

the old equation to determine which equation is the best model for the given data. Most importantly, however, we should remember that a multiple regression model does not imply causation any more than a regression model with one explanatory variable does. Just as we know that an increase in a child's age does not *cause* the child's reading level to increase, we should not presume that adding more variables to a regression equation accounts for all the possible influences on the response variable, or that the variables included are even *causing* the changes in the response variable at all.

12.4 Section Exercises

Interpreting ANOVA Tables

The following ANOVA tables were generated using Microsoft Excel. Based on the ANOVA table given, is there sufficient evidence at the 0.05 level of significance to conclude that the linear relationship between the explanatory variables and the response variable is statistically significant?

1.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	2677.934566	1338.967283	167.6684672	5.21957E-10
Residual	13	103.8154339	7.985802609		
Total	15	2781.75			

2.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	4.272742323	2.136371162	8.88737913	0.004285466
Residual	12	2.88459101	0.240382584		
Total	14	7.157333333			

3.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	1691672.173	563890.7244	11.50954064	0.080992754
Residual	2	97986.66027	48993.33013		
Total	5	1789658.833			

4.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	8.553919903	4.276959952	3.555389644	0.061293877
Residual	12	14.43541343	1.202951119		
Total	14	22.98933333			

For each of the following computer outputs, write the multiple regression equation. Define each of the variables used in the multiple regression equation. Could one of the explanatory variables be eliminated from the model? If so, which one?

5. The following table was generated from a sample of 16 statistics students regarding each student's number of absences in the class, average number of hours spent studying for the class per week, and final grade in the class.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	63.88047809	4.914092338	12.99944602	7.95988E-09
Absences	-1.416204746	0.422223932	-3.354155552	0.005179159
Hours Studied per Week	4.168586523	0.73626183	5.661826206	7.77404E-05

6. The following table was generated from a sample of 16 babies regarding the weight of the mother at birth, the weight of the father at birth, and the weight of the baby at birth. All weights are measured in pounds.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-2.882473037	0.9737576	-2.960154597	0.01105403
Mother's Weight	0.356055126	0.090343376	3.941131517	0.001689591
Father's Weight	1.02230869	0.099308161	10.29430688	1.28376E-07

7. The following table was generated from a sample of 15 babies regarding the number of weeks of gestation, the number of prenatal doctor visits the mother had, and the birth weight of the baby in pounds.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-19.56998222	5.716106062	-3.423656245	0.005044399
Weeks of Gestation	0.684805526	0.12588231	5.44004576	0.000150053
Number of Prenatal Visits	0.063077042	0.126763644	0.497595684	0.627761713

8. The following table was generated from a sample of 20 college freshmen regarding their high school GPAs, their ACT scores, and their GPAs after their first year at college.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-0.616112354	0.345507101	-1.783211841	0.092415337
High School GPA	0.975068267	0.186279465	5.234437779	6.7297E-05
ACT Score	0.015000337	0.019048101	0.787497806	0.441832554

Constructing and Analyzing a Multiple Regression Model

For each data set, use available technology, such as Microsoft Excel, to determine if a statistically significant linear relationship exists between the explanatory variables and the response variable at the 0.05 level of significance. If the linear relationship is significant, identify the multiple regression equation that best fits the data.

9. The following data were collected to explore how tuition costs, student-to-faculty ratios, and extra time to graduation index affect freshman enrollment. The student-to-faculty ratio is expressed as the number of students per faculty member. The extra time to graduation index is expressed as the average number of semesters students take above the normal time to completion.

Effects on Freshman Enrollment			
Tuition Cost (in Dollars), x_1	Student-to-Faculty Ratio, x_2	Extra Time to Grad Index, x_3	Freshman Enrollment, y
5495	18	1.5	3219
7912	18	2	2781
5812	6	1.7	3556
7180	4	2.1	2987
8883	15	3	1794
6495	19	1.8	3014

10. The following data were collected to explore how the average number of hours a junior high student is unsupervised per night and the average number of hours per night the student watches television affect his class average.

Effects on Junior High Students' Class Averages		
Hours Unsupervised, x_1	Hours Watching TV, x_2	Class Average, y
0	2	96
0	3	91
0.5	3	88
1	2	92
1	1	94
1.5	2	91
2	4	87
3	3	85
3	4	81
4	5	80
4.5	4	77
5.5	5	72

11. The following data were collected to explore how people's ages and their average daily calcium intakes (measured in mg per day) affect their bone mineral density.

Effects on Bone Density		
Age (in Years), x_1	Daily Calcium Intake (in mg per Day), x_2	Bone Density (in mg/cm ²), y
34	791	950
43	832	875
47	865	804
52	901	723
58	923	691

12. Common predictors for freshman grade point average are ACT scores and high school GPA. Let's explore how well two unusual factors, a student's age when entering college and the number of parking tickets the student receives during the student's first semester at school, predict a student's GPA after one semester in college.

Effects on GPA		
Student's Age, x_1	Number of Parking Tickets, x_2	GPA, y
20	0	3.6
21	0	3.9
17	0	2.4
18	0	3.1
19	1	3.5
35	1	4.0
23	1	3.6
19	2	2.8
18	2	3.0
18	2	2.2
24	3	3.9
19	3	3.1
18	5	2.1
17	7	2.8
18	8	1.7

Challenge Questions

Answer each of the following questions.

13. Which value in the ANOVA table tells you the number of explanatory variables in the multiple regression model?
14. Should you always use as many explanatory variables as possible?
15. What is the disadvantage of the multiple coefficient of determination?