**Visual Display of Data in *Stata***

Please remember to use your Lab ID# as your name and your nine digit UCLA ID# as your password when you log-in.

In this lab we will plot histograms with various bin widths, add appropriate scales and titles, and look at other plots to help us better understand the demographic data collected in a recent census of Los Angeles County.

. use http://www.stat.ucla.edu/labs/datasets/smcensus.dta

**Visual Display of Data**

Suppose we are curious about the ages of those in our sample.

. summarize age

This tells us that the youngest respondent was 16 while the oldest was 87, with a mean of 43 years, give or take 18 years. The command

. summarize age, detail

gives us additional information about the distribution of ages.

*Question 1: What is the median age of the respondents? What is the age span for the middle fifty percent of those in our sample?*

If we plotted the data by hand, we might use a horizontal scale from 0 to 100. In *Stata*, we type:

. graph age, histogram xscale(0, 100)

Now try the command:

. graph age, histogram xscale(0, 100) norm
Question 2: What does Stata do when we add norm to the command

Adding Labels and Titles We can get a better sense of the age plot, if we add more bins, but not too many bins or the plot will look like a city skyline.

```
. graph age, histogram xscale(0, 100) bin(5) norm
```

Question 3: Now change the number of bins from 5 to 10 and 20. Which graph do you find most pleasing and informational?

Labeling the horizontal axis may help us locate the center of the distribution. We can also include additional information by adding two titles in four different locations: top, bottom, right or left.

```
. graph age, histogram xscale(0, 100) bin(10) xlabel(0, 10, 20, 30, 40, 50, 60, 70, 80, 90) t1title(Ages of the Respondents in our Sample)
```

If we are interested in comparing the ages of men and women, we want to look at two histograms. We need to sort the data by gender before we plot.

```
. sort gender

. graph age, histogram by(gender)
```

Question 4: Reproduce the graph comparing ages of men and women with appropriate labels, number of bins, and a title. Describe what you see.

Decoding Data Now we will consider the rent paid by those in our sample.

Type:

```
. summarize rent, detail
```
Before we plot the data, we can remove the zeros for those who paid no rent by storing the zeros as missing values. Type:

```
. mvdecode rent, mv(0)
```

*Question 5: How many people paid no rent?*

```
. summarize rent, detail
```

*Question 6: How do the median and mean rents compare? What does this tell us about our data?*

Next we plot our rent data. In this command we simplified the xlabel command.

```
. graph rent, histogram xscale (0, 700) xlabel(0 100 to 800) t1title(Monthly Rents)
```

*Question 7: Reproduce this graph again, changing the number of bins to ten. Does this change the distribution?*

Now we will consider several other questions about the rent data.

*Question 8: Do women and men in our sample pay comparable rents? What command is needed to find out? What are the findings?*

We may wonder if there is a relationship between age and rent? To display paired data, we type:

```
. graph rent age
```

*Question 9: Do people pay more rent as they get older?*
Question 10: Is there a relationship between the income someone earns and the rent that they pay? What command is needed to find out? What are the findings?
Assignment In this lab, we looked at age, rent, and gender. Select different variables such as, marital status, race, or hours worked. Repeat the process of summarizing and plotting the data. Select a carefully prepared plot with appropriate titles and scaling that tells you something of interest about the people in our sample. Include a brief written interpretation of your findings.