

17.801, Political Science Lab
Spring 2001
Problem set # 3: Bivariate and multiple regression

Handed out: March 6, 2001
Due back: March 20, 2001

When you hand back in your problem set, please estimate the number of hours (rounded to the nearest quarter hour) it took you to complete it.

Part I.

Do the following Review Exercises in Freedman, **3rd edition**:

Chapter 9 (pp. 153–157), # 4, 7, 10
Chapter 10 (pp. 176–178), # 1, 3, 6, 7
Chapter 11 (pp. 198–201), # 1, 2, 6, 9
Chapter 12 (pp. 213–216), # 3, 4, 8, 11

Part II.

General directions. The following problems present you with real-life research situations and ask you to make judgements about either the data you have and what they tell you *or* the data you would need to answer the question presented you. There are no trick questions here.

Each of the questions asks you to write something to explain what you did. Please take the written assignments seriously, because you will be graded on quality of writing and substance. Accompanying most of the questions you should hand in a log file that shows the results you are talking about and a “do” file that could reproduce those results if necessary.

- A. The MIT administration is interested in studying why freshmen make the dormitory choices they do. There is a debate over whether freshmen care more about the size of the rooms in various dormitories (larger rooms are better) or about the physical condition of the building (newer is better). Being empirical scientists at MIT, they decide to settle this matter using a multiple regression. They gather data about dormitory preferences, average room sizes, and age of the buildings. The (mostly fictional) data are presented on the accompanying table.

Using these data, do the following.

1. Run the multiple regression necessary to answer the above debate.

2. Run two separate bivariate regressions on the same data, first with room size as the independent variable and then building age. Compare the coefficients you get with the multiple regression with those you get from the two bivariate regressions.
 3. Write a (clearly worded) paragraph that explains why the coefficients in the two analyses are so different. Cite the parts of the assigned readings that address this issue. Explain which sets of coefficients you trust and why.
- B. You are interested in the relationship between the amount of money spent by congressional candidates and the votes they receive on election day. You decide to pursue this topic by studying the 1998 U.S. Senate election. The data you gather are the number of votes received by incumbents and challengers running in 1998, plus the amount of money spent in 1998 by these candidates. The data are in /mit/17.801/Examples/senate98.dta. (These are real data.)
- a. What is the effect of challenger spending on the *number of votes* received by the incumbent? In answering this question, specify the most appropriate transformation (if any) of the dependent and independent variables. Turn in the log file that shows the regression you ran and a paragraph that summarizes your answer.
 - b. What is the effect of challenger spending on the *percentage of the two-party vote* received by the incumbent? In answering this question, specify the most appropriate transformation (if any) of the dependent and independent variables. Turn in the table that shows the regression you ran and a paragraph that summarizes your answer.
 - c. From these data, what is the short answer to the question, “what is the effect of challenger spending on Senate election outcomes?”?

Data set for problem II-AA: Dormitory preference. (Mostly fictitious data)

Dormitory name	# of first choices in freshman housing lottery, 1999	Year building was built	Average sq. ft. of dormitory room space per resident
Baker	120	1949	145
Bexley	31	1920	107
Burton-Connor	50	1940	135
East Campus	59	1930	127
MacGregor	125	1970	150
McCormick	81	1965	200
New	56	1976	175
Next	98	1981	185
Random	27	1910	97
Senior	75	1916	125