Problem Set 4
Limited Dependent Variables

1. Download dataset at http://www.law.upenn.edu/fac/jklick/PS4.txt which is a 1% random sample from PS3.txt
2. Run regression (linear probability model) of diabetes on bmi and income.
3. Run regression (linear probability model) of diabetes on bmierr and income. Remember bmierr is bmi plus a mean zero, variance 100 random variable drawn from a normal distribution. Note the change in coefficients. Intuitively, why does using an x with random measurement error cause problems that are not present when y is measured with random error? What would happen to your estimated relationship between diabetes and bmi if bmi were measured with an error of mean zero and variance of 10, relative to what you found for the relationship with bmierr? Explain your intuition.
4. Re-do #2 with a logit model. Re-do with a probit model.
5. Graph the predicted values from #2, and from the two models in #4 with bmi on the x axis.
6. bmipriv reports a BMI of 35 for anyone whose true BMI is greater than 35 to protect those individuals’ privacy. Regress, using the appropriate model, bmipriv on income and education. Now regress bmi on income and education. Provide some intuition for the different magnitudes you estimate for the effects in these two regressions.
7. bminormal is bmi for the subset of the population whose bmi lies in the range from 18.5 to 25. For everyone else, bmi is listed as missing. Regress, using the appropriate model, bminormal on income and educa. Note that BMI is approximately normally distributed. Provide some intuition for the difference in your estimates using bminormal and bmi.