

## Problem set 8, due Wednesday 11/11/2015

### **1. Find and extract the data**

Go to <https://www.openintro.org/stat/extras.php>

From there, click on --> “Data and Analysis” --> “100 Data Sets (zip)”

Then, unzip/extract the file, which should become a folder with subfolders that contain the advertised 100 data sets, in Windows and Mac formats.

### **2. Find the documentation, and choose a data set**

In order to make sense of the files, you should open the "guides" folder that comes with the data, and then open the file “openintro.pdf”. This will tell you what variables in each data set mean. With this in hand, choose a data set for this assignment – you can use any of them except for the “census” data that we used in class.

### **3. Open the data in Stata**

One of the Henderson labs has Stata. From Stata, go to --> “File” --> “Import”, and then see which of those options can open the data files that you have correctly.

### **4. Explore the data**

Learn about your data using commands like “des”, “sum”, “browse”. Convert variables into more convenient formats using commands like “destring”, “generate”, “replace”. Create graphs using commands like “hist”, “graph box”, “scatter”. Construct confidence intervals around means using the “mean” command. Experiment with other commands.

### **5. Submit report including commands, narrative, and graphs**

Please submit your assignment by e-mail. In the text, send your command history (edited to remove missteps etc.), tables and a brief description of your findings. Include your graphs as .TIF attachments. You don't need to do anything too ambitious here; the main point is just to open up the data in Stata and start playing with it. So let's say at least two graphs, one summary table of the data, one confidence interval around a mean (again, a table), and the rest is up to you.