Interpreting the Coefficients of the Multiple Regression Model

In interpreting the coefficients of the model, we ask the question, *Do the signs and magnitudes of the estimated coefficients appear to be reasonable*?

In the simple linear regression model, the estimated coefficient, b_1, is the slope of the line. It is interpreted to be the average change in the dependent variable associated with a one-unit change in the independent variable. This interpretation remains basically valid for the multiple regression model as well. For the pizza delivery model, the coefficient b_1 is the estimated change in delivery time for a one-unit increase in the number of pizzas, given that distance traveled is constant. Is it reasonable to believe that each additional pizza would add approximately 1.6 minutes to the delivery time? While the delivery time varies, 1.6 minutes seems sensible.

The coefficient b_2 is the estimated change in delivery time for a one-unit increase in distance (measured in miles), given a specific number of pizzas (i.e., the number of pizzas to be delivered is constant). That is, for each additional mile added to the delivery, we should expect the average delivery time to increase by approximately 1.57 minutes. All other conditions being equal, do we find that the signs of the coefficients are reasonable? If we add more pizzas and distance traveled to the delivery route, it seems reasonable to expect the delivery time to increase. Thus, the positive signs on the coefficients seem to make sense. Were the signs of the coefficients unexpected, a reasonable question would be, *Is the estimate accurate?* Are there other factors that have not been considered that could reasonably change the signs of the coefficients? We will consider this question later in Section 14.6.

14.1 Exercises

Basic Concepts

1. Explain why a simple linear regression model might not always suffice when attempting to establish a relationship between variables in a business environment.
2. What is the general multiple regression model?
3. What are the assumptions about the error term in a multiple regression model? Are these different from the assumptions required for the simple linear model?
4. What method is used to find the estimated regression equation? Is this method different from the one used to find the simple linear regression equation?
5. What is the greatest challenge in building a multiple regression model?
6. What are some questions that should be asked once a multiple regression model is estimated? Give at least four.
7. In the simple linear regression model, what is the interpretation of b_1? Does this interpretation change in the multiple regression model?
8. When interpreting the coefficient of an independent variable in a multiple regression model, what assumption are we making regarding the other independent variables?
9. What two aspects of the model coefficients are usually analyzed first when studying a multiple regression model?

Exercises

10. Consider the following computer output of a multiple regression analysis relating annual salary to years of education and years of work experience.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.566946595
R Square	0.321428441
Adjusted R Square	0.29192533
Standard Error	10909.996
Observations	49

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2593556200	1296778100	10.89473033	0.000133875
Residual	46	5475288584	119028012.7		
Total	48	8068844784			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	11214.19915	5625.172956	1.993574106	0.052147881	-108.6867382	22537.08504
Education (Years)	2854.891271	689.6666061	4.139523715	0.000146836	1466.664395	4243.118147
Experience (Years)	839.6360369	261.7094444	3.208275646	0.002433357	312.842248	1366.429826

a. Identify the estimated values of the coefficients b_0, b_1, and b_2.

b. Write the estimated multiple regression equation.

c. Can you think of other independent variables that may be useful in predicting annual salary?

11. The manager of a publishing company would like to conduct cost analysis on the most recent books the company has published. He would like to estimate a multiple regression model to relate the cost of printing (per book) to the number of pages in the book and the number of copies printed. A computer output of the multiple regression model for the manager's data is given in the following table.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.987606014
R Square	0.975365639
Adjusted R Square	0.972467479
Standard Error	0.445885396
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	133.8201656	66.91008281	336.5464936	2.12863E-14
Residual	17	3.379834375	0.198813787		
Total	19	137.2			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	6.134155476	3.993435752	1.536059638	0.142925974	-2.291257484	14.55956844
Number of Pages	0.010801	0.004147682	2.604105041	0.018522101	0.002050156	0.019551845
Number of Copies	-0.009954478	0.005271436	-1.888380579	0.07616193	-0.021076236	0.00116728

a. Identify the estimated regression coefficients.

b. Write the estimated multiple regression equation.

c. Do the magnitudes and signs of the coefficients seem reasonable? Explain.

d. What other variables do you think could be useful in explaining printing cost per book?

12. A nutritionist wishes to study body weight based on height, age, average calories consumed per day, and the average number of minutes spent exercising per day.

- **a.** Write the multiple regression model the nutritionist is interested in in terms of weight, height, age, calories, and exercise. Assume the coefficients have not yet been estimated.
- **b.** Identify the independent variables in the multiple regression model.
- **c.** Predict the sign of the coefficient for each of the independent variables in the model. Explain your answers.
- **d.** Can you think of any other variables that might be useful for the nutritionist to take into account before performing the regression analysis?

13. Suppose the CEO of an electronics company wants to study the effects of various business practices on annual revenue.

- **a.** Make a list of independent variables the CEO might be interested in studying.
- **b.** Suppose the CEO has narrowed his list of factors down, and decided he wants to mainly study the effects of research and development expenditures, advertising expenditures, and the average annual salary paid to employees. Write the multiple regression model in terms of the dependent and independent variables, assuming the coefficients have not yet been estimated.
- **c.** Make a guess of the sign of the coefficient of research and development expenditures. Explain your prediction.
- **d.** Why should the CEO be cautious when using this model for revenue estimation and prediction?

14. Consider the following estimated multiple regression equation relating the number of study hours and GPA to a student's ACT score.

$$\text{Estimated ACT Score} = 8.35 + 1.53(\text{Study Hours}) + 0.30(\text{GPA})$$

- **a.** Identify the values of b_0, b_1, and b_2.
- **b.** Interpret the value of b_0 in terms of the problem.
- **c.** Interpret the value of b_1 in terms of the problem.
- **d.** Interpret the value of b_2 in terms of the problem.

15. Consider the following estimated regression model relating annual salary to years of education and work experience, which was presented in Exercise 10.

$$\text{Estimated Salary} = 11214.20 + 2854.89(\text{Education}) + 839.64(\text{Experience})$$

- **a.** Consider the coefficient for the education variable. Do the sign and magnitude of the coefficient seem to make sense? Explain.
- **b.** Consider the coefficient for the experience variable. Do the sign and magnitude of the coefficient seem to make sense? Explain.
- **c.** Interpret the regression coefficient for years of experience.
- **d.** Suppose an employee with 8 years of education (note that education years are the number of years after 8th grade) has been with the company for 5 years. According to this model, what is his estimated annual salary?
- **e.** How would you expect his salary to change if he stays at the company for another year?
- **f.** Suppose two employees at the company have been working there for five years. One has a bachelor's degree (8 years of education) and one has a master's degree (10 years of education). Which employee would you expect to earn a higher salary? How much more money would you expect him to make?

16. Suppose the owner of a car dealership wishes to study how certain factors affect the number of new cars sold. Specifically, he wishes to construct a multiple regression model relating car price, average income per capita in the surrounding area, and the interest rate to the quantity of new cars sold. After compiling historical data, he obtains the following summary output for the multiple regression model. In the original data, car price and average income per capita were in thousands of dollars, and interest rates were reported as percentages.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.581495041
R Square	0.338136482
Adjusted R Square	0.305043307
Standard Error	227.5372802
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	2361116.922	787038.9742	10.21771025	1.57438E-05
Residual	60	4621616.515	77026.94192		
Total	63	6982733.438			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-308.6097834	971.3332832	-0.317717707	0.751802188	-2251.565634	1634.346068
Price	-18.29845156	6.196928308	-2.952826085	0.004489897	-30.69415375	-5.902749358
Income	458.1641011	137.8510094	3.323618036	0.001518198	182.4210273	733.907175
Interest Rate	-26.29956832	11.93107528	-2.204291541	0.031350669	-50.16527221	-2.43386442

a. Write the estimated multiple regression equation.

b. Consider the coefficient of the price variable. Do you think the magnitude and sign of this coefficient make sense? Explain your answer.

c. Consider the coefficient of the income variable. Do you think the magnitude and sign of this coefficient make sense? Explain your answer.

d. Consider the coefficient of the interest rate variable. Do you think the magnitude and sign of this coefficient make sense? Explain your answer.

e. How many additional cars would the dealership owner expect to sell if per capita income for the area increased by $1000?

f. How would the dealership owner expect the quantity of cars sold to change if the interest rate increased by 2 percentage points?

14.2 The Coefficient of Determination and Adjusted R^2

Just as for simple linear regression, we will discuss methods that can be used to evaluate the overall effectiveness of multiple regression models. For the pizza delivery model in the previous section, one of the questions to ask is, how do we determine whether the model explains a substantial portion of the variation in the delivery times? The overall effectiveness and usefulness of multiple regression models can be addressed using the coefficient of determination (R^2) and the adjusted R^2 (R_a^2) statistics.

Coefficient of Determination (R^2)

Recall our discussion about the coefficient of determination (R^2) in the previous chapter. In Section 13.3, we defined the R^2 statistic as the statistic that directly measures the degree to