PS207 Homework #4:
Selection Models

There is a Stata dataset on my website called ps207hw4. Download it and do the following two problems. Give me a couple of paragraphs writeup on each problem, as well as the log file. Obviously this is made-up data, so you can't do too much in the way of substantive interpretation, but describe the results as best you can.

Problem 1

Suppose you are interested in estimating the following regression:

$$Y = \alpha_1 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \varepsilon_1$$

Run this regression and describe the results.

After running this regression, someone points out that there could be a sample selection problem in this case, with this selection mechanism:

$$Z = \alpha_2 + \beta_4 X4 + \beta_5 X5 + \varepsilon_2$$

Estimate and describe the appropriate model for this sample selection problem. Does this change your earlier regression results? How? Was sample selection a problem here?

Problem 2

$Y2$ is a binary version of Y from problem 1. Suppose you wish to estimate the same outcome equation as in problem 1, but replacing Y with $Y2$. Estimate a probit for this model and describe the results.

Of course, sample selection may still be a problem. Now estimate a model that takes sample selection into account, and describe the results as you did in problem 1.