The coefficient, b_3, is the estimated change in *List Price* for an additional bedroom for a given square footage and age. Intuitively, we would expect this coefficient to be positive since additional finishing costs are associated with adding a bedroom. The coefficient for b_3 is 1284.92, which means that the price of a home will increase by $1284.92 for each bedroom, all else remaining equal. It is difficult to evaluate whether $1284.92 is a reasonable magnitude, so we will rely on statistical inference to evaluate the coefficients in the next section.

14.1 Exercises

Basic Concepts

1. What is the multiple regression model?
2. 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?
3. What method is used to find the estimated multiple regression equation? Is this method different from the one used to find the simple linear regression equation?
4. What is the greatest challenge in building a multiple regression model?
5. What are some questions that should be asked once a multiple regression model is estimated?
6. What two aspects of the model coefficients are usually analyzed first when studying a multiple regression model?
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?

Exercises

9. 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?

d. Use the model in part **b**. to predict the annual salary of someone with 12 years of education and 2 years of work experience.

10. The manager of a publishing company would like to conduct cost analysis on the most recent books the company has published. The manager 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?

11. 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?

12. 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?

13. Compare the following two estimated multiple regression equations that relate housing prices in thousands of dollars to the number of bedrooms in the house (*Bedrooms*), size of the lot (*Acreage*), and size of the house (*Square Footage*).

$$\hat{y} = -69{,}280.13 + 142{,}935.73\ Bedrooms + 369{,}879.29\ Acreage,$$

$$\hat{y} = -28{,}520.81 - 34{,}641.71\ Bedrooms + 194{,}986.08\ Acreage + 240.21\ Square\ Footage.$$

What happened to the coefficient on the *Bedrooms* term when the additional variable *Square Footage* was added? Does this make sense? Explain.

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

$$ACT\ Score = 8.35 + 0.015\ Study\ Hours + 0.30\ GPA$$

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 9.

$$Salary = 11{,}214.20 + 2854.89\ Education + 839.64\ Experience$$

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 their estimated annual salary?

e. How would you expect an employee's salary to change if they stay 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? What is the difference in salary between the two employees?