17.801 Spring 2001 Problem Set # 2: Descriptive Statistics and Bivariate Regressions

Handed out: February 27, 2001 Due back: March 6, 2001

Part A.

Do the following review exercises in Freedman, et al., **3rd edition**. When an exercise calls on you to explain something, do so legibly, using complete sentences, paragraphs, etc. Typing is preferred if the explanation is more than a sentence in length.

Chapter 3 review exercises (pp. 50–55), # 1, 2, 4, 8, 9, 10 Chapter 4 review exercises (pp. 74–76), # 1, 2, 3, 5, 6, 7, 10 Chapter 8 review exercises (pp. 134–139), # 1, 2, 3, 6, 7, 11

Part B.

The data on the following page is a random sample taken from responses to the Spring1995 Subject Evaluation Guide at MIT.

- 1. What is the correlation between the overall rating of the subject and the subject's enrollment?
- 2. What is the correlation between the overall rating of the subject and the *log* of the subject's enrollment?
- 3. Discuss which correlation is the better summary of the relationship between subject size and class quality.

For this part, show the calculations that produce your results for B-1 and B-2.

Part C.

Create a STATA data set of the data you hand-analyzed in Part B.

- 1. What does STATA calculate as the correlation between overall rating of the subject and the subject's enrollment (to 3 decimal places)?
- 2. What does STATA calculate as the correlation between the overall rating of the subject and the *log* of the subject's enrollment (to 3 decimal places)?
- 3. Generate two graphs: (a) Overall subject rating against enrollment and (b) Overall subject rating against the *log* of enrollment. Does this graph cause you to change your answer to B-3. Why or why not? (Hint 1: When I say "generate a graph of y against x," y is the dependent variable and x is the independent variable. Hint 2: In drawing graph (b), use the ",xlog" option to the graph command instead of generating the log of enrollment and plotting against that.)

For this part, turn in a log file that shows (1) the variable names in the data set, (2) the averages of the variables, and (3) the correlations among the relevant variables. Also, turn in printed versions of the graphs in C-3.

Estimate the amount of time it took to do this problem set: _____. ___ hours.

Subject Evaluation Data

The following are (real) data, randomly-sampled from the Spring 1995 Subject Evaluation Guide. Students were asked to rate, on a 7-point scale (1 = poor, 7 = excellent), the overall quality of the subject. Subject enrollment is what was reported by the Registrar.

ID #	Rating	Enrollment
1	6	56
2	5.9	52
3	4.7	18
4	5.4	45
5	6.5	13
6	3.8	501
7	5.8	10
8	6	50
9	6	17
10	5.5	25
11	4.9	22
12	5.2	29
13	6	39
14	5.3	14
15	6.3	22
16	6.6	15
17	5.3	116
18	6.4	13
19	5.4	30
20	4.2	65
21	6.1	20
22	5.5	28
23	6.2	22
24	5	7
25	5.7	25
26	6.6	19