Economics 466: Economics of Population
Fall 2008
Problem Set #1
Distributed: Wednesday, September 10
Due: Monday, September 15
(Submit electronically to the course web site on CTools by 11:45 p.m. on Monday)

Use the 2000 United States census data introduced in the lab on September 10. Open the file "**`ipums-us2000.1%.dta`**" using STATA. Create a log file called **`e466.ps1.name.log`**, where name is your UM unique name. Create a Word file in which you will write up the answers to each question. You should upload both the Word file and the log file to CTools.

1. Describing households:

Find the household which has household identification number (serial) 293000. Describe features of the household based on the following variables:

a) household size
b) race of individuals in the household
c) state that the household resides in
d) describe the members of the household in terms of age, gender, and relationship to head
e) what level of education does the resident head of the household have?
f) Which Stata commands can you use to answer the above questions?

2. Using the simple summary command **`tabulate`** and the variable **`marst`**:
   a) Produce a table with the distribution of marital status for the population aged 20-24 for the entire U.S. population. Paste this into your Word file.
   b) Produce a single table that compares the distribution of marital status for males and females aged 20-24 for the entire U.S. population. Paste this into your Word file.
   c) Write a few sentences discussing the differences in the distribution of marital status between males and females. Use complete sentences and write clearly.
   d) What proportion of females aged 20-24 in Michigan are single?
   e) What proportion of single women in Michigan are aged 20-24?
   f) Which state has the highest proportion of males aged 20-24 who are single?

3. Age structure:
   a) What proportion of the U.S. population is exactly age 60?
   b) What proportion of the U.S. population is under age 60?
   c) What proportion of the Florida population is under age 60?
   d) What proportion of the California population is under age 60?
   e) What Stata commands did you use?

4. Education of age groups:
   a) Produce a single table showing the distribution of education for males and females aged 25-29 in the United States. The table should show just the percentages in each schooling category for each sex, without frequencies. Use the variable educus. Paste this table into your Word file. Write the Stata command that you used to produce this table.
   b) Write a few sentences discussing some of the differences in the distribution of schooling between males and females.
   c) What proportion of the U.S. population aged 30-34 has at least a high school diploma?
   d) What proportion of the U.S. population aged 70-74 has at least a high school diploma?

Economics 466: Economics of Population
Fall 2008
Problem Set #2
Distributed: Wednesday, September 17, 2008
Due: Monday, September 22, 2008
(Submit electronically to the course web site on CTools by 11:55 p.m. Monday)

Use the 1990 U.S. census data introduced in the lab on Wednesday. Open the file "`ipums_us1990.2%.dta`" that we worked with in the lab. Create a log file called `econ466ps2.log`. Create a do file called `econ466ps2.do` to run the necessary commands. Submit your log file, do file, and a Word file with answers and discussion of the following questions.

1. Generating and modifying variables with **`generate`**, **`replace`**, and **`egen`**:
   a) Use the generate command to generate a simpler age variable that has five-year age groups: 0-4, 5-9, 10-14, etc., with the highest category being 80+. Assign a variable label and value labels, and use the tabulate command to show that your variable assigns the correct category to each age group. Call this variable agefiveyr. What percentage of the U.S. population was in the 20-24 age group in 1990?
   b) Use the replace command to change the "NIU" (not in universe) code for the children ever born (chborn) variable from "99" to "." (the Stata missing value code). Do a tabulate command with the missing option to show that this worked correctly. Paste this table into your Word file.
   c) Generate a variable called "grade12" to indicate whether an individual has at least 12 years of education. Set grade12 equal to 0 for individuals with less than 12 years of education (using the variable educus) and equal to 1 for individuals with ***at least*** grade 12 (including the group "grade 12 no diploma"). Use this variable in a tabulate command to show the percentage of the population aged 30-49 with less than grade 12 education in each state. Which state has the highest percentage of 30-49 year-olds with less than grade 12 education, and what is that percentage? Which has the lowest percentage?
   d) Use the egen command to construct a total household income variable. Call it inctothh. What is the mean of this variable for the entire United States? What is the mean total household income for Michigan? Using your inctothh variable and the persons variable, generate a variable for household income per person. Call it incpchh. What is the mean household income per capita for Michigan?

2. Using **`tabsum:`**
   Use the tabsum command to answer the following questions:
   a) What is the mean number of children ever born to women in each five-year age group from 15-19 through 65-69? Generate a nice table and paste it into your Word file. Write a paragraph that discusses the relationship between age and the number of children ever born. What do you think is going on at the older ages?
   b) Add the urban variable to your tabsum, showing how the number of children varies with age and rural/urban location. Choose the order of the variables and the tabsum options that make this a readable table with just the means for each cell. Copy this table into your Word file. Write a few sentences discussing what the table shows.

3. Income and fertility:
   a) Using your total household income variable (inctothh) from Question 1, generate another new variable that breaks this into five income categories: 1) $0-$8,000; 2) $8,001-$20,000; 3) $20,001-$35,000; 4) $35,001-$50,000; 5) over $50,000. Tabulate this variable. Copy a table into your Word file showing the proportion of individuals in each of these income categories.
   b) Use a tabsum to calculate the mean number of children ever born to women aged 40-49 in each of the five income categories. Paste this into your Word file. Discuss your results.
   c) Combine your analysis of income categories with your analysis of age groups. Use tabsum to show the mean number of children ever born to women in each income category for each five-year age group between 20-24 and 45-59. Copy your table into your Word file. Discuss your results.

Load your Word file, log file, and do file to CTools.

Economics 466: Economics of Population
Fall 2008
Problem Set #3
Distributed: Wednesday, September 24, 2008
Due: Monday, September 29, 2008
(Submit to CTools by 11:55 p.m. Monday)

This problem set uses the combined U.S. census data sets for 1960, 1970, 1980, and 1990. You should do all your work in a do file. You should submit a Word file with your answers, your log file, and your do file to the CTools web site. Also, remember to use person weights (wtper_new) in all of your calculations unless otherwise specified.

1. Append the U.S. census data sets for 1960, 1970, 1980, and 1990 into a single data set, dropping the observations for individuals under age 15.
   a) Produce a table showing how many observations you have from each census year (unweighted)? Now use weights and show how many weighted observations you have for each year.
   b) Use the replace command to change the "NIU" (not in universe) code for the children ever born (chborn) variable from "99" to "." (the Stata missing value code). Do a tabulate command with the missing option to show that this worked correctly. Use the same procedure to set a missing value for the total income (inctot) variable.
   c) Use the egen command to construct a total household income variable. Since you have appended the years, and since the same serial number may be used in different years, you need to use "by (year serial)" with the egen command. Call this variable inctothh. Divide this by household size (persons) to create a per capita household income variable called incpchh. Use tabsum to calculate the mean per capita household income by census year, using only household heads for the calculation. Do this with and without using the household weight variable (wthh). How much difference does it make whether you use sample weights?
   d) Based on your table in part (c), what can you say about the change in per capita household income in the U.S. from 1960 to 1990? Are there reasons why the answer from your table might be misleading?

2. Analyze the age profile of children ever born in each of the four censuses:
   a) Use tabsum to produce a table that shows the mean number of children ever born to ever-married women at each age from age 15 to age 69 for each of the census years (use person weights). Do the same thing without the restriction to women having been married. What difference does it make whether you restrict the sample to ever-married women?
   b) Copy your table into Excel and produce a graph showing the mean number of children ever born for ever-married women at each age in each census. Give the figure a number and a title and paste it into your Word document. Write a well-written paragraph discussing what you see in the graph.

3. Generate two new variables to indicate "working" and "in the labor force" following the steps from the lab.
   a) Use tabsum to produce a table that shows the percentage of women working at each age from age 15 to age 69 for each of the census years (use person weights). Do the same thing for men. Do the same thing using labor force participation rather than the proportion working. How much difference does it make whether you use the proportion working or the proportion in the labor force.
   b) Copy your tables into Excel and produce a graph showing the percentage of women working at each age in each census. Make another graph for men. Give each figure a number and a title and paste it into your Word document. Write a well-written paragraph discussing what you see in the graphs. How are your results in Question 3 related to your results about fertility in Question 2.

Economics 466: Economics of Population
University of Michigan
Fall 2008
Problem Set #4 – "Write Your Own Problem Set"/Term Paper Proposal

Guidelines for Term Paper

Two-Page Proposal Due: Wednesday October 8, 11:55 p.m., posted to CTools as PDF file
First Draft Due: Wednesday Nov. 12, 11:55 p.m., posted to CTools, hard copy handed in by Thursday noon
Final Draft Due: Wednesday Dec. 17 at 12:30 p.m., posted to CTools, hard copy handed in by 5:00 p.m.
Paper Length: 15 pages, including tables and figures
Grading: Grade on paper will be 30% of course grade. Grade on first draft is 30% of grade on paper.

Your paper must be based on statistical analysis of census data, using the IPUMS-International census files available in the AFS class folder. Your paper should address some topic that can be analyzed using one or more of these census data sets. You are encouraged to focus on questions related to the topics that are being covered in class, such as demographic outcomes, education, labor force participation, and inequality. You should feel free to be creative, however, taking advantage of the wide variety of variables that are included in these data sets. You are encouraged to use two data sets, either looking at change in one country between two censuses or comparing two different countries. You may also want to use more than two data sets.

The paper should include carefully constructed figures and tables that present the results of your analysis. These figures and tables should have titles and should be thoroughly explained and discussed in the text of the paper. The paper should include an explanation of the data sets you are using. You should be careful about explaining the variables you are working with, referring back to the original questionnaire if necessary to clarify the meaning of particular variables. (See more instructions on the back of this page).

The paper is not intended to be a review of other papers or books written on a topic. In order to develop a clear research question and to decide on the key variables involved in answering that question you will want to read some of the literature in your proposed research area. However, any literature review should be a small part of the paper itself (2 pages is the maximum). The paper should draw heavily on the data. Clear descriptive analysis of the data is more important than trying to explain the causal mechanisms driving your results. An important purpose of the project is to develop skill in the statistical analysis of data and the presentation of your results in a clear and effective written form. Although it is not required that you include regressions in your paper, it is strongly encouraged that you do so. Ideally your paper will include a mix of descriptive statistics using tables and graphs, along with multivariate regressions that provide a clearer picture of the relationships you are studying.

The first step in the process is to write a two-page proposal describing the paper you plan to write. This proposal should describe the questions you are interested in analyzing. It should indicate the data sets you plan to use for the analysis, including a description of some of the key variables you plan to use. Your proposal should include at least one table and one figure that begin to explore the questions you are interested in. Your proposal should consist of 1-2 pages of well-written text along with another 1-2 pages that present at least one table and graph. The table and graph should be thoughtfully constructed, should have numbers and titles, and should be discussed in the text.

You may want to think of your paper proposal an exercise in writing your own problem set. Take a topic you are interested in and try analyzing it using one or more of the census data sets. Pose some simple questions and create a table or graph to answer the question. Then think about where you might take this for a more detailed analysis.

The Structure and Presentation of Term Papers

Be sure to take advantage of the writing resources provided through the UM Sweetland Writing Center. This includes workshops and individual consultations with both peer tutors and professional writing consultants. The Sweetland web site has many useful resources, such as style sheets for bibliographies and general writing tips.

Sweetland Center web site: http://www.lsa.umich.edu/swc

Structure and Organization:

Your paper should be either double-spaced or 1.5 spaced, with about 1" to 1.5" margins. Be sure to put page numbers in your paper. Your paper should have a title page with an informative title. Use headings and sub-headings to break up your paper. It will make the paper easier to read and help the reader follow the logic of your analysis.

If you are citing articles or books, be sure to list the references at the end of the paper using some standard bibliographic style. The Sweetland Center web site has standard style sheets available to use as guides.

Clearly describe the data sets you are working with, including the year the data was collected, the size of the sample, etc. Describe the variables clearly – don't just use the variable name in the data set. Write the paper so that someone can understand it without knowing what "inctot" refers to.

Tables and Figures:

Put a number and title on each figure and table. Refer to these in the text. For example, you should say things like "Figure 1 shows the relationship between … ". Use the papers you are reading for the course for examples.

Make sure your figures and tables are effectively presented. Don't just paste in tables and figures from Stata without modification. You need to convert the Stata output into tables that make sense for the reader. Tables should contain key information, such as sample size, important sample restrictions (ages included, whether males, females, or both, etc.). Use the papers assigned for the lectures for good examples of tables and figures.

Use a reasonable number of decimal places in your tables and text. Don't just use the full number of decimals from the Stata output. Usually one or two decimal places are sufficient, depending on the variable.

Data Analysis and discussion:

Be sure to consult the census questionnaire and the IPUMS-I web site to be sure you understand the variables you are working with. Explain important features of the variables in your paper, such as the definition of work, the time period for an income variable, or the age restriction on an employment or fertility question. In general you should use the sample weights provided in the data in your analysis. You should make this clear in your paper.

Make sure it is always clear who is being included in a particular analysis. For example: are you including observations with zero in your measures of income or hours worked? Is your mean schooling measure for the entire population, or just adults? Are you using individuals or households as the unit of analysis? Does your table include both men and women?

Economics 466: Economics of Population
Fall 2008
Problem Set #5
Distributed: Wednesday, October 22, 2008
Due: Monday, October 27, 2008
(Submit electronically to the course web site on CTools by 11:55 p.m. Monday)

Copy the 2001 South African census data ipums_southafrica2001.20%.dta to your home AFS space to use for this problem set. You should do all your work in a do file. Submit your log file, your do file, and a PDF version of your Word file to the CTools web site.

1. Create a new "years of education" variable (call it edyears) using this code:
   ```
   recode educza ( 102=1) (103=2) (104=3) (105=4) (106=5) (107=6) (108=7) (109=8)
   (110=9) (111=10) (112=11) (113=12) (121=12) (122=12) (130=13) (140=14) (210=15)
   (220=16) (230=16) (240=18) (999=.), gen(edyears)
   ```
   Do a tabulation to make sure you created the variable correctly. Follow the steps done in the lab to create a set of variables for the mother and father. For the mother and father create variables for age, race, and years of education (and anything else you find interesting).
2. What is the mean years of schooling of the mothers of children aged 15? What is the mean years of schooling of their fathers?
3. Use tabsum to produce a table that shows the mean years of schooling of 15 year-olds (using your edyears variable) for each year of schooling of the mother. Do the same thing for 12 year-olds. Make a nice Excel graph showing the schooling of both 12 year-olds and 15 year-olds by the schooling of the mother. Discuss your results.
4. Use tabsum to produce a table that shows the mean years of schooling of the father (using your edyears variable) for each year of schooling of the mother. Discuss your results.
5. Estimate a regression with schooling of the child as the dependent variable and the schooling of the mother as the independent variable, using only 15 year-olds. Discuss your results.
6. Estimate a regression with schooling of the child as the dependent variable and the schooling of the father as the (only) independent variable, using only 15-year-olds. Discuss your results. How do they compare to the results for mother's education in Question 4?
7. Now include both schooling of the mother and schooling of the father as independent variables in the same regression. How do the coefficients on mother's and father's schooling compare to those estimated in questions 4 and 5? How would you explain the change? What happens to the R^2?
8. Use the egen command to construct a total household income variable. Call it inctothh. Divide this by household size (persons) to create a per capita household income variable called incpchh. Take the natural log to create the log of per capita household income (call this logy). Add your per capita income variable (incpchh) to the regression you estimated in Question 6 (you should have mother's education, father's education, and income as independent variables). Discuss the new results. How does the inclusion of income affect the coefficient on mother's education and father's education? What happens to the R^2?
9. Now use log income instead of the incpchh variable (your regression should now have mother's education, father's education, and logy as independent variables). How does this regression differ from the regression using incpchh? What happens to the coefficients on parental education? What happens to the R^2?

Economics 466: Economics of Population
Fall 2008
Problem Set #6
Distributed: Wednesday, October 29, 2008
Due: Monday, November 3, 2008
(Submit electronically to the course web site on CTools by 11:55 p.m. Monday)

---

1. Analyze the impact of race and education on earnings in the United States. Use the 1% 2000 U.S. census data that was used in the lab. Submit your do file, your log file, and a PDF version of your Word file to the CTools web site.

   a) Create a new education variable, edyears, using the following algorithm:

   ```
   recode educus (0=.) (100=0) (101=0) (200=3) (300=7) (400=9) (500=10) ///
   (600=11) (701=11) (702=12) (821=13) (822=14) (824=16) (825=18)
   (826=18) (827=20), gen(edyears)
   ```

   Do a tabulation to make sure this worked correctly.

   b) Generate a dummy variable for white, with white=1 for whites (raceus==100) and white=0 for everyone else. Do a tab of white by raceus to make sure this worked.

   c) Generate the log of earned income (incearn). Call this logy. How does logy differ for whites and non-whites, limiting the analysis to individuals aged 30-59?

   d) Generate a variable for age squared. Estimate a regression of logy on the white dummy variable, age, and age squared. How do you interpret the coefficient on the white dummy?

   e) Estimate a regression of logy on edyears, age, and age squared. How do you interpret the coefficient on edyears? Also discuss the *t*-statistic and the R^2.

   f) Now estimate a regression of logy on white, edyears, age, and age squared. What happens to the R^2 compared to (d) and (e). How does the coefficient on white compare to the answer in (d)? How does the coefficient on edyears compare to the answer in (e)? Discuss.

   g) Compare the returns to schooling for whites and non-whites by running two separate regressions. What's the difference between these estimates on the returns to schooling?

   h) Following the steps done in the lab, create an interaction variable whitexeduc equal to white * edyears. Estimate a regression of logy on white, edyears, whitexeduc, age, and age squared. Interpret the coefficient on the interaction variable. How does this compare to your answer in (g)? What can you say about whether there is a statistically significant difference in the returns to schooling for white and non-whites?

Economics 466: Economics of Population
Fall 2008
Problem Set #7
Distributed: Wednesday, November 19, 2008
Due: Monday, November 24, 2008
(Submit electronically to the course web site on CTools by 11:55 p.m. Monday)

Improve your paper!

Your goal in this problem set is to begin revising your paper by focusing on one or two tables or graphs that you feel should be added to the paper or that exist in your first draft but would benefit from improvements. You should use our comments on your paper to guide your decision about what to work on.

Option 1: Take an existing graph or table (or a set of graphs and tables) for which we made suggestions for improvement. Or choose some yourself that you think can be substantially improved. Include the original version in your submission. Explain how the tables/graphs can be improved. Produce new graphs and/or tables. Include new text (no more than two or three paragraphs) that you can include in the paper that discusses the new graphs/tables. Work carefully on your writing. Aim for very polished graphs/tables with very polished writing.

Option 2: Develop some new graph or table that extends your analysis in some way. Include new text (no more than two or three paragraphs) that you can include in the paper that discusses the new graphs/tables. Work carefully on your writing. Aim for very polished graphs/tables with very polished writing.

Submit a PDF of your new graphs/tables and accompanying text. Also submit a nice clean do file that does the analysis.