Chapter 13
Multiple Regression Analysis

LEARNING OBJECTIVES

This chapter presents the potential of multiple regression analysis as a tool in business decision making and its applications, thereby enabling you to:

1. Explain how, by extending the simple regression model to a multiple regression model with two independent variables, it is possible to determine the multiple regression equation for any number of unknowns.
2. Examine significance tests of both the overall regression model and the regression coefficients.
3. Calculate the residual, standard error of the estimate, coefficient of multiple determination, and adjusted coefficient of multiple determination of a regression model.
4. Use a computer to find and interpret multiple regression outputs.

CHAPTER OUTLINE

KEY TERMS

Adjusted R^2
Coefficient of Multiple Determination (R^2)
Dependent Variable
Independent Variable
Least Squares Analysis
Multiple Regression
Outliers
Partial Regression Coefficient
R^2
Residual
Response Plane
Response Surface
Response Variable
Standard Error of the Estimate (s_e)
Sum of Squares of Error

STUDY QUESTIONS

1. In multiple regression, an _________________ statistic is used to test for the overall effectiveness of the model.

2. The significance of individual regression coefficients in a multiple regression model is tested using a ______ ratio.

3. The value, s_e, represents the _________________________________.

4. The coefficient of multiple determination is denoted by ______.

5. Residuals can sometimes be used to locate _____________________ or values that are apart from the mainstream of the data.

6. Because R^2 may sometimes yield an inflated figure, statisticians have developed a(n) ___________________________ to take into consideration both the additional information of each new independent variable and the changed degrees of freedom.

7. Examine the computer output below taken from a multiple regression analysis with three independent variables.

   The regression equation is: $\hat{y} = 28.4 + 1.30\, x_1 - 0.25\, x_2 + 2.20\, x_3$

Predictor	Coef	Stdev	*t*-ratio	*p*
Constant	28.410	56.520	0.50	0.631
x_1	1.295	2.052	0.63	0.548
x_2	-0.247	1.816	-0.14	0.895
x_3	2.202	.028	0.73	0.491

$s_e = 32.68$ $R^2 = 8.1\%$ R^2 (adj) = 0.0%

Analysis of Variance

Source	df	SS	MS	*F*	*p*
Regression	3	659	220	0.21	0.889
Error	7	7474	1068		
Total	10	8134			

The overall test of significance yields ______________. This test is (is not) significant ______________. The coefficient of multiple determination is ______________. The standard error of the estimate is ______________. The *t*-ratios and their associated probabilities indicate that ______________________________________ are significant predictor variables. The regression coefficient of the x_3 variable is _____________. The value of the adjusted R^2 is ______________. This indicates ______________________.

ANSWERS TO STUDY QUESTIONS

1. *F*

2. *t*

3. Standard Error of the Estimate

4. R^2

5. Outliers

6. Adjusted R^2

7. $F = 0.21$ with $p = .889$, not, .081, 32.68, None of the Independent Variables, 2.20, .000, There is Virtually No Predictability in this Model

SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 13

13.1 The regression model is:

$\hat{y}$ **= 25.03 - 0.0497 x_1 + 1.928 x_2**

Predicted value of y for x_1 = 200 and x_2 = 7 is:

$\hat{y}$ = 25.03 - 0.0497(200) + 1.928(7) = **28.586**

13.3 The regression model is:

$\hat{y}$ = **121.62 - 0.174 x_1 + 6.02 x_2 + 0.00026 x_3 + 0.0041 x_4**

There are **four** independent variables. If x_2, x_3, and x_4 are held constant, the predicted y will decrease by - 0.174 for every unit increase in x_1. Predicted y will increase by 6.02 for every unit increase in x_2 as x_1, x_3, and x_4 are held constant. Predicted y will increase by 0.00026 for every unit increase in x_3 holding x_1, x_2, and x_4 constant. If x_4 is increased by one unit, the predicted y will increase by 0.0041 if x_1, x_2, and x_3 are held constant.

13.5 The regression model is:

Per Capita = -7,629.627 + 116.2549 Paper Consumption
- 120.0904 Fish Consumption + 45.73328 Gasoline Consumption.

For every unit increase in paper consumption, the predicted per capita consumption increases by 116.2549 if fish and gasoline consumption are held constant. For every unit increase in fish consumption, the predicted per capita consumption decreases by 120.0904 if paper and gasoline consumption are held constant. For every unit increase in gasoline consumption, the predicted per capita consumption increases by 45.73328 if paper and fish consumption are held constant.

13.7 There are 9 predictors in this model. The F test for overall significance of the model is 1.99 with a probability of .0825. This model is not significant at α = .05. Only one of the t values is statistically significant. Predictor x_1 has a t of 2.73 which has an associated probability of .011 and this is significant at α = .05.

13.9 The regression model is:

Per Capita Consumption = -7,629.627 + 116.2549 Paper Consumption - 120.0904 Fish Consumption + 45.73328 Gasoline Consumption

This model yields an F = 14.319 with p-value = .0023. Thus, there is overall significance at α = .01. One of the three predictors is significant. Gasoline Consumption has a t = 2.67 with p-value of .032 which is statistically significant at α = .05. The p-values of the t statistics for the other two predictors are insignificant indicating that a model with just Gasoline Consumption as a single predictor might be nearly as strong.

13.11 The regression model is:

$$\hat{y} = \mathbf{3.981 + 0.07322}\ \boldsymbol{x_1} - \mathbf{0.03232}\ \boldsymbol{x_2} - \mathbf{0.003886}\ \boldsymbol{x_3}$$

The overall F for this model is 100.47 with is significant at α = .001. Only one of the predictors, x_1, has a significant t value (t = 3.50, p-value of .005). The other independent variables have non significant t values (x_2: t = -1.55, p-value of .15 and x_3: t = -1.01, p-value of .332). Since x_2 and x_3 are non significant predictors, the researcher should consider the using a simple regression model with only x_1 as a predictor. The R^2 would drop some but the model would be much more parsimonious.

13.13 There are **3 predictors in this model and 15 observations**.

The regression equation is:

$\hat{y}$ **=** $\mathbf{657.053 + 5.7103\, x_1 - 0.4169\, x_2 - 3.4715\, x_3}$

F = 8.96 with a p-value of .0027
x_1 is significant at $\alpha = .01$ (t = 3.19, p-value of .0087)
x_3 is significant at $\alpha = .05$ (t = - 2.41, p-value of .0349)

The model is significant overall.

13.15 s_e = 9.722, R^2 = .515 but the adjusted R^2 is only .404. The difference in the two is due to the fact that two of the three predictors in the model are non-significant. The model fits the data only modestly. The adjusted R^2 indicates that 40.4% of the variance of y is accounted for by this model and 59.6% is unaccounted for by the model.

13.17 s_e = 6.544. R^2 = .005. This model has no predictability.

13.19 For the regression equation for the model using both x_1 and x_2, s_e = 6.333, R^2 = .963 and adjusted R^2 = .957. Overall, this is a very strong model. For the regression model using only x_1 as a predictor, the standard error of the estimate is 6.124, R^2 = .963 and the adjusted R^2 = .960. The value of R^2 is the same as it was with the two predictors. However, the adjusted R^2 is slightly higher with the one-predictor model because the non-significant variable has been removed. In conclusion, by using the one predictor model, we get virtually the same predictability as with the two predictor model and it is more parsimonious.

13.21 The Histogram indicates that there may be some problem with the error terms being normally distributed as does the Normal Probability Plot of the Residuals in which the plotted points are not completely lined up on the line. The Residuals vs. Fits plot reveals that there may be some lack of homogeneity of error variance.

13.23 There are two predictors in this model. The equation of the regression model is:

$$\hat{y} = 203.3937 + 1.1151\, x_1 - 2.2115\, x_2$$

The F test for overall significance yields a value of 24.55 with an associated p-value of .0000013 which is significant at $\alpha = .00001$. Both variables yield t values that are significant at a 5% level of significance. x_2 is significant at $\alpha = .001$. The R^2 is a rather modest 66.3% and the standard error of the estimate is 51.761.

13.25 The regression model is:

$$\hat{Y} = 362.3054 - 4.745518\, x_1 - 13.89972\, x_2 + 1.874297\, x_3$$

$F = 16.05$ with $p = .001$, $s_e = 37.07$, $R^2 = .858$, adjusted $R^2 = .804$. For x_1, $t = -4.35$ with $p = .002$; for x_2, $t = -0.73$ with $p = .483$, for x_3, $t = 1.96$ with $p = .086$. Thus, only one of the three predictors, x_1, is a significant predictor in this model. This model has very good predictability ($R^2 = .858$). The gap between R^2 and adjusted R^2 underscores the fact that there are two non-significant predictors in this model.

13.27 The regression model was:

Employment = 71.03 + 0.4620 NavalVessels + 0.02082 Commercial

F = 1.22 with p = .386 (not significant)

R^2 = .379 and adjusted R^2 = .068

The low value of adjusted R^2 indicates that the model has very low predictability. Both t values are not significant ($t_{\text{NavalVessels}}$ = 0.67 with p = .541 and $t_{\text{Commercial}}$ = 1.07 with p = .345). Neither predictor is a significant predictor of employment.

13.29 The regression model was:

Corn = -2718 + 6.26 Soybeans – 0.77 Wheat

F = 14.25 with a p-value of .003 which is significant at α = .01
s_e = 862.4, R^2 = 80.3%, adjusted R^2 = 74.6%
One of the two predictors, Soybeans, yielded a t value that was significant at α = .01 while the other predictor, Wheat was not significant (t = -0.75 with a p-value of .476).

13.31 The regression equation is:

$\hat{y}$ = 87.89 – 0.256 x_1 – 2.714 x_2 + 0.0706 x_3
F = 47.57 with a p-value of .000 significant at α = .001.
s_e = 0.8503, R^2 = .941, adjusted R^2 = .921.
All three predictors produced significant t tests with two of them (x_2 and x_3) significant at .01 and the other, x_1 significant at α = .05. This is a very strong model.