Problem set 10, due Monday 11/23/2015

1. Get the data

a) Go to https://www.openintro.org/stat/extras.php

b) Download the "100 data sets", and extract the files.

2. Regress SAT scores on high school GPA

a) Open the "satGPA" data set in Stata. Regress SAT scores (dependent variable) on high school GPA (independent variable).

b) Still in Stata, draw a scatter plot with linear fit that illustrates your regression results.

c) Open the same data set in Excel, perform the same regression, and draw the same graph. Find the OLS estimate of the slope, intercept, standard error of the slope estimate, 95% confidence interval of the slope, and p value of the slope coefficient.

d) In words, write briefly what your results mean.

3. Regress race finishing time on age

a) Open the "run10" data set in Stata. Destring the "age" variable using the command `destring age, replace ignore("NA")`. Regress race time (dependent variable) on age (independent variable).

b) Still in Stata, draw a scatter plot with linear fit that illustrates your regression results.

c) Open the same data set in Excel, perform the same regression, and draw the same graph. Find the OLS estimate of the slope, intercept, standard error of the slope estimate, 95% confidence interval of the slope, and p value of the slope coefficient.

d) In words, write briefly what your results mean.

4. Submit your assignment

Submit your assignment by e-mail *before class* on Monday, 11/23. In the text of your e-mail, include your regression output tables from Stata, and the interpretation of your results. Include your Excel work as an attachment.