

Questions 1-7 refer to the following information:

A graduate-level Statistics course has 40 enrolled students. At the end of the semester, the professor of the course decides to examine some factors that may predict a student's success (or lack thereof) on the final exam by running a multiple regression. The professor sets up the initial regression model with just two predictor variables:

- Absences – the number of class sessions that each student missed throughout the semester
- Midterm – the score that each student obtained on the course's midterm exam (out of 100)

The dependent variable for the model is the score that each student obtained on the final exam (out of 100). The average score on the final exam was 81 with a standard deviation of 12.252.

Running this initial regression yields the following results:

Regression Statistics			
Multiple R	0.587		
R Square	0.345		
Adjusted R Square	0.309		
Standard Error	10.182		
Observations	40		

Summary Table	Unstandardized Coefficient		Standardized Coefficient
	B	Standard Error	Beta
Intercept	51.782	11.128	
Absences	-2.146	0.903	-0.322
Midterm	0.419	0.131	0.432

ANOVA			
	<i>df</i>	<i>F</i>	<i>Significance F</i>
Regression	2	9.734	0.00040
Residual	37		
Total	39		

1) What is the statistical significance of this regression model?

2) If a student missed just one class session and scored an 80 on the midterm, what would you estimate she would score on the final exam?

3) Two students, Jennifer and Sheera, had the same score on the midterm, but Jennifer was absent for three more class sessions than Sheera. Provide a 95% confidence interval for the amount that you would expect Jennifer's score to be below Sheera's.

4) One of the professor's teaching assistant's claims that every missed class will lower a student's score on the final exam by at least four points. Does the regression provide enough evidence to reject this claim with 99% certainty?

5) Two students, James and Winston, had the same number of absences throughout the semester but Winston's score on the midterm exam was exactly one standard deviation higher than James'. How many points higher than James would you predict Winston to score on the final exam?

6) The professor decides to test a new model with one additional predictor variable included. The new variable, Study Hours, is the number of hours that each student reported studying for the final exam in the week leading up to the test. The results of the new regression model are below:

Regression Statistics			
Multiple R	0.594		
R Square	0.353		
Adjusted R Square	0.299		
Standard Error	10.258		
Observations	40		

Summary Table	Unstandardized Coefficient		Standardized Coefficient
	B	Standard Error	Beta
Intercept	46.056	14.087	
Absences	-2.119	0.911	-0.318
Midterm	0.442	0.137	0.456
Study Hours	0.523	0.779	0.093

ANOVA			
	<i>df</i>	<i>F</i>	<i>Significance F</i>
Regression	3	6.543	0.00120
Residual	36		
Total	39		

Has the inclusion of the new predictor variable improved or worsened the predictive power of the regression?

7) Suppose that students in the course come from four academic disciplines: Psychology, Education, Business, and Communications. The professor wants to measure whether there are differences in students' final exam scores across academic disciplines. To that end, she decides to include Academic Discipline among the predictor variables in the regression. In order to do so, how many new variables will she need to add to the regression model?

Questions 8 & 9 refer to the following information:

The owner of a neighborhood bar collects data on the nightly sales of alcoholic drinks over the course of nine weeks. During those nine weeks, each of the bar's three part-time bartenders worked twenty-one nights. The data collected on nightly drink sales by bartender are below:

Bartender	Beverly	Jill	Kevin
Average	\$1,161.43	\$961.38	\$867.67
Std. Dev.	\$481.97	\$368.98	\$268.57

8) Using the following website, <http://www.danielsoper.com/statcalc3/calc.aspx?id=43>, does ANOVA indicate with 95% certainty that there is a difference in effectiveness among the three bartenders?

9) Because he is interested in evaluating alcohol sales by the days of the week as well, the bar's owner decides to run a 2-way ANOVA using the bartenders and the days of the week as the two factors. The results of the new ANOVA are below:

ANOVA				
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>F</i>	<i>P-value</i>
Day of the Week	2278505.968	6.000	8.029	0.000008
Bartender	945682.317	2.000	9.998	0.000281
Interaction	4546478.127	12.000	8.011	0.000001
Within	1986420.667	42.000		

Do these results indicate that some bartenders are more or less effective on certain days of the week?

10) The CEO of an equipment manufacturing company collects the following data about the company's monthly sales and promotional activities throughout the year:

	Sales Revenue (thousands)	Trade Shows Attended	Marketing Spend (thousands)
Jan	\$98	3	\$7
Feb	\$110	5	\$12
Mar	\$114	5	\$13
Apr	\$134	5	\$16
May	\$144	7	\$13
Jun	\$133	5	\$12
Jul	\$130	4	\$14
Aug	\$148	8	\$14
Sep	\$135	7	\$10
Oct	\$112	4	\$10
Nov	\$103	3	\$9
Dec	\$120	4	\$16

The CEO wants to run a regression using this data to predict how promotional activities affect sales- but also wants to understand how the marketing spend from the previous month may affect the current month's sales. With the inclusion of this additional factor, what is the Adjusted R^2 for the regression model?