17.871, Political Science Lab
Spring 2002
Problem set # 3: Bivariate regression

Handed out: February 28, 2002
Due back: March 14, 2002

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
>   Chapter 10 (pp. 176–178), # 1, 3, 7
>   Chapter 11 (pp. 198–201), # 1, 2, 6, 9
>   Chapter 12 (pp. 213–216), # 4, 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 data file in /mit/17.871/Examples/abortion.dta contains two variables about religious practice and attitudes about abortion, from the 2000 American National Election Study. The coding is given at the end of the problem set. Do the following:

>   1.   Run a simple linear regression that explains abortion attitudes in terms of frequency of attendance at religious services. (By "simple" I mean don't do any recoding of the variables. Take them as given.) Report the coefficients. Interpret what the coefficients mean.

>   2.   Run a second regression that uses at least one transformation of the data to take into account non-linearities. Interpret these results.

3.   Draw a single graph that illustrates the estimated effect of religious practice on abortion attitudes from the previous two steps. In substantive terms, compare the results.

B.   You have been hired by the American Library Association to help lobby in favor of funding to buy more library books. You gather data from the U.S. Department of Education about reading test scores (measured by the National Assessment of Educational Progress, "NAEP") and the number of school library books in each state (measured in terms of books per 100 students). This data is in the file /mit/17.871/Examples/reading.dta. Using these data, do the following

1.   Run a regression that describes the relationship between reading scores and library books. Report any transformation of the variables you use to get the relationship "right."

2.   Examine the residuals from the regression you performed in the previous step. Discuss any patterns in the residuals that suggest systematic errors in the regression that might bear further analysis.

3.   The average reading score is 210.85. Using the regression you performed in step 1, suggest how you might raise the average reading score by 5 points by buying more library books.

C.   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/17.871/Examples/senate98.dta. (These are real data.)

1.   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 variable. Turn in the log file that shows the regression you ran and a paragraph that summarizes your answer.

2.   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 variable. Turn in the table that shows the regression you ran and a paragraph that summarizes your answer.

3.   From these data, what is the short answer to the question, "what is the effect of challenger spending on Senate election outcomes?"?

Coding of variables in abortion.dta

abortion:

"Which one of the opinions on this page best agrees with your view [on abortion]?

1.   By law, abortion should never be permitted.
2.   The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.
3.   The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.
4.   By law, a woman should always be able to obtain an abortion as a matter of personal choice.

attend:

How often do you attend religious services? [This scale is created based on three questions which asks (1) whether you ever go and (2) how often you go, if you answer "yes" to (1), and (3) whether you go weekly or more often than weekly depending on an answer to (2).]

0   More often than once a week
1   Once a week
2   Almost every week
3   Once or twice a month
4   A few times a year
5   Never