

Self-Test

- Before taking the self-test, refer to the learning objectives at the beginning of the chapter, the notes in the margins, and the glossary at the end of the chapter.
 - Use the key at the back of the book (see Appendix H) to correct your answers.
 - Restudy pages that correspond to any questions that you answered incorrectly or material you feel uncertain about.
1. One of the assumptions in regression analysis is that
 - a. the errors have a mean of 1.
 - b. the errors have a mean of 0.
 - c. the observations (Y) have a mean of 1.
 - d. the observations (Y) have a mean of 0.
 2. A graph of the sample points that will be used to develop a regression line is called
 - a. a sample graph.
 - b. a regression diagram.
 - c. a scatter diagram.
 - d. a regression plot.
 3. When using regression, an error is also called
 - a. an intercept.
 - b. a prediction.
 - c. a coefficient.
 - d. a residual.
 4. In a regression model, Y is called
 - a. the independent variable.
 - b. the dependent variable.
 - c. the regression variable.
 - d. the predictor variable.
 5. A quantity that provides a measure of how far each sample point is from the regression line is
 - a. the SSR.
 - b. the SSE.
 - c. the SST.
 - d. the MSR.
 6. The percentage of the variation in the dependent variable that is explained by a regression equation is measured by
 - a. the coefficient of correlation.
 - b. the MSE.
 - c. the coefficient of determination.
 - d. the slope.
 7. In a regression model, if every sample point is on the regression line (all errors are 0), then
 - a. the correlation coefficient would be 0.
 - b. the correlation coefficient would be -1 or 1 .
 - c. the coefficient of determination would be -1 .
 - d. the coefficient of determination would be 0.
 8. When using dummy variables in a regression equation to model a qualitative or categorical variable, the number of dummy variables should equal
 - a. the number of categories.
 - b. 1 more than the number of categories.
 - c. 1 less than the number of categories.
 - d. the number of other independent variables in the model.
 9. A multiple regression model differs from a simple linear regression model because the multiple regression model has more than one
 - a. independent variable.
 - b. dependent variable.
 - c. intercept.
 - d. error.
 10. The overall significance of a regression model is tested using an F test. The model is significant if
 - a. the F value is low.
 - b. the significance level of the F value is low.
 - c. the r^2 value is low.
 - d. the slope is lower than the intercept.
 11. A new variable should not be added to a multiple regression model if that variable causes
 - a. r^2 to decrease.
 - b. the adjusted r^2 to decrease.
 - c. the SST to decrease.
 - d. the intercept to decrease.
 12. A good regression model should have
 - a. a low r^2 and a low significance level for the F test.
 - b. a high r^2 and a high significance level for the F test.
 - c. a high r^2 and a low significance level for the F test.
 - d. a low r^2 and a high significance level for the F test.

Discussion Questions and Problems

Discussion Questions

- 4-1 What is the meaning of least squares in a regression model?
- 4-2 Discuss the use of dummy variables in regression analysis.
- 4-3 Discuss how the coefficient of determination and the coefficient of correlation are related and how they are used in regression analysis.

4-4 Explain how a scatter diagram can be used to identify the type of regression to use.

4-5 Explain how the adjusted r^2 value is used in developing a regression model.

4-6 Explain what information is provided by the F test.

4-7 What is the SSE? How is this related to the SST and the SSR?

4-8 Explain how a plot of the residuals can be used in developing a regression model.

Problems

- 4-9 John Smith has developed the following forecasting model:

$$\hat{Y} = 36 + 4.3X_1$$

where

\hat{Y} = demand for K10 air conditioners

X_1 = the outside temperature ($^{\circ}\text{F}$)

- (a) Forecast the demand for K10 when the temperature is 70°F .
 (b) What is the demand for a temperature of 80°F ?
 (c) What is the demand for a temperature of 90°F ?

- QM**: 4-10 The operations manager of a musical instrument distributor feels that demand for a particular type of guitar may be related to the number of YouTube views for a music video by the popular rock group Marble Pumpkins during the preceding month. The manager has collected the data shown in the following table:

YouTube VIEWS (1,000s)	GUITAR SALES
30	8
40	11
70	12
60	10
80	15
50	13

- (a) Graph these data to see whether a linear equation might describe the relationship between the views on YouTube and guitar sales.
 (b) Using the equations presented in this chapter, compute the SST, SSE, and SSR. Find the least-squares regression line for these data.
 (c) Using the regression equation, predict guitar sales if there were 40,000 views last month.
- QM**: 4-11 Using the data in Problem 4-10, test to see if there is a statistically significant relationship between sales and YouTube views at the 0.05 level of significance. Use the formulas in this chapter and Appendix D.
- QM**: 4-12 Using computer software, find the least-squares regression line for the data in Problem 4-10. Based on the F test, is there a statistically significant relationship between the demand for guitars and the number of YouTube views?
- QM**: 4-13 Students in a management science class have just received their grades on the first test. The instructor has provided information about the first test grades in some previous classes, as well as the final averages for the same students. Some of these grades have been sampled and are as follows:

STUDENT	1	2	3	4	5	6	7	8	9
1st test grade	98	77	88	80	96	61	66	95	69
Final average	93	78	84	73	84	64	64	95	76

- (a) Develop a regression model that could be used to predict the final average in the course based on the first test grade.
 (b) Predict the final average of a student who made an 83 on the first test.
 (c) Give the values of r and r^2 for this model. Interpret the value of r^2 in the context of this problem.
- QM**: 4-14 Using the data in Problem 4-13, test to see if there is a statistically significant relationship between the grade on the first test and the final average at the 0.05 level of significance. Use the formulas in this chapter and Appendix D.
- QM**: 4-15 Using computer software, find the least-squares regression line for the data in Problem 4-13. Based on the F test, is there a statistically significant relationship between the first test grade and the final average in the course?
- : 4-16 Steve Caples, a real estate appraiser in Lake Charles, Louisiana, has developed a regression model to help appraise residential housing in the Lake Charles area. The model was developed using recently sold homes in a particular neighborhood. The price (Y) of the house is based on the square footage (X) of the house. The model is

$$\hat{Y} = 33,478 + 62.4X$$

The coefficient of correlation for the model is 0.63.

- (a) Use the model to predict the selling price of a house that is 1,860 square feet.
 (b) A house with 1,860 square feet recently sold for \$165,000. Explain why this is not what the model predicted.
 (c) If you were going to use multiple regression to develop an appraisal model, what other quantitative variables might be included in the model?
 (d) What is the coefficient of determination for this model?
- : 4-17 Accountants at the firm Walker and Walker believed that several traveling executives submitted unusually high travel vouchers when they returned from business trips. The accountants took a sample of 200 vouchers submitted from the past year; they then developed the following multiple regression equation relating expected travel cost (Y) to number of days on the road (X_1) and distance traveled (X_2) in miles:

$$\hat{Y} = \$90.00 + \$48.50X_1 + \$0.40X_2$$

The coefficient of correlation computed was 0.68.

- (a) If Thomas Williams returns from a 300-mile trip that took him out of town for 5 days, what is the expected amount that he should claim as expenses?
- (b) Williams submitted a reimbursement request for \$685; what should the accountant do?
- (c) Comment on the validity of this model. Should any other variables be included? Which ones? Why?

4-18 Thirteen students entered the undergraduate business program at Rollins College 2 years ago. The following table indicates what their grade-point averages (GPAs) were after being in the program for 2 years and what each student scored on the SAT exam (maximum 2400) when he or she was in high school. Is there a meaningful relationship between grades and SAT scores? If a student scores a 1200 on the SAT, what do you think his or her GPA will be? What about a student who scores 2400?

STUDENT	SAT SCORE	GPA	STUDENT	SAT SCORE	GPA
A	1263	2.90	H	1443	2.53
B	1131	2.93	I	2187	3.22
C	1755	3.00	J	1503	1.99
D	2070	3.45	K	1839	2.75
E	1824	3.66	L	2127	3.90
F	1170	2.88	M	1098	1.60
G	1245	2.15			

4-19 Bus and subway ridership in Washington, D.C., during the summer months is believed to be heavily tied to the number of tourists visiting the city. During the past 12 years, the following data have been obtained:

YEAR	NUMBER OF	
	TOURISTS (1,000,000s)	RIDERSHIP (100,000s)
1	7	15
2	2	10
3	6	13
4	4	15
5	14	25
6	15	27
7	16	24
8	12	20
9	14	27
10	20	44
11	15	34
12	7	17

- (a) Plot these data and determine whether a linear model is reasonable.
- (b) Develop a regression model.
- (c) What is expected ridership if 10 million tourists visit the city?

(d) If there are no tourists at all, explain the predicted ridership.

4-20 Use computer software to develop a regression model for the data in Problem 4-19. Explain what this output indicates about the usefulness of this model.

4-21 The following data give the starting salary for students who recently graduated from a local university and accepted jobs soon after graduation. The starting salary, grade-point average (GPA), and major (business or other) are provided.

SALARY	\$29,500	\$46,000	\$39,800	\$36,500
GPA	3.1	3.5	3.8	2.9
Major	Other	Business	Business	Other
SALARY	\$42,000	\$31,500	\$36,200	
GPA	3.4	2.1	2.5	
Major	Business	Other	Business	

- (a) Using a computer, develop a regression model that could be used to predict starting salary based on GPA and major.
- (b) Use this model to predict the starting salary for a business major with a GPA of 3.0.
- (c) What does the model say about the starting salary for a business major compared to a nonbusiness major?
- (d) Do you believe this model is useful in predicting the starting salary? Justify your answer, using information provided in the computer output.

4-22 The following data give the selling price, square footage, number of bedrooms, and age of houses that have sold in a neighborhood in the past 6 months. Develop three regression models to predict the selling price based upon each of the other factors individually. Which of these is best?

SELLING PRICE (\$)	SQUARE FOOTAGE	BEDROOMS	AGE (YEARS)
84,000	1,670	2	30
79,000	1,339	2	25
91,500	1,712	3	30
120,000	1,840	3	40
127,500	2,300	3	18
132,500	2,234	3	30
145,000	2,311	3	19
164,000	2,377	3	7
155,000	2,736	4	10
168,000	2,500	3	1
172,500	2,500	4	3
174,000	2,479	3	3
175,000	2,400	3	1
177,500	3,124	4	0
184,000	2,500	3	2
195,500	4,062	4	10
195,000	2,854	3	3

PROBLEM 4-23 Use the data in Problem 4-22 and develop a regression model to predict selling price based on the square footage and number of bedrooms. Use this to predict the selling price of a 2,000-square-foot house with three bedrooms. Compare this model with the models in Problem 4-22. Should the number of bedrooms be included in the model? Why or why not?

PROBLEM 4-24 Use the data in Problem 4-22 and develop a regression model to predict selling price based on the square footage, number of bedrooms, and age. Use this to predict the selling price of a 10-year-old, 2,000-square-foot house with three bedrooms.

PROBLEM 4-25 The total expenses of a hospital are related to many factors. Two of these factors are the number of beds in the hospital and the number of admissions. Data were collected on 14 hospitals, as shown in the following table:

HOSPITAL	NUMBER OF BEDS	ADMISSIONS (100s)	TOTAL EXPENSES (\$1,000,000s)
1	215	77	57
2	336	160	127
3	520	230	157
4	135	43	24
5	35	9	14
6	210	155	93
7	140	53	45
8	90	6	6
9	410	159	99
10	50	18	12
11	65	16	11
12	42	29	15
13	110	28	21
14	305	98	63

Find the best regression model to predict the total expenses of a hospital. Discuss the accuracy of this model. Should both variables be included in the model? Why or why not?

PROBLEM 4-26 A sample of 20 automobiles was taken, and the miles per gallon (MPG), horsepower, and total weight were recorded. Develop a linear regression model to predict MPG, using horsepower as the only independent variable. Develop another model with weight as the independent variable. Which of these two models is better? Explain.

MPG	HORSEPOWER	WEIGHT
44	67	1,844
44	50	1,998
40	62	1,752

MPG	HORSEPOWER	WEIGHT
37	69	1,980
37	66	1,797
34	63	2,199
35	90	2,404
32	99	2,611
30	63	3,236
28	91	2,606
26	94	2,580
26	88	2,507
25	124	2,922
22	97	2,434
20	114	3,248
21	102	2,812
18	114	3,382
18	142	3,197
16	153	4,380
16	139	4,036

PROBLEM 4-27 Use the data in Problem 4-26 to develop a multiple linear regression model. How does this compare with each of the models in Problem 4-26?

PROBLEM 4-28 Use the data in Problem 4-26 to find the best quadratic regression model. (There is more than one to consider.) How does this compare to the models in Problems 4-26 and 4-27?

PROBLEM 4-29 A sample of nine public universities and nine private universities was taken. The total cost for the year (including room and board) and the median SAT score (maximum total is 2400) at each school were recorded. It was felt that schools with higher median SAT scores would have a better reputation and would charge more tuition as a result of that. The data are in the following table. Use regression to help answer the following questions based on the sample data. Do schools with higher SAT scores charge more in tuition and fees? Are private schools more expensive than public schools when SAT scores are taken into consideration? Discuss how accurate you believe these results are, using information related to the regression models.

CATEGORY	TOTAL COST (\$)	MEDIAN SAT
Public	21,700	1990
Public	15,600	1620
Public	16,900	1810
Public	15,400	1540
Public	23,100	1540
Public	21,400	1600
Public	16,500	1560
Public	23,500	1890

(Continued on next page)