Dealing with data by hand or even with a calculator can be tedious. Working with appropriate statistical software enables us to explore the data and to deepen our understanding of statistics. In this lab we will consider demographic data from Los Angeles County. *Stata* encourages us to focus on the story that the data are telling us.

**Opening Stata**

After you have logged on, using your lab ID as your name and your nine digit UCLA ID as your password, you need to double click to open the *Stata* icon. You will see four windows: Review, Variables, Stata results, and Command. You may want to resize the windows to fit the screen. To change the font size, you need to go to Preferences or Prefs.

To begin, type

```
. describe
```

in the Command window and press enter. All commands (written in bold type following the dot) will be typed in the Command window. *Stata* is line command oriented which makes it fast with lots of memory free for data. You must be careful to type each command exactly as written, but without the dot.

*Question 1: Since there is no data, what is written in the Stata results window?*

Now we will open a file that contains data that we will use several times in our course.

```
. use http://www.stat.ucla.edu/labs/datasets/smallcen.dta
```
Green words mean that the typing is fine while red words mean that the typing needs to be modified. Type

```
. describe
```
or take a short cut by using the page up option on the keyboard.

**Question 2: What are these data?**

Type return to scroll line by line and space bar to scroll down page by page.

Since we want to look at the individual cases, we type

```
. list
```

**Question 3: What do we know about these individuals?**

Scrolling through 2500 observations is tedious. To break the scroll, we use `.q` for quit or press Apple and period at the same time.

It is hard to concentrate with so many variables and cases, so let us focus on gender and monthly income. You can look at the data by typing

```
. list gender income
```

**Question 4: What does the 0 represent?**

Let us look at the income for the youngest ten people. Type

```
. sort age
```
And then type

```
. list age income in 1/10
```
Question 5: What is the typical income?

Describe what we see when we type
\[ \text{summarize income} \]

We want to look at mean income for those who have an income. To remove 0’s we code them as missing values and type
\[ \text{mvdecode income, mv(0)} \]

Question 6: Now what do we see when we type summarize income?

We want to compare the incomes of men and women. Type
\[ \begin{align*} 
\text{summarize income if gender==1} \\
\text{summarize income if gender==2} 
\end{align*} \]
where 1 is the code for males and 2 is the code for females.

Question 7: What do we see?

Now we type:
\[ \begin{align*} 
\text{sort gender} \\
\text{by gender: summarize income} 
\end{align*} \]

Question 8: How does this display differ from our previous displays?
Finally, we want to compare the incomes of men and women visually. If we want to use boxplots to display the income by gender, we type

```
. search boxplot
```

and then

```
. help box
```

to find out how to display these data. Type

```
. graph income, box by (gender)
```

*Question 9: Write a paragraph to describe the differences in the income earned by men and women.*

Here is a list of the commands we used in the Getting Started with *Stata* Lab. Use the space next to each command to make notes on what that command does.

```
describe
list
sort
summarize
graph
search
help
```

Here is a list of conditions:

```
in
if
```
To leave Stata, type

```
. clear
```

and then type

```
. exit
```

**Assignment**

Your TA will tell you when your assignment for the Getting Started with Stata Lab is due. So far, we have looked at income differences with respect to gender. Select a different variable, such as marital status, race, educational level, type of household. Repeat the process of summarizing and plotting the data. Write a brief summary which compares income with respect to the variable of your choice.