Chapter 15
Multiple Regression

Learning Objectives

1. Understand how multiple regression analysis can be used to develop relationships involving one dependent variable and several independent variables.

2. Be able to interpret the coefficients in a multiple regression analysis.

3. Know the assumptions necessary to conduct statistical tests involving the hypothesized regression model.

4. Understand the role of computer packages in performing multiple regression analysis.

5. Be able to interpret and use computer output to develop the estimated regression equation.

6. Be able to determine how good a fit is provided by the estimated regression equation.

7. Be able to test for the significance of the regression equation.

8. Understand how multicollinearity affects multiple regression analysis.

9. Know how residual analysis can be used to make a judgement as to the appropriateness of the model, identify outliers, and determine which observations are influential.

10. Understand how logistic regression is used for regression analyses involving a binary dependent variable.

Solutions:

1. a. b_1 = .5906 is an estimate of the change in y corresponding to a 1 unit change in x_1 when x_2 is held constant.

   b_2 = .4980 is an estimate of the change in y corresponding to a 1 unit change in x_2 when x_1 is held constant.

2. a. The estimated regression equation is

   $\hat{y} = 45.06 + 1.94x_1$

   An estimate of y when $x_1 = 45$ is

   $\hat{y} = 45.06 + 1.94(45) = 132.36$

   b. The estimated regression equation is

   $\hat{y} = 85.22 + 4.32x_2$

   An estimate of y when $x_2 = 15$ is

   $\hat{y} = 85.22 + 4.32(15) = 150.02$

   c. The estimated regression equation is

   $\hat{y} = -18.37 + 2.01x_1 + 4.74x_2$

   An estimate of y when $x_1 = 45$ and $x_2 = 15$ is

   $\hat{y} = -18.37 + 2.01(45) + 4.74(15) = 143.18$

3. a. b_1 = 3.8 is an estimate of the change in y corresponding to a 1 unit change in x_1 when x_2, x_3, and x_4 are held constant.

   b_2 = -2.3 is an estimate of the change in y corresponding to a 1 unit change in x_2 when x_1, x_3, and x_4 are held constant.

   b_3 = 7.6 is an estimate of the change in y corresponding to a 1 unit change in x_3 when x_1, x_2, and x_4 are held constant.

   b_4 = 2.7 is an estimate of the change in y corresponding to a 1 unit change in x_4 when x_1, x_2, and x_3 are held constant.

4. a. $\hat{y} = 25 + 10(15) + 8(10) = 255$; sales estimate: $255,000

   b. Sales can be expected to increase by $10 for every dollar increase in inventory investment when advertising expenditure is held constant. Sales can be expected to increase by $8 for every dollar increase in advertising expenditure when inventory investment is held constant.

5. a. The Minitab output is shown below:

```
The regression equation is
Revenue = 88.6 + 1.60 TVAdv

Predictor        Coef     SE Coef          T        P
Constant       88.638       1.582      56.02    0.000
TVAdv          1.6039      0.4778       3.36    0.015

S = 1.215       R-Sq = 65.3%     R-Sq(adj) = 59.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      16.640      16.640     11.27    0.015
Residual Error     6       8.860       1.477
Total              7      25.500
```

b. The Minitab output is shown below:

```
The regression equation is
Revenue = 83.2 + 2.29 TVAdv + 1.30 NewsAdv

Predictor        Coef     SE Coef          T        P
Constant       83.230       1.574      52.88    0.000
TVAdv          2.2902      0.3041       7.53    0.001
NewsAdv        1.3010      0.3207       4.06    0.010

S = 0.6426      R-Sq = 91.9%     R-Sq(adj) = 88.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      23.435      11.718     28.38    0.002
Residual Error     5       2.065       0.413
Total              7      25.500

Source       DF      Seq SS
TVAdv         1      16.640
NewsAdv       1       6.795
```

c. No, it is 1.60 in part (a) and 2.29 above. In part (b) it represents the marginal change in revenue due to an increase in television advertising with newspaper advertising held constant.

d. Revenue = 83.2 + 2.29(3.5) + 1.30(1.8) = $93.56 or $93,560

6. a. The Minitab output is shown below:

```
The regression equation is
Speed = 49.8 + 0.0151 Weight

Predictor       Coef      SE Coef          T        P
Constant       49.78        19.11       2.61    0.021
Weight      0.015104     0.006005       2.52    0.025


S = 7.000       R-Sq = 31.1%     R-Sq(adj) = 26.2%
```

```
Analysis of Variance

Source          DF        SS        MS        F       P
Regression       1    309.95    309.95     6.33   0.025
Residual Error  14    686.00     49.00
Total           15    995.95
```

b. The Minitab output is shown below:

```
The regression equation is
Speed = 80.5 - 0.00312 Weight + 0.105 Horsepwr

Predictor        Coef     SE Coef         T        P
Constant       80.487       9.139      8.81    0.000
Weight      -0.003122    0.003481     -0.90    0.386
Horsepwr      0.10471     0.01331      7.86    0.000

S = 3.027       R-Sq = 88.0%     R-Sq(adj) = 86.2%

Analysis of Variance

Source             DF        SS        MS        F        P
Regression          2    876.80    438.40    47.83    0.000
Residual Error     13    119.15      9.17
Total              15    995.95
```

c. $\hat{y}$ = 80.5 - 0.00312(2910) + 0.105(296) = 102.5

7. a. The Minitab output is shown below:

```
The regression equation is
Price = 356 - 0.0987 Capacity + 123 Comfort

Predictor        Coef     SE Coef         T        P
Constant        356.1       197.2      1.81    0.114
Capacity     -0.09874     0.04588     -2.15    0.068
Comfort        122.87       21.80      5.64    0.001

S = 51.14       R-Sq = 83.2%     R-Sq(adj) = 78.4%

Analysis of Variance

Source             DF        SS        MS        F        P
Regression          2     90548     45274    17.31    0.002
Residual Error      7     18304      2615
Total               9    108852
```

b. b_1 = -.0987 is an estimate of the change in the price with respect to a 1 cubic inch change in capacity with the comfort rating held constant. b_2 = 123 is an estimate of the change in the price with respect to a 1 unit change in the comfort rating with the capacity held constant.

c. $\hat{y}$ = 356 - .0987(4500) + 123 (4) = 404

8. a. The Minitab output is shown below:

```
The regression equation is
Return = 247 - 32.8 Safety + 34.6 ExpRatio

Predictor        Coef     SE Coef          T        P
Constant        247.4       110.4       2.24    0.039
Safety         -32.84       13.95      -2.35    0.031
ExpRatio        34.59       14.13       2.45    0.026

S = 16.98       R-Sq = 58.2%     R-Sq(adj) = 53.3%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      6823.2      3411.6     11.84    0.001
Residual Error    17      4899.7       288.2
Total             19     11723.0
```

b. $\hat{y} = 247 - 32.8(7.5) + 34.6(2) = 70.2$

9. a. The Minitab output is shown below:

```
The regression equation is
%College = 26.7 - 1.43 Size + 0.0757 SatScore

Predictor        Coef     SE Coef          T        P
Constant        26.71       51.67       0.52    0.613
Size          -1.4298      0.9931      -1.44    0.170
SatScore      0.07574     0.03906       1.94    0.072

S = 12.42       R-Sq = 38.2%     R-Sq(adj) = 30.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      1430.4       715.2      4.64    0.027
Residual Error    15      2312.7       154.2
Total             17      3743.1
```

b. $\hat{y} = 26.7 - 1.43(20) + 0.0757(1000) = 73.8$

Estimate is 73.8%

10. a. The Minitab output is shown below:

```
The regression equation is
Revenue = 33.3 + 7.98 Cars

Predictor       Coef     SE Coef          T        P
Constant       33.34       83.08       0.40    0.695
Cars          7.9840      0.6323      12.63    0.000

S = 226.7       R-Sq = 92.5%     R-Sq(adj) = 91.9%
```

```
Analysis of Variance

Source          DF        SS        MS       F       P
Regression       1   8192067   8192067  159.44   0.000
Residual Error  13    667936     51380
Total           14   8860003
```

b. An increase of 1000 cars in service will result in an increase in revenue of \$7.98 million.

c. The Minitab output is shown below:

```
The regression equation is
Revenue = 106 + 8.94 Cars - 0.191 Location

Predictor       Coef     SE Coef         T        P
Constant      105.97       85.52      1.24    0.239
Cars          8.9427      0.7746     11.55    0.000
Location     -0.1914      0.1026     -1.87    0.087

S = 207.7       R-Sq = 94.2%     R-Sq(adj) = 93.2%

Analysis of Variance

Source          DF        SS        MS       F       P
Regression       2   8342186   4171093   96.66   0.000
Residual Error  12    517817     43151
Total           14   8860003
```

11. a. SSE = SST - SSR = 6,724.125 - 6,216.375 = 507.75

b. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{6{,}216.375}{6{,}724.125} = .924$

c. $R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.924)\frac{10-1}{10-2-1} = .902$

d. The estimated regression equation provided an excellent fit.

12. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{14{,}052.2}{15{,}182.9} = .926$

b. $R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.926)\frac{10-1}{10-2-1} = .905$

c. Yes; after adjusting for the number of independent variables in the model, we see that 90.5% of the variability in y has been accounted for.

13. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{1760}{1805} = .975$

b. $R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.975)\frac{30-1}{30-4-1} = .971$

c. The estimated regression equation provided an excellent fit.

14. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{12,000}{16,000} = .75$

b. $R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-.25\frac{9}{7} = .68$

c. The adjusted coefficient of determination shows that 68% of the variability has been explained by the two independent variables; thus, we conclude that the model does not explain a large amount of variability.

15. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{23.435}{25.5} = .919$

$R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.919)\frac{8-1}{8-2-1} = .887$

b. Multiple regression analysis is preferred since both R^2 and R_a^2 show an increased percentage of the variability of y explained when both independent variables are used.

16. Note: the Minitab output is shown with the solution to Exercise 6.

a. No; R-Sq = 31.1%

b. Multiple regression analysis is preferred since both R-Sq and R-Sq(adj) show an increased percentage of the variability of y explained when both independent variables are used.

17. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{1430.4}{3743.1} = .382$

$R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.382)\frac{18-1}{18-2-1} = .30$

b. The fit is not very good

18. Note: The Minitab output is shown with the solution to Exercise 10.

a. R-Sq = 94.2% R-Sq(adj) = 93.2%

b. The fit is very good.

19. a. $MSR = SSR/p = 6{,}216.375/2 = 3{,}108.188$

$$MSE = \frac{SSE}{n-p-1} = \frac{507.75}{10-2-1} = 72.536$$

b. $F = MSR/MSE = 3{,}108.188/72.536 = 42.85$

Using F table (2 degrees of freedom numerator and 7 denominator), p-value is less than .01

Because p-value $\leq \alpha = .05$, the overall model is significant.

c. $t = .5906/.0813 = 7.26$

Using t table (7 degrees of freedom), area in tail is less than .005; p-value is less than .01

Because p-value $\leq \alpha$, β_1 is significant.

d. $t = .4980/.0567 = 8.78$

Using t table (7 degrees of freedom), area in tail is less than .005; p-value is less than .01

Because p-value $\leq \alpha$, β_2 is significant.

20. A portion of the Minitab output is shown below.

```
The regression equation is
Y = - 18.4 + 2.01 X1 + 4.74 X2

Predictor        Coef     SE Coef          T        P
Constant       -18.37       17.97      -1.02    0.341
X1             2.0102      0.2471       8.13    0.000
X2             4.7378      0.9484       5.00    0.002

S = 12.71       R-Sq = 92.6%     R-Sq(adj) = 90.4%

Analysis of Variance

Source             DF          SS          MS         F        P
Regression          2     14052.2      7026.1     43.50    0.000
Residual Error      7      1130.7       161.5
Total               9     15182.9
```

a. Since the p-value corresponding to $F = 43.50$ is $.000 < \alpha = .05$, we reject H_0: $\beta_1 = \beta_2 = 0$; there is a significant relationship.

b. Since the p-value corresponding to $t = 8.13$ is $.000 < \alpha = .05$, we reject H_0: $\beta_1 = 0$; β_1 is significant.

c. Since the p-value corresponding to $t = 5.00$ is $.002 < \alpha = .05$, we reject H_0: $\beta_2 = 0$; β_2 is significant.

21. a. In the two independent variable case the coefficient of x_1 represents the expected change in y corresponding to a one unit increase in x_1 when x_2 is held constant. In the single independent variable case the coefficient of x_1 represents the expected change in y corresponding to a one unit increase in x_1.

b. Yes. If x_1 and x_2 are correlated one would expect a change in x_1 to be accompanied by a change in x_2.

22. a. SSE = SST - SSR = 16000 - 12000 = 4000

$$s^2 = \frac{\text{SSE}}{n - p - 1} = \frac{4000}{7} = 571.43$$

$$\text{MSR} = \frac{\text{SSR}}{p} = \frac{12000}{2} = 6000$$

b. F = MSR/MSE = 6000/571.43 = 10.50

Using F table (2 degrees of freedom numerator and 7 denominator), p-value is less than .01

Because p-value $\leq \alpha$, we reject H_0. There is a significant relationship among the variables.

23. a. F = 28.38

Using F table (2 degrees of freedom numerator and 7 denominator), p-value is less than .01

Actual p-value = .002

Because p-value $\leq \alpha$, there is a significant relationship.

b. t = 7.53

Using t table (7 degrees of freedom), area in tail is less than .005; p-value is less than .01

Actual p-value = .001

Because p-value $\leq \alpha, \beta_1$ is significant and x_1 should not be dropped from the model.

c. t = 4.06

Actual p-value = .010

Because p-value $\leq \alpha$, β_2 is significant and x_2 should not be dropped from the model.

24. Note: The Minitab output is shown in part (b) of Exercise 6

a. F = 47.83

p-value = .000

Because p-value $\leq \alpha$, we can reject H_0.

b. For Weight:

H_0: $\beta_1 = 0$ H_a: $\beta_1 \neq 0$

p-value = .386

Because p-value $> \alpha$, we cannot reject H_0

For Horsepower:

H_0: $\beta_2 = 0$ H_a: $\beta_2 \neq 0$

p-value = .000

Because p-value $\leq \alpha$, we can reject H_0

25. a. The Minitab output is shown below:

```
The regression equation is
Rating = 0.345 + 0.255 TradeEx + 0.132 Use + 0.459 Range

Predictor        Coef     SE Coef         T        P
Constant       0.3451      0.5307      0.65    0.540
TradeEx       0.25482     0.08556      2.98    0.025
Use            0.1325      0.1404      0.94    0.382
Range          0.4585      0.1232      3.72    0.010

S = 0.2431      R-Sq = 88.6%     R-Sq(adj) = 82.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     2.74541     0.91514     15.49    0.003
Residual Error     6     0.35459     0.05910
Total              9     3.10000
```

b. Because the p-value = .003 < α = .05, there is a significant relationship.

c. For TradeEx: Because the p-value = .025 < α = .05, TradeEx is significant.

For Use: Because the p-value = .382 > α = .05, Use is not significant.

For Range: Because the p-value = .010 < α = .05, Range is significant.

The Minitab output after removing Use is shown below:

```
The regression equation is
Rating = 0.672 + 0.264 TradeEx + 0.485 Range

Predictor        Coef     SE Coef         T        P
Constant       0.6718      0.3989      1.68    0.136
TradeEx       0.26406     0.08432      3.13    0.017
Range          0.4853      0.1189      4.08    0.005

S = 0.2412      R-Sq = 86.9%     R-Sq(adj) = 83.1%
```

```
Analysis of Variance

Source            DF        SS        MS        F       P
Regression         2    2.6928    1.3464    23.15   0.001
Residual Error     7    0.4072    0.0582
Total              9    3.1000
```

The coefficient of determination for the estimated regression equation developed in part (a) is .886. After the removal of Use, the coefficient of determination is .869. There is very little difference in the fit provided by the two estimated regression equations. But, because Use is not significant, this result is as expected.

26. Note: The Minitab output is shown with the solution to Exercise 10.

a. Since the p-value corresponding to $F = 96.66$ is $0.000 < \alpha = .05$, there is a significant relationship among the variables.

b. For Cars: Since the p-value $= 0.000 < \alpha = 0.05$, Cars is significant

c. For Location: Since the p-value $= 0.087 > \alpha = 0.05$, Location is not significant

27. a. $\hat{y} = 29.1270 + .5906(180) + .4980(310) = 289.8150$

b. The point estimate for an individual value is $\hat{y} = 289.8150$, the same as the point estimate of the mean value.

28. a. Using Minitab, the 95% confidence interval is 132.16 to 154.16.

b. Using Minitab, the 95% prediction interval is 111.13 to 175.18.

29. a. $\hat{y} = 83.2 + 2.29(3.5) + 1.30(1.8) = 93.555$ or \$93,555

Note: In Exercise 5b, the Minitab output also shows that $b_0 = 83.230$, $b_1 = 2.2902$, and $b_2 = 1.3010$; hence, $\hat{y} = 83.230 + 2.2902x_1 + 1.3010x_2$. Using this estimated regression equation, we obtain

$\hat{y} = 83.230 + 2.2902(3.5) + 1.3010(1.8) = 93.588$ or \$93,588

The difference (\$93,588 - \$93,555 = \$33) is simply due to the fact that additional significant digits are used in the computations. From a practical point of view, however, the difference is not enough to be concerned about. In practice, a computer software package is always used to perform the computations and this will not be an issue.

The Minitab output is shown below:

```
    Fit  Stdev.Fit         95% C.I.          95% P.I.
 93.588      0.291   ( 92.840, 94.335)  ( 91.774, 95.401)
```

Note that the value of FIT ($\hat{y}$) is 93.588.

b. Confidence interval estimate: 92.840 to 94.335 or \$92,840 to \$94,335

c. Prediction interval estimate: 91.774 to 95.401 or $91,774 to $95,401

30. a. Since weight is not statistically significant (see Exercise 24), we will use an estimated regression equation which uses only Horsepower to predict the speed at 1/4 mile. The Minitab output is shown below:

```
The regression equation is
Speed = 72.6 + 0.0968 Horsepwr

Predictor        Coef     SE Coef          T        P
Constant       72.650       2.655      27.36    0.000
Horsepwr     0.096756    0.009865       9.81    0.000

S = 3.006       R-Sq = 87.3%     R-Sq(adj) = 86.4%

Analysis of Variance

Source           DF          SS          MS         F         P
Regression        1      869.43      869.43     96.21     0.000
Residual Error   14      126.52        9.04
Total            15      995.95

Unusual Observations
Obs   Horsepwr    Speed       Fit    SE Fit    Residual    St Resid
  2        290  108.000   100.709     0.814       7.291       2.52R
  6        450  116.200   116.190     2.036       0.010       0.00 X

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

The output shows that the point estimate is a speed of 101.290 miles per hour.

b. The 95% confidence interval is 99.490 to 103.089 miles per hour.

c. The 95% prediction interval is 94.596 to 107.984 miles per hour.

31. a. Using Minitab the 95% confidence interval is 58.37% to 75.03%.

b. Using Minitab the 95% prediction interval is 35.24% to 90.59%.

32. a. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ where

$x_2 = 0$ if level 1 and 1 if level 2

b. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2(0) = \beta_0 + \beta_1 x_1$

c. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2(1) = \beta_0 + \beta_1 x_1 + \beta_2$

d. $\beta_2 = E(y \mid \text{level 2}) - E(y \mid \text{level 1})$

β_1 is the change in $E(y)$ for a 1 unit change in x_1 holding x_2 constant.

33. a. two

b. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ where

x_2	x_3	Level
0	0	1
1	0	2
0	1	3

c. $E(y \mid \text{level 1}) = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(0) = \beta_0 + \beta_1 x_1$

$E(y \mid \text{level 2}) = \beta_0 + \beta_1 x_1 + \beta_2(1) + \beta_3(0) = \beta_0 + \beta_1 x_1 + \beta_2$

$E(y \mid \text{level 3}) = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(0) = \beta_0 + \beta_1 x_1 + \beta_3$

$\beta_2 = E(y \mid \text{level 2}) - E(y \mid \text{level 1})$

$\beta_3 = E(y \mid \text{level 3}) - E(y \mid \text{level 1})$

β_1 is the change in $E(y)$ for a 1 unit change in x_1 holding x_2 and x_3 constant.

34. a. $15,300

b. Estimate of sales = 10.1 - 4.2(2) + 6.8(8) + 15.3(0) = 56.1 or $56,100

c. Estimate of sales = 10.1 - 4.2(1) + 6.8(3) + 15.3(1) = 41.6 or $41,600

35. a. Let Type = 0 if a mechanical repair
Type = 1 if an electrical repair

The Minitab output is shown below:

```
The regression equation is
Time = 3.45 + 0.617 Type

Predictor        Coef     SE Coef          T        P
Constant       3.4500      0.5467       6.31    0.000
Type           0.6167      0.7058       0.87    0.408

S = 1.093       R-Sq = 8.7%      R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         1       0.913       0.913       0.76    0.408
Residual Error     8       9.563       1.195
Total              9      10.476
```

b. The estimated regression equation did not provide a good fit. In fact, the p-value of .408 shows that the relationship is not significant for any reasonable value of α.

c. Person = 0 if Bob Jones performed the service and Person = 1 if Dave Newton performed the service. The Minitab output is shown below:

```
The regression equation is
Time = 4.62 - 1.60 Person

Predictor        Coef     SE Coef          T        P
Constant       4.6200      0.3192      14.47    0.000
Person        -1.6000      0.4514      -3.54    0.008

S = 0.7138      R-Sq = 61.1%     R-Sq(adj) = 56.2%

Analysis of Variance

Source           DF          SS          MS         F        P
Regression        1      6.4000      6.4000     12.56    0.008
Residual Error    8      4.0760      0.5095
Total             9     10.4760
```

d. We see that 61.1% of the variability in repair time has been explained by the repair person that performed the service; an acceptable, but not good, fit.

36. a. The Minitab output is shown below:

```
The regression equation is
Time = 1.86 + 0.291 Months + 1.10 Type - 0.609 Person

Predictor        Coef     SE Coef          T        P
Constant       1.8602      0.7286       2.55    0.043
Months        0.29144     0.08360       3.49    0.013
Type           1.1024      0.3033       3.63    0.011
Person        -0.6091      0.3879      -1.57    0.167

S = 0.4174      R-Sq = 90.0%     R-Sq(adj) = 85.0%

Analysis of Variance

Source           DF          SS          MS         F        P
Regression        3      9.4305      3.1435     18.04    0.002
Residual Error    6      1.0455      0.1743
Total             9     10.4760
```

b. Since the p-value corresponding to F = 18.04 is .002 < α = .05, the overall model is statistically significant.

c. The p-value corresponding to t = -1.57 is .167 > α = .05; thus, the addition of Person is not statistically significant. Person is highly correlated with Months (the sample correlation coefficient is -.691); thus, once the effect of Months has been accounted for, Person will not add much to the model.

37. a. Let Position = 0 if a guard
Position = 1 if an offensive tackle.

b. The Minitab output is shown below:

```
The regression equation is
Rating = 11.2 + 0.732 Position + 0.0222 Weight - 2.28 Speed

Predictor        Coef     SE Coef          T        P
Constant       11.223       4.523       2.48    0.022
Position       0.7324      0.2893       2.53    0.019
Weight        0.02219     0.01039       2.14    0.045
Speed         -2.2775      0.9290      -2.45    0.023

S = 0.6936      R-Sq = 47.5%     R-Sq(adj) = 40.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      9.1562      3.0521      6.35    0.003
Residual Error    21     10.1014      0.4810
Total             24     19.2576
```

c. Since the p-value corresponding to $F = 6.35$ is $.003 < \alpha = .05$, there is a significant relationship between rating and the independent variables.

d. The value of R-Sq (adj) is 40.1%; the estimated regression equation did not provide a very good fit.

e. Since the p-value for Position is $t = 2.53 < \alpha = .05$, position is a significant factor in the player's rating.

f. $\hat{y} = 11.2 + .732(1) + .0222(300) - 2.28(5.1) = 6.96$

38. a. The Minitab output is shown below:

```
The regression equation is
Risk = - 91.8 + 1.08 Age + 0.252 Pressure + 8.74 Smoker

Predictor        Coef     SE Coef          T        P
Constant       -91.76       15.22      -6.03    0.000
Age            1.0767      0.1660       6.49    0.000
Pressure      0.25181     0.04523       5.57    0.000
Smoker          8.740       3.001       2.91    0.010

S = 5.757       R-Sq = 87.3%     R-Sq(adj) = 85.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      3660.7      1220.2     36.82    0.000
Residual Error    16       530.2        33.1
Total             19      4190.9
```

b. Since the p-value corresponding to $t = 2.91$ is $.010 < \alpha = .05$, smoking is a significant factor.

c. Using Minitab, the point estimate is 34.27; the 95% prediction interval is 21.35 to 47.18. Thus, the probability of a stroke (.2135 to .4718 at the 95% confidence level) appears to be quite high. The physician would probably recommend that Art quit smoking and begin some type of treatment designed to reduce his blood pressure.

39. a. The Minitab output is shown below:

```
The regression equation is
Y = 0.20 + 2.60 X

Predictor        Coef     SE Coef          T        P
Constant        0.200       2.132       0.09    0.931
X              2.6000      0.6429       4.04    0.027

S = 2.033       R-Sq = 84.5%     R-Sq(adj) = 79.3%

Analysis of Variance

Source           DF          SS          MS         F        P
Regression        1      67.600      67.600     16.35    0.027
Residual Error    3      12.400       4.133
Total             4      80.000
```

b. Using Minitab we obtained the following values:

x_i	y_i	$\hat{y}_i$	Standardized Residual
1	3	2.8	.16
2	7	5.4	.94
3	5	8.0	-1.65
4	11	10.6	.24
5	14	13.2	.62

The point (3,5) does not appear to follow the trend of remaining data; however, the value of the standardized residual for this point, -1.65, is not large enough for us to conclude that (3, 5) is an outlier.

c. Using Minitab, we obtained the following values:

x_i	y_i	Studentized Deleted Residual
1	3	.13
2	7	.91
3	5	- 4.42
4	11	.19
5	14	.54

$t_{.025} = 4.303$ ($n - p - 2 = 5 - 1 - 2 = 2$ degrees of freedom)

Since the studentized deleted residual for (3, 5) is $-4.42 < -4.303$, we conclude that the 3rd observation is an outlier.

40. a. The Minitab output is shown below:

```
The regression equation is
Y = -53.3 + 3.11 X

 Predicator           Coef    SE Coef         T       p
 Constant          -53.280      5.786     -9.21   0.003
 X                  3.1100     0.2016     15.43   0.001

S = 2.851   R-sq = 98.8%   R-sq (adj) = 98.3%

Analysis of Variance

SOURCE             DF        SS        MS        F       p
Regression          1    1934.4    1934.4   238.03   0.001
Residual Error      3      24.4       8.1
Total               4    1598.8
```

b. Using the Minitab we obtained the following values:

x_i	y_i	Studentized Deleted Residual
22	12	-1.94
24	21	-.12
26	31	1.79
28	35	.40
40	70	-1.90

$t_{.025} = 4.303$ ($n - p - 2 = 5 - 1 - 2 = 2$ degrees of freedom)

Since none of the studentized deleted residuals are less than -4.303 or greater than 4.303, none of the observations can be classified as an outlier.

c. Using Minitab we obtained the following values:

x_i	y_i	h_i
22	12	.38
24	21	.28
26	31	.22
28	35	.20
40	70	.92

The critical value is

$$\frac{3(p+1)}{n} = \frac{3(1+1)}{5} = 1.2$$

Since none of the values exceed 1.2, we conclude that there are no influential observations in the data.

d. Using Minitab we obtained the following values:

x_i	y_i	D_i
22	12	.60
24	21	.00
26	31	.26
28	35	.03
40	70	11.09

Since $D_5 = 11.09 > 1$ (rule of thumb critical value), we conclude that the fifth observation is influential.

41. a. The Minitab output appears in the solution to part (b) of Exercise 5; the estimated regression equation is:

Revenue = 83.2 + 2.29 TVAdv + 1.30 NewsAdv

b. Using Minitab we obtained the following values:

$\hat{y}_i$	Standardized Residual
96.63	-1.62
90.41	-1.08
94.34	1.22
92.21	- .37
94.39	1.10
94.24	- .40
94.42	-1.12
93.35	1.08

With the relatively few observations, it is difficult to determine if any of the assumptions regarding the error term have been violated. For instance, an argument could be made that there does not appear to be any pattern in the plot; alternatively an argument could be made that there is a curvilinear pattern in the plot.

c. The values of the standardized residuals are greater than -2 and less than +2; thus, using test, there are no outliers. As a further check for outliers, we used Minitab to compute the following studentized deleted residuals:

Observation	Studentized Deleted Residual
1	-2.11
2	-1.10
3	1.31
4	- .33
5	1.13
6	- .36
7	-1.16
8	1.10

$t_{.025} = 2.776$ ($n - p - 2 = 8 - 2 - 2 = 4$ degrees of freedom)

Since none of the studentized deleted residuals is less than -2.776 or greater than 2.776, we conclude that there are no outliers in the data.

d. Using Minitab we obtained the following values:

Observation	h_i	D_i
1	.63	1.52
2	.65	.70
3	.30	.22
4	.23	.01
5	.26	.14
6	.14	.01
7	.66	.81
8	.13	.06

The critical average value is

$$\frac{3(p+1)}{n} = \frac{3(2+1)}{8} = 1.125$$

Since none of the values exceed 1.125, we conclude that there are no influential observations. However, using Cook's distance measure, we see that $D_1 > 1$ (rule of thumb critical value); thus, we conclude the first observation is influential. Final Conclusion: observations 1 is an influential observation.

42. a. The Minitab output is shown below:

```
The regression equation is
Speed = 71.3 + 0.107 Price + 0.0845 Horsepwr

Predictor        Coef     SE Coef          T        P
Constant       71.328       2.248      31.73    0.000
Price         0.10719     0.03918       2.74    0.017
Horsepwr     0.084496    0.009306       9.08    0.000

S = 2.485       R-Sq = 91.9%     R-Sq(adj) = 90.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      915.66      457.83     74.12    0.000
Residual Error    13       80.30        6.18
Total             15      995.95

Source       DF      Seq SS
Price         1      406.39
Horsepwr      1      509.27

Unusual Observations
Obs      Price     Speed       Fit     SE Fit   Residual   St Resid
  2       93.8   108.000   105.882      2.007      2.118     1.45 X

X denotes an observation whose X value gives it large influence.
```

b. The standardized residual plot is shown below. There appears to be a very unusual trend in the standardized residuals.

```
        -
        -                         xx         x
     1.2+
        -                              x
SRES1   -            x
        -                  x
        -            x
     0.0+          x   x
        -                                                      x
        -          x   x x
        -
        -
    -1.2+    x
        -
        -                                   x
        -                               x
        -
          --------+---------+---------+---------+---------+--------FITS1
               90.0      96.0     102.0     108.0     114.0
```

c. The Minitab output shown in part (a) did not identify any observations with a large standardized residual; thus, there does not appear to be any outliers in the data.

d. The Minitab output shown in part (a) identifies observation 2 as an influential observation.

43. a. The Minitab output is shown below:

```
The regression equation is
%College = - 26.6 + 0.0970 SatScore

Predictor        Coef     SE Coef          T        P
Constant       -26.61       37.22      -0.72    0.485
SatScore      0.09703     0.03734       2.60    0.019

S = 12.83       R-Sq = 29.7%     R-Sq(adj) = 25.3%
Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      1110.8      1110.8      6.75    0.019
Residual Error    16      2632.3       164.5
Total             17      3743.1

Unusual Observations
Obs   SatScore   %College    Fit     SE Fit    Residual    St Resid
  3        716      40.00  42.86      10.79       -2.86     -0.41 X

X denotes an observation whose X value gives it large influence.
```

b. The Minitab output shown in part a identifies observation 3 as an influential observation.

c. The Minitab output appears in the solution to Exercise 9; the estimates regression equation is

%College = 26.7 - 1.43 Size + 0.0757 SATScore

d. The following Minitab output was also provided as part of the regression output for part c.

Unusual Observations

Obs.	Size	%College	Fit	Stdev.Fit	Residual	St.Resid
3	30.0	40.0	38.04	10.97	1.96	0.34 X

X denotes an obs. whose X value gives it large influence.

Observation 3 is still identified as an influential observation.

44. a. $$E(y) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$$

b. It is an estimate of the probability that a customer that does not have a Simmons credit card will make a purchase.

c. A portion of the Minitab binary logistic regression output follows:

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-0.9445	0.3150	-3.00	0.003			
Card	1.0245	0.4235	2.42	0.016	2.79	1.21	6.39

```
Log-Likelihood = -64.265
Test that all slopes are zero: G = 6.072, DF = 1, P-Value = 0.014
```

Thus, the estimated logit is $\hat{g}(x)$ = -0.9445 + 1.0245x

d. For customers that do not have a Simmons credit card (x = 0)

$\hat{g}(0)$ = -0.9445 + 1.245(0) = 0.9445

and

$$\hat{y} = \frac{e^{\hat{g}(0)}}{1 + e^{\hat{g}(0)}} = \frac{e^{-0.9445}}{1 + e^{-0.9445}} = \frac{0.3889}{1 + 0.3889} = 0.28$$

For customers that have a Simmons credit card (x = 1)

$\hat{g}(1)$ = -0.9445 + 1.245(1) = 0.0800

and

$$\hat{y} = \frac{e^{\hat{g}(1)}}{1 + e^{\hat{g}(1)}} = \frac{e^{0.08}}{1 + e^{0.08}} = \frac{1.0833}{1 + 1.0833} = 0.52$$

e. Using the Minitab output shown in part (c), the estimated odds ratio is 2.79. We can conclude that the estimated odds of making a purchase for customers who have a Simmons credit card are 2.79 times greater than the estimated odds of making a purchase for customers that do not have a Simmons credit card.

45. a. $\text{odds} = \dfrac{.3413}{1-.3413} = .4584$

b. $\text{odds}_1 = \dfrac{.5790}{1-.5790} = 1.3753$

odds_0 = .4584 (from part (a))

$\text{odds ratio} = \dfrac{\text{odds}_1}{\text{odds}_0} = \dfrac{1.3753}{.4584} = 3.00$

c. The odds ratio for x_2 computed holding annual spending constant at \$2000 is also 3.00. This shows that the odds ratio for x_2 is independent of the value of x_1.

46. a. $E(y) = \dfrac{e^{\beta_0+\beta x}}{1+e^{\beta_0+\beta x}}$

b. A portion of the Minitab binary logistic regression output follows:

```
Logistic Regression Table
                                                         Odds        95% CI
Predictor        Coef    SE Coef         Z      P       Ratio     Lower    Upper
Constant      -2.6335     0.7985     -3.30  0.001
Balance       0.22018    0.09002      2.45  0.014        1.25      1.04     1.49

Log-Likelihood = -25.813
Test that all slopes are zero: G = 9.460, DF = 1, P-Value = 0.002
```

Thus, the estimated logistic regression equation is

$$E(y) = \frac{e^{-2.6355+0.22018x}}{1+e^{-2.6355+0.22018x}}$$

c. Significant result: the *p*-value corresponding to the *G* test statistic is 0.0002.

d. For an average monthly balance of \$1000, $x = 10$

$$E(y) = \frac{e^{-2.6355+0.22018x}}{1+e^{-2.6355+0.22018x}} = \frac{e^{-2.6355+0.22018(10)}}{1+e^{-2.6355+0.22018(10)}} = \frac{e^{-0.4317}}{1+e^{-0.4317}} = \frac{0.6494}{1.6494} = 0.39$$

Thus, an estimate of the probability that customers with an average monthly balance of \$1000 will sign up for direct payroll deposit is 0.39.

e. Repeating the calculations in part (d) using various values for x, a value of $x = 12$ or an average monthly balance of approximately \$1200 is required to achieve this level of probability.

f. Using the Minitab output shown in part (b), the estimated odds ratio is 1.25. Because values of x are measured in hundreds of dollars, the estimated odds of signing up for payroll direct deposit for customers that have an average monthly balance of \$600 is 1.25 times greater than the estimated odds of signing up for payroll direct deposit for customers that have an average monthly balance of \$500. Moreover, this interpretation is true for any one hundred dollar increment in the average monthly balance.

47. a. $E(y)=\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2}}$

b. For a given GPA, it is an estimate of the probability that a student who did not attend the orientation program will return to Lakeland for the sophomore year.

c. A portion of the Minitab binary logistic regression output follows:

```
Logistic Regression Table
                                                    Odds       95% CI
Predictor       Coef    SE Coef        Z      P    Ratio    Lower    Upper
Constant      -6.893      1.747    -3.94 0.000
GPA           2.5388     0.6729     3.77 0.000     12.66     3.39    47.35
Program       1.5608     0.5631     2.77 0.006      4.76     1.58    14.36

Log-Likelihood = -40.169
Test that all slopes are zero: G = 47.869, DF = 2, P-Value = 0.000
```

Thus, the estimated logit is $\hat{g}(x_1,x_2)=-6.893+2.5388x_1+1.5608x_2$

d. Significant result: the *p*-value corresponding to the *G* test statistic is 0.0000.

e. Both variables are significant at $\alpha = .01$: the *p*-value for x_1 is 0.000 and the *p*-value for x_2 is 0.006

f. For x_1 =2.5 and x_2 = 0

$\hat{g}$ (2.5, 0) = -6.893 + 2.5388(2.5) + 1.5608(0) = -0.5460

and

$$\hat{y}=\frac{e^{\hat{g}(2.5,0)}}{1+e^{\hat{g}(2.5,0)}}=\frac{e^{-0.5460}}{1+e^{-0.5460}}=\frac{0.5793}{1+0.5793}=0.37$$

For x_1 =2.5 and x_2 = 1

$\hat{g}$ (2.5, 1) = -6.893 + 2.5388(2.5) + 1.5608(1) = 1.0148

and

$$\hat{y}=\frac{e^{\hat{g}(2.5,1)}}{1+e^{\hat{g}(2.5,1)}}=\frac{e^{1.0148}}{1+e^{1.0148}}=\frac{2.7588}{1+2.7588}=0.73$$

g. From the Minitab output in part (c) we see that the estimated odds ratio is 4.76 for the orientation program. This means that the odds of students who attended the orientation program continuing are 4.76 times greater than for students who did not attend the program.

h. We recommend making the orientation program required. From part (e), we see that the odds of continuing are much higher for students who have attended the orientation program.

48. a. $E(y) = \dfrac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$

b. A portion of the Minitab binary logistic regression output follows:

```
Logistic Regression Table
                                                          Odds        95% CI
Predictor       Coef    SE Coef         Z      P     Ratio     Lower    Upper
Constant      -2.805      1.432     -1.96 0.050
Price         1.1492     0.5143      2.23 0.025       3.16      1.15     8.65

Log-Likelihood = -8.200
Test that all slopes are zero: G = 9.465, DF = 1, P-Value = 0.002
```

Thus, the estimated logit is $\hat{g}(x) = \text{-}2.805 + 1.1492x$

c. For chocolates that have a price per serving of $4.00

$\hat{g}(4) = \text{-}2.805 + 1.1492(4) = 1.7918$

and

$$\hat{y} = \frac{e^{\hat{g}(4)}}{1 + e^{\hat{g}(4)}} = \frac{e^{1.7918}}{1 + e^{1.7918}} = \frac{6.0002}{1 + 6.0002} = 0.86$$

d. Using the Minitab output shown in part (b), the estimated odds ratio is 3.16. We can conclude that the estimated odds of having a quality rating of very good or excellent for a chocolate that has a price of $4.00 per serving is 3.16 times greater than the estimated odds for a chocolate with a price of $3.00 per serving. Moreover, this interpretation is true for any one dollar difference in the price per serving.

49. a. The expected increase in final college grade point average corresponding to a one point increase in high school grade point average is .0235 when SAT mathematics score does not change. Similarly, the expected increase in final college grade point average corresponding to a one point increase in the SAT mathematics score is .00486 when the high school grade point average does not change.

b. $\hat{y}$ = -1.41 + .0235(84) + .00486(540) = 3.19

50. a. Job satisfaction can be expected to decrease by 8.69 units with a one unit increase in length of service if the wage rate does not change. A dollar increase in the wage rate is associated with a 13.5 point increase in the job satisfaction score when the length of service does not change.

b. $\hat{y}$ = 14.4 - 8.69(4) + 13.5(6.5) = 67.39

51. a. The computer output with the missing values filled in is as follows:

```
The regression equation is

Y = 8.103 + 7.602 X1 + 3.111 X2


Predictor        Coef     SE Coef         T
Constant        8.103       2.667      3.04
X1              7.602       2.105      3.61
X2              3.111       0.613      5.08

S = 3.35     R-sq = 92.3%     R-sq (adj) = 91.0%

Analysis of Variance

SOURCE             DF          SS           MS         F
Regression          2        1612          806     71.82
Residual Error     12      134,67      11.2225
Total              14     1746.67
```

b. Using t table (12 degrees of freedom), area in tail corresponding to $t = 3.61$ is less than .005; p-value is less than .01

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

Using t table (12 degrees of freedom), area in tail corresponding to $t = 5.08$ is less than .005; p-value is less than .01

Because p-value $\leq \alpha$, reject $H_0 : \beta_2 = 0$

c. See computer output.

d. $R_a^2 = 1 - (1 - .923)\frac{14}{12} = .91$

52. a. The regression equation is

```
Y = -1.41 + .0235 X1 + .00486 X2

Predictor         Coef       SE Coef           T
Constant       -1.4053        0.4848       -2.90
X1            0.023467      0.008666        2.71
X2              .00486      0.001077        4.51

S = 0.1298   R-sq = 93.7%     R-sq (adj) = 91.9%

Analysis of Variance

SOURCE            DF          SS           MS         F
Regression         2     1.76209         .881     52.44
Residual Error     7       .1179        .0168
Total              9     1.88000
```

b. Using F table (2 degrees of freedom numerator and 7 degrees of freedom denominator), p-value is less than .01

Because p-value $\leq \alpha$, there is a significant relationship.

c. $R^2 = \frac{\text{SSR}}{\text{SST}} = .937$

$R_a^2 = 1 - (1 - .937)\frac{9}{7} = .919$

good fit

d. $t_{.025} = 2.365$ (7 DF)

for β_1: p-value is between .02 and .05; reject H_0: $\beta_1 = 0$

for β_2: p-value is less than .01; reject H_0: $\beta_2 = 0$

53. a. The regression equation is

```
Y = 14.4 - 8.69 X1 + 13.52 X2

Predictor         Coef       SE Coef            T
Constant        14.448         8.191         1.76
X1               -8.69         1.555        -5.59
X2              13.517         2.085         6.48

S = 3.773     R-sq = 90.1%      R-sq (adj) = 86.1%

Analysis of Variance

SOURCE             DF          SS           MS          F
Regression          2      648.83      324.415      22.79
Residual Error      5       71.17       14.234
Total               7      720.00
```

b. $F_{.05} = 5.79$ (5 DF)

$F = 22.79 > F_{.05}$; significant relationship.

c. $R^2 = \frac{\text{SSR}}{\text{SST}} = .901$

$R_a^2 = 1 - (1 - .901)\frac{7}{5} = .861$

good fit

d. $t_{.025} = 2.571$ (5 DF)

for β_1: $t = -5.59 < -2.571$; reject $H_0 : \beta_1 = 0$

for β_2: $t = 6.48 > 2.571$; reject $H_0 : \beta_2 = 0$

54. a. The Minitab output is shown below:

```
The regression equation is
Score = 50.6 + 1.56 RecRes

Predictor        Coef     SE Coef          T        P
Constant       50.610       2.064      24.52    0.000
RecRes         1.5621      0.4226       3.70    0.002

S = 6.339       R-Sq = 43.1%     R-Sq(adj) = 40.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      548.92      548.92     13.66    0.002
Residual Error    18      723.29       40.18
Total             19     1272.21
```

b. The fit provided by the estimated regression equation is not that good; only 43.1% of the variability in Score is explained by the RecRes.

c. The Minitab output is shown below:

```
The regression equation is
Score = 33.5 + 1.90 RecRes + 2.61 Afford

Predictor        Coef     SE Coef          T        P
Constant       33.485       3.228      10.37    0.000
RecRes         1.8998      0.2603       7.30    0.000
Afford         2.6108      0.4545       5.74    0.000

S = 3.803       R-Sq = 80.7%     R-Sq(adj) = 78.4%
Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2     1026.29      513.14     35.47    0.000
Residual Error    17      245.92       14.47
Total             19     1272.21
```

Because the p-value = .000 < α = .05, there is a significant overall relationship. Moreover, the fit provided by the two-independent variable estimated regression is much better.

55. a. The Minitab output is shown below:

```
The regression equation is
Speed = 97.6 + 0.0693 Price - 0.00082 Weight + 0.0590
Horsepwr - 2.48 Zero60

Predictor        Coef     SE Coef          T        P
Constant        97.57       11.79       8.27    0.000
Price         0.06928     0.03805       1.82    0.096
Weight      -0.000816    0.002593      -0.31    0.759
Horsepwr      0.05901     0.01543       3.82    0.003
Zero60        -2.4836      0.9601      -2.59    0.025

S = 2.127       R-Sq = 95.0%     R-Sq(adj) = 93.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4      946.18      236.55     52.28    0.000
Residual Error    11       49.77        4.52
Total             15      995.95
```

b. Since the p-value corresponding to the F test is 0.000, the relationship is significant.

c. Since the p-values corresponding to the t test for both Horsepwr (p-value = .003) and Zero60 (p-value = .025) are less than .05, both of these independent variables are significant.

d. The Minitab output is shown below:

```
The regression equation is
Speed = 103 + 0.0558 Horsepwr - 3.19 Zero60

Predictor        Coef     SE Coef          T        P
Constant      103.103       9.448      10.91    0.000
Horsepwr      0.05582     0.01452       3.84    0.002
Zero60        -3.1876      0.9658      -3.30    0.006

S = 2.301       R-Sq = 93.1%     R-Sq(adj) = 92.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      927.12      463.56     87.54    0.000
Residual Error    13       68.84        5.30
Total             15      995.95


Source       DF      Seq SS
Horsepwr      1      869.43
Zero60        1       57.68

Unusual Observations
Obs   Horsepwr     Speed        Fit     SE Fit   Residual   St Resid
  2        290   108.000    103.352      1.015      4.648      2.25R
 12        155    84.600     82.747      1.773      1.853     1.26 X

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

e. The standardized residual plot is shown below:

```
          -
  SRES    -                                     x
          -
          -
       1.5+
          -   x                                                   x
          -
          -
          -                   2     x            x
       0.0+              x    x                 2
          -                  x
          -
          -                      xx
          -
      -1.5+
          -                                    x    x
          -
            ----+---------+---------+---------+---------+---------+--FIT
             84.0      90.0      96.0     102.0     108.0     114.0
```

There is an unusual trend in the plot and one observation appears to be an outlier.

f. The Minitab output indicates that observation 2 is an outlier

g. The Minitab output indicates that observation 12 is an influential observation.

56. The Minitab output is shown below:

```
The regression equation is
CityMPG = 24.1 - 2.10 Displace

Predictor        Coef     SE Coef          T        P
Constant       24.121       2.381      10.13    0.000
Displace      -2.1019      0.4850      -4.33    0.000

S = 1.446       R-Sq = 36.3%     R-Sq(adj) = 34.3%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      39.276      39.276     18.78    0.000
Residual Error    33      69.009       2.091
Total             34     108.286
```

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

b. The Minitab output is shown below:

```
The regression equation is
CityMPG = 26.4 - 2.44 Displace - 1.20 Drive4

Predictor        Coef     SE Coef          T        P
Constant       26.383       2.377      11.10    0.000
Displace      -2.4387      0.4683      -5.21    0.000
Drive4        -1.2015      0.4722      -2.54    0.016

S = 1.339       R-Sq = 47.0%     R-Sq(adj) = 43.7%
```

```
Analysis of Variance

Source           DF        SS        MS       F      P
Regression        2    50.889    25.445   14.19  0.000
Residual Error   32    57.397     1.794
Total            34   108.286
```

c. Because the p-value = .016 < α = .05, Drive4 is significant.

d. The Minitab output is shown below:

```
The regression equation is
CityMPG = 33.3 - 4.15 Displace - 1.24 Drive4 + 2.16 EightCyl

Predictor        Coef     SE Coef         T        P
Constant       33.333       3.345      9.97    0.000
Displace      -4.1492      0.7584     -5.47    0.000
Drive4        -1.2368      0.4310     -2.87    0.007
EightCyl       2.1615      0.7919      2.73    0.010

S = 1.222       R-Sq = 57.3%     R-Sq(adj) = 53.1%

Analysis of Variance

Source           DF        SS        MS       F      P
Regression        3    62.011    20.670   13.85  0.000
Residual Error   31    46.275     1.493
Total            34   108.286
```

e. Because the p-value = .000 < α = .05, there is a significant overall relationship.

For Displace: Because the p-value = .000 < α = .05, Displace is significant.

For Drive4: Because the p-value = .007 > α = .05, Drive4 is significant.

For EightCyl: Because the p-value = .010 < α = .05, EightCyl is significant.

57. a. The Minitab output is shown below:

```
Resale% = 38.8 +0.000766 Price

Predictor        Coef     SE Coef         T        P
Constant       38.772       4.348      8.92    0.000
Price       0.0007656   0.0001900      4.03    0.000

S = 5.421       R-Sq = 36.7%     R-Sq(adj) = 34.4%


Analysis of Variance

Source           DF        SS        MS       F      P
Regression        1    477.25    477.25   16.24  0.000
Residual Error   28    822.92     29.39
Total            29   1300.17
```

Since the p-value corresponding to F = 16.24 is .000 < α = .05, there is a significant relationship between Resale% and Price.

b. R-Sq = 36.7%; not a very good fit.

c. Let Type1 = 0 and Type2 = 0 if a small pickup; Type1 = 1 and Type2 = 0 if a full-size pickup; and Type1 = 0 and Type2 = 1 if a sport utility.

The Minitab output using Type1, Type2, and Price is shown below:

```
The regression equation is
Resale% = 42.6 + 9.09 Type1 + 7.92 Type2 +0.000341 Price

Predictor        Coef     SE Coef          T        P
Constant       42.554       3.562      11.95    0.000
Type1           9.090       2.248       4.04    0.000
Type2           7.917       2.163       3.66    0.001
Price       0.0003415   0.0001800       1.90    0.069

S = 4.298       R-Sq = 63.1%     R-Sq(adj) = 58.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      819.77      273.26     14.79    0.000
Residual Error    26      480.40       18.48
Total             29     1300.17
```

d. Since the p-value corresponding to $F = 14.79$ is $.000 < \alpha = .05$, there is a significant relationship between Resale% and the independent variables. Note that individually, Price is not significant at the .05 level of significance. If we rerun the regression using just Type1 and Type2 the value of R-Sq (adj) decreases to 54.4%, a drop of only 4%. Thus, it appears that for these data, the type of vehicle is the strongest predictor of the resale value.