Statistics M11/Economics 40 LAB ASSIGNMENT #1 DUE JAN 26, 2001 LEW

INTRODUCTION

Retrieve the dataset for Lab #1 by issuing this entire command (quotes and all) in the command window:

use "http://www.stat.ucla.edu/~vlew/stat11/labs/lahmda.dta"

and press enter. Here is a description of the dataset from Stata

```
Contains data from lahmda.dta
  obs:         1,652                          2000 Los Angeles Economic and
                                                Demographic Data
 vars:            11                          9 Jan 2001 18:34
 size:        72,688 (98.5% of memory free)
-------------------------------------------------------------------------------
   1. tractid   float  %9.0g                  Census Tract Identifier
   2. inclevel  str8   %9s                    Income Level of Tract
   3. medpct    float  %9.0g                  Percentage of Los Angeles
                                                Median Income
   4. lamedian  long   %12.0g                 Los Angeles Median Income
   5. medfam00  long   %12.0g                 2000 Median Family Income in
                                                Dollars
   6. medfam90  long   %12.0g                 1990 Median Family Income in
                                                Dollars
   7. totpop    int    %8.0g                  Total Population of the Tract
   8. pctminor  float  %9.0g                  Percentage Minority in the Tract
   9. minorpop  int    %8.0g                  Minority Population of the Tract
  10. owners    int    %8.0g                  Number of Homeowners in the
                                                tract
  11. renters   int    %8.0g                  Number of renters in the tract
-------------------------------------------------------------------------------
Sorted by:  tractid
```

Note the "sorted by" above, what this means is that the 1,652 observations in this dataset are found in order of this variable. If you issue the command:

list tractid

You will "see" what I mean by being sorted. You will learn to re-sort the dataset in this lab.

Some additional Stata commands for you to learn.

1. help or search

Remember, in Stata if you want to learn more about something, you can issue a

help sort

so if this lab is unclear, issue some help commands or a

search sort

2. *codebook* or *describe*

Will give you information about the variables in the dataset (and if you wanted to learn more about them, issue a 'help codebook' or 'help describe'). The list of variables above is the result of issuing the command *describe* from the command window.

3. summarize and summarize, detail

If you issue the command:

summarize

Stata will give you basic statistics for all of the NUMERIC variables in the dataset. Try a

summarize, detail

to see what additional information you get.

4. tabulate

For non-numeric (or character data) you need to issue a tabulate to see what's there. We have one of these in the dataset, it's called inclevel

tabulate inclevel

You should see something like this:

```
. tabulate inclevel

     Income |
   Level of |
      Tract |      Freq.     Percent        Cum.
------------+-----------------------------------
        Low |        147        8.90        8.90
     Middle |        544       32.93       41.83
   Moderate |        377       22.82       64.65
    Unknown |         19        1.15       65.80
      Upper |        565       34.20      100.00
------------+-----------------------------------
      Total |       1652      100.00
```

All this means is that of the 1,652 areas in Los Angeles. 147 were classified as "low income", 544 were "middle" and you can interpret the rest.

5. sort

To sort data, issue the command

sort variablename

So if you want the dataset sorted by median family income in 2000, issue this command:

```
sort medfam00
```

It will seem as if nothing happens, so issue something like:

```
list tractid medfam00
```

to see if the dataset has changed as a result of your sort.

6. graph

There are graph commands in Stata, to create a box plot, for medfam00 for example, issue the graph command

```
graph medfam00, box
```

To create a histogram, leave off the box option

```
graph  medfam00
```

To modify the histogram by adding extra bins, issue this command:

```
graph medfam00, bin(15)
```

by default, Stata gives your histogram 5 bins, but it might be misleading.

To put multiple boxes on the same graph, you need to first sort the data, so you might issue the command:

```
sort inclevel
```

This will sort the dataset in ascending order on the variable income level. Issue a

```
list  tractid inclevel
```

to see how the dataset has been changed. Then to get more than one box on the page, issue the command:

```
graph medfam00, box by(inclevel)
```

This should be enough help for you to finish this lab

ASSIGNMENT: Answer the following 5 questions about the data.

1. Provide a five number summary for variables medpct, medfam90, medfam00, and pctminor in the dataset. Clearly identify the numbers used the five-number summary.

2. Which 5 census tracts had the highest median family income in 2000? Which 5 census tracts had the lowest median family income in 2000? Which 5 census tracts had the highest median family income in 1990? Which 5 census tracts had the lowest median family income in 1990?

3. Construct a box plot of percentage minority by income level of tract. From the results, what can you say about the relationship between the income of the area and percentage minority?

4. Construct a histogram for medfam00 and medfam90 and choose something other than the default 5 bins. Indicate how many bins you used. Are there any noticeable differences between the two variables?

5. Is this data from an experiment or observational study? Explain your reasoning.

Include the Stata results for questions 1 and 2, answers to questions 1-5, and graphs for 3 and 4 in your lab assignment. Clearly identify your answers. Staple the pages, put your name on your and your TA's name or your section (1A, 1B or 1C). If we cannot clearly identify your section when grades are being recorded, we will deduct one point from your lab score. This lab is worth 6 points total.

THIS LAB IS DUE ON JANUARY 26, 2001 BEFORE THE END OF LECTURE.