**Chapter 16**

**Regression Analysis: Model Building**

**Learning Objectives**

1. Learn how the general linear model can be used to model problems involving curvilinear relationships.

2. Understand the concept of interaction and how it can be accounted for in the general linear model.

3. Understand how an *F* test can be used to determine when to add or delete one or more variables.

4. Develop an appreciation for the complexities involved in solving larger regression analysis problems.

5. Understand how variable selection procedures can be used to choose a set of independent variables for an estimated regression equation.

6. Learn how analysis of variance and experimental design problems can be analyzed using a regression model.

7. Know how the Durbin-Watson test can be used to test for autocorrelation.

**Solutions:**

1. a. The Minitab output is shown below:

The regression equation is

Y = - 6.8 + 1.23 X

Predictor Coef SE Coef T p

Constant -6.77 14.17 -0.48 0.658

X 1.2296 0.4697 2.62 0.059

S = 7.269 R-sq = 63.1% R-sq(adj) = 53.9%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 362.13 362.13 6.85 0.059

Residual Error 4 211.37 52.84

Total 5 573.50

b. Since the *p*-value corresponding to *F* = 6.85 is 0.59 > ** the relationship is not significant.

c.

The scatter diagram suggests that a curvilinear relationship may be appropriate.

d. The Minitab output is shown below:

The regression equation is

Y = - 169 + 12.2 X - 0.177 XSQ

Predictor Coef SE Coef T p

Constant -168.88 39.79 -4.24 0.024

X 12.187 2.663 4.58 0.020

XSQ -0.17704 0.04290 -4.13 0.026

S = 3.248 R-sq = 94.5% R-sq(adj) = 90.8%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 541.85 270.92 25.68 0.013

Residual Error 3 31.65 10.55

Total 5 573.50

e. Since the *p*-value corresponding to *F* = 25.68 is .013 < ** the relationship is significant.

f. = -168.88 + 12.187(25) - 0.17704(25)2 = 25.145

2. a. The Minitab output is shown below:

The regression equation is

Y = 9.32 + 0.424 X

Predictor Coef SE Coef T p

Constant 9.315 4.196 2.22 0.113

X 0.4242 0.1944 2.18 0.117

S = 3.531 R-sq = 61.4% R-sq(adj) = 48.5%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 59.39 59.39 4.76 0.117

Residual Error 3 37.41 12.47

Total 4 96.80

The high *p*-value (.117) indicates a weak relationship; note that 61.4% of the variability in *y* has been explained by *x*.

b. The Minitab output is shown below:

The regression equation is

Y = - 8.10 + 2.41 X - 0.0480 XSQ

Predictor Coef SE Coef T p

Constant -8.101 4.104 -1.97 0.187

X 2.4127 0.4409 5.47 0.032

XSQ -0.04797 0.01050 -4.57 0.045

S = 1.279 R-sq = 96.6% R-sq(adj) = 93.2%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 93.529 46.765 28.60 0.034

Residual Error 2 3.271 1.635

Total 4 96.800

At the .05 level of significance, the relationship is significant; the fit is excellent.

c. = -8.101 + 2.4127(20) - 0.04797(20)2 = 20.965

3. a. The scatter diagram shows some evidence of a possible linear relationship.

b. The Minitab output is shown below:

The regression equation is

Y = 2.32 + 0.637 X

Predictor Coef SE Coef T p

Constant 2.322 1.887 1.23 0.258

X 0.6366 0.3044 2.09 0.075

S = 2.054 R-sq = 38.5% R-sq(adj) = 29.7%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 18.461 18.461 4.37 0.075

Residual Error 7 29.539 4.220

Total 8 48.000

c. The following standardized residual plot indicates that the constant variance assumption is not satisfied.



d. The logarithmic transformation does not appear to eliminate the wedged-shaped pattern in the above residual plot. The reciprocal transformation does, however, remove the wedge-shaped pattern. Neither transformation provides a good fit. The Minitab output for the reciprocal transformation and the corresponding standardized residual pot are shown below.

The regression equation is

1/Y = 0.275 - 0.0152 X

Predictor Coef SE Coef T p

Constant 0.27498 0.04601 5.98 0.000

X -0.015182 0.007421 -2.05 0.080

S = 0.05009 R-sq = 37.4% R-sq(adj) = 28.5%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 0.010501 0.010501 4.19 0.080

Residual Error 7 0.017563 0.002509

Total 8 0.028064



4. a. The Minitab output is shown below:

The regression equation is

Y = 943 + 8.71 X

Predictor Coef SE Coef T p

Constant 943.05 59.38 15.88 0.000

X 8.714 1.544 5.64 0.005

S = 32.29 R-sq = 88.8% R-sq(adj) = 86.1%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 33223 33223 31.86 0.005

Residual Error 4 4172 1043

Total 5 37395

b. *p*-value = .005 < ** = .01; reject *H*0

5. The Minitab output is shown below:

The regression equation is

Y = 433 + 37.4 X - 0.383 XSQ

Predictor Coef SE Coef T p

Constant 432.6 141.2 3.06 0.055

X 37.429 7.807 4.79 0.017

XSQ -0.3829 0.1036 -3.70 0.034

S = 15.83 R-sq = 98.0% R-sq(adj) = 96.7%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 36643 18322 73.15 0.003

Residual Error 3 751 250

Total 5 37395

b. Since the linear relationship was significant (Exercise 4), this relationship must be significant. Note also that since the *p*-value of .003 < ** = .05, we can reject *H*0.

c. The fitted value is 1302.01, with a standard deviation of 9.93. The 95% confidence interval is 1270.41 to 1333.61; the 95% prediction interval is 1242.55 to 1361.47.

6. a. The scatter diagram is shown below:

b. No; the relationship appears to be curvilinear.

c. Several possible models can be fitted to these data, as shown below:

= 2.90 - 0.185*x* + .00351*x*2 

7. a.

b. The Minitab output follows.

The regression equation is

Mortgage ($) = - 198 + 1.07 Rent ($)

Predictor Coef SE Coef T P

Constant -198.0 187.7 -1.05 0.322

Rent ($) 1.0699 0.2148 4.98 0.001

S = 78.7819 R-Sq = 75.6% R-Sq(adj) = 72.6%

Analysis of Variance

Source DF SS MS F P

Regression 1 153962 153962 24.81 0.001

Residual Error 8 49653 6207

Total 9 203614

Unusual Observations

Rent Mortgage

Obs ($) ($) Fit SE Fit Residual St Resid

1 840 539.0 700.8 25.5 -161.8 -2.17R

R denotes an observation with a large standardized residual.



c. The Minitab output follows.

The regression equation is

Mortgage ($) = 3966 - 8.26 Rent ($) + 0.00513 RentSq

Predictor Coef SE Coef T P

Constant 3966 1335 2.97 0.021

Rent ($) -8.261 2.982 -2.77 0.028

RentSq 0.005131 0.001637 3.13 0.017

S = 54.3363 R-Sq = 89.8% R-Sq(adj) = 86.9%

Analysis of Variance

Source DF SS MS F P

Regression 2 182947 91474 30.98 0.000

Residual Error 7 20667 2952

Total 9 203614

Source DF Seq SS

Rent ($) 1 153962

RentSq 1 28986

Unusual Observations

Rent Mortgage

Obs ($) ($) Fit SE Fit Residual St Resid

1 840 539.0 646.8 24.6 -107.8 -2.23R

R denotes an observation with a large standardized residual.

d. The estimated regression equation in part (d) provides a better fit.

8. a. The scatter diagram is shown below:

A simple linear regression model does not appear to be appropriate. There appears to be a curvilinear relationship between the two variables.

b. The Minitab output is shown below:

The regression equation is

Price = 33829 - 4571 Rating + 154 RatingSq

Predictor Coef SE Coef T P

Constant 33829 13657 2.48 0.029

Rating -4571 1688 -2.71 0.019

RatingSq 153.55 51.72 2.97 0.012

S = 668.312 R-Sq = 70.2% R-Sq(adj) = 65.3%

Analysis of Variance

Source DF SS MS F P

Regression 2 12643604 6321802 14.15 0.001

Residual Error 12 5359686 446641

Total 14 18003290

c. The Minitab output is shown below:

The regression equation is

logPrice = - 10.2 + 10.4 logRating

Predictor Coef SE Coef T P

Constant -10.152 1.890 -5.37 0.000

logRating 10.422 1.544 6.75 0.000

S = 0.283438 R-Sq = 77.8% R-Sq(adj) = 76.1%

Analysis of Variance

Source DF SS MS F P

Regression 1 3.6595 3.6595 45.55 0.000

Residual Error 13 1.0444 0.0803

Total 14 4.7038

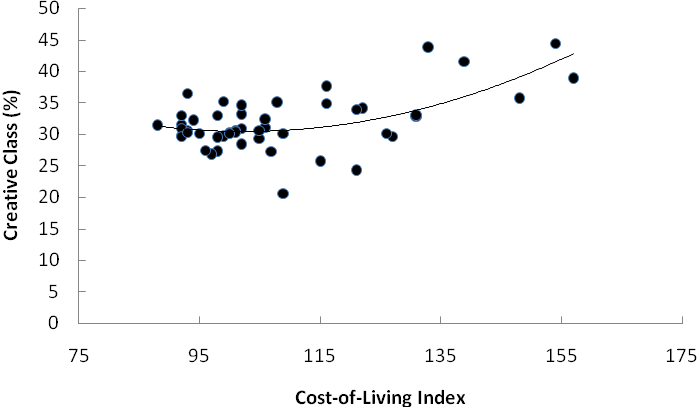
d. The model in part (c) is preferred because it provides a better fit.

9. a.



A simple linear regression model appears to be appropriate.

b.



Note the line drawn through the data. This line indicates a possible curvilinar relationship between these two variables.

c. In the Minitab output that follows IndexSq denotes the square of the Cost-of-Living Index.

The regression equation is

Creative Class (%) = 49.2 - 0.673 Cost-of-Living Index + 0.00282 IndexSq

+ 0.404 Income

Predictor Coef SE Coef T P

Constant 49.24 17.25 2.85 0.006

Cost-of-Living Index -0.6725 0.2888 -2.33 0.024

IndexSq 0.002821 0.001223 2.31 0.026

Income 0.40418 0.06772 5.97 0.000

S = 2.70934 R-Sq = 64.4% R-Sq(adj) = 62.0%

Analysis of Variance

Source DF SS MS F P

Regression 3 609.96 203.32 27.70 0.000

Residual Error 46 337.67 7.34

Total 49 947.62

At the .05 level of significance there is overall significance. And, each of the three independent variables (Cost-of-Living Index, IndexSq, and Income) is significant.

d. Cost-of-Living Index = 99, IndexSq = 9801, and Income = 42.984

Estimate = 49.24 - .6725(99)+.002821(9801) + .40418(42.984)= 27.7%

The primary concern of using this estimate is that the estimated regression equation was developed for metropolitan areas with a population of 1,000,000 or more. But, the population for Tucson is so close to 1,000,000 that the estimated regression equation should still provide a good estimate. Note to Instructor: the actual value of the percentage of the workforce in creative fields for Tucson reported by Kiplinger was 31.1%.

10. a. SSR = SST - SSE = 1030

MSR = 1030 MSE = 520/25 = 20.8 *F*  = 1030/20.8 = 49.52

Using Excel or Minitab, the *p*-value corresponding to *F* = 49.52 is .000.

Because *p*-value ≤ *α*, *x*1 is significant.

b. 

Using Excel or Minitab, the *p*-value corresponding to *F* = 48.3 is .000.

Because *p*-value ≤ *α*, the addition of variables *x*2 and *x*3 is significant.

11. a. SSE = SST - SSR = 1805 - 1760 = 45

MSR = 1760/4 = 440 MSE =45/25 = 1.8

*F* = 440/1.8 = 244.44

Using Excel or Minitab, the *p*-value corresponding to *F* = 244.44 is .000.

Because *p*-value ≤ *α*, the overall relationship is significant.

b. SSE(*x*1, *x*2, *x*3, *x*4) = 45

c. SSE(*x*2, *x*3) = 1805 - 1705 = 100

d. 

Using Excel or Minitab, the *p*-value corresponding to *F* = 15.28 is .000.

Because *p*-value ≤ *α*, *x*1 and *x*4 contribute significantly to the model.

12. a. A portion of the Minitab output follows:

The regression equation is

Scoring Avg. = 46.3 + 14.1 Putting Avg.

Predictor Coef SE Coef T P

Constant 46.277 6.026 7.68 0.000

Putting Avg. 14.103 3.356 4.20 0.000

S = 0.510596 R-Sq = 38.7% R-Sq(adj) = 36.5%

Analysis of Variance

Source DF SS MS F P

Regression 1 4.6036 4.6036 17.66 0.000

Residual Error 28 7.2998 0.2607

Total 29 11.9035

b. A portion of the Minitab output follows:

The regression equation is

Scoring Avg. = 59.0 - 10.3 Greens in Reg. + 11.4 Putting Avg. - 1.81 Sand Saves

Predictor Coef SE Coef T P

Constant 59.022 5.774 10.22 0.000

Greens in Reg. -10.281 2.877 -3.57 0.001

Putting Avg. 11.413 2.760 4.14 0.000

Sand Saves -1.8130 0.9210 -1.97 0.060

S = 0.407808 R-Sq = 63.7% R-Sq(adj) = 59.5%

Analysis of Variance

Source DF SS MS F P

Regression 3 7.5795 2.5265 15.19 0.000

Residual Error 26 4.3240 0.1663

Total 29 11.9035

c. SSE(reduced) = 7.2998

SSE(full) = 4.3240

MSE(full) = .1663



The *p*-value associated with *F* = 8.95 (2 degrees of freedom numerator and 26 denominator) is .001. With a *p*-value < *α* =.05, the addition of the two independent variables is statistically significant.

13. a. A portion of the Minitab output follows:

The regression equation is

Earnings ($1000) = 14528 - 7640 Putting Avg.

Predictor Coef SE Coef T P

Constant 14528 4410 3.29 0.003

Putting Avg. -7640 2456 -3.11 0.004

S = 373.671 R-Sq = 25.7% R-Sq(adj) = 23.0%

Analysis of Variance

Source DF SS MS F P

Regression 1 1350901 1350901 9.67 0.004

Residual Error 28 3909645 139630

Total 29 5260546

b. A portion of the Minitab output follows:

The regression equation is

Earnings ($1000) = 5214 + 6873 Greens in Reg. - 5623 Putting Avg.

+ 2217 Sand Saves

Predictor Coef SE Coef T P

Constant 5214 3757 1.39 0.177

Greens in Reg. 6873 1871 3.67 0.001

Putting Avg. -5623 1795 -3.13 0.004

Sand Saves 2216.6 599.2 3.70 0.001

S = 265.305 R-Sq = 65.2% R-Sq(adj) = 61.2%

Analysis of Variance

Source DF SS MS F P

Regression 3 3430493 1143498 16.25 0.000

Residual Error 26 1830053 70387

Total 29 5260546

c. SSE(reduced) = 3,909,645

SSE(full) = 1,830,053

MSE(full) = 70,387



The *p*-value associated with *F* = 16.25 (2 degrees of freedom numerator and 26 denominator) is .000. With a *p*-value < *α* =.05, the addition of the two independent variables is statistically significant.

d. A portion of the Minitab output follows:

The regression equation is

Earnings ($1000) = 36697 - 501 Scoring Avg.

Predictor Coef SE Coef T P

Constant 36697 5909 6.21 0.000

Scoring Avg. -501.20 82.53 -6.07 0.000

S = 284.751 R-Sq = 56.8% R-Sq(adj) = 55.3%

Analysis of Variance

Source DF SS MS F P

Regression 1 2990221 2990221 36.88 0.000

Residual Error 28 2270325 81083

Total 29 5260546

Because the equation developed in part (b) provides a better fit, it is preferred over the equation developed in part (d).

14. a. The Minitab output is shown below:

Risk = - 111 + 1.32 Age + 0.296 Pressure

Predictor Coef SE Coef T P

Constant -110.94 16.47 -6.74 0.000

Age 1.3150 0.1733 7.59 0.000

Pressure 0.29640 0.05107 5.80 0.000

S = 6.908 R-Sq = 80.6% R-Sq(adj) = 78.4%

Analysis of Variance

Source DF SS MS F P

Regression 2 3379.6 1689.8 35.41 0.000

Residual Error 17 811.3 47.7

Total 19 4190.9

Source DF Seq SS

Age 1 1772.0

Pressure 1 1607.7

Unusual Observations

Obs Age Risk Fit SE Fit Residual St Resid

17 66.0 8.00 25.05 1.67 -17.05 -2.54R

R denotes an observation with a large standardized residual

b. The Minitab output is shown below:

Risk = - 123 + 1.51 Age + 0.448 Pressure + 8.87 Smoker -

0.00276 AgePress

Predictor Coef SE Coef T P

Constant -123.16 56.94 -2.16 0.047

Age 1.5130 0.7796 1.94 0.071

Pressure 0.4483 0.3457 1.30 0.214

Smoker 8.866 3.074 2.88 0.011

AgePress -0.002756 0.004807 -0.57 0.575

S = 5.881 R-Sq = 87.6% R-Sq(adj) = 84.3%

Analysis of Variance

Source DF SS MS F P

Regression 4 3672.11 918.03 26.54 0.000

Residual Error 15 518.84 34.59

Total 19 4190.95

Source DF Seq SS

Age 1 1771.98

Pressure 1 1607.66

Smoker 1 281.10

AgePress 1 11.37

Unusual Observations

Obs Age Risk Fit SE Fit Residual St Resid

17 66.0 8.00 20.91 2.01 -12.91 -2.34R

R denotes an observation with a large standardized residual

c. 

The *p*-value associated with *F* = 4.23 (2 numerator and 15 denominator DF) is .000

Because *p*-value ≤ *α* = .05, the addition of the two terms is significant.

15. a. A portion of the Minitab output follows:

The regression equation is

ERA = - 0.253 + 0.453 H/9

Predictor Coef SE Coef T P

Constant -0.2535 0.7351 -0.34 0.732

H/9 0.45271 0.08347 5.42 0.000

S = 0.466619 R-Sq = 38.0% R-Sq(adj) = 36.7%

Analysis of Variance

Source DF SS MS F P

Regression 1 6.4044 6.4044 29.41 0.000

Residual Error 48 10.4512 0.2177

Total 49 16.8556

b. A portion of the Minitab output follows:

The regression equation is

ERA = - 2.56 + 0.512 H/9 + 0.980 HR/9 + 0.340 BB/9

Predictor Coef SE Coef T P

Constant -2.5639 0.5383 -4.76 0.000

H/9 0.51213 0.05506 9.30 0.000

HR/9 0.9799 0.1657 5.91 0.000

BB/9 0.34000 0.05067 6.71 0.000

S = 0.285210 R-Sq = 77.8% R-Sq(adj) = 76.4%

Analysis of Variance

Source DF SS MS F P

Regression 3 13.1137 4.3712 53.74 0.000

Residual Error 46 3.7419 0.0813

Total 49 16.8556

c. SSE(reduced) = 10.4512

SSE(full) = 3.7419

MSE(full) = .0813



The *p*-value associated with *F* = 41.26 (2 degrees of freedom numerator and 46 denominator) is .000. With a *p*-value < *α* =.05, the addition of the two independent variables is statistically significant.

16. a. The sample correlation coefficients are as follows:

Weeks Age Educ Married Head Tenure Manager

Age 0.577

0.000

Educ 0.007 0.100

0.962 0.490

Married -0.130 -0.209 -0.151

0.370 0.145 0.296

Head -0.205 0.027 -0.156 -0.449

0.153 0.854 0.280 0.001

Tenure 0.398 0.459 0.174 -0.057 -0.046

0.004 0.001 0.228 0.692 0.750

Manager -0.198 0.097 0.160 0.073 -0.200 -0.113

0.167 0.504 0.266 0.616 0.164 0.435

Sales -0.134 0.137 0.124 -0.148 -0.013 0.097 -0.156

0.354 0.343 0.393 0.306 0.926 0.504 0.279

Cell Contents: Pearson correlation

P-Value

The independent variable most correlated with Weeks is Age. The Minitab output corresponding to using Age as the independent variable is shown below:

The regression equation is

Weeks = - 8.9 + 1.51 Age

Predictor Coef SE Coef T P

Constant -8.86 11.01 -0.80 0.425

Age 1.5092 0.3080 4.90 0.000

S = 19.5342 R-Sq = 33.3% R-Sq(adj) = 32.0%

Analysis of Variance

Source DF SS MS F P

Regression 1 9161.4 9161.4 24.01 0.000

Residual Error 48 18316.1 381.6

Total 49 27477.5

b. The Minitab Stepwise Regression output is shown below.

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is Weeks on 7 predictors, with N = 50

Step 1 2 3 4

Constant -8.86002 -9.09741 -0.10922 -0.06890

Age 1.51 1.57 1.61 1.73

T-Value 4.90 5.30 5.74 6.51

P-Value 0.000 0.000 0.000 0.000

Manager -20.1 -24.6 -28.7

T-Value -2.26 -2.88 -3.53

P-Value 0.029 0.006 0.001

Head -14.3 -15.1

T-Value -2.61 -2.95

P-Value 0.012 0.005

Sales -17.4

T-Value -2.79

P-Value 0.008

S 19.5 18.7 17.7 16.5

R-Sq 33.34 39.87 47.64 55.38

R-Sq(adj) 31.95 37.31 44.22 51.41

Mallows C-p 22.5 17.8 11.8 5.9

The results suggest a model using four independent variables: Age, Manager, Head, and Sales. The corresponding Minitab output is shown below:

The regression equation is

Weeks = - 0.07 + 1.73 Age - 28.7 Manager - 15.1 Head - 17.4 Sales

Predictor Coef SE Coef T P

Constant -0.069 9.843 -0.01 0.994

Age 1.7252 0.2651 6.51 0.000

Manager -28.672 8.117 -3.53 0.001

Head -15.086 5.121 -2.95 0.005

Sales -17.421 6.236 -2.79 0.008

S = 16.5069 R-Sq = 55.4% R-Sq(adj) = 51.4%

Analysis of Variance

Source DF SS MS F P

Regression 4 15216.0 3804.0 13.96 0.000

Residual Error 45 12261.5 272.5

Total 49 27477.5

c. The results using Minitab’s Forward Selection procedure are the same as the results using Minitab’s Stepwise procedure in part (b).

d. The results using Minitab’s Backward Elimination procedure are shown below:

Backward elimination. Alpha-to-Remove: 0.05

Response is Weeks on 7 predictors, with N = 50

Step 1 2 3 4

Constant 22.85070 13.62308 13.06817 -0.06890

Age 1.51 1.52 1.64 1.73

T-Value 4.96 5.04 6.18 6.51

P-Value 0.000 0.000 0.000 0.000

Educ -0.61

T-Value -0.66

P-Value 0.516

Married -10.7 -9.9 -9.8

T-Value -1.79 -1.69 -1.69

P-Value 0.081 0.098 0.099

Head -19.8 -19.0 -19.4 -15.1

T-Value -3.39 -3.35 -3.44 -2.95

P-Value 0.002 0.002 0.001 0.005

Tenure 0.43 0.37

T-Value 0.91 0.82

P-Value 0.366 0.418

Manager -26.7 -27.7 -29.0 -28.7

T-Value -3.21 -3.40 -3.64 -3.53

P-Value 0.003 0.001 0.001 0.001

Sales -18.6 -19.0 -19.0 -17.4

T-Value -2.96 -3.06 -3.07 -2.79

P-Value 0.005 0.004 0.004 0.008

S 16.3 16.2 16.2 16.5

R-Sq 59.14 58.72 58.08 55.38

R-Sq(adj) 52.33 52.96 53.32 51.41

Mallows C-p 8.0 6.4 5.1 5.9

These results also suggest using the model with four independent variables: Age, Head, Manager, and Sales.

e. The results using Mintab’s Best-Subset procedure are shown below:

M M

a T a

r e n S

E r H n a a

A d i e u g l

Mallows g u e a r e e

Vars R-Sq R-Sq(adj) C-p S e c d d e r s

1 33.3 32.0 22.5 19.534 X

1 15.8 14.0 40.6 21.954 X

2 39.9 37.3 17.8 18.749 X X

2 38.2 35.6 19.5 19.005 X X

3 47.6 44.2 11.8 17.686 X X X

3 46.8 43.3 12.7 17.831 X X X

4 55.4 51.4 5.9 16.507 X X X X

4 49.1 44.6 12.3 17.628 X X X X

5 58.1 53.3 5.1 16.179 X X X X X

5 56.0 51.0 7.3 16.582 X X X X X

6 58.7 53.0 6.4 16.241 X X X X X X

6 58.3 52.5 6.8 16.318 X X X X X X

7 59.1 52.3 8.0 16.350 X X X X X X X

The results suggest a model using five independent variables: Age, Married, Head, Manager, and Sales.

The corresponding Minitab output is shown below:

The regression equation is

Weeks = 13.1 + 1.64 Age - 9.76 Married - 19.4 Head - 29.0 Manager - 19.0 Sales

Predictor Coef SE Coef T P

Constant 13.07 12.40 1.05 0.298

Age 1.6369 0.2651 6.18 0.000

Married -9.764 5.794 -1.69 0.099

Head -19.405 5.636 -3.44 0.001

Manager -28.986 7.958 -3.64 0.001

Sales -18.967 6.181 -3.07 0.004

S = 16.1794 R-Sq = 58.1% R-Sq(adj) = 53.3%

Analysis of Variance

Source DF SS MS F P

Regression 5 15959.5 3191.9 12.19 0.000

Residual Error 44 11518.0 261.8

Total 49 27477.5

17. The output obtained using Minitab’s Best Subset Regression is shown below:

Response is Scoring Avg.

G

D r

r e P

i e u D

v n t S r

e s t a i

i n v

A i n d e

v n g G

e S r

r R A a e

a e v v e

Mallows g g g e n

Vars R-Sq R-Sq(adj) C-p S e . . s s

1 38.7 36.5 28.3 0.51060 X

1 33.0 30.7 33.3 0.53350 X

2 58.3 55.2 12.9 0.42897 X X

2 53.9 50.5 16.8 0.45059 X X

3 63.7 59.5 10.2 0.40781 X X X

3 60.3 55.7 13.2 0.42659 X X X

4 72.0 67.5 4.8 0.36514 X X X X

4 64.7 59.0 11.3 0.41015 X X X X

5 72.9 67.2 6.0 0.36672 X X X X X

The Best Subset Regression output indicates that a model using four independent variables, Drive Average, Greens in Reg., Putting Average, and DriveGreens, may be a good choice. The Minitab output for this model is shown below:

The regression equation is

Scoring Avg. = - 88.1 + 0.591 Drive Average + 209 Greens in Reg.

+ 9.74 Putting Avg. - 0.868 DriveGreens

Predictor Coef SE Coef T P

Constant -88.10 42.20 -2.09 0.047

Drive Average 0.5907 0.1692 3.49 0.002

Greens in Reg. 209.19 62.85 3.33 0.003

Putting Avg. 9.736 2.575 3.78 0.001

DriveGreens -0.8677 0.2478 -3.50 0.002

S = 0.365139 R-Sq = 72.0% R-Sq(adj) = 67.5%

Analysis of Variance

Source DF SS MS F P

Regression 4 8.5703 2.1426 16.07 0.000

Residual Error 25 3.3332 0.1333

Total 29 11.9035

18. a. Because the independent variable most highly correlated with RPG is OBP, it will provide the best one-variable estimated regression equation. The Minitab output using OBP to predict RPG is shown below:

The regression equation is

RPG = - 4.05 + 27.6 OBP

Predictor Coef SE Coef T P

Constant -4.049 1.006 -4.02 0.001

OBP 27.555 3.103 8.88 0.000

S = 0.956308 R-Sq = 81.4% R-Sq(adj) = 80.4%

Analysis of Variance

Source DF SS MS F P

Regression 1 72.108 72.108 78.85 0.000

Residual Error 18 16.461 0.915

Total 19 88.569

b. The output using Minitab’s Stepwise Regression procedure using Alpha-to-Enter = 0.05 and Alpha-to-Remove = 0.05 is shown below:

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is RPG on 12 predictors, with N = 20

Step 1 2 3

Constant -4.0491 -1.5951 -0.9808

OBP 27.6 17.2 25.1

T-Value 8.88 5.10 6.88

P-Value 0.000 0.000 0.000

HR 0.071 0.069

T-Value 4.16 5.06

P-Value 0.001 0.000

AVG -12.6

T-Value -3.23

P-Value 0.005

S 0.956 0.693 0.556

R-Sq 81.41 90.78 94.43

R-Sq(adj) 80.38 89.70 93.38

Mallows C-p 66.6 27.0 12.8

Using less sensitive values for Alpha-to-Enter and Alpha-to-Remove will provide a model with additional independent variables. For example, the output using Minitab’s Stepwise Regression procedure using Alpha-to-Enter = 0.10 and Alpha-to-Remove = 0.10 is shown below:

Alpha-to-Enter: 0.1 Alpha-to-Remove: 0.1

Response is RPG on 12 predictors, with N = 20

Step 1 2 3 4 5

Constant -4.0491 -1.5951 -0.9808 -0.6161 -0.9088

OBP 27.6 17.2 25.1 26.6 32.2

T-Value 8.88 5.10 6.88 7.64 9.40

P-Value 0.000 0.000 0.000 0.000 0.000

HR 0.071 0.069 0.068 0.109

T-Value 4.16 5.06 5.34 6.26

P-Value 0.001 0.000 0.000 0.000

AVG -12.6 -16.5 -21.5

T-Value -3.23 -3.96 -5.65

P-Value 0.005 0.001 0.000

3B 0.182 0.244

T-Value 1.88 2.99

P-Value 0.079 0.010

BB -0.0223

T-Value -2.92

P-Value 0.011

S 0.956 0.693 0.556 0.516 0.421

R-Sq 81.41 90.78 94.43 95.49 97.20

R-Sq(adj) 80.38 89.70 93.38 94.29 96.20

Mallows C-p 66.6 27.0 12.8 10.0 4.5

The following output using Minitab’s Best Subset procedure also confirms that a variety of models will provide a good fit.

R O S A

Mallows 2 3 H B B S S C B L V

Vars R-Sq R-Sq(adj) C-p S H B B R I B O B S P G G

1 81.4 80.4 66.6 0.95631 X

1 78.9 77.7 77.9 1.0192 X

2 90.8 89.7 27.0 0.69299 X X

2 88.4 87.0 37.8 0.77872 X X

3 94.4 93.4 12.8 0.55552 X X X

3 94.4 93.3 13.0 0.55820 X X X

4 95.8 94.6 8.8 0.50014 X X X X

4 95.5 94.3 10.0 0.51589 X X X X

5 97.2 96.2 4.5 0.42096 X X X X X

5 97.2 96.2 4.6 0.42336 X X X X X

6 97.6 96.6 4.5 0.40042 X X X X X X

6 97.5 96.4 5.1 0.41198 X X X X X X

7 98.2 97.2 3.9 0.36245 X X X X X X X

7 98.2 97.1 4.1 0.36664 X X X X X X X

8 98.3 97.1 5.3 0.36471 X X X X X X X X

8 98.3 97.0 5.7 0.37269 X X X X X X X X

9 98.4 97.0 7.1 0.37506 X X X X X X X X X

9 98.4 96.9 7.3 0.38077 X X X X X X X X X

10 98.4 96.7 9.0 0.39477 X X X X X X X X X X

10 98.4 96.7 9.0 0.39496 X X X X X X X X X X

11 98.4 96.3 11.0 0.41758 X X X X X X X X X X X

11 98.4 96.2 11.0 0.41848 X X X X X X X X X X X

12 98.4 95.7 13.0 0.44629 X X X X X X X X X X X X

It would be hard to make an argument that there is one best model given these results. The five variable model identified using Minitab’s Stepwise Regression procedure with Alpha-to-Enter = 0.10 and Alpha-to-Remove = 0.10 seems like a reasonable choice. The Minitab regression output corresponding to this model is shown below:

The regression equation is

RPG = - 0.909 + 32.2 OBP + 0.109 HR - 21.5 AVG + 0.244 3B - 0.0223 BB

Predictor Coef SE Coef T P

Constant -0.9088 0.6169 -1.47 0.163

OBP 32.184 3.423 9.40 0.000

HR 0.10877 0.01739 6.26 0.000

AVG -21.511 3.810 -5.65 0.000

3B 0.24388 0.08168 2.99 0.010

BB -0.022306 0.007638 -2.92 0.011

S = 0.420960 R-Sq = 97.2% R-Sq(adj) = 96.2%

Analysis of Variance

Source DF SS MS F P

Regression 5 86.088 17.218 97.16 0.000

Residual Error 14 2.481 0.177

Total 19 88.569

The corresponding standardized residual plot follows:



The standardized residual plot does not indicate any reason to question the usual assumptions regarding the error term. Thus, the estimated regression equation using OBP, HR, AVG, 3B, and BB appears to be a good choice.

19. See the solution to Exercise 14 in this chapter. The Minitab output using the best subsets regression procedure is shown below:

Response is Risk

P A

r g

e S e

s m P

s o r

A u k e

g r e s

Vars R-Sq R-Sq(adj) C-p S e e r s

1 63.3 61.3 28.5 9.2430 X

1 46.3 43.3 49.1 11.182 X

2 80.6 78.4 9.5 6.9083 X X

2 79.5 77.1 10.8 7.1058 X X

3 87.3 85.0 3.3 5.7566 X X X

3 86.2 83.7 4.7 6.0051 X X X

4 87.6 84.3 5.0 5.8813 X X X X

This output suggests that the model involving Age, Pressure, and Smoker is the preferred model; the Minitab output for this model is shown below:

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

Source DF Seq SS

Age 1 1772.0

Pressure 1 1607.7

Smoker 1 281.1

Unusual Observations

Obs Age Risk Fit SE Fit Residual St Resid

17 66.0 8.00 21.11 1.94 -13.11 -2.42R

20. The dummy variables are defined as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| *x*1 | *x*2 | *x*3 | Treatment |
| 0 | 0 | 0 | A |
| 1 | 0 | 0 | B |
| 0 | 1 | 0 | C |
| 0 | 0 | 1 | D |

*E*(*y*) = *β*0 + *β*1 *x*1 + *β*2 *x*2 + *β*3 *x*3

21. The dummy variables are defined as follows:

|  |  |  |
| --- | --- | --- |
| *x*1 | *x*2 | Treatment |
| 0 | 0 | 1 |
| 1 | 0 | 2 |
| 0 | 1 | 3 |

*x*3 = 0 if block 1 and 1 if block 2

*E*(*y*) = *β*0 + *β*1*x*1 + *β*2*x*2 + *β*3*x*3

22. Factor A

*x*1 = 0 if level 1 and 1 if level 2

Factor B

|  |  |  |
| --- | --- | --- |
| *x*2 | *x*3 | Level |
| 0 | 0 | 1 |
| 1 | 0 | 2 |
| 0 | 1 | 3 |

*E*(*y*) = *β*0 + *β*1 *x*1 + *β*2 *x*2 + *β*3 *x*1*x*2 + *β*4 *x*1*x*3

23. a. The dummy variables are defined as follows:

|  |  |  |
| --- | --- | --- |
| D1 | D2 | Mfg. |
| 0 | 0 | 1 |
| 1 | 0 | 2 |
| 0 | 1 | 3 |

*E*(*y*) = *β*0 + *β*1 D1 + *β*2 D2

b. The Minitab output is shown below:

The regression equation is

TIME = 23.0 + 5.00 D1 - 2.00 D2

Predictor Coef SE Coef T p

Constant 23.000 1.106 20.80 0.000

D1 5.000 1.563 3.20 0.011

D2 -2.000 1.563 -1.28 0.233

S = 2.211 R-sq = 70.3% R-sq(adj) = 63.7%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 104.000 52.000 10.64 0.004

Residual Error 9 44.000 4.889

Total 11 148.000

c. *H*0 : *β*1 = *β*2 = 0

d. The *p*-value of .004 is less than ** = .05; therefore, we can reject *H*0 and conclude that the mean time to mix a batch of material is most the same for each manufacturer.

24. a. The dummy variables are defined as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| D1 | D2 | D3 | Paint |
| 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 2 |
| 0 | 1 | 0 | 3 |
| 0 | 0 | 1 | 4 |

The Minitab output is shown below:

The regression equation is

TIME = 133 + 6.00 D1 + 3.00 D2 + 11.0 D3

Predictor Coef SE Coef T p

Constant 133.000 2.941 45.22 0.000

D1 6.000 4.159 1.44 0.168

D2 3.000 4.159 0.72 0.481

D3 11.000 4.159 2.64 0.018

S = 6.576 R-sq = 32.3% R-sq(adj) = 19.6%

Analysis of Variance

SOURCE DF SS MS F p

Regression 3 330.00 110.00 2.54 0.093

Residual Error 16 692.00 43.25

Total 19 1022.00

The appropriate hypothesis test is:

*H*0 : *β*1 = *β*2 = *β*3 = 0

The *p*-value of .093 is greater than ** = .05; therefore, at the 5% level of significance we can not reject *H*0.

b. Note: Estimating the mean drying for paint 2 using the estimated regression equations developed in part (a) may not be the best approach because at the 5% level of significance, we cannot reject *H*0. But, if we want to use the output, we would proceed as follows.

D1 = 1 D2 = 0 D3 = 0

TIME = 133 + 6(1) + 3(0) +11(0) = 139

25. X1 = 0 if computerized analyzer, 1 if electronic analyzer

X2 and X3 are defined as follows:

|  |  |  |
| --- | --- | --- |
| X2 | X3 | Car |
| 0 | 0 | 1 |
| 1 | 0 | 2 |
| 0 | 1 | 3 |

The complete data set and the Minitab output are shown below:

Y X1 X2 X3

50 0 0 0

55 0 1 0

63 0 0 1

42 1 0 0

44 1 1 0

46 1 0 1

The regression equation is

Y = 52.0 - 12.0 X1 + 3.50 X2 + 8.50 X3

Predictor Coef SE Coef T p

Constant 52.000 2.646 19.65 0.003

X1 -12.000 2.646 -4.54 0.045

X2 3.500 3.240 1.08 0.393

X3 8.500 3.240 2.62 0.120

s = 3.240 R-sq = 93.2% R-sq(adj) = 83.1%

Analysis of Variance

SOURCE DF SS MS F p

Regression 3 289.00 96.33 9.17 0.100

Residual Error 2 21.00 10.50

Total 5 310.00

To test for any significant difference between the two analyzers we must test *H*0: *β*1 Since the *p*-value corresponding to *t* = -4.54 is .045 < ** = .05, we reject *H*0: *β*0 the time to do a tuneup is not the same for the two analyzers.

26. Size = 0 if a small advertisement and 1 if a large advertisement

DesignB and DesignC are defined as follows:

|  |  |  |
| --- | --- | --- |
| DesignB | DesignC | Advertisement Design |
| 0 | 0 | A |
| 1 | 0 | B |
| 0 | 1 | C |

LargeDesignB denotes the interaction between Large and DesignB

LargeDesignC denotes the interaction between Large and DesignC

The complete data set and the Minitab output are shown below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Number | Size | DesignB | DesignC | LargeDesignB | LargeDesignC |
| 8 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 1 | 0 | 0 | 0 |
| 14 | 0 | 1 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1 | 0 | 0 |
| 18 | 0 | 0 | 1 | 0 | 0 |
| 12 | 1 | 0 | 0 | 0 | 0 |
| 8 | 1 | 0 | 0 | 0 | 0 |
| 26 | 1 | 1 | 0 | 1 | 0 |
| 30 | 1 | 1 | 0 | 1 | 0 |
| 18 | 1 | 0 | 1 | 0 | 1 |
| 14 | 1 | 0 | 1 | 0 | 1 |

The regression equation is

Number = 10.0 + 0.00 Size + 8.00 DesignB + 4.00 DesignC + 10.0 LargeDesignB

+ 2.00 LargeDesignC

Predictor Coef SE Coef T P

Constant 10.000 2.828 3.54 0.012

Size 0.000 4.000 0.00 1.000

DesignB 8.000 4.000 2.00 0.092

DesignC 4.000 4.000 1.00 0.356

LargeDesignB 10.000 5.657 1.77 0.128

LargeDesignC 2.000 5.657 0.35 0.736

S = 4 R-Sq = 82.4% R-Sq(adj) = 67.6%

Analysis of Variance

Source DF SS MS F P

Regression 5 448.00 89.60 5.60 0.029

Residual Error 6 96.00 16.00

Total 11 544.00

Overall the model is significant because the *p*-value corresponding to *F* = 5.60 < *α* = .05.

Individually, none of the variables are significant using *α* = .05

The Minitab output using only Design B follows:

The regression equation is

Number = 12.5 + 10.5 DesignB

Predictor Coef SE Coef T P

Constant 12.500 1.768 7.07 0.000

DesignB 10.500 3.062 3.43 0.006

S = 5 R-Sq = 54.0% R-Sq(adj) = 49.4%

Analysis of Variance

Source DF SS MS F P

Regression 1 294.00 294.00 11.76 0.006

Residual Error 10 250.00 25.00

Total 11 544.00

Thus, DesignB is significant using *α* = .05. However, the model involving just the interaction between Large and DesignB also provides some interesting results:

The regression equation is

Number = 13.6 + 14.4 LargeDesignB

Predictor Coef SE Coef T P

Constant 13.600 1.409 9.66 0.000

LargeDesignB 14.400 3.450 4.17 0.002

S = 4.45421 R-Sq = 63.5% R-Sq(adj) = 59.9%

Analysis of Variance

Source DF SS MS F P

Regression 1 345.60 345.60 17.42 0.002

Residual Error 10 198.40 19.84

Total 11 544.00

Here we see that the interaction term is significant. Thus, one might consider that the differences are due to both design and a large advertisement. But, it is hard to reach any definite conclusions given the size of the data set. A larger sample size is really needed to make any stronger conclusions about the relationships among the variables.

27. a. The Minitab output is shown below:

The regression equation is

Price ($) = 82.0 + 0.402 Period

Predictor Coef SE Coef T P

Constant 81.9894 0.2805 292.30 0.000

Period 0.40244 0.02342 17.19 0.000

S = 0.603819 R-Sq = 94.3% R-Sq(adj) = 93.9%

Analysis of Variance

Source DF SS MS F P

Regression 1 107.70 107.70 295.39 0.000

Residual Error 18 6.56 0.36

Total 19 114.26

Unusual Observations

Obs Period Price ($) Fit SE Fit Residual St Resid

5 5.0 82.840 84.002 0.187 -1.162 -2.02R

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 0.798118

b. The Durbin-Watson statistic is .798118. At the .05 level of significance, *d*L = 1.20 and *d*U =1.41. Because *d* < *d*L, there is significant positive autocorrelation.

28. From Minitab, *d* = 1.60. At the .05 level of significance, *d*L = 1.04 and *d*U = 1.77. Since *d*L  *d* *d*U, the test is inconclusive.

29. a. A scatter diagram follows.

A simple linear regression model does not appear to be appropriate; there appears to be a curvilinear relationship.

b. A portion of the Minitab output follows.

The regression equation is

Yield = 1.02 + 0.461 Years - 0.0103 YearsSq

Predictor Coef SE Coef T P

Constant 1.0170 0.4354 2.34 0.025

Years 0.46064 0.08136 5.66 0.000

YearsSq -0.010253 0.002587 -3.96 0.000

S = 0.958250 R-Sq = 66.8% R-Sq(adj) = 65.0%

Analysis of Variance

Source DF SS MS F P

Regression 2 68.301 34.151 37.19 0.000

Residual Error 37 33.975 0.918

Total 39 102.276

c. A portion of the Minitab output follows.

The regression equation is

Yield = 0.828 + 1.56 lnYears

Predictor Coef SE Coef T P

Constant 0.8279 0.3790 2.18 0.035

lnYears 1.5626 0.1781 8.77 0.000

S = 0.943120 R-Sq = 67.0% R-Sq(adj) = 66.1%

Analysis of Variance

Source DF SS MS F P

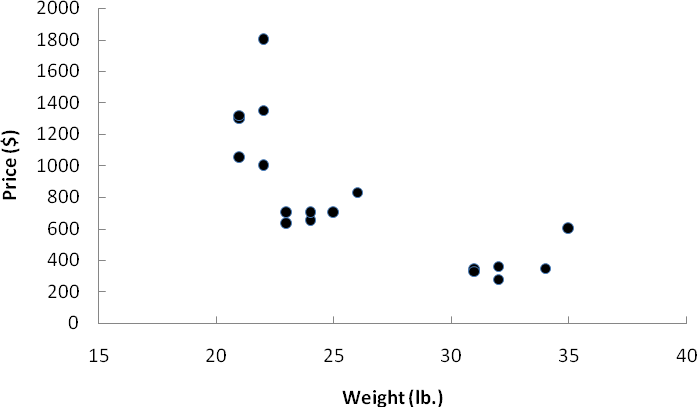
Regression 1 68.476 68.476 76.98 0.000

Residual Error 38 33.800 0.889

Total 39 102.276

With R-Sq = 67.0%, the fit provided by the logarithmic model is slightly better than the fit provided by the second-order model with R-Sq (adj) = 65.0%. But, looking at the estimated regression lines corresponding to both models shown on the following scatter diagrams, the logarithmic model appears to provide a better description of the relationship between years to maturity and yield at higher values of years to maturity. For instance, the second-order model implies that yield is starting to move downward whereas the logarithmic model shows yields increasing but at a much slower rate.

30. a.



There appears to be a curvilinear relationship between weight and price.

b. A portion of the Minitab output follows:

The regression equation is

Price = 11376 - 728 Weight + 12.0 WeightSq

Predictor Coef SE Coef T P

Constant 11376 2565 4.43 0.000

Weight -728.3 193.7 -3.76 0.002

WeightSq 11.974 3.539 3.38 0.004

S = 242.804 R-Sq = 77.0% R-Sq(adj) = 74.1%

Analysis of Variance

Source DF SS MS F P

Regression 2 3161747 1580874 26.82 0.000

Residual Error 16 943263 58954

Total 18 4105011

The results obtained support the conclusion that there is a curvilinear relationship between weight and price.

c. A portion of the Minitab output follows:

The regression equation is

Price = 1284 - 572 Type\_Fitness - 907 Type\_Comfort

Predictor Coef SE Coef T P

Constant 1283.75 95.22 13.48 0.000

Type\_Fitness -571.8 153.5 -3.72 0.002

Type\_Comfort -907.1 145.5 -6.24 0.000

S = 269.328 R-Sq = 71.7% R-Sq(adj) = 68.2%

Analysis of Variance

Source DF SS MS F P

Regression 2 2944410 1472205 20.30 0.000

Residual Error 16 1160601 72538

Total 18 4105011

Type of bike appears to be a significant factor in predicting price. But, the estimated regression equation developed in part (b) appears to provide a slightly better fit.

d. A portion of the Minitab output follows. In this output WxF denotes the interaction between the weight of the bike and the dummy variable Type\_Fitness and WxC denotes the interaction between the weight of the bike and the dummy variable Type\_Comfort.

The regression equation is

Price = 5924 - 215 Weight - 6343 Type\_Fitness - 7232 Type\_Comfort + 261 WxF

+ 266 WxC

Predictor Coef SE Coef T P

Constant 5924 1547 3.83 0.002

Weight -214.56 71.42 -3.00 0.010

Type\_Fitness -6343 2596 -2.44 0.030

Type\_Comfort -7232 2518 -2.87 0.013

WxF 261.3 111.8 2.34 0.036

WxC 266.41 93.98 2.83 0.014

S = 224.438 R-Sq = 84.0% R-Sq(adj) = 77.9%

Analysis of Variance

Source DF SS MS F P

Regression 5 3450170 690034 13.70 0.000

Residual Error 13 654841 50372

Total 18 4105011

By taking into account the type of bike, the weight, and the interaction between these two factors this estimated regression equation provides an excellent fit.

31. a. The Minitab output is shown below:

The regression equation is

Delay = 80.4 + 11.9 Industry - 4.82 Public - 2.62 Quality - 4.07 Finished

Predictor Coef SE Coef T P

Constant 80.429 5.916 13.60 0.000

Industry 11.944 3.798 3.15 0.003

Public -4.816 4.229 -1.14 0.263

Quality -2.624 1.184 -2.22 0.033

Finished -4.073 1.851 -2.20 0.035

S = 10.9235 R-Sq = 38.3% R-Sq(adj) = 31.2%

Analysis of Variance

Source DF SS MS F P

Regression 4 2587.7 646.9 5.42 0.002

Residual Error 35 4176.3 119.3

Total 39 6764.0

b. The low value of the adjusted coefficient of determination (31.2%) does not indicate a good fit.

c. The scatter diagram is shown below:

The scatter diagram suggests a curvilinear relationship between these two variables.

d. The output from Minitab’s best subsets procedure is shown below, where FinishedSq is the square of Finished.

Response is Delay

F

i

I F n

n Q i i

d P u n s

u u a i h

s b l s e

t l i h d

Mallows r i t e S

Vars R-Sq R-Sq(adj) C-p S y c y d q

1 15.9 13.7 34.0 12.234 X

1 9.9 7.6 39.0 12.662 X

2 37.8 34.4 17.8 10.667 X X

2 26.9 22.9 26.9 11.561 X X

3 51.1 47.1 8.7 9.5813 X X X

3 42.2 37.3 16.1 10.425 X X X

4 59.1 54.4 4.1 8.8960 X X X X

4 51.6 46.1 10.3 9.6712 X X X X

5 59.1 53.1 6.0 9.0149 X X X X X

The estimated regression equation using Industry, Quality, Finished, and FinishedSq has an adjusted coefficient of determination of 54.4%.

32. The computer output is shown below:

The regression equation is

Delay = 63.0 + 11.1 Industry

Predictor Coef SE Coef T p

Constant 63.000 3.393 18.57 0.000

Industry 11.074 4.130 2.68 0.011

S = 12.23 R-sq = 15.9% R-sq(adj) = 13.7%

Analysis of Variance

SOURCE DF SS MS F p

Regression 1 1076.1 1076.1 7.19 0.011

Residual Error 38 5687.9 149.7

Total 39 6764.0

Durbin-Watson statistic = 1.55

At the .05 level of significance, *d*L = 1.44 and *d*U = 1.54. Since *d* = 1.55 > *d*U, there is no significant positive autocorrelation.

33. a. The Minitab output is shown below:

The regression equation is

Delay = 70.6 + 12.7 Industry - 2.92 Quality

Predictor Coef SE Coef T p

Constant 70.634 4.558 15.50 0.000

Industry 12.737 3.966 3.21 0.003

Quality -2.919 1.238 -2.36 0.024

S = 11.56 R-sq = 26.9% R-sq(adj) = 22.9%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 1818.6 909.3 6.80 0.003

Residual Error 37 4945.4 133.7

Total 39 6764.0

Durbin-Watson statistic = 1.43

b. The residual plot as a function of the order in which the data are presented is shown below:



There is no obvious pattern in the data indicative of positive autocorrelation.

c. At the .05 level of significance, *d*L = 1.39 and *d*U = 1.60. Since *d*L ≤ *d* ≤ *d*U, the test is inconclusive.

34. The dummy variables are defined as follows:

|  |  |  |
| --- | --- | --- |
| D1 | D2 | Type |
| 0 | 0 | Non |
| 1 | 0 | Light |
| 0 | 1 | Heavy |

The Minitab output is shown below:

The regression equation is

Score = 4.25 + 1.00 D1 + 1.50 D2

Predictor Coef SE Coef T p

Constant 4.2500 0.3819 11.13 0.000

D1 1.0000 0.5401 1.85 0.078

D2 1.5000 0.5401 2.78 0.011

S = 1.080 R-sq = 27.6% R-sq(adj) = 20.7%

Analysis of Variance

SOURCE DF SS MS F p

Regression 2 9.333 4.667 4.00 0.034

Residual Error 21 24.500 1.167

Total 23 33.833

Since the *p*-value = .034 is less than  = .05, there are significant differences between comfort levels for the three types of browsers.

35. Let SizeMidsize = 1 if a midsize car, 0 otherwise; and SizeLarge = 1 if a large car, 0 otherwise.

A portion of the Minitab output follows.

The regression equation is

Hwy MPG = 31.8 - 2.14 SizeMidsize - 6.05 SizeLarge

Predictor Coef SE Coef T P

Constant 31.8140 0.4641 68.55 0.000

SizeMidsize -2.1445 0.6670 -3.21 0.001

SizeLarge -6.0512 0.8284 -7.30 0.000

S = 5.27077 R-Sq = 14.9% R-Sq(adj) = 14.3%

Analysis of Variance

Source DF SS MS F P

Regression 2 1486.86 743.43 26.76 0.000

Residual Error 306 8500.99 27.78

Total 308 9987.85

Significant overall difference; *p*-value corresponding to *F* = 26.76 is .000 <= .05. And, the *p*-values for both of the independent variables are significant as well because both *p*-values values are less than = .05.