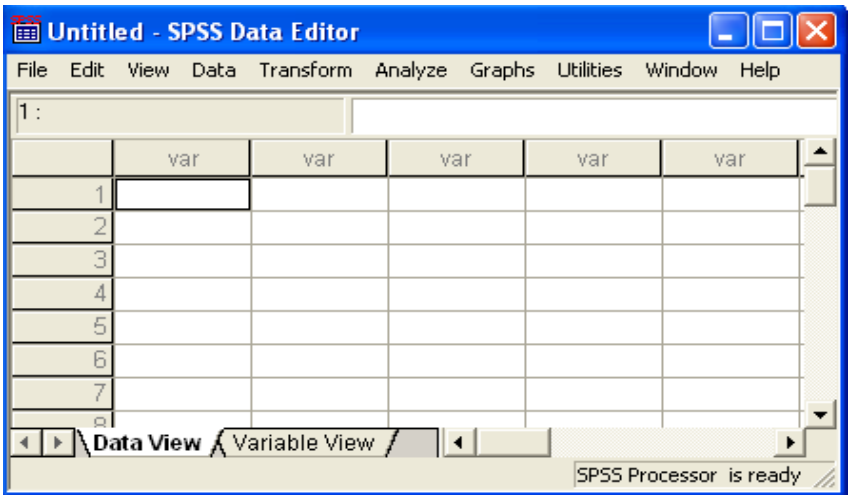
Exercise 14. Analyzing Census 2000 PUMS Data

Purpose: The goal of this exercise is to demonstrate how the *PUMS* data downloaded from the IPUMS web site can be brought into *SPSS* for processing. It is not the intent of this exercise to get too deeply into *SPSS* since that package can be a class in itself. However, you can read through the following pages to get an overview of how one might use *SPSS* on *PUMS* data.

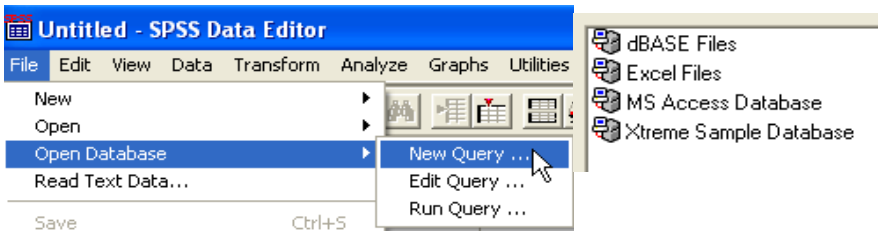
Loading Data into SPSS

1. Locate and open the *SPSS* program.

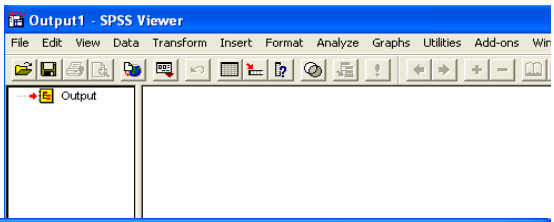
2. When the *Data Editor* spreadsheet opens, cancel the option to open an existing file.
You will need to populate the database either manually or from existing sources.



SPSS can read several existing formats with the aid of a database wizard that is located in the *File* menu under *Open Database* > *New Query*.

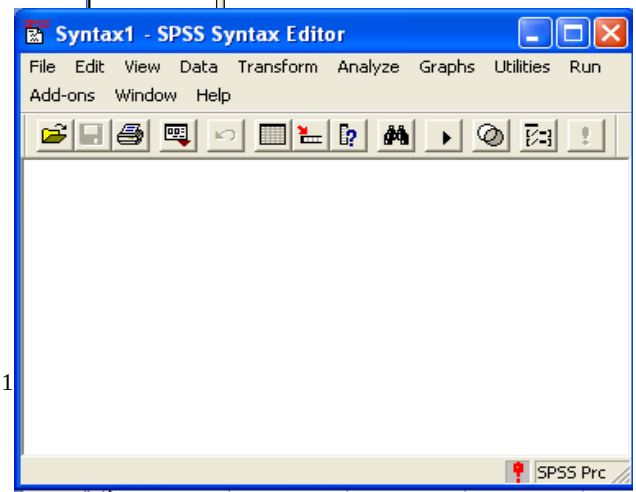


However, your data is in text format and the field widths and column positions must be defined.



The *SPSS* program commonly uses several windows that depend on what tasks you need to perform. One of these is a *Data Editor* window shown at the top. Another is the *Output* window at right. Each window has accompanying tools.

In some cases one must revert to the old form of *SPSS* that is driven with various commands. For example, the *PUMS* data extracts from *IPUMS* are generated in a text

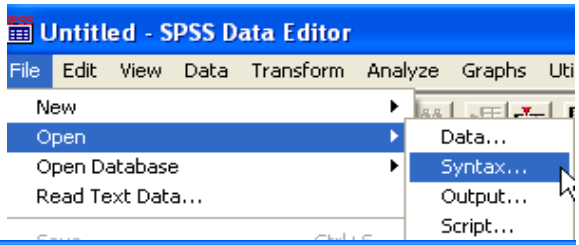


format that necessitates the use of *SPSS* syntax. Fortunately *IPUMS* also generates a syntax file to define and input the needed data. When read, the imported data can be saved as an *SPSS sav* file for future use. Data in this native format need only be opened.

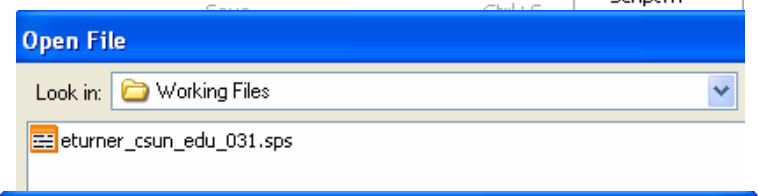
For program execution, *SPSS* provides a *Syntax* window (shown right) where necessary commands are specified in text form. Each ends with a period.

SPSS also provides a *Script* window for running customized programs.

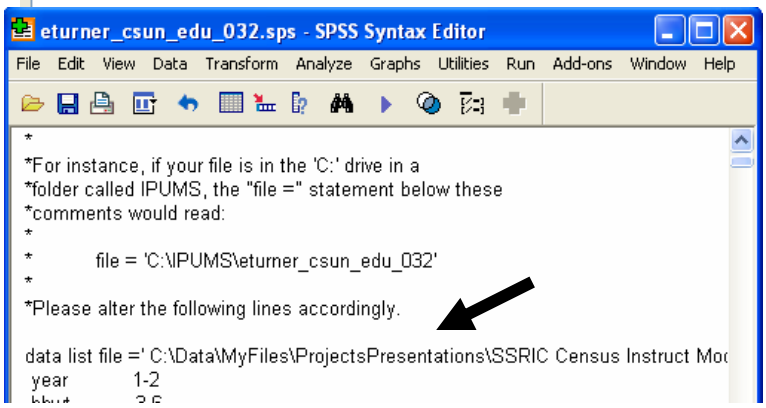
2. Select *File* > *Open* > *Syntax*.

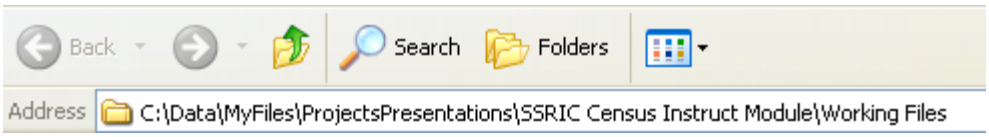


3. From the *Open File* window locate the *SPSS* syntax file (has .sps suffix) you downloaded from *IPUMS* earlier. Shown is *eturner_csun_edu_031.sps*. Click *Open*.



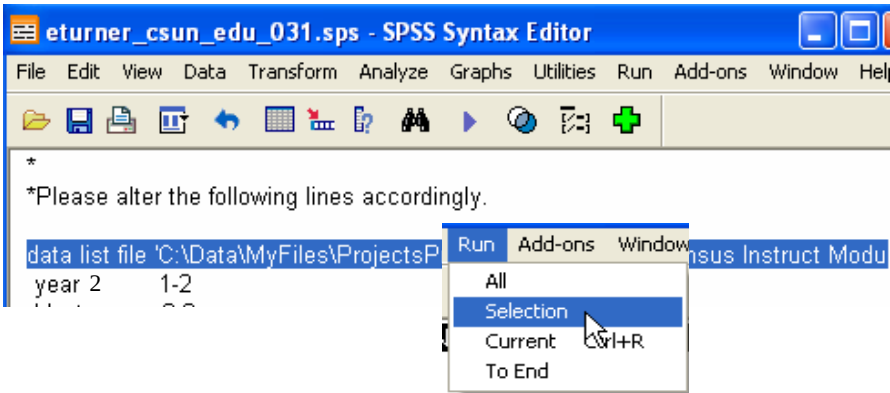
The *SPSS* command syntax will be listed in the *Syntax Editor* window shown right. From here you must change the path and file name so that the data can be loaded. Below, *Windows* was used to specify the path to the file location. The path can be copied and pasted into the first line in the *Syntax Editor*. Note the revised path below:





4. Highlight just the first line and then select *Run* > *Selection.*

SPSS should be able to locate the data and happily list out in an *Output* window the variables from the

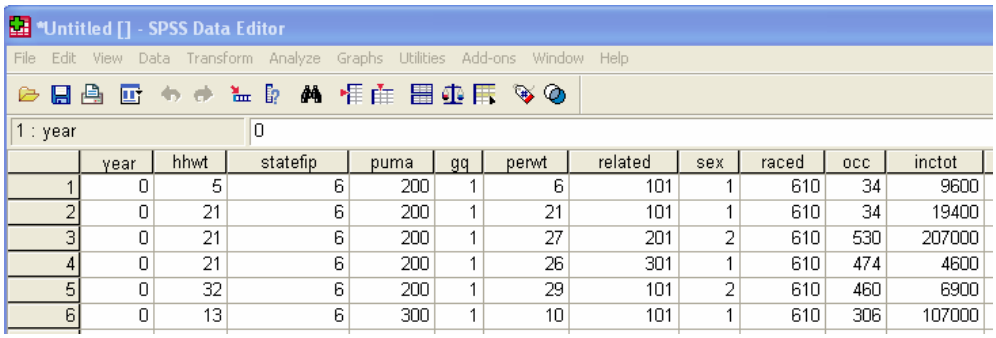


executed *Data List* command.

5. If all seems OK, select *Run* > *All* option.

SPSS should populate the spreadsheet. This may take awhile since PUMS files may be quite large, up to a million records in some cases. Note that occasionally you have to embed an *SPSS* command such as *Frequencies* at the end of the syntax file to make the command file finish executing properly.

At right is the populated *Data Editor* window.



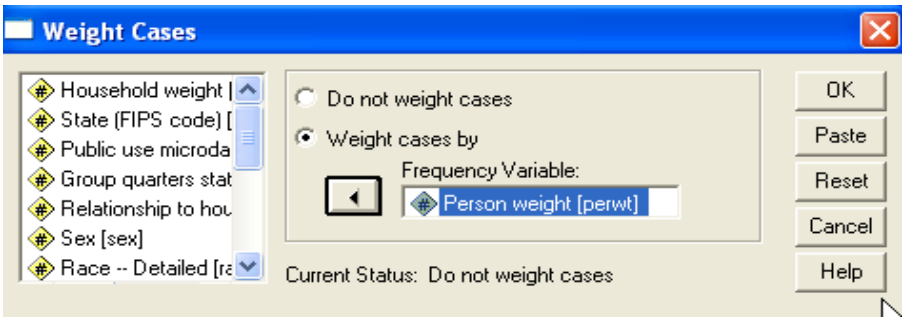
	year	hhwt	statefip	puma	gq	perwt	related	sex	raced	occ	inctot
1	0	5	6	200	1	6	101	1	610	34	9600
2	0	21	6	200	1	21	101	1	610	34	19400
3	0	21	6	200	1	27	201	2	610	530	207000
4	0	21	6	200	1	26	301	1	610	474	4600
5	0	32	6	200	1	29	101	2	610	460	6900
6	0	13	6	300	1	10	101	1	610	306	107000

6. Now select *File* > *Save As*. When the window below opens, name your file and save it in your work space. This creates an *SPSS.sav* file that you can use to directly open your data in the future. At this point you are ready to begin processing your data.

Analysis

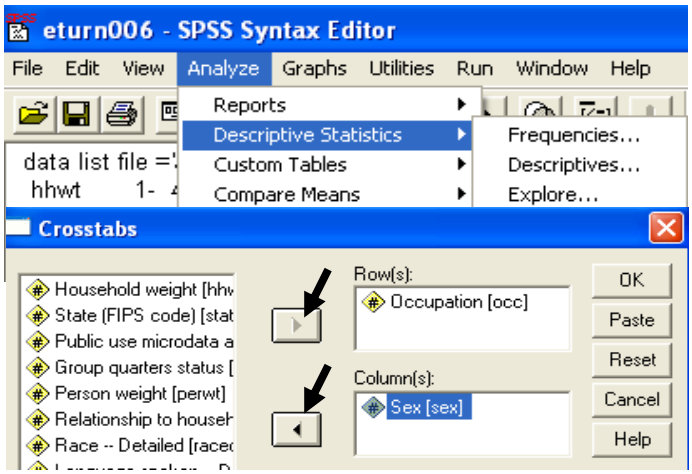
Our first goal is to determine how Asian Indian men and women have taken different jobs in different states. Select the FIPS codes for two states.

1. In the *Data Editor* window select *Data* > *Weight Cases*. When the window at right opens, click the *Weight Cases By* button and choose the *PERWT* variable. Click *OK*.



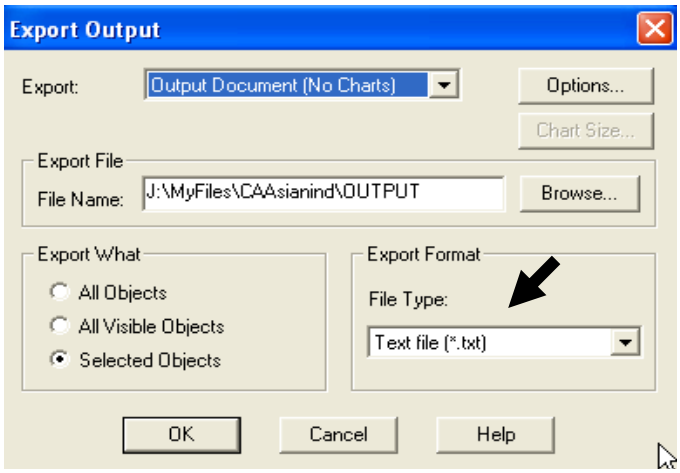
2. Select *Analyze* > *Descriptive Statistics* > *Crosstabs*.

3. When the *Crosstabs* window opens select *Occupation* from the list on the left and click on the *Rows* arrow to enter it in the top window. Do the same for *Sex* and click the *Columns* arrow in

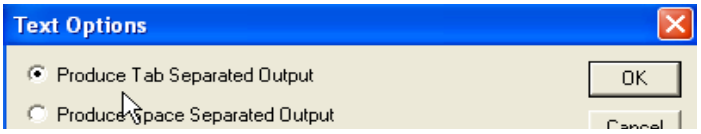


the lower window. Click *OK* and the table will be generated.

4. Look over your table in the *Output* window to make sure things are OK. Now select *File* > *Export*. When the *Export Output* window shown right opens, change the *Export Format* to *Text* file and click on the *Options* button.



5. Select the *Produce Tab Separated Ouput* and click *OK* two times.



6. You can now go to *Excel* and import your file. Compute the percentage of the employed population in each occupation and sort them in descending order. Note that you DO NOT want to include occupations coded with a 0 in your calculations since they are non-employed persons such as children.

What are the top ten categories for males and females?

Major Asian Indian Occupations
By Number and Percent
for Males and Females in California, 1990

Males			**Females**		
Code	**No.**	**Pct.**	**Code**	**No.**	**Pct.**
22	3619	6.5	276	1895	5.2
276	2497	4.5	95	1159	3.2
243	2290	4.1	313	1064	2.9

55	1693	3.0	447	1063	2.9
17	1571	2.8	274	933	2.6
64	1390	2.5	23	921	2.5
84	1322	2.4	379	903	2.5
53	1311	2.4	337	840	2.3
23	1257	2.3	785	754	2.1
804	1211	2.2	385	749	2.1

See the IPUMS documentation for Occupation Codes

Males

017 Managers, food serving and lodging establishments
022 Managers and administrators, n.e.c.
023 Accountants and auditors
053 Civil Engineers
055 Electrical and electronic Engineers
064 Computer systems analysts and scientists
084 Physicians
243 Supervisors and proprietors, sales occupations
276 Cashiers
804 Truck drivers

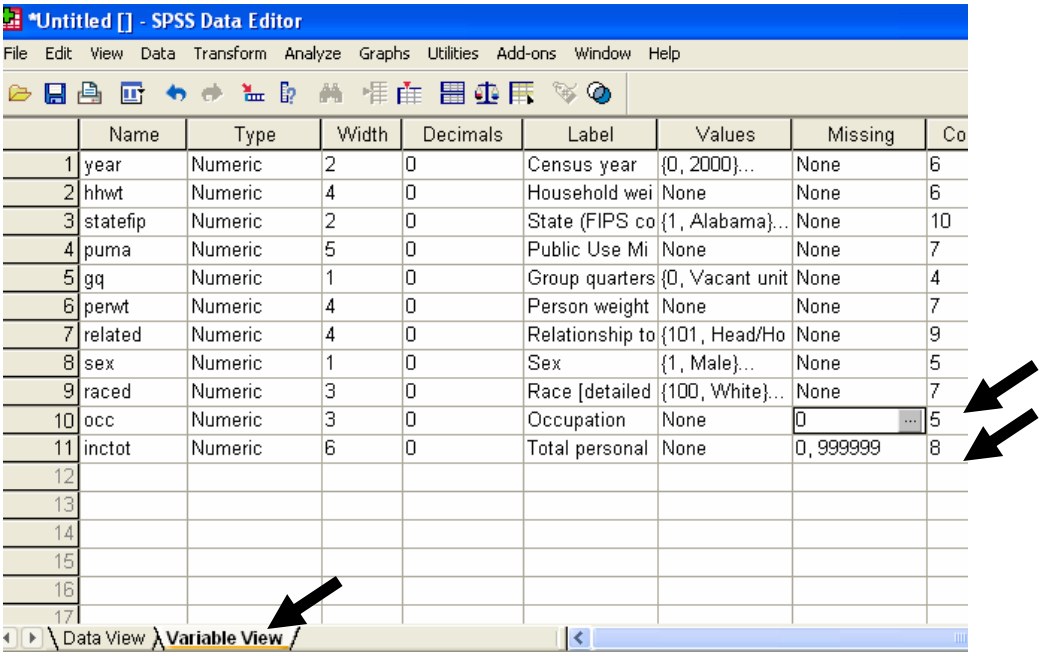
Females

023 Accountants and auditors
095 Registered nurses
274 Sales workers, other commodities
276 Cashiers
313 Secretaries
337 Bookkeepers, accounting, and auditing clerks
379 General office clerks
385 Data-entry keyers
447 Nursing aides, orderlies, and attendants
785 Assemblers

Personal Incomes for Men and Women

Another interesting issue that can be investigated in *PUMS* is the equity in income between men and women. Many believe that women are paid less than men for the same work and so you can check this by comparing the income from wages and salaries for similar groups of Asian Indian men and women. You can use the *Basic Tables* option to generally check on this issue.

1. First go to the *SPSS Data Editor* and select the *Variable View* tab at the bottom of the window. Locate the *Missing* column and the cell for *inctot*. Set the *Missing* values to *Discrete* with values of *0* and *999999* so that those values will not be included in calculations. Also set the *Missing* value for *Occupation* to *0*.

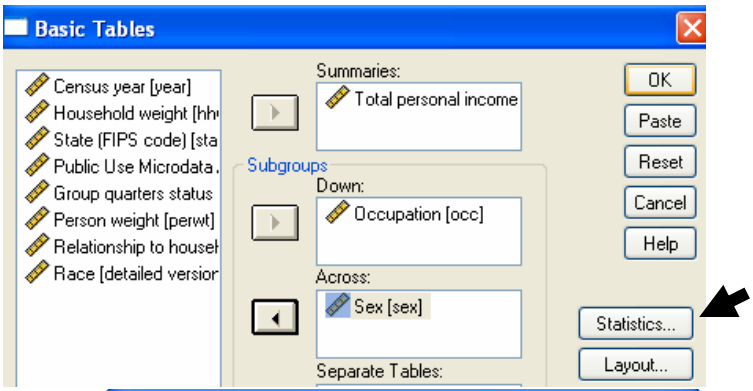


	Name	Type	Width	Decimals	Label	Values	Missing	Co
1	year	Numeric	2	0	Census year	{0, 2000}...	None	6
2	hhwt	Numeric	4	0	Household wei	None	None	6
3	statefip	Numeric	2	0	State (FIPS co	{1, Alabama}...	None	10
4	puma	Numeric	5	0	Public Use Mi	None	None	7
5	gq	Numeric	1	0	Group quarters	{0, Vacant unit	None	4
6	perwt	Numeric	4	0	Person weight	None	None	7
7	related	Numeric	4	0	Relationship to	{101, Head/Ho	None	9
8	sex	Numeric	1	0	Sex	{1, Male}...	None	5
9	raced	Numeric	3	0	Race [detailed	{100, White}...	None	7
10	occ	Numeric	3	0	Occupation	None	0	5
11	inctot	Numeric	6	0	Total personal	None	0, 999999	8
12								
13								
14								
15								
16								
17								

Then click *OK*. Click the *Variable View* tab.

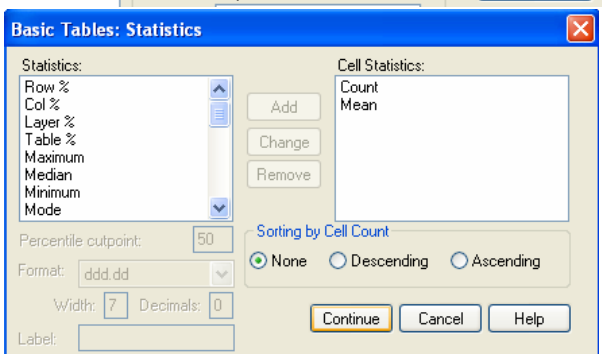
2. From the *Analyze* menu select *Tables* > *Basic Tables...*

3. Select the *Total income* item and click the arrow for *Summaries*. Select *Occupation* for the *Down Subgroup* and *Sex* for the *Across Subgroup*.



4. Select the *Statistics* button and the window right will open. Select the *Count* and *Mean* options.

Click the *Descending* button if you wish the *Occupations* sorted by counts. Then click *Continue*.



5. In the *Basic Tables* window click *OK* to start the processing.

In the *Output* window you will get a listing of the mean personal incomes for men and women in various occupations. Below, the *M/F Income Ratio* was created in *Excel* after exporting the table in the *Output* window.

CA Asian Indian Wage and Salary Personal Income
by Selected Occupation, 1990

	Male		Female		
Occup	**Count**	**Mean$**	**Count**	**Mean$**	**M/F Ratio**
0	27578	.	36149	.	
17	1571	32256	407	17655	1.8
22	3619	61210	727	25654	2.4
23	1257	30217	921	20871	1.4
53	1311	39181	79	35764	1.1
55	1693	45417	103	33861	1.3
64	1390	40599	313	31415	1.3
84	1322	100290	739	72692	1.4
95	74	12197	1159	30273	0.4
243	2290	24723	733	20166	1.2
274	854	13067	933	5093	2.6
276	2497	9151	1895	8050	1.1
313	68	10221	1064	18189	0.6
337	562	20708	840	14691	1.4
379	456	19526	903	11035	1.8
385	194	16746	749	12052	1.4
447	218	12186	1063	14766	0.8
785	556	16389	754	10116	1.6
804	1211	16381	13	12000	1.4

Using the percent employed in the top occupations compare the incomes of Asian Indian males and females. Do women and men earn similar incomes?

Exercises

1. Using the *Tables* analysis, compute the average Asian Indian income for males and females by PUMA. This could be mapped to see what areas in California pay higher wages to members of this group.

2. Compute the male/female income ratio for the PUMAs to see where women are paid better.

3. Go to the *IPUMS* web site and select the above data plus language spoken at home. See if different linguistic groups reflect different occupational niches among Asian Indians. Do some specialize in professional services, business, engineering, or health care?

4. Go to the *IPUMS* web site and select the above variables for a different state or a different ethnic group. Make sure your choice has a sufficient number of the group.