Sociology 213 Social Demography
Lab #2 directions

Useful URLs:
IPUMS : http://www.ipums.org/usa/
Stata basic overview and frequently asked questions:
http://www.ats.ucla.edu/stat/stata/notes/default.htm
Stata home page: www.stata.com

Today's lab:
In this lab we will learn how to extract individual-level data from past censuses using the IPUMS database. I will demonstrate how to get the extracted data into Stata and how to do a basic descriptive analysis on data from the 1880 census.

To use our time efficiently, I have divided today's lab into two parts. I will demonstrate part 1 first, then you get to repeat what I did on your own computer afterwards. As I demonstrate, take notes on the lab instructions, as the written instructions will only be complete if combined with my example. After I demonstrate, I will go around and help anyone who is having difficulty. If you finish the first part before the rest of the class does, click on "quick reference" at the top of the IPUMS web page and read about the variables that are on the IPUMS data.

Outline & goals for this lab:
Part 1
0. Let's go ahead and register as IPUMS users. www.ipums.org

1. General Comments on infile and editing the data you downloaded from
http://fisher.lib.virginia.edu/census/
--what the _M_ would be etc.

2. Do-files in Stata. What are they, how to make them, and how to use them. Example using s1880.do

3. How to make an extract in IPUMS using the interactive web page. Example making a simple data set for 1890.

4. How to download the dataset you made, and how to use pkunzip to unpack it. Example using tedmo22.dat.Z (1880 data from IPUMS)

Part 2
5. Another way to import data into Stata: infile using a dictionary for fixed-format data. Example using tedmo022.dat and tedmo022.dct.

6. What do you do when you have the data? A basic overview of exploratory data analysis in Stata, via interactive commands and a do-file using tedmo022.dta (bw1880.dta)

Lab Assignment:
1) Repeat the lab exercise for today (i.e. steps 0-5 above)
2a) Make a IPUMS data set and bring it into Stata (**year indicated below**)
Variables that should be included (if they are available during that year):
[from the personal record]
Rectype
Year
Datanum
Serial
Pernum
Perwt
Relate
Age
Sex
Race (general)
Bpl (general)
Mpbl (general), if available
Occ
Occ1950
Occscore
[from the household record]
region
state (fips code)
county
metarea (detailed)

+ Two other variables of your choice
[warning: start w/ this small data set first, can get very large]

Put a copy of this data set in **/afs/isis.unc.edu/html/courses/2003spring/soci/213/001/ipums**

2b) 1) Make a table of the distribution of white and black residents by state (example to follow in lab)
2) Make a table of the average occscore for black residents by state (example to follow)
Print the log file with both of these tables on it (see example below) and turn it in.

3) Make another extract of the black and white population using the historical aggregate data (http://fisher.lib.virginia.edu/census/) following the procedure from the last lab. Note that the definition of "race" changes over time. Do your best to get information on the size of the black and the white population of each race. Bring the data into Stata using infile, save it as sxxxx.dta, (where xxxx is the year) and put a copy of the data in
/afs/isis.unc.edu/html/courses/2003spring/soci/213/001/agg

demographer id	IPUMS (2)	Agg. Data (3)
1	1890	1890
2	1900	1900
3	1910	1910
4	1920	1920
5	1940, white	1930
6	1940, non-white	1940
7	1960, white	1950
8	1960, non-white	1960

Part 1

0. Let's go ahead and register as IPUMS users. www.ipums.org

1. General Comments on infile and editing the data you downloaded from http://fisher.lib.virginia.edu/census/
--what the _M_ would be etc.

2. Do-files in Stata. What are they, how to make them, and how to use them. Example using s1880.do

See the do-file example s1880.do

----example do file for s1880.do---
```
capture log close
log using s1880, replace
clear
infile str20 state total white colored bis using s1880.txt

save s1880, replace

lab var total "total population"
lab var white "white population"
lab var colored "non-white pop"
lab var bis "born in this state"

gen pctbis=bis/total
lab var pctbis "pct born in state"

des
```

sort pctbis
list state total pctbis

Stata
Run the file s1880.do: "do s1880"

You now have state data on population by race for 1870 and 1880.

3. How to make an extract in IPUMS using the interactive web page. Example making a simple data set for 1890.

A. How to make an extract

In Netscape
1890 census data for blacks
http://www.ipums.org/usa/data.html

let's look at my recent extract
registered users
tedmouw@email.unc.edu
sample density page: small, regular
sample selection page: rectangular, stata, full, thematic
sample: 1890
variables.
Show how to select by race.
[Notes on downloading data: when your extract runs, you will get an email. Go to download data in Netscape. Back click on the data to save it. "Save link as"

4. How to download the dataset you made, and how to unpack it. Example using tedmo22.dat.Z (1880 data from IPUMS)

Open up windows explorer and double click on tedmo22.dat.Z
Make sure to extract the file to the same directory (sometimes this is not the default).

If you finish early, browse the variables on IPUMS.

Part 2
5. Another way to import data into Stata: infile using a dictionary for fixed-format data. Example using tedmo022.dat and tedmo022.dct.

How to download and input the IPUMS data into Stata.

Downloading the data from IPUMS…follow directions on the web page…remember to back click and save link as.

To get the data dictionary in IPUMS: in netscape, click on raceg
Download data and dct file.

Edit the tedmo022.dct file in the Stata do-file editor---*note that IPUMS did not put str in front of my string variables*---

```
Clear
Set mem 50m
Infix using tedmo022.dct
des
compress
des
save ipums1880, replace
```

6. What do you do when you have the data? A basic overview of exploratory data analysis in Stata, via interactive commands and a do-file using tedmo022.dta (ipums1880.dta)

```
log using i1880, replace
des
tab race
Label define raceg 1 “white” 2 “black” 3 “am Indian” 4 “Chinese”
Label val raceg raceg

Help tab
Tab raceg

Tab insane
Tab sex insane
Tab sex insane, row nofreq

Tab raceg
Tab state raceg
Tab state raceg, col nofreq
Tab raceg, sum(occscore)
Tab raceg if race==2, sum(occscore)


Save ipums1880, replace
```

Now lets review the commands and make it into a do file:
What is the command for review? Help review or lookup review

Make a do-file in the do-file editor with some of these commands:
Example