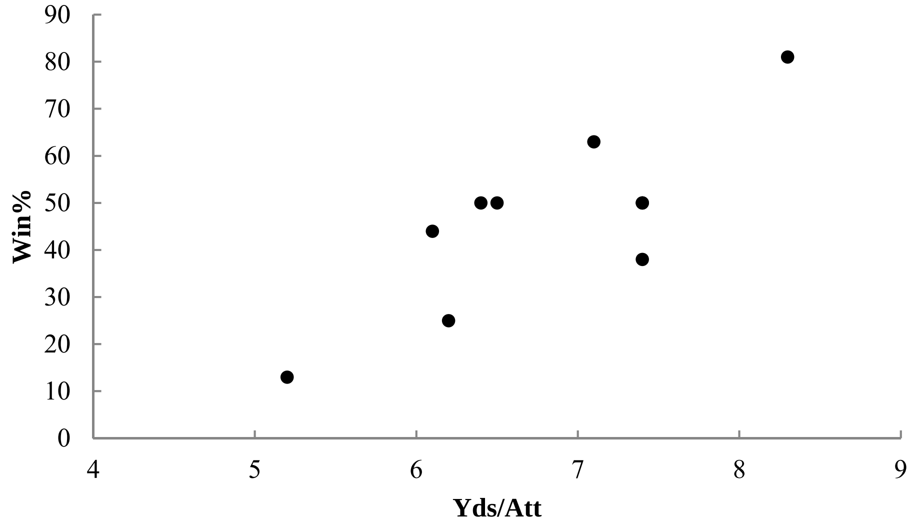
$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-300}{1000} = -.3$$

$$b_0 = \bar{y} - b_1\bar{x} = 17 - (-.3)(35) = 27.5$$

$$\hat{y} = 27.5 - .3x$$

d. $\hat{y} = 27.5 - .3x = 27.5 - .3(25) = 20$

6. a.



b. The scatter diagram indicates a positive linear relationship between x = average number of passing yards per attempt and y = the percentage of games won by the team.

c. $\bar{x} = \Sigma x_i / n = 680/10 = 6.8 \quad \bar{y} = \Sigma y_i / n = 464/10 = 46.4$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 121.6 \quad \Sigma(x_i - \bar{x})^2 = 7.08$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{121.6}{7.08} = 17.1751$$

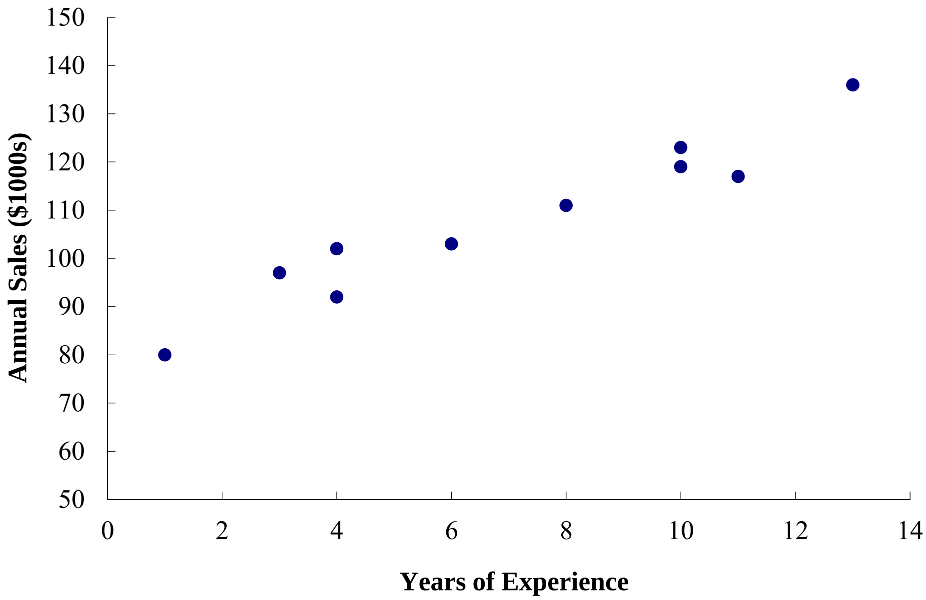
$$b_0 = \bar{y} - b_1\bar{x} = 46.4 - (17.1751)(6.8) = -70.391$$

$$\hat{y} = -70.391 + 17.1751x$$

d. The slope of the estimated regression line is approximately 17.2. So, for every increase of one yard in the average number of passes per attempt, the percentage of games won by the team increases by 17.2%.

e. With an average number of passing yards per attempt of 6.2, the predicted percentage of games won is $\hat{y}$ = -70.391 + 17.175(6.2) = 36%. With a record of 7 wins and 9 loses, the percentage of wins that the Kansas City Chiefs won is 43.8 or approximately 44%. Considering the small data size, the prediction made using the estimated regression equation is not too bad.

7. a.



b. Let x = years of experience and y = annual sales ($1000s)

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{70}{10} = 7 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{1080}{10} = 108$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 568 \qquad \Sigma(x_i - \bar{x})^2 = 142$$

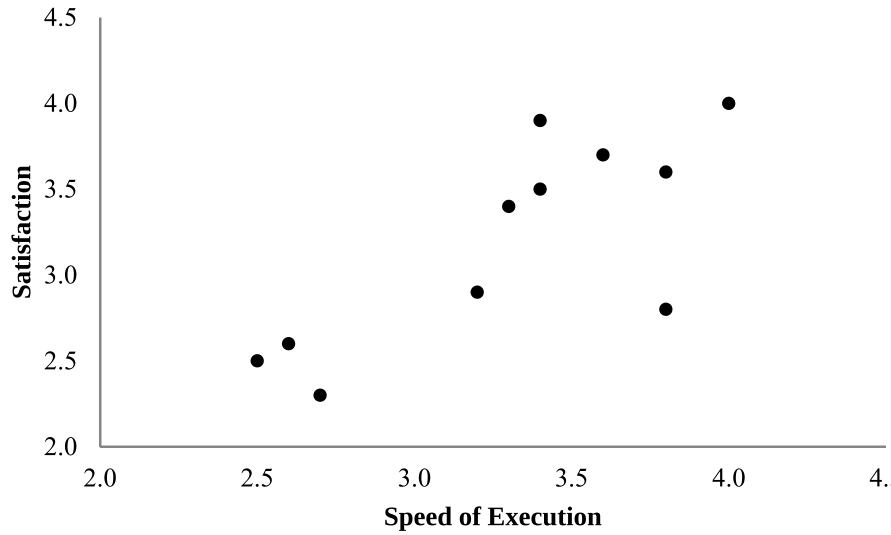
$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{568}{142} = 4$$

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

$$\hat{y} = 80 + 4x$$

c. $\hat{y} = 80 + 4x = 80 + 4(9) = 116$ or $116,000

8. a.



b. The scatter diagram indicates a positive linear relationship between x = speed of execution rating and y = overall satisfaction rating for electronic trades.

c. $\bar{x} = \Sigma x_i / n = 36.3/11 = 3.3 \quad \bar{y} = \Sigma y_i / n = 35.2/11 = 3.2$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 2.4 \quad \Sigma(x_i - \bar{x})^2 = 2.6$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2.4}{2.6} = .9077$$

$$b_0 = \bar{y} - b_1\bar{x} = 3.2 - (.9077)(3.3) = .2046$$

$$\hat{y} = .2046 + .9077x$$

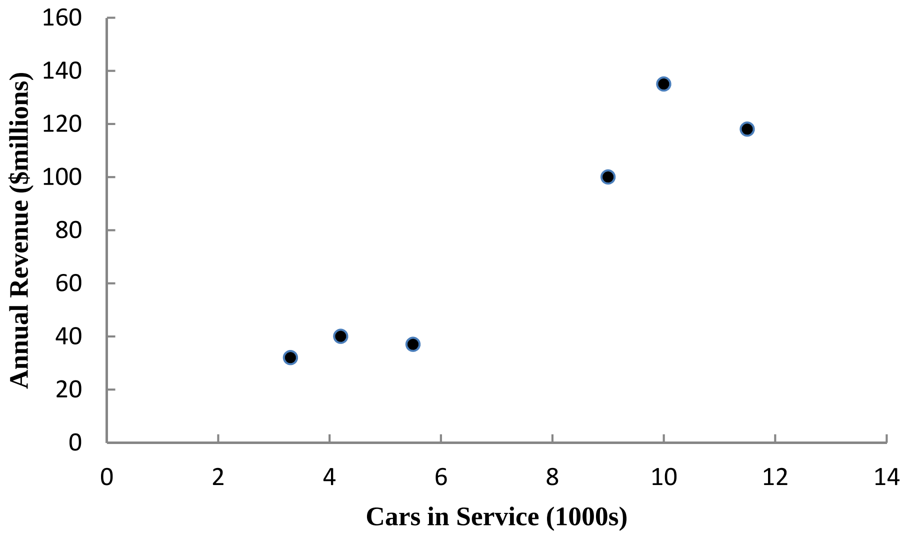
d. The slope of the estimated regression line is approximately .9077. So, a one unit increase in the speed of execution rating will increase the overall satisfaction rating by approximately .9 points.

e. The average speed of execution rating for the other brokerage firms is 3.4. Using this as the new value of x for Zecco.com, we can use the estimated regression equation developed in part (c) to estimate the overall satisfaction rating corresponding to x = 3.4.

$$\hat{y} = .2046 + .9077x = .2046 + .9077(3.4) = 3.29$$

Thus, an estimate of the overall satisfaction rating when x = 3.4 is approximately 3.3.

9. a.



b. The scatter diagram indicates a positive linear relationship between x = cars in service (1000s) and y = annual revenue ($millions).

c. $\bar{x} = \Sigma x_i / n = 43.5 / 6 = 7.25 \quad \bar{y} = \Sigma y_i / n = 462 / 6 = 77$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 734.6 \quad \Sigma(x_i - \bar{x})^2 = 56.655$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{734.6}{56.655} = 12.9662$$

$b_0 = \bar{y} - b_1\bar{x} = 77 - (12.9662)(7.25) = -17.005$

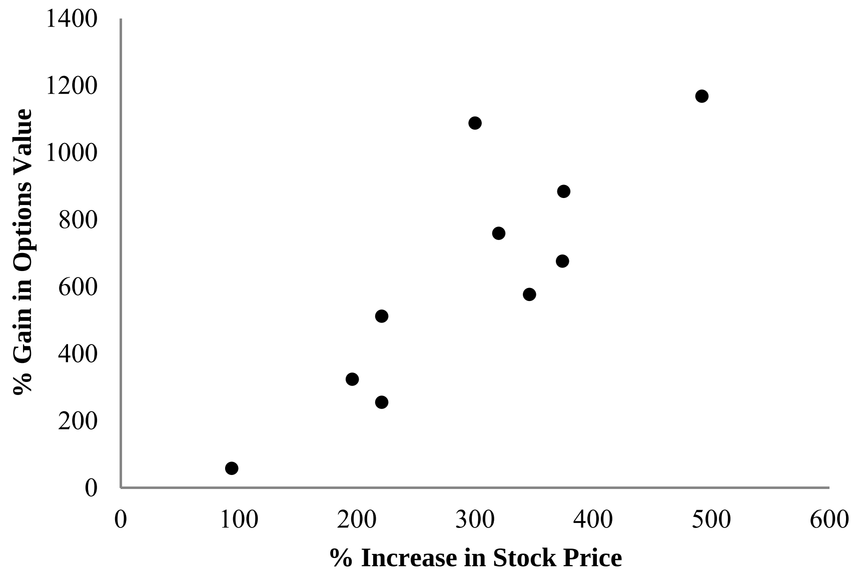
$\hat{y} = -17.005 + 12.966x$

d. For every additional 1000 cars placed in service annual revenue will increase by 12.966 ($millions) or $12,966,000. Therefor every additional car placed in service will increase annual revenue by $12,966.

e. $\hat{y} = -17.005 + 12.966x = -17.005 + 12.966(11) = 125.621$

A prediction of annual revenue for Fox Rent A Car is approximately $126 million.

10. a.



b. The scatter diagram indicates a positive linear relationship between x = percentage increase in the stock price and y = percentage gain in options value. In other words, options values increase as stock prices increase.

c. $\bar{x} = \Sigma x_i / n = 2939/10 = 293.9 \quad \bar{y} = \Sigma y_i / n = 6301/10 = 630.1$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 314{,}501.1 \quad \Sigma(x_i - \bar{x})^2 = 115{,}842.9$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{314{,}501.1}{115{,}842.9} = 2.7149$$

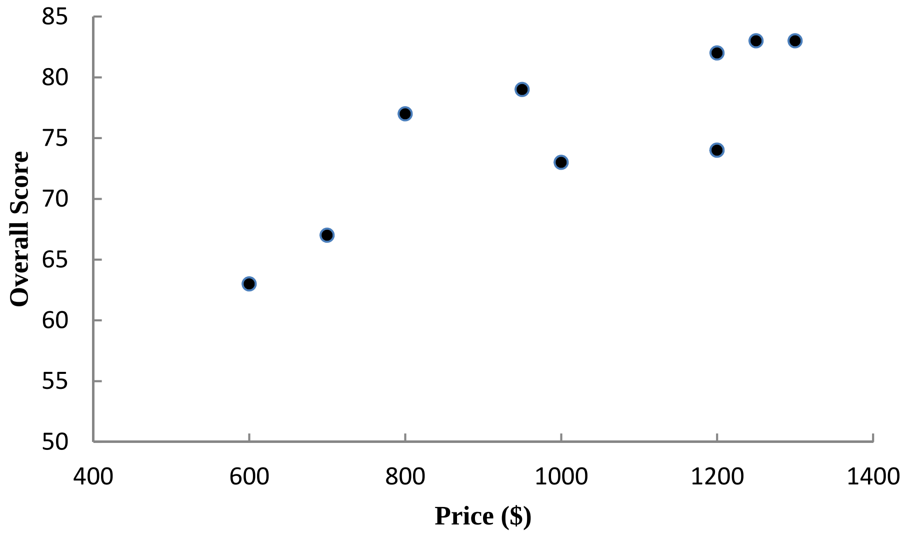
$b_0 = \bar{y} - b_1\bar{x} = 630.1 - (2.1749)(293.9) = -167.81$

$\hat{y} = -167.81 + 2.7149x$

d. The slope of the estimated regression line is approximately 2.7. So, for every percentage increase in the price of the stock the options value increases by 2.7%.

e. The rewards for the CEO do appear to be based upon performance increases in the stock value. While the rewards may seem excessive, the executive is being rewarded for his/her role in increasing the value of the company. This is why such compensation schemes are devised for CEOs by boards of directors. A compensation scheme where an executive got a big salary increase when the company stock went down would be bad. And, if the stock price for a company had gone down during the periods in question, the value of the CEOs options would also go down.

11. a.



b. The scatter diagram indicates a positive linear relationship between x = price ($) and y = overall score.

c. $\bar{x} = \Sigma x_i / n = 10{,}200/10 = 1020 \quad \bar{y} = \Sigma y_i / n = 755/10 = 75.5$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 11{,}900 \quad \Sigma(x_i - \bar{x})^2 = 561{,}000$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{11{,}900}{561{,}000} = .021212$$

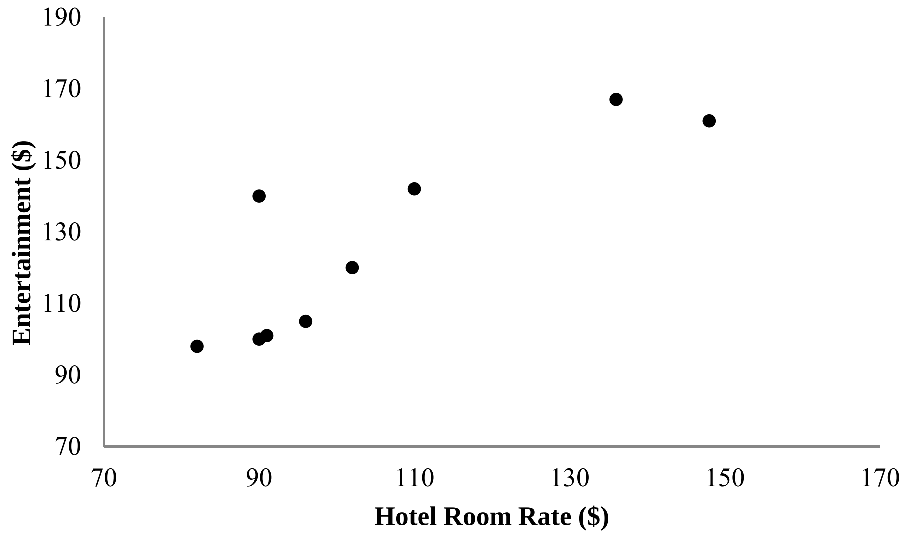
$b_0 = \bar{y} - b_1\bar{x} = 75.5 - (.021212)(1020) = 53.864$

$\hat{y} = 53.864 + .0212x$

d. The slope of .0212 means that spending an additional $100 in price will increase the overall score by approximately 2 points.

e. A prediction of the overall score is $\hat{y} = 53.864 + .0212x = 53.864 + .0212(700) = 68.7$

12. a.



b. The scatter diagram indicates a positive linear relationship between x = hotel room rate and the amount spent on entertainment.

c. $\bar{x} = \Sigma x_i / n = 945/9 = 105 \quad \bar{y} = \Sigma y_i / n = 1134/9 = 126$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 4237 \quad \Sigma(x_i - \bar{x})^2 = 4100$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{4237}{4100} = 1.0334$$

$b_0 = \bar{y} - b_1\bar{x} = 126 - (1.0334)(105) = 17.49$

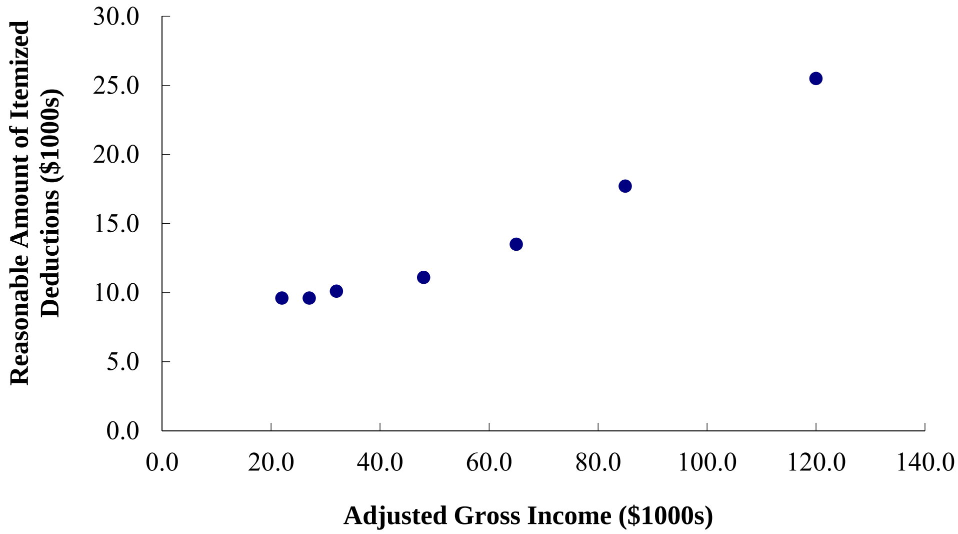
$\hat{y} = 17.49 + 1.0334x$

d. With a value of x = \$128, the predicted value of y for Chicago is

$\hat{y} = 17.49 + 1.0334x = 17.49 + 1.0334(128) = 150$

Note: In The Wall Street Journal article the entertainment expense for Chicago was \$146. Thus, the estimated regression equation provided a good estimate of entertainment expenses for Chicago.

13. a.



b. Let x = adjusted gross income and y = reasonable amount of itemized deductions

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{399}{7} = 57 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{97.1}{7} = 13.8714$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 1233.7 \qquad \Sigma(x_i - \bar{x})^2 = 7648$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{1233.7}{7648} = 0.1613$$

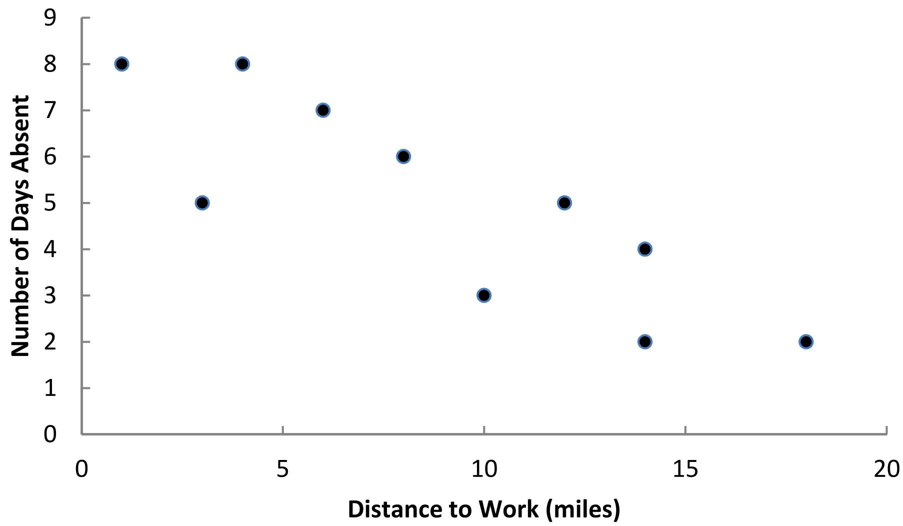
$$b_0 = \bar{y} - b_1\bar{x} = 13.8714 - (0.1613)(57) = 4.6773$$

$$\hat{y} = 4.68 + 0.16x$$

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

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

14. a.



The scatter diagram indicates a negative linear relationship between x = distance to work and y = number of days absent.

b. $\bar{x} = \Sigma x_i / n = 90/10 = 9 \quad \bar{y} = \Sigma y_i / n = 50/10 = 5$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -95 \quad \Sigma(x_i - \bar{x})^2 = 276$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-95}{276} = -.3442$$

$b_0 = \bar{y} - b_1\bar{x} = 5 - (-.3442)(9) = 8.0978$

$\hat{y} = 8.0978 - .3442x$

c. A prediction of the number of days absent is $\hat{y} = 8.0978 - .3442(5) = 6.4$ or approximately 6 days.

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

$\hat{y}_i = 0.2 + 2.6x_i \qquad \bar{y} = 8$

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

$\text{SSE} = \sum(y_i - \hat{y}_i)^2 = 12.40 \qquad \text{SST} = \sum(y_i - \bar{y})^2 = 80$

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. $r_{xy} = \sqrt{.845} = +.9192$

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

$\hat{y}_i = 68 - 3x \qquad \bar{y} = 35$

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

$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 230 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 1850$

Thus, SSR = SST - SSE = 1850 - 230 = 1620

b. r^2 = SSR/SST = 1620/1850 = .876

The least squares line provided an excellent fit; 87.6% of the variability in y has been explained by the estimated regression equation.

c. $r_{xy} = \sqrt{.876} = -.936$

Note: the sign for r is negative because the slope of the estimated regression equation is negative. (b_1 = -3)

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

$\hat{y}_i = 7.6 + .9x \qquad \bar{y} = 16.6$

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

$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 127.3 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 281.2$

Thus, SSR = SST - SSE = 281.2 – 127.3 = 153.9

r^2 = SSR/SST = 153.9/281.2 = .547

We see that 54.7% of the variability in y has been explained by the least squares line.

$r_{xy} = \sqrt{.547} = +.740$

18. a. $\bar{x} = \Sigma x_i / n = 600/6 = 100 \quad \bar{y} = \Sigma y_i / n = 330/6 = 55$

$\text{SST} = \Sigma(y_i - \bar{y})^2 = 1800 \quad \text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 287.624$

SSR = SST – SSR = 1800 – 287.624 = 1512.376

b. $r^2 = \dfrac{\text{SSR}}{\text{SST}} = \dfrac{1512.376}{1800} = .84$

c. $r = \sqrt{r^2} = \sqrt{.84} = .917$

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

$\hat{y} = 80 + 4x$ $\quad\quad$ $\bar{y} = 108$

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

$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 170 \quad\quad \text{SST} = \Sigma(y_i - \bar{y})^2 = 2442$

Thus, SSR = SST - SSE = 2442 - 170 = 2272

b. r^2 = SSR/SST = 2272/2442 = .93

We see that 93% of the variability in y has been explained by the least squares line.

c. $r_{xy} = \sqrt{.93} = +.96$

20. a. $\bar{x} = \Sigma x_i / n = 160/10 = 16 \quad \bar{y} = \Sigma y_i / n = 55{,}500/10 = 5550$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -31{,}284 \quad \Sigma(x_i - \bar{x})^2 = 21.74$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-31{,}284}{21.74} = -1439$$

$b_0 = \bar{y} - b_1\bar{x} = 5550 - (-1439)(16) = 28{,}574$

$\hat{y} = 28{,}574 - 1439x$

b. SST = 52,120,800 SSE = 7,102,922.54

SSR = SST – SSR = 52,120,800 - 7,102,922.54 = 45,017,877

r^2 = SSR/SST = 45,017,877/52,120,800 = .864

The estimated regression equation provided a very good fit.

c. $\hat{y} = 28{,}574 - 1439x = 28{,}574 - 1439(15) = 6989$

Thus, an estimate of the price for a bike that weighs 15 pounds is $6989.

21. a. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{3450}{6} = 575 \quad \bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{33{,}700}{6} = 5616.67$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 712{,}500 \quad \Sigma(x_i - \bar{x})^2 = 93{,}750$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{712{,}500}{93{,}750} = 7.6$$

$b_0 = \bar{y} - b_1\bar{x} = 5616.67 - (7.6)(575) = 1246.67$

$\hat{y} = 1246.67 + 7.6x$

b. $7.60

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

$$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 233{,}333.33 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 5{,}648{,}333.33$$

Thus, SSR = SST - SSE = 5,648,333.33 - 233,333.33 = 5,415,000

r^2 = SSR/SST = 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. SSE = 1043.03

$$\bar{y} = \Sigma y_i / n = 462 / 6 = 77 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 10{,}568$$

SSR = SST – SSR = 10,568 – 1043.03 = 9524.97

$$r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{9524.97}{10{,}568} = .9013$$

b. The estimated regression equation provided a very good fit; approximately 90% of the variability in the dependent variable was explained by the linear relationship between the two variables.

c. $r = \sqrt{r^2} = \sqrt{..9013} = .95$

This reflects a strong linear relationship between the two variables.

23. a. s^2 = MSE = SSE / (n - 2) = 12.4 / 3 = 4.133

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

c. $\Sigma(x_i - \bar{x})^2 = 10$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{2.033}{\sqrt{10}} = 0.643$$

d. $t = \dfrac{b_1}{s_{b_1}} = \dfrac{2.6}{.643} = 4.044$

Using t table (3 degrees of freedom), area in tail is between .01 and .025

p-value is between .02 and .05

Using Excel or Minitab, the p-value corresponding to t = 4.04 is .0272.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

e. MSR = SSR / 1 = 67.6

F = MSR / MSE = 67.6 / 4.133 = 16.36

Using F table (1 degree of freedom numerator and 3 denominator), p-value is between .025 and .05

Using Excel or Minitab, the p-value corresponding to F = 16.36 is .0272.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	67.6	1	67.6	16.36	.0272
Error	12.4	3	4.133		
Total	80.0	4			

24. a. s^2 = MSE = SSE/(n - 2) = 230/3 = 76.6667

b. $s = \sqrt{\text{MSE}} = \sqrt{76.6667} = 8.7560$

c. $\Sigma(x_i - \bar{x})^2 = 180$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{8.7560}{\sqrt{180}} = 0.6526$$

d. $t = \dfrac{b_1}{s_{b_1}} = \dfrac{-3}{.653} = -4.59$

Using t table (3 degrees of freedom), area in tail is less than .01; p-value is less than .02

Using Excel or Minitab, the p-value corresponding to t = -4.59 is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

e. MSR = SSR/1 = 1620

F = MSR/MSE = 1620/76.6667 = 21.13

Using F table (1 degree of freedom numerator and 3 denominator), p-value is less than .025

Using Excel or Minitab, the p-value corresponding to F = 21.13 is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	1620	1	1620	21.13	.0193
Error	230	3	76.6667		
Total	1850	4			

25. a. s^2 = MSE = SSE/(n - 2) = 127.3/3 = 42.4333

$$s = \sqrt{\text{MSE}} = \sqrt{42.4333} = 6.5141$$

b. $\Sigma(x_i - \bar{x})^2 = 190$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{6.5141}{\sqrt{190}} = 0.4726$$

$$t = \frac{b_1}{s_{b_1}} = \frac{.9}{.4726} = 1.90$$

Using t table (3 degrees of freedom), area in tail is between .05 and .10

p-value is between .10 and .20

Using Excel or Minitab, the p-value corresponding to t = 1.90 is .1530.

Because p-value $> \alpha$, we cannot reject H_0: $\beta_1 = 0$; x and y do not appear to be related.

c. MSR = SSR/1 = 153.9 /1 = 153.9

F = MSR/MSE = 153.9/42.4333 = 3.63

Using F table (1 degree of freedom numerator and 3 denominator), p-value is greater than .10

Using Excel or Minitab, the p-value corresponding to F = 3.63 is .1530.

Because p-value $> \alpha$, we cannot reject H_0: $\beta_1 = 0$; x and y do not appear to be related.

26. a. In the statement of exercise 18, $\hat{y} = 23.194 + .318x$

In solving exercise 18, we found SSE = 287.624

s^2 = MSE = SSE/(n-2) =287.624 / 4 = 71.906

$$s = \sqrt{\text{MSE}} = \sqrt{71.906} = 8.4797$$

$$\sum (x - \bar{x})^2 = 14{,}950$$

$$s_{b_1} = \frac{s}{\sqrt{\sum (x - \bar{x})^2}} = \frac{8.4797}{\sqrt{14{,}950}} = .0694$$

$$t = \frac{b_1}{s_{b_1}} = \frac{.318}{.0694} = 4.58$$

Using t table (4 degrees of freedom), area in tail is between .005 and .01

p-value is between .01 and .02

Using Excel, the p-value corresponding to t = 4.58 is .010.

Because p-value $\leq \alpha$, we reject H_0: β_1 = 0; there is a significant relationship between price and overall score

b. In exercise 18 we found SSR = 1512.376

MSR = SSR/1 = 1512.376/1 = 1512.376

F = MSR/MSE = 1512.376/71.906 = 21.03

Using F table (1 degree of freedom numerator and 4 denominator), p-value is between .025 and .01

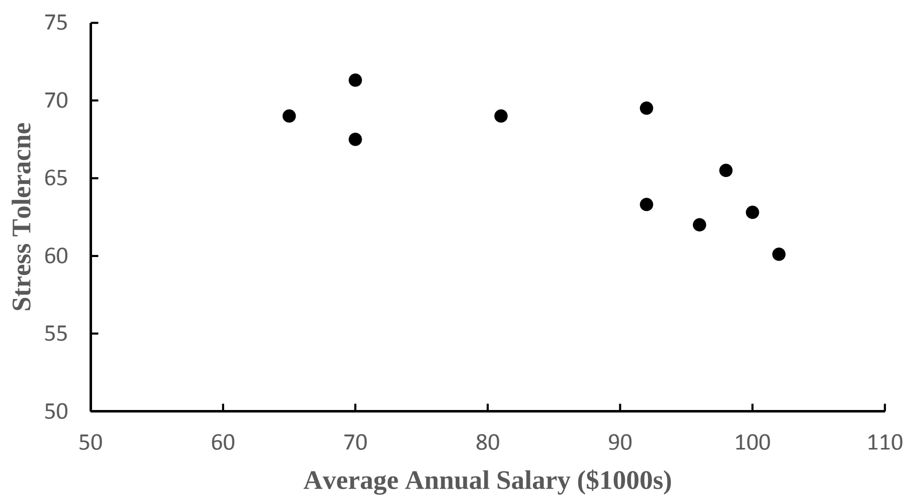
Using Excel, the p-value corresponding to F = 11.74 is .010.

Because p-value $\leq \alpha$, we reject H_0: β_1 = 0

c.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	1512.376	1	1512.376	21.03	.010
Error	287.624	4	71.906		
Total	1800	5			

27. a.



The scatter diagram suggests a negative linear relationship between the two variables.

b. Let x = stress tolerance and y = average annual salary ($)

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{866}{10} = 86.6 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{660}{10} = 66$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -367.2 \quad \Sigma(x_i - \bar{x})^2 = 1742.4$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-367.2}{1742.4} = -.2107$$

$$b_0 = \bar{y} - b_1\bar{x} = 66 - (-.2107)(86.6) = 84.2466$$

$$\hat{y} = 84.2466 - .2107x$$

c. SSE = $\Sigma(y_i - \hat{y}_i)^2 = 51.7949$ SST $= \Sigma(y_i - \bar{y})^2 = 129.18$

Thus, SSR = SST - SSE = 129.18 – 51.7949 = 77.3851

MSR = SSR/1 = 77.3851

MSE = SSE/(n - 2) = 129.18/8 = 6.4744

F = MSR / MSE = 77.3851/6.4744 = 11.9525

Using F table (1 degree of freedom numerator and 8 denominator), p-value is less than .01

Using Excel, the p-value corresponding to F = 11.9525 is .0086.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Average annual salary and stress tolerance are related.

d. r^2 = SSR/SST = 77.3851/129.18 = .5990

The estimated regression equation provided a reasonably good fit; we should feel comfortable using the estimated regression equation to estimate the stress level tolerance given the average annual salary as long as the value of the average annual salary is within the range of the current data.

e. The relationship between the average annual salary and stress tolerance is counterintuitive because one would think that jobs that pay more are most likely going to require more time and will likely involve a more stressful environment. One possibility is that the limited size of the data set is masking a much different relationship that might be more evident with a larger sample of occupations. And, the stress tolerance rating used in this study may not necessarily be a good indicator of the actual stress.

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

SSE $= \Sigma(y_i - \hat{y}_i)^2 = 1.4379$ SST $= \Sigma(y_i - \bar{y})^2 = 3.5800$

Thus, SSR = SST - SSE = 3.5800 – 1.4379 = 2.1421

s^2 = MSE = SSE / (n - 2) = 1.4379 / 9 = .1598

$$s = \sqrt{\text{MSE}} = \sqrt{.1598} = .3997$$

We can use either the t test or F test to determine whether speed of execution and overall satisfaction are related.

We will first illustrate the use of the t test.

$\Sigma(x_i - \overline{x})^2 = 2.6$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \overline{x})^2}} = \frac{.3997}{\sqrt{2.6}} = .2479$$

$$t = \frac{b_1}{s_{b_1}} = \frac{.9077}{.2479} = 3.66$$

Using t table (9 degrees of freedom), area in tail is less than .005; p-value is less than .01

Using Excel or Minitab, the p-value corresponding to t = 3.66 is .000.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Because we can reject H_0: $\beta_1 = 0$ we conclude that speed of execution and overall satisfaction are related.

Next we illustrate the use of the F test.

MSR = SSR / 1 = 2.1421

F = MSR / MSE = 2.1421 / .1598 = 13.4

Using F table (1 degree of freedom numerator and 9 denominator), p-value is less than .01

Using Excel or Minitab, the p-value corresponding to F = 13.4 is .000.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Because we can reject H_0: $\beta_1 = 0$ we conclude that speed of execution and overall satisfaction are related.

The ANOVA table is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	2.1421	1	2.1421	13.4	.000
Error	1.4379	9	.1598		
Total	3.5800	10			

29. SSE $= \Sigma(y_i - \hat{y}_i)^2 = 233{,}333.33$ SST $= \Sigma(y_i - \overline{y})^2 = 5{,}648{,}333.33$

Thus, SSR = SST – SSE= 5,648,333.33 –233,333.33 = 5,415,000

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

MSR = SSR/1 = 5,415,000

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

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	5,415,000.00	1	5,415,000	92.83	.0006
Error	233,333.33	4	58,333.33		
Total	5,648,333.33	5			

Using F table (1 degree of freedom numerator and 4 denominator), p-value is less than .01

Using Excel or Minitab, the p-value corresponding to $F = 92.83$ is .0006.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$. Production volume and total cost are related.

30. $\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 1043.03$ $\quad$ $\text{SST} = \Sigma(y_i - \bar{y})^2 = 10{,}568$

Thus, SSR = SST – SSE = 10,568 – 1043.03 = 9524.97

s^2 = MSE = SSE/(n-2) = 1043.03/4 = 260.7575

$$s = \sqrt{260.7575} = 16.1480$$

$$\sum(x_i - \bar{x})^2 = 56.655$$

$$s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{16.148}{\sqrt{56.655}} = 2.145$$

$$t = \frac{b_1}{s_{b_1}} = \frac{12.966}{2.145} = 6.045$$

Using t table (4 degrees of freedom), area in tail is less than .005
p-value is less than .01

Using Excel, the p-value corresponding to $t = 6.045$ is .004.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

There is a significant relationship between cars in service and annual revenue.

31. SST = 52,120,800 $\quad$ SSE = 7,102,922.54

SSR = SST – SSR = 52,120,800 - 7,102,922.54 = 45,017,877

MSR = SSR/1 = 45,017,877

MSE = SSE/(n - 2) = 7,102,922.54/8 = 887,865.3

F = MSR / MSE = 45,017,877/887,865.3 = 50.7

Using F table (1 degree of freedom numerator and 8 denominator), p-value is less than .01

Using Excel, the p-value corresponding to $F = 32.015$ is .000.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

Weight and price are related.

32. a. $s = 2.033$

$\bar{x} = 3 \quad \Sigma(x_i - \bar{x})^2 = 10$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 2.033\sqrt{\frac{1}{5} + \frac{(4-3)^2}{10}} = 1.11$$

b. $\hat{y}^* = .2 + 2.6\, x^* = .2 + 2.6(4) = 10.6$

$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$

$10.6 \pm 3.182\,(1.11) = 10.6 \pm 3.53$

or 7.07 to 14.13

c. $$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 2.033\sqrt{1 + \frac{1}{5} + \frac{(4-3)^2}{10}} = 2.32$$

d. $\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

$10.6 \pm 3.182\,(2.32) = 10.6 \pm 7.38$

or 3.22 to 17.98

33. a. $s = 8.7560$

b. $\bar{x} = 11 \quad \Sigma(x_i - \bar{x})^2 = 180$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 8.7560\sqrt{\frac{1}{5} + \frac{(8-11)^2}{180}} = 4.3780$$

$\hat{y}^* = 0.2 + 2.6x^* = 0.2 + 2.6(4) = 10.6$

$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$

$44 \pm 3.182\,(4.3780) = 44 \pm 13.93$

or 30.07 to 57.93

c. $$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 8.7560\sqrt{1 + \frac{1}{5} + \frac{(8-11)^2}{180}} = 9.7895$$

d. $\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

$44 \pm 3.182(9.7895) = 44 \pm 31.15$

or 12.85 to 75.15

34. $s = 6.5141$

$\overline{x} = 10 \quad \Sigma(x_i - \overline{x})^2 = 190$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{\Sigma(x_i - \overline{x})^2}} = 6.5141\sqrt{\frac{1}{5} + \frac{(12-10)^2}{190}} = 3.0627$$

$\hat{y}^* = 7.6 + .9x^* = 7.6 + .9(12) = 18.40$

$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$

$18.40 \pm 3.182(3.0627) = 18.40 \pm 9.75$

or 8.65 to 28.15

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{\Sigma(x_i - \overline{x})^2}} = 6.5141\sqrt{1 + \frac{1}{5} + \frac{(12-10)^2}{190}} = 7.1982$$

$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

$18.40 \pm 3.182(7.1982) = 18.40 \pm 22.90$

or -4.50 to 41.30

The two intervals are different because there is more variability associated with predicting an individual value than there is a mean value.

35. a. $\hat{y}^* = 2090.5 + 581.1x^* = 2090.5 + 581.1(3) = 3833.8$

b. $s = \sqrt{\text{MSE}} = \sqrt{21,284} = 145.89\, s = 145.89$

$\overline{x} = 3.2 \quad \Sigma(x_i - \overline{x})^2 = 0.74$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{\Sigma(x_i - \overline{x})^2}} = 145.89\sqrt{\frac{1}{6} + \frac{(3-3.2)^2}{0.74}} = 68.54$$

$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$

$3833.8 \pm 2.776\,(68.54) = 3833.8 \pm 190.27$

or \$3643.53 to \$4024.07

c. $s_{\text{pred}} = s\sqrt{1+\frac{1}{n}+\frac{(x^*-\overline{x})^2}{\Sigma(x_i-\overline{x})^2}} = 145.89\sqrt{1+\frac{1}{6}+\frac{(3-3.2)^2}{0.74}} = 161.19$

$\hat{y}^* \pm t_{\alpha/2}s_{\text{pred}}$

3833.8 ± 2.776 (161.19) = 3833.8 ± 447.46

or \$3386.34 to \$4281.26

d. As expected, the prediction interval is much wider than the confidence interval. This is due to the fact that it is more difficult to predict the starting salary for one new student with a GPA of 3.0 than it is to estimate the mean for all students with a GPA of 3.0.

36. a. $s_{\hat{y}^*} = s\sqrt{\frac{1}{n}+\frac{(x^*-\overline{x})^2}{\Sigma(x_i-\overline{x})^2}} = 4.6098\sqrt{\frac{1}{10}+\frac{(9-7)^2}{142}} = 1.6503$

$\hat{y}^* \pm t_{\alpha/2}s_{\hat{y}^*}$

$\hat{y}^* = 80 + 4x^* = 80 + 4(9) = 116$

116 ± 2.306(1.6503) = 116 ± 3.8056

or 112.19 to 119.81 (\$112,190 to \$119,810)

b. $s_{\text{pred}} = s\sqrt{1+\frac{1}{n}+\frac{(x^*-\overline{x})^2}{\Sigma(x_i-\overline{x})^2}} = 4.6098\sqrt{1+\frac{1}{10}+\frac{(9-7)^2}{142}} = 4.8963$

$\hat{y}^* \pm t_{\alpha/2}s_{\text{pred}}$

116 ± 2.306(4.8963) = 116 ± 11.2909

or 104.71 to 127.29 (\$104,710 to \$127,290)

c. As expected, the prediction interval is much wider than the confidence interval. This is due to the fact that it is more difficult to predict annual sales for one new salesperson with 9 years of experience than it is to estimate the mean annual sales for all salespersons with 9 years of experience.

37. a. $\overline{x} = 57 \quad \Sigma(x_i-\overline{x})^2 = 7648$

$s^2 = 1.88 \quad s = 1.37$

$s_{\hat{y}^*} = s\sqrt{\frac{1}{n}+\frac{(x^*-\overline{x})^2}{\Sigma(x_i-\overline{x})^2}} = 1.37\sqrt{\frac{1}{7}+\frac{(52.5-57)^2}{7648}} = 0.52$

$\hat{y}^* \pm t_{\alpha/2}s_{\hat{y}^*}$

$\hat{y}^* = 4.68 + 0.16\,x^* = 4.68 + 0.16(52.5) = 13.08$

13.08 ± 2.571 (.52) = 13.08 ± 1.34

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

b. $s_{\text{pred}} = 1.47$

13.08 ± 2.571 (1.47) = 13.08 ± 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 = MSE = 58,333.33 $s = 241.52$

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 241.52\sqrt{1 + \frac{1}{6} + \frac{(500 - 575)^2}{93,750}} = 267.50$$

$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

5046.67 ± 4.604 (267.50) = 5046.67 ± 1231.57

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. With $x^* = 89$, $\hat{y}^* = 17.49 + 1.0334x^* = 17.49 + 1.0334(89) = \$109.46$

b. s^2 = MSE = SSE/(n – 2) = 1541.4/7 = 220.2

$s = \sqrt{220.2} = 14.391$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 14.8391\sqrt{\frac{1}{9} + \frac{(89 - 105)^2}{4100}} = 6.1819$$

$\hat{y}^* \pm t_{.025} s_{\hat{y}^*}$ = 109.46 ± 2.365(6.1819) = 109.46 ± 14.6202

or $94.84 to $124.08

c. $\hat{y}^* = 17.49 + 1.0334x = 17.49 + 1.0334(128) = \$149.77$

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 14.8391\sqrt{1 + \frac{1}{9} + \frac{(128 - 105)^2}{4100}} = 16.525$$

$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

$149.77 \pm 2.365(16.525) = 149.77 \pm 39.08$

or \$110.69 to \$188.85

40\. a. 9

b. $\hat{y} = 20.0 + 7.21x$

c. 1.3626

d. SSE = SST - SSR = 51,984.1 - 41,587.3 = 10,396.8

MSE = 10,396.8/7 = 1,485.3

F = MSR / MSE = 41,587.3 /1,485.3 = 28.00

Using F table (1 degree of freedom numerator and 7 denominator), p-value is less than .01

Using Excel or Minitab, the p-value corresponding to F = 28.00 is .0011.

Because p-value $\leq \alpha = .05$, we reject H_0: $B_1 = 0$.

Selling price is related to annual gross rents.

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

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

b. $t = \dfrac{b_1 - B_1}{s_{b_1}} = \dfrac{.8951 - 0}{.149} = 6.01$

Using the t table (8 degrees of freedom), area in tail is less than .005
p-value is less than .01

Using Excel or Minitab, the p-value corresponding to t = 6.01 is .0003.

Because p-value $\leq \alpha = .05$, we reject H_0: $B_1 = 0$

Maintenance expense is related to usage.

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

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

b. 30

c. F = MSR / MSE = 6828.6/82.1 = 83.17

Using F table (1 degree of freedom numerator and 28 denominator), p-value is less than .01

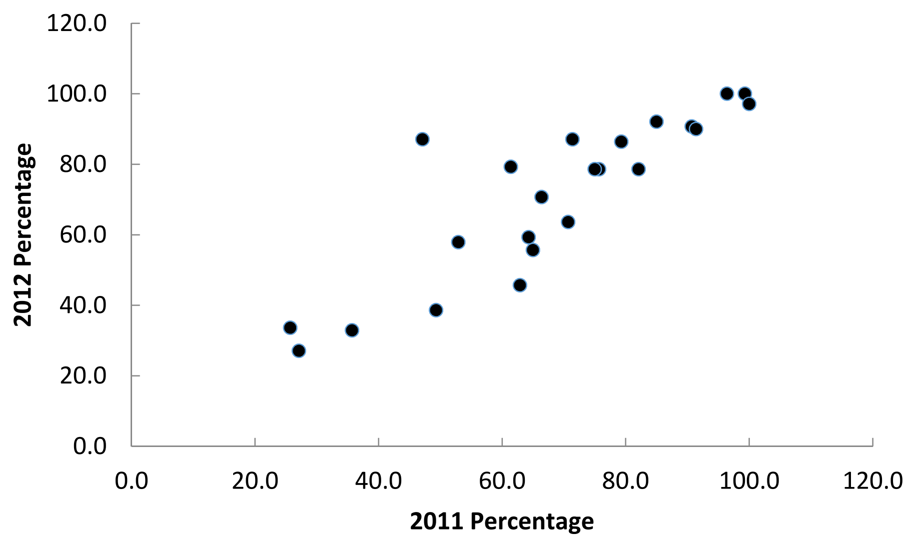
Using Excel or Minitab, the p-value corresponding to $F = 83.17$ is .000.

Because p-value $< \alpha = .05$, we reject H_0: $B_1 = 0$.

Annual sales is related to the number of salespersons.

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

43. a.



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

c. The Excel output is shown below.

Regression Statistics	
Multiple R	0.8702
R Square	0.7572
Adjusted R Square	0.7456
Standard Error	11.5916
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8798.2391	8798.2391	65.4802	6.85277E-08
Residual	21	2821.6609	134.3648		
Total	22	11619.9			

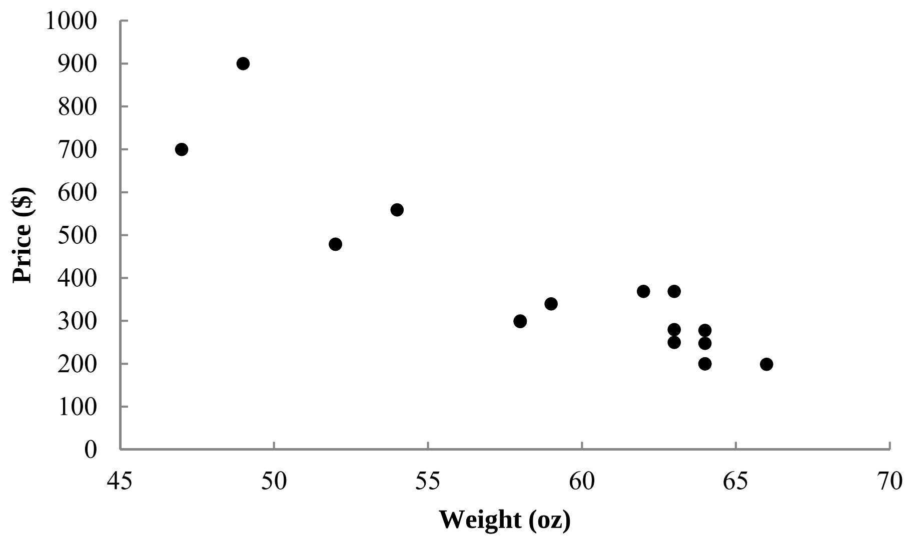
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	7.3880	8.2125	0.8996	0.3785
2011 Percentage	0.9276	0.1146	8.0920	6.85277E-08

$\hat{y}$ = 7.3880 + 0.9276(2011 Percentage)

d. Significant relationship: p-value = 0.000 < α = .05.

e. r^2 = .7572; a good fit.

44. a. Scatter diagram:



b. There appears to be a negative linear relationship between the two variables. The heavier helmets tend to be less expensive.

c. The Minitab output is shown below:

```
                    Analysis of Variance

Source          DF  Adj SS  Adj MS  F-Value  P-Value
Regression       1  462761  462761    54.90    0.000
  Weight         1  462761  462761    54.90    0.000
Error           16  134865    8429
  Lack-of-Fit    8  122784   15348    10.16    0.002
  Pure Error     8   12080    1510
Total           17  597626


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
91.8098  77.43%     76.02%      68.22%
```

```
Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant      2044      226     9.03    0.000
Weight      -28.35     3.83    -7.41    0.000  1.00


Regression Equation

Price = 2044 - 28.35 Weight


Fits and Diagnostics for Unusual Observations

                              Std
Obs  Price    Fit  Resid  Resid
  7  900.0  655.2  244.8   3.03  R

R  Large residual
```

d. Significant relationship: p-value = .000 < α = .05

e. r^2 = 0.774; A good fit

45. a. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{70}{5} = 14 \qquad \bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{76}{5} = 15.2$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 200 \qquad \Sigma(x_i - \bar{x})^2 = 126$

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

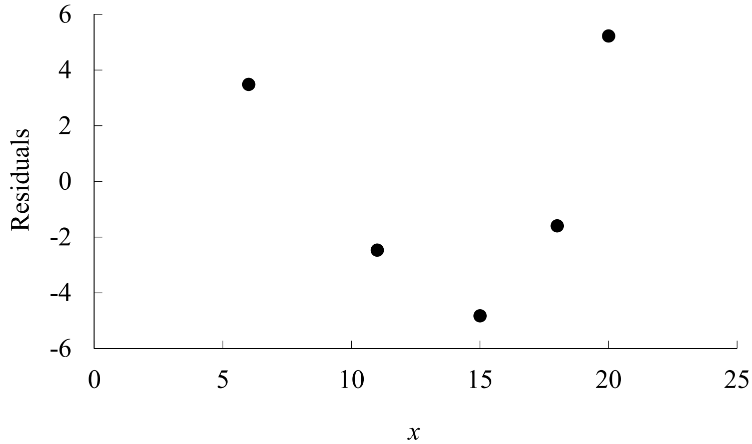
$b_0 = \bar{y} - b_1\bar{x} = 15.2 - (1.5873)(14) = -7.0222$

$\hat{y} = -7.02 + 1.59x$

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. $s^2 = 23.78$

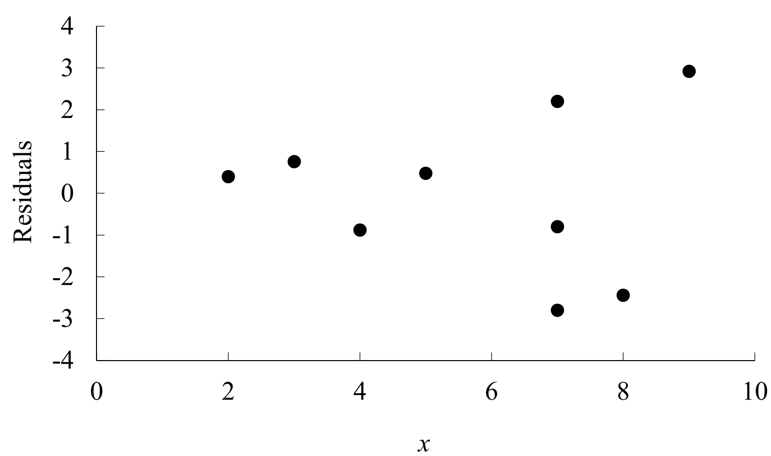
$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} = \frac{1}{5} + \frac{(x_i - 14)^2}{126}$$

The standardized residuals are 1.32, -.59, -1.11, -.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. $\hat{y} = 2.32 + .64x$

b.



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. SST = 1002   SSE = 310.28   SSR = 691.72

MSR = SSR / 1 = 691.72

MSE = SSE / (n - 2) = 310.28/ 5 = 62.0554

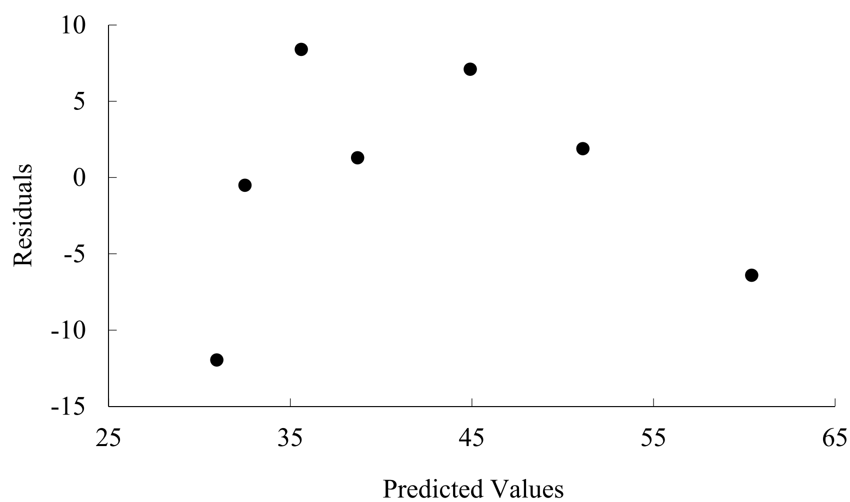
F = MSR / MSE = 691.72/ 62.0554= 11.15

Using F table (1 degree of freedom numerator and 5 denominator), p-value is between .01 and .025

Using Excel or Minitab, the p-value corresponding to F = 11.15 is .0206.

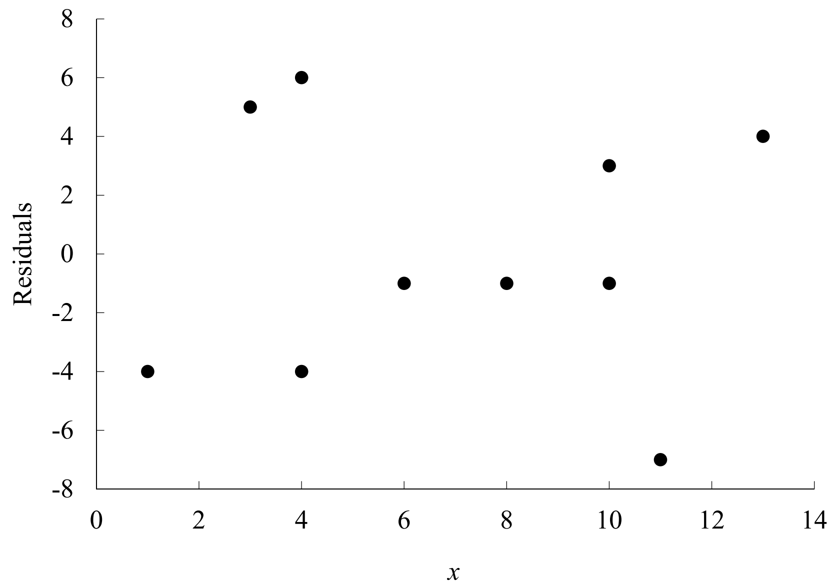
Because p-value $\leq \alpha$ = .05, we conclude that the two variables are related.

c.



d. 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. A portion of the Excel output follows:

Regression Statistics	
Multiple R	0.8696
R Square	0.7561
Adjusted R Square	0.7257
Standard Error	78.7819
Observations	10

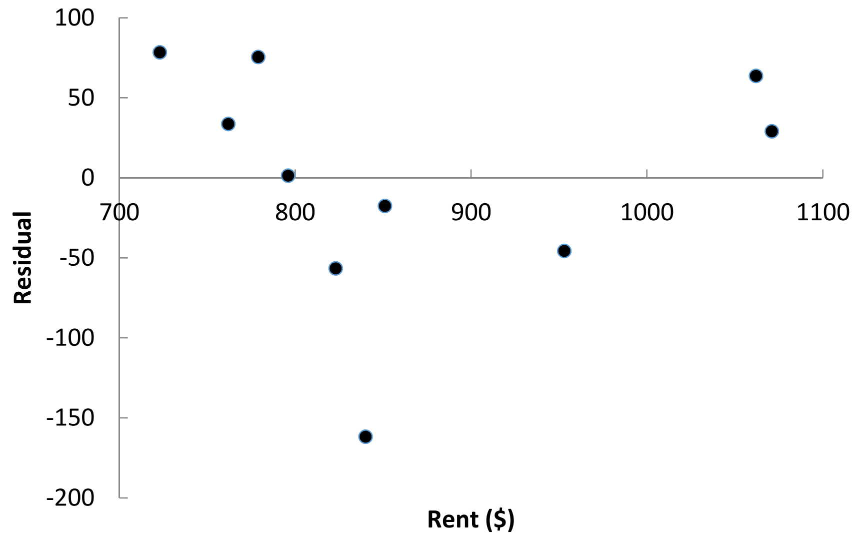
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	153961.6801	153961.6801	24.8062	0.0011
Residual	8	49652.7199	6206.5900		
Total	9	203614.4			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-197.9583	187.6950	-1.0547	0.3224
Rent ($)	1.0699	0.2148	4.9806	0.0011

$\hat{y}$ = -197.9583 + 1.0699 Rent ($)

b.



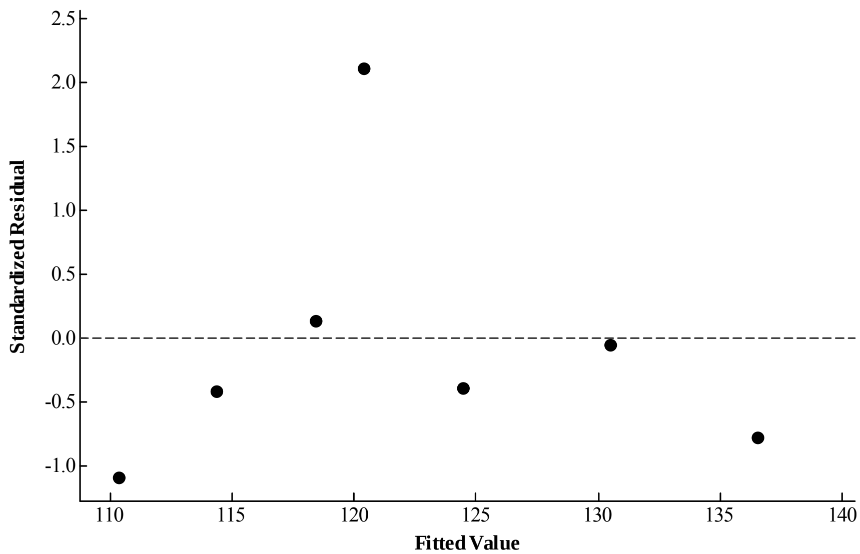
c. The residual plot leads us to question the assumption of a linear relationship between the average asking rent and the monthly mortgage. Therefore, even though the relationship is very significant (p-value = .0011), using the estimated regression equation to make predictions of the monthly mortgage beyond the range of the data is not recommended.

50. a. The Minitab output follows:

```
Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
Regression   1   497.2   497.2     3.12    0.137
  x          1   497.2   497.2     3.12    0.137
Error        5   795.7   159.1
Total        6  1292.9


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
12.6151  38.45%     26.15%       0.00%


Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant   66.1     32.1     2.06    0.094
x         0.402    0.228     1.77    0.137  1.00


Regression Equation

y = 66.1 + 0.402 x
```

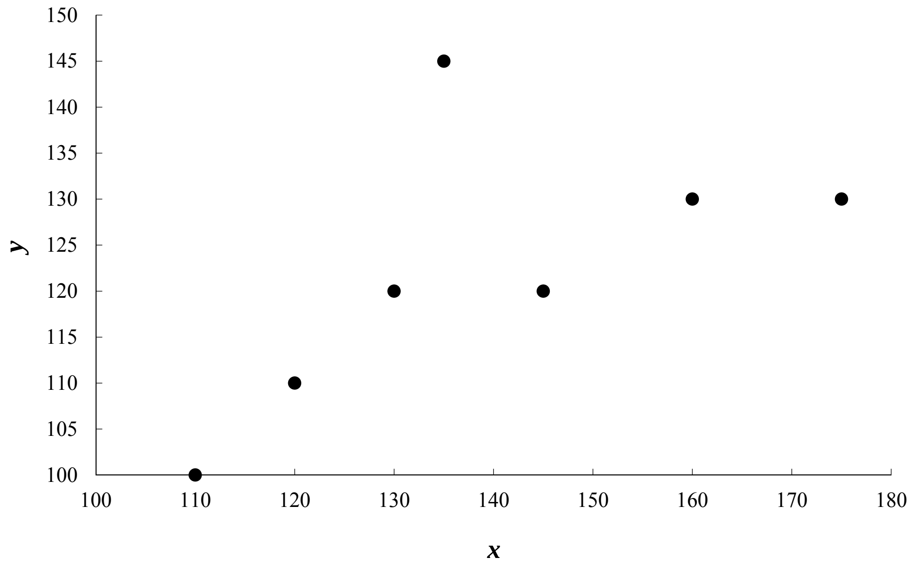
```
Fits and Diagnostics for Unusual Observations

                                 Std
Obs        y     Fit  Resid  Resid
  1  145.00  120.42  24.58   2.11  R

R  Large 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:

```
Analysis of Variance

Source          DF   Adj SS  Adj MS  F-Value  P-Value
Regression       1   40.779  40.779     4.03    0.091
  x              1   40.779  40.779     4.03    0.091
Error            6   60.721  10.120
  Lack-of-Fit    5   52.721  10.544     1.32    0.576
  Pure Error     1    8.000   8.000
Total            7  101.500


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
3.18123  40.18%     30.21%       0.00%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   13.00     2.40     5.43    0.002
x          0.425    0.212     2.01    0.091  1.00


Regression Equation

y = 13.00 + 0.425 x
```

```
Fits and Diagnostics for Unusual Observations

Obs      y    Fit  Resid  Std Resid
  7  24.00  18.10   5.90       2.00  R
  8  19.00  22.35  -3.35      -2.16  R  X

R  Large residual
X  Unusual X
```

The standardized residuals are: -1.00, -.41, .01, -.48, .25, .65, -2.00, -2.16

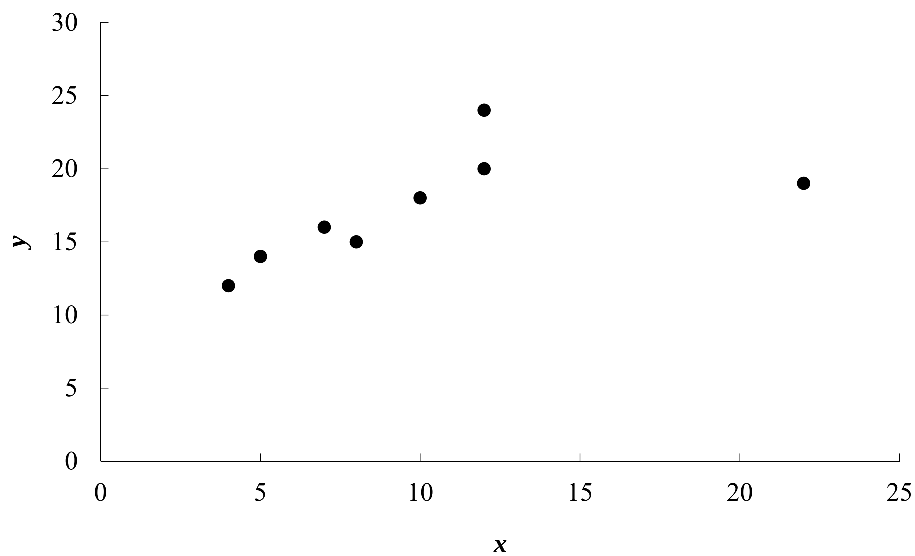
The last two observations in the data set appear to be outliers since the standardized residuals for these observations are 2.00 and -2.16, respectively.

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

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

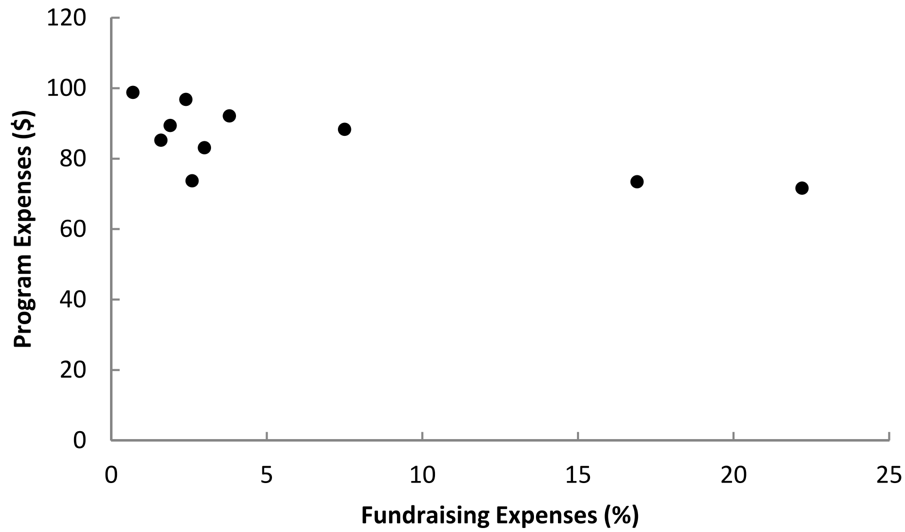
MINITAB identifies an observation as having high leverage if $h_i > 6/n$; for these data, $6/n = 6/8 = .75$. Since the leverage for the observation $x = 22$, $y = 19$ is .76, Minitab would identify observation 8 as a high leverage point. Thus, we conclude that observation 8 is an influential observation.

c.



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

52. a.



The scatter diagram does indicate potential influential observations. For example, the 22.2% fundraising expense for the American Cancer Society and the 16.9% fundraising expense for the St. Jude Children's Research Hospital look like they may each have a large influence on the slope of the estimated regression line. And, with a fundraising expense of on 2.6%, the percentage spend on programs and services by the Smithsonian Institution (73.7%) seems to be somewhat lower than would be expected; thus, this observeraton may need to be considered as a possible outlier

b. A portion of the Minitab output follows:

```
Analysis of Variance

Source                      DF  Adj SS  Adj MS  F-Value  P-Value
Regression                   1   408.4  408.35     7.31    0.027
  Fundraising Expenses (%)   1   408.4  408.35     7.31    0.027
Error                        8   446.9   55.86
Total                        9   855.2


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
7.47387  47.75%     41.22%      29.38%


Coefficients

Term                        Coef  SE Coef  T-Value  P-Value   VIF
Constant                   90.98     3.18    28.64    0.000
Fundraising Expenses (%)  -0.917    0.339    -2.70    0.027  1.00


Regression Equation

Program Expenses (%) = 90.98 - 0.917 Fundraising Expenses (%)
```

```
Fits and Diagnostics for Unusual Observations

      Program
     Expenses
Obs       (%)    Fit   Resid  Std Resid
  3     73.70  88.60  -14.90      -2.13  R
  5     71.60  70.62    0.98       0.21     X

R  Large residual
X  Unusual X


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

c. The slope of the estimtaed regression equation is -0.917. Thus, for every 1% increase in the amount spent on fundraising the percentage spent on program expresses will decrease by .917%; in other words, just a little under 1%. The negative slope and value seem to make sense in the context of this problem situation.

d. The Minitab output in part (b) indicates that there are two unusual observations:

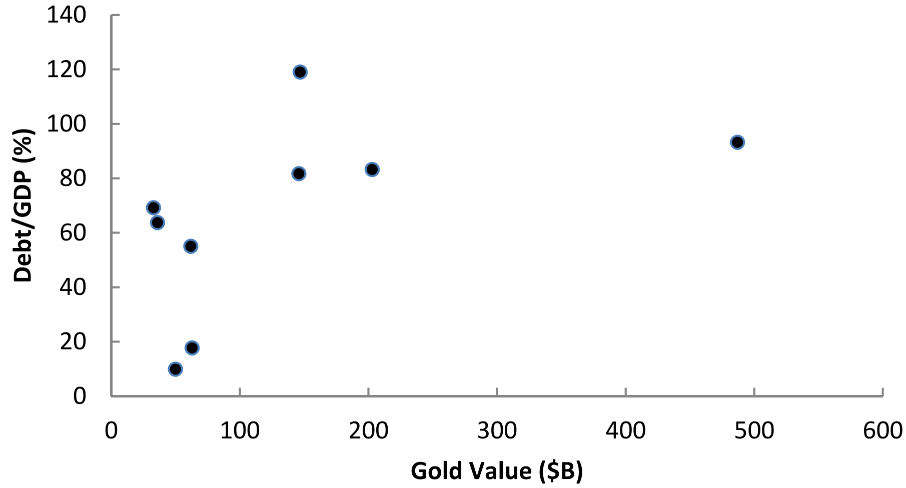
- Observation 3 (Smithsonian Institution) is an outlier because it has a large standardized residual.
- Observation 5 (American Cancer Society) is an influential observation becasuse has high leverage.

Although fundraising expenses for the Smithsonian Institution are on the low side as compared to most of the other super-sized charities, the percentage spent on program expenses appears to be much lower than one would expect. It appears that the Smithsonian's administrative expenses are too high. But, thinking about the expenses of running a large museum like the Smithsonian, the percetage spent on administrative expenses may not be unreasonable and is just due to the fact that operating costs for a museum are in general higher than for some other types of organizations. The very large value of fundraising expenses for the American Cancer Society suggests that this obervation has a large influence on the estiamted regresion equation. The following Minitab output shows the results if this observatoin is deleted from the original data.

```
The regression equation is
Program Expenses (%) = 91.3 - 1.00 Fundraising Expenses (%)


Predictor                      Coef   SE Coef      T      P
Constant                     91.256     3.654  24.98  0.000
Fundraising Expenses (%)    -1.0026    0.5590  -1.79  0.116


S = 7.96708   R-Sq = 31.5%   R-Sq(adj) = 21.7%
```

The y-intercept has changed slightly, but the slope has changed from -.917 to -1.00.

53. a.



b. There appears to be a positive relationship between the two variables. But, observation 9 (U.S.) appears to be an observation with high leverage and may be very influential in terms of fitting a linear model to the data.

c. The Minitab output follows.

```
Analysis of Variance

Source        DF  Adj SS  Adj MS  F-Value  P-Value
Regression     1    2522    2522     2.46    0.161
  Gold Value   1    2522    2522     2.46    0.161
Error          7    7186    1027
Total          8    9708


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
32.0394  25.98%     15.40%       0.00%


Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant      49.1     15.1     3.25    0.014
Gold Value  0.1230   0.0785     1.57    0.161  1.00


Regression Equation

Debt = 49.1 + 0.1230 Gold Value


Fits and Diagnostics for Unusual Observations

Obs  Debt    Fit  Resid  Std Resid
  9  93.2  109.0  -15.8      -1.27  X
```

```
X  Unusual X
```

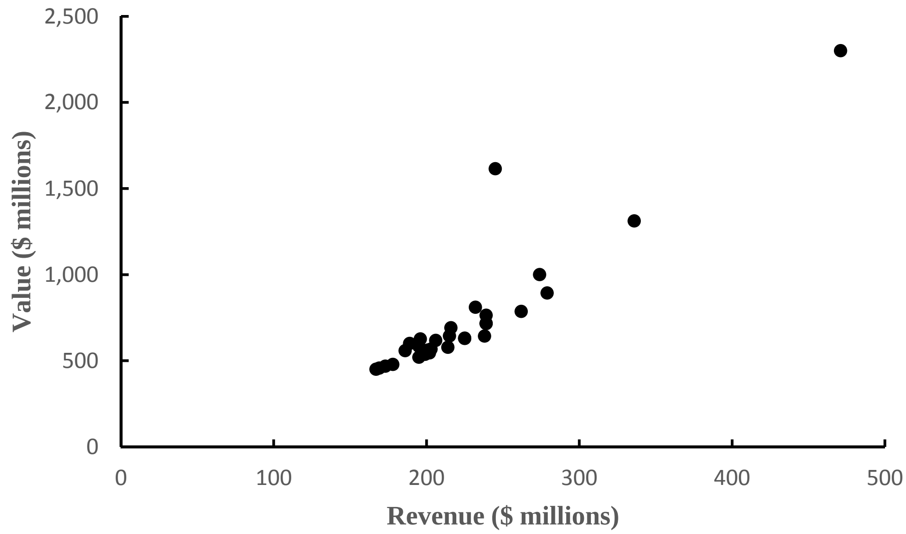
d. The Minitab output identifies observation 9 as an observation whose x value gives it large leverage.

e. Looking at the scatter diagram in part (a) it looks like observation 9 will have a lot of influence on the estimated regression equation. To investigate this we can simply drop the observation from the data set and fit a new estimated regression equation. The Minitab output we obtained follows.

```
Analysis of Variance

Source        DF  Adj SS  Adj MS  F-Value  P-Value
Regression     1    3324  3324.2     3.60    0.107
  Gold Value   1    3324  3324.2     3.60    0.107
Error          6    5542   923.6
Total          7    8866


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
30.3907  37.49%     27.08%       0.00%


Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant      30.8     19.8     1.55    0.172
Gold Value   0.342    0.180     1.90    0.107  1.00


Regression Equation

Debt = 30.8 + 0.342 Gold Value
```

Note that the slope of the estimated regression equation is now .342 as compared to a value of .123 when this observation is included. Thus, we see that this observation has a big impact on the value of the slope of the fitted line and hence we would say that it is an influential observation.

54. a.



The scatter diagram does indicate potential outliers and/or influential observations. For example, the New York Yankees have both the hightest revenue and value, and appears to be an influential observation. The Los Angeles Dodgers have the second highest value and appears to be an outlier.

b. A portion of the Excel output follows:

Regression Statistics	
Multiple R	0.9062
R Square	0.8211
Adjusted R Square	0.8148
Standard Error	165.6581
Observations	30

ANOVA

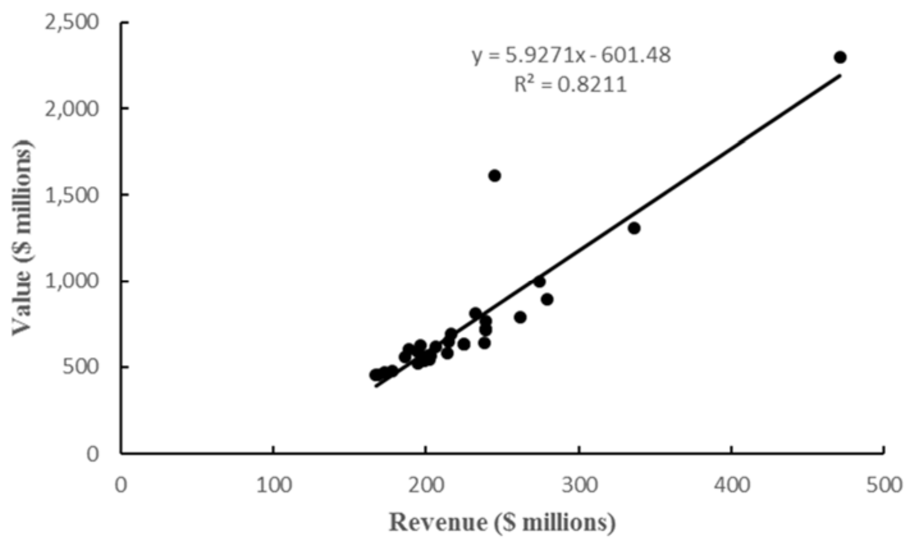
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3527616.598	3527616.6	128.5453	5.616E-12
Residual	28	768392.7687	27442.599		
Total	29	4296009.367			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-601.4814	122.4288	-4.9129	3.519E-05	-852.2655	-350.6973
Revenue ($ millions)	5.9271	0.5228	11.3378	5.616E-12	4.8562	6.9979

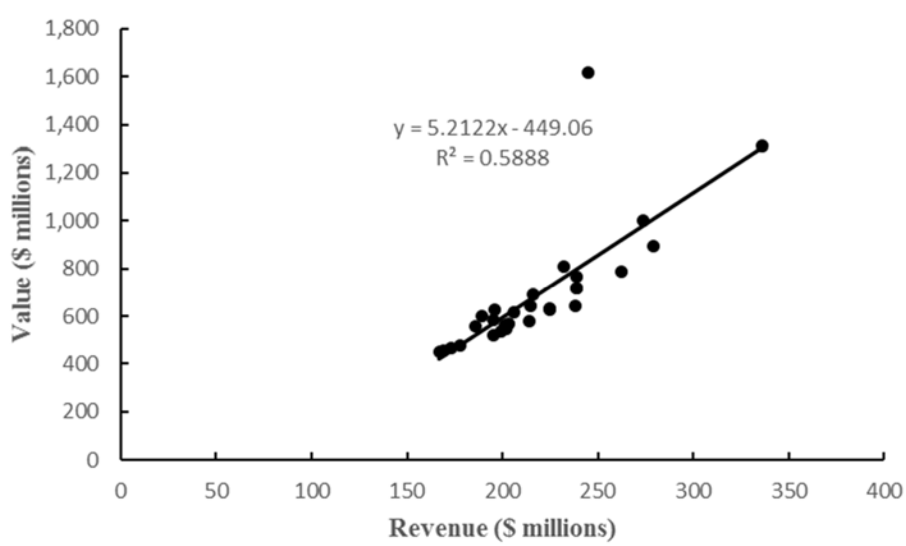
Thus, the estimated regression equation that can be used to predict the team's value given the value of annual revenue is $\hat{y}$ = -601.4814 + 5.9271 Revenue.

c. The Standard Residual value for the Los Angeles Dodgers is 4.7 and should be treated as an outlier. To determine if the New York Yankees point is an influential observation we can remove the observation and compute a new estimated regression equation. The results show that the estimated regresssion equation is $\hat{y}$ = -449.061 + 5.2122 Revenue. The following two scatter diagrams illustrate the small change in the estimated regression equation after removing the observation for the New York Yankees. These scatter diagrams show that the effect of the New York Yankees observation on the regression results is not that dramatic.

Scatter Diagram Including the New York Yankees Observation



Scatter Diagram Excluding the New York Yankees Observation

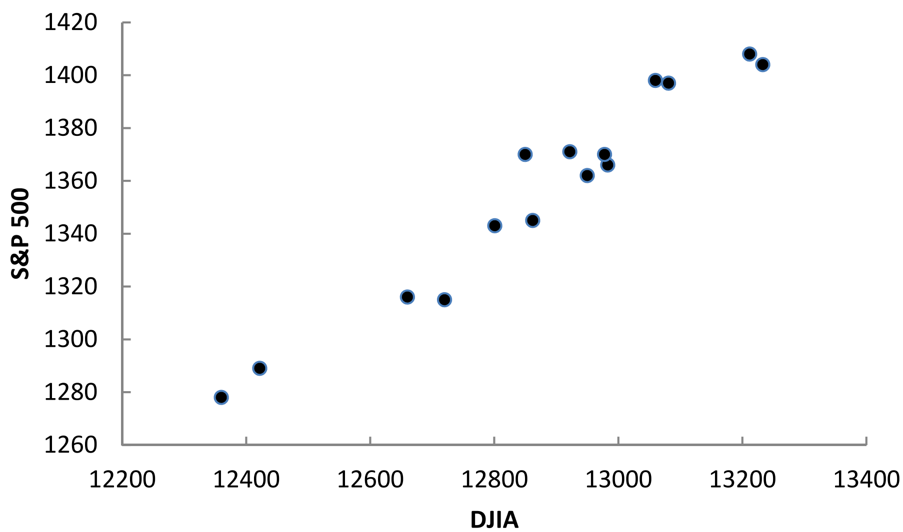


55. No. Regression or correlation analysis can never prove that two variables are causally related.

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. The purpose of testing whether $\beta_1 = 0$ is to determine whether or not there is a significant relationship between x and y. However, rejecting $\beta_1 = 0$ does not necessarily imply a good fit. For example, if $\beta_1 = 0$ is rejected and r^2 is low, there is a statistically significant relationship between x and y but the fit is not very good.

58. a.



b. A portion of the Minitab output is shown below:

```
Analysis of Variance

Source      DF  Adj SS   Adj MS  F-Value  P-Value
Regression   1   22146  22145.6   239.89    0.000
  DJIA       1   22146  22145.6   239.89    0.000
Error       13    1200     92.3
Total       14   23346


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
9.60811  94.86%     94.46%      93.61%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    -669      131    -5.12    0.000
DJIA      0.1573   0.0102    15.49    0.000  1.00
```

```
Regression Equation

S&P = -669 + 0.1573 DJIA
```

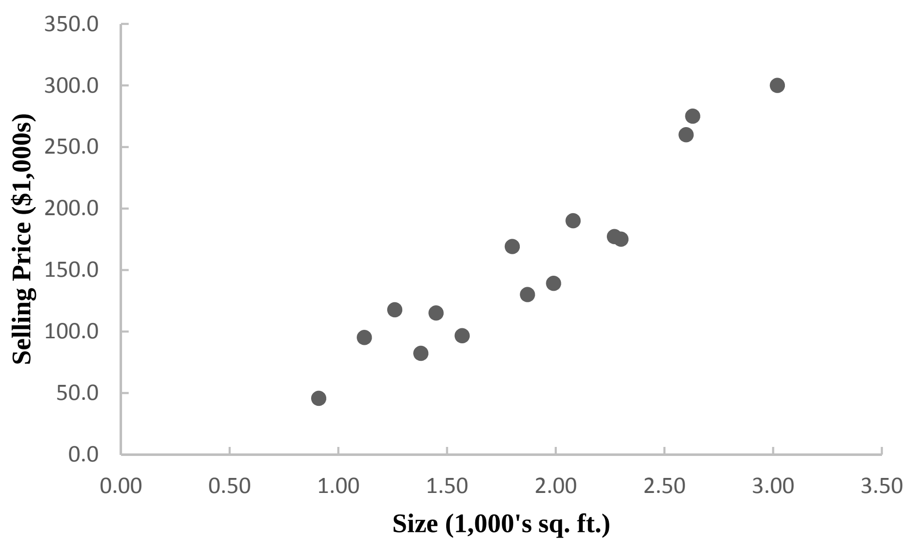
c. Using the F test, the p-value corresponding to $F = 239.89$ is .000. Because the p-value $\leq \alpha = .05$, we reject $H_0: \beta_1 = 0$; there is a significant relationship.

d. With R-Sq = 94.9%, the estimated regression equation provided an excellent fit.

e. $\hat{y} = -669.0 + .15727(\text{DJIA}) = -669.0 + .15727(13{,}500) = 1454$

f. The DJIA is not that far beyond the range of the data. With the excellent fit provided by the estimated regression equation, we should not be too concerned about using the estimated regression equation to predict the S&P500.

59. a.



The scatter diagram suggests that there is a linear relationship between size and selling price and that as size increases, selling price increases.

b. The Excel output appears below:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9474
R Square	0.8975
Adjusted R Square	0.8896
Standard Error	24.6040
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	68897.0802	68897.0802	113.8124	8.45157E-08
Residual	13	7869.6358	605.3566		
Total	14	76766.7160			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-59.0156	21.2877	-2.7723	0.0158	-105.0049	-13.0262	-105.0049	-13.0262
Size (1000's sq. ft.)	115.0915	10.7882	10.6683	0.0000	91.7850	138.3979	91.7850	138.3979

The estimated regression equation is: $\hat{y}$ = -59.016 + 115.091x

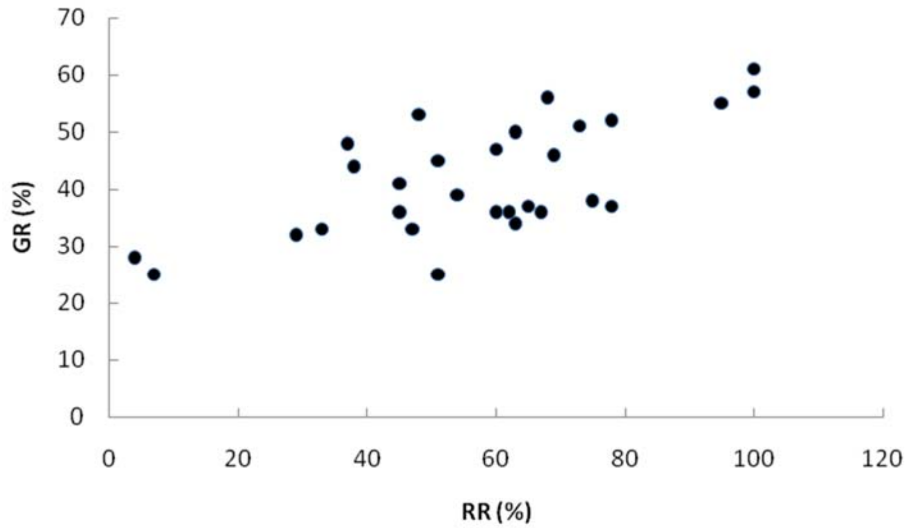
c. Significant relationship: p-value = .000 < α = .05

d. $\hat{y}$ = -59.016 + 115.091(square feet) = -59.016 + 115.091(2.0) = 171.166 or approximately $171,166.

e. The estimated regression equation should provide a good estimate because r^2 = 0.897.

f. This estimated equation might not work well for other cities. Housing markets are also driven by other factors that influence demand for housing, such as job market and quality-of-life factors. For example, because of the existence of high tech jobs and its proximity to the ocean, the house prices in Seattle, Washington might be very different from the house prices in Winston, Salem, North Carolina.

60. a.



The scatter diagram indicates a positive linear relationship between the two variables. Online universities with higher retention rates tend to have higher graduation rates.

b. The Minitab output follows:

```
Analysis of Variance

Source          DF  Adj SS   Adj MS  F-Value  P-Value
Regression       1  1224.3  1224.29    22.02    0.000
  RR(%)          1  1224.3  1224.29    22.02    0.000
Error           27  1501.0    55.59
  Lack-of-Fit   21   979.5    46.64     0.54    0.865
  Pure Error     6   521.5    86.92
Total           28  2725.3


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
7.45610  44.92%     42.88%      38.68%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   25.42     3.75     6.79    0.000
RR(%)     0.2845   0.0606     4.69    0.000  1.00


Regression Equation

GR(%) = 25.42 + 0.2845 RR(%)


Fits and Diagnostics for Unusual Observations
```

```
Obs  GR(%)    Fit   Resid  Std Resid
  2  25.00  39.93  -14.93      -2.04  R
  3  28.00  26.56    1.44       0.22     X

R  Large residual
X  Unusual X


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

c. Because the p-value = .000 < α =.05, the relationship is significant.

d. The estimated regression equation is able to explain 44.9% of the variability in the graduation rate based upon the linear relationship with the retention rate. It is not a great fit, but given the type of data, the fit is reasonably good.

e. In the Minitab output in part (b), South University is identified as an observation with a large standardized residual. With a retention rate of 51% it does appear that the graduation rate of 25% is low as compared to the results for other online universities. The president of South University should be concerned after looking at the data. Using the estimated regression equation, we estimate that the gradation rate at South University should be 25.4 + .285(51) = 40%.

f. In the Minitab output in part (b), the University of Phoenix is identified as an observation whose x value gives it large influence. With a retention rate of only 4%, the president of the University of Phoenix should be concerned after looking at the data.

61. The Minitab output is shown below:

```
Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
Regression   1   860.1  860.05    47.62    0.000
  Usage      1   860.1  860.05    47.62    0.000
Error        8   144.5   18.06
Total        9  1004.5


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
4.24962  85.62%     83.82%      75.21%


Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant  10.53     3.74     2.81    0.023
Usage     0.953    0.138     6.90    0.000  1.00


Regression Equation

Expense = 10.53 + 0.953 Usage
```

```
Variable  Setting
Usage          30


    Fit   SE Fit        95% CI              95% PI
39.1312  1.49251  (35.6894, 42.5729)  (28.7447, 49.5176)
```

a. $\hat{y} =$ 10.53 + .953 Usage

b. Since the p-value corresponding to F = 47.62 = .000 < α = .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 $\hat{y} =$ 10.53 + .953(30) = 39.12 or $3912.

62. a. The Minitab output is shown below:

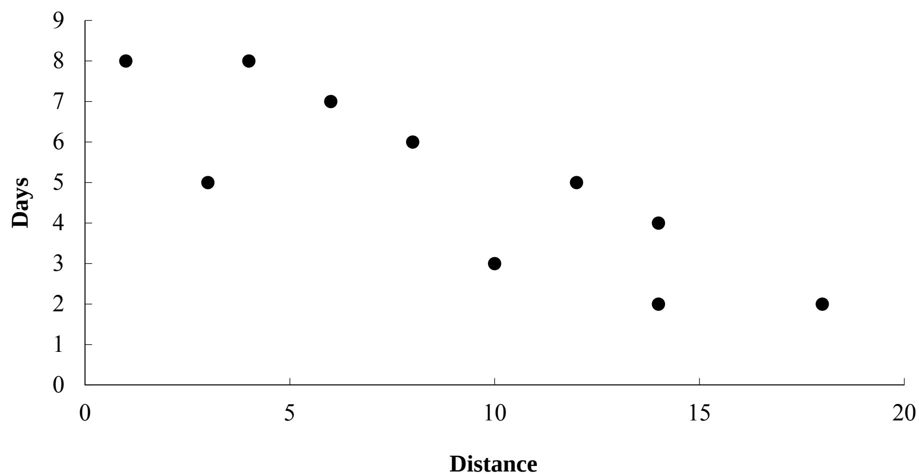
```
Analysis of Variance

Source         DF  Adj SS  Adj MS  F-Value  P-Value
Regression      1  25.130  25.130    11.33    0.028
  Speed         1  25.130  25.130    11.33    0.028
Error           4   8.870   2.217
  Lack-of-Fit   2   4.870   2.435     1.22    0.451
  Pure Error    2   4.000   2.000
Total           5  34.000


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.48909  73.91%     67.39%      36.69%


Coefficients

Term         Coef  SE Coef  T-Value  P-Value   VIF
Constant    22.17     1.65    13.42    0.000
Speed     -0.1478   0.0439    -3.37    0.028  1.00


Regression Equation

Defects = 22.17 - 0.1478 Speed


Variable  Setting
Speed          50


    Fit    SE Fit         95% CI              95% PI
14.7826  0.896327  (12.2940, 17.2712)  (9.95703, 19.6082)
```

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 17.2712.

63. 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:

```
Analysis of Variance

Source          DF   Adj SS   Adj MS  F-Value  P-Value
Regression       1   32.699   32.699    19.67    0.002
  Distance       1   32.699   32.699    19.67    0.002
Error            8   13.301    1.663
  Lack-of-Fit    7   11.301    1.614     0.81    0.698
  Pure Error     1    2.000    2.000
Total            9   46.000


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.28941  71.09%     67.47%      57.04%


Coefficients

Term         Coef  SE Coef  T-Value  P-Value   VIF
```

```
Constant    8.098    0.809    10.01    0.000
Distance  -0.3442   0.0776    -4.43    0.002  1.00


Regression Equation

Days = 8.098 - 0.3442 Distance

Variable  Setting
Distance        5

    Fit    SE Fit         95% CI              95% PI
6.37681  0.512485  (5.19502, 7.55860)  (3.17717, 9.57646)
```

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

There is a significant relationship between the number of days absent and the distance to work.

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 5.19502 to 7.5586 or approximately 5.2 to 7.6 days.

64. a. The Minitab output is shown below:

```
Analysis of Variance

Source         DF  Adj SS  Adj MS  F-Value  P-Value
Regression      1  312050  312050    54.75    0.000
  Age           1  312050  312050    54.75    0.000
Error           8   45600    5700
  Lack-of-Fit   3    6150    2050     0.26    0.852
  Pure Error    5   39450    7890
Total           9  357650


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
75.4983  87.25%     85.66%      79.52%


Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant  220.0     58.5     3.76    0.006
Age       131.7     17.8     7.40    0.000  1.00


Regression Equation

Cost = 220.0 + 131.7 Age
```

```
Variable  Setting
Age             4

    Fit   SE Fit        95% CI              95% PI
746.667  29.7769  (678.001, 815.332)  (559.515, 933.818)
```

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

Maintenance cost and age of bus are related.

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

d. The 95% prediction interval is 559.515 to 933.818 or $559.52 to $933.82

65. a. The Minitab output is shown below:

```
Analysis of Variance

Source         DF  Adj SS   Adj MS  F-Value  P-Value
Regression      1  3249.7  3249.72    57.42    0.000
  Hours         1  3249.7  3249.72    57.42    0.000
Error           8   452.8    56.60
  Lack-of-Fit   7   340.3    48.61     0.43    0.828
  Pure Error    1   112.5   112.50
Total           9  3702.5


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
7.52312  87.77%     86.24%      82.23%


Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant   5.85     7.97     0.73    0.484
Hours     0.830    0.109     7.58    0.000  1.00


Regression Equation

Points = 5.85 + 0.830 Hours

Variable  Setting
Hours          95

Fit   SE Fit        95% CI              95% PI
84.6533  3.66780  (76.1953, 93.1112)  (65.3529, 103.954)
```

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

Total points earned is related to the hours spent studying.

c. 84.65 points

d. The 95% prediction interval is 65.3529 to 103.954

66. a. The Minitab output is shown below:

```
              Analysis of Variance

Source          DF  Adj SS  Adj MS  F-Value  P-Value
Regression       1   50.26  50.255     7.08    0.029
  S&P 500        1   50.26  50.255     7.08    0.029
Error            8   56.78   7.098
  Lack-of-Fit    7   45.26   6.466     0.56    0.776
  Pure Error     1   11.52  11.520
Total            9  107.04


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
2.66413  46.95%     40.32%       5.96%


Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant  0.275    0.900     0.31    0.768
S&P 500   0.950    0.357     2.66    0.029  1.00


Regression Equation

Horizon = 0.275 + 0.950 S&P 500
```

The market beta for Horizon is $b_1 = .95$

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. Xerox has higher risk with a market beta of 1.22.

67. a. The Minitab output is shown below:

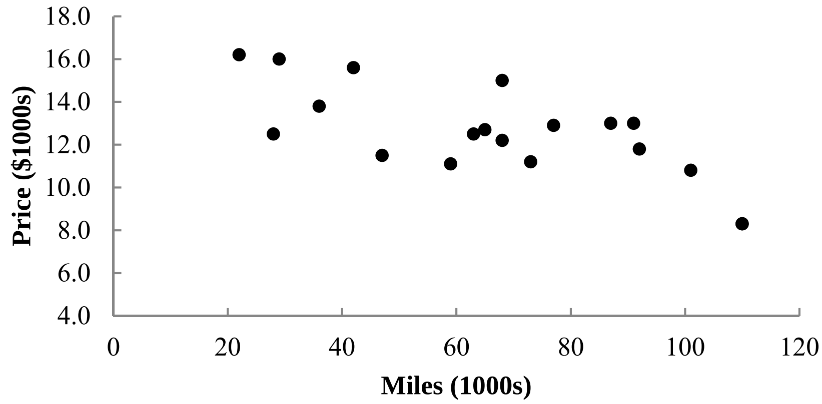
```
Analysis of Variance

Source                   DF  Adj SS   Adj MS  F-Value  P-Value
Regression                1  0.2175  0.21749     4.99    0.038
  Adjusted_Gross Income   1  0.2175  0.21749     4.99    0.038
Error                    18  0.7845  0.04358
Total                    19  1.0020


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.208768  21.71%     17.36%       6.61%


Coefficients

Term                         Coef   SE Coef  T-Value  P-Value   VIF
Constant                   -0.471     0.584    -0.81    0.431
Adjusted_Gross Income  0.000039  0.000017     2.23    0.038  1.00


Regression Equation

Percent_Audited = -0.471 + 0.000039 Adjusted_Gross Income


Variable               Setting
Adjusted_Gross Income    35000


     Fit     SE Fit         95% CI                  95% PI
0.882770  0.0523186  (0.772853, 0.992687)  (0.430602, 1.33494)
```

b. Since the p-value = 0.038 is less than $\alpha = .05$, 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 .772853 to .992687.

68. a.



b. There appears to be a negative relationship between the two variables that can be approximated by a straight line. An argument could also be made that the relationship is perhaps curvilinear because at some point a car has so many miles that its value becomes very small.

c. The Minitab output is shown below.

```
Analysis of Variance

Source            DF  Adj SS  Adj MS  F-Value  P-Value
Regression         1  47.158  47.158    19.85    0.000
  Miles (1000s)    1  47.158  47.158    19.85    0.000
Error             17  40.389   2.376
  Lack-of-Fit     15  36.469   2.431     1.24    0.535
  Pure Error       2   3.920   1.960
Total             18  87.547


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.54138  53.87%     51.15%      41.30%


Coefficients

Term               Coef  SE Coef  T-Value  P-Value   VIF
Constant         16.470    0.949    17.36    0.000
Miles (1000s)  -0.0588   0.0132    -4.46    0.000  1.00


Regression Equation

Price ($1000s) = 16.470 - 0.0588 Miles (1000s)
```

d. Significant relationship: p-value = 0.000 < α = .05.

e. r^2 = .5387; a reasonably good fit considering that the condition of the car is also an important factor in what the price is.

f. The slope of the estimated regression equation is -.0558. Thus, a one-unit increase in the value of x coincides with a decrease in the value of y equal to .0558. Because the data were recorded in thousands, every additional 1000 miles on the car's odometer will result in a $55.80 decrease in the predicted price.

g. The predicted price for a 2007 Camry with 60,000 miles is $\hat{y}$ = 16.47 -.0588(60) = 12.942 or $12,942. Because of other factors, such as condition and whether the seller is a private party or a dealer, this is probably not the price you would offer for the car. But, it should be a good starting point in figuring out what to offer the seller.