distance that is driven. The interpretation of R_a^2 is the same—96.08% of the variation in delivery time is explained by the two independent variables in the model.

c. With both number of pizzas and distance in the model, the value of R^2 increased by 0.0374, indicating that adding the variable distance to the model helped explain more variability in delivery times. The value of R_a^2 increased by slightly more (0.0375). Using both variables in the model explained nearly 4% more variability in delivery time.

As the number of independent variables increases, the difference between the R^2 and adjusted R^2 values also increases. R_a^2 is commonly used as a method of comparison between multiple regression models when one is attempting to find the model that best fits the data. Unlike the R^2 value, the adjusted coefficient of determination may actually become smaller when another independent variable is added to the model. Thus, the adjusted R^2 value is most useful when comparing multiple regression models with different numbers of independent variables.

14.2 Exercises

Basic Concepts

1. What is the purpose of the R^2 and adjusted R^2 statistics?

2. What values can the coefficient of determination take?

3. If a particular regression model explains 68% of variation in the dependent variable, what is the value of R^2?

4. If the coefficient of determination has a value of zero, is it possible for a regression coefficient to have a value other than zero? Explain why.

5. If the coefficient of determination has a value of one, what is the relationship between the sum of squares of regression and the total sum of squares? Explain why.

6. Does a large value of R^2 always indicate that the fitted model is useful? Explain.

7. Explain the difference between R^2 and adjusted R^2.

8. Explain why the adjusted R^2 statistic is sometimes a better measure to use to evaluate the fit of a regression model.

9. Will there ever be a situation in which the adjusted R^2 statistic is greater than R^2 statistic? Explain your answer.

Exercises

10. Consider the following ANOVA table for a multiple regression model relating housing prices (in thousands of dollars) to the number of bedrooms in the house and the size of the lot on which the house was built (in square feet). There were 88 total observations.

$$\text{Estimated Price} = 63.26 + 57.31(\text{Bedrooms}) + 0.0029(\text{Lot Size})$$

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	309148.8902	154574.4451	21.58486376	2.62677E-08
Residual	85	608705.618	7161.242565		
Total	87	917854.5083			

a. Identify the values of SSR, SSE, and TSS from the table.

b. What is the coefficient of determination for this model? Interpret this value in terms of the problem.

c. What is R_a^2 ? Interpret this value.

d. Compare the R^2 and R_a^2 values. Which value should be used to evaluate the fit of the multiple regression model? Explain why.

11. Suppose an additional variable, Square Feet, was added to the housing price model from Exercise 10. The summary output is given below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.819976968
R Square	0.672362228
Adjusted R Square	0.660660879
Standard Error	59.83347988
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	617130.7018	205710.2339	57.46023188	2.69597E-20
Residual	84	300723.8065	3580.045315		
Total	87	917854.5083			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-21.7703086	29.47504196	-0.738601446	0.462207782	-80.38466199	36.84404478
Bedrooms	13.85252186	9.010145446	1.537435988	0.127945059	-4.065140472	31.7701842
Lot Size	0.002067707	0.000642126	3.220095719	0.001822929	0.000790769	0.003344644
Square Feet	0.122778185	0.013237407	9.275092996	1.65802E-14	0.096454149	0.149102222

a. What is R_a^2 for this model?

b. How does the adjusted R^2 value for this model compare to the adjusted R^2 value for the model in Exercise 10?

c. Do you think adding the additional independent variable, Square Feet, improved the model? Explain your answer.

12. The owner of a new pizzeria in town wants to study the relationship between weekly revenues and advertising expenditures. Both measures were recorded in thousands of dollars. The computer output for the simple linear regression model is given below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.858179902
R Square	0.736472743
Adjusted R Square	0.692551534
Standard Error	1.058296197
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18.78005496	18.78005496	16.76804334	0.006394067
Residual	6	6.719945042	1.11999084		
Total	7	25.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	74.69887795	7.104358625	10.51451396	4.34789E-05	57.31513863	92.08261726
Advertising Expenditures	1.854820243	0.452960815	4.094880138	0.006394067	0.746465058	2.963175428

a. Write the estimated regression equation.

b. What is the coefficient of determination for this model? Interpret this value.

c. What is the value of the adjusted R^2 statistic? Is this statistic useful for the pizzeria owner as he studies this model? Explain.

d. Do you believe this model is useful in explaining revenues based on advertising expenditures? Explain your answer.

e. How could the restaurant owner improve this model? Are there other independent variables that he should consider including?

13. The owner of the pizzeria discussed in Exercise 12 wishes to build on the model relating revenues to advertising expenditures by breaking the advertising expenditures into three categories: television advertising, newspaper advertising, and direct mail advertising.

a. Write the new regression model in terms of television, newspaper, and mail expenditures. Assume the coefficients have not yet been estimated.

b. Consider the following summary output for the new model. Write the estimated multiple regression equation.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.967040091
R Square	0.935166537
Adjusted R Square	0.88654144
Standard Error	0.64289449
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	23.8467467	7.948915566	19.23217829	0.007708883
Residual	4	1.653253302	0.413313326		
Total	7	25.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	73.93199827	4.523870838	16.34264127	8.20538E-05	61.37171922	86.49227731
Television	2.383047934	0.318133378	7.490719616	0.001698799	1.499768074	3.266327793
Newspaper	1.454439994	0.355820285	4.087569076	0.015004989	0.466524505	2.442355483
Mail	1.815990841	0.276487962	6.568064755	0.002780349	1.048337191	2.58364449

c. Interpret the coefficient for television advertising expenditures. Remember that revenues and expenditures are in thousands of dollars.

d. What is the adjusted coefficient of determination? Interpret this value.

e. How does the coefficient of determination of this model compare to the coefficient of determination for the simple linear regression model in Exercise 12? Does this appear to be a more useful model? Explain.

f. What is the value of the R^2 statistic for this model? Should we use the R^2 value or the adjusted R^2 value when evaluating the usefulness of this model? Explain why.