Sociology 3306 (Sections 570)
Investigating the Social World: Quantitative Research
Using SPSS (Statistical Package for the Social Sciences)
Assignment 1
Due February 5th (Monday, at the beginning of class)

SPSS consultant hours (lab east end of Cafeteria):
Monday 1:30 p.m. – 5:30 p.m.
Wednesday: 9:30 p.m. – 1:30 p.m.
Friday: 9:30 a.m.– 1:30 p.m.

Introduction:
The ability to work with SPSS (and other software packages) is a fundamental skill for sociologists and necessary in completing many of the assignments that we will be completing throughout the term. For this reason, we will be spending some time in the computing lab familiarizing ourselves with this software.

Virtually all of the computers in the Student Computer Lab (east end of Cafeteria) have an up to date version of SPSS (Statistical Package for the Social Sciences). Students will also be able to work with SPSS via remote access if they have access to internet from home. I recommend that you work with the version of SPSS as available in the computer labs or via remote access. There is a big advantage in doing these assignments in the computing lab as there is a consultant available when working with SPSS.

The purpose of this 1st assignment is to introduce you to this software and to some rather elementary data manipulations and statistical computations that are possible using SPSS. In addition, this assignment is meant to introduce you to the major datasets relied upon in the current course, including the 2009 General Social Survey (Victimization), the 2006 Canadian Census Public use File (Individuals), the 2010 Canadian Community Health Survey (CCHC) and the 1994 National Longitudinal Survey on Children and Youth (NLSCY). SPSS is probably the most widely used statistical package in sociology departments across Canada, largely due to its user friendly character. Once you become proficient on SPSS, you should not have too many difficulties in moving on to other statistical software, such as SAS (widely used outside of academia) or Stata (widely used by social scientists interested in apply certain advanced statistical procedures not available in SPSS). There are innumerable software packages used in neighboring social sciences. For example, the equivalent to SPSS in geography is GIS (Geographic Information Systems) which is particularly useful in manipulating data for various geographic units and mapping datasets.

Many of the examples provided in Sociology 2205 (introductory stats) involved relatively small samples (or few cases) in the explanation of some basic statistical procedures. Yet obviously in reality, much social research involves virtually 1000’s (or even 10,000’s) of cases. Consider a national survey of several 30,000 Canadians, involving the collection of detailed information on a wide range of variables. It is clearly not feasible to analyze such information with a hand calculator; hence the utility of software such as SPSS. Consider for that matter the Canadian census (Statistics Canada up in Ottawa has in-house a dataset with over 33 million cases as gathered in 2006). The Public Use File from the Census that

you work with a simple random sample of 800,000 persons drawn from this larger dataset.
For the purpose of the current course, I have selected 4 data sets, as aforementioned. These surveys are large, with respective samples of about 19,000 (GSS), 130,000 persons (CCHS), 800,000 cases (Public Use file: Census), and 23,000 cases (NLSCY) respectively. Information on these datasets can be obtained via the datasets page of my website. The actual data can be obtained via the “Scotty” system at Kings. Details follow.

****
Accessing Scotty:

Instructions for connecting can be found at the following link:

http://www.kings.uwo.ca/its/support/

This allows you to gain access to the Scotty Network from your home computer using the link:

http://scotty.kc.uwo.ca

Note: None of the SPSS datafiles will work on your home computer if you not do not a SPSS license installed on your machine locally. Hence, you must log in via this network or work in one of our computing labs (W022 or if sometimes available W143). If you want to print up your work, you can directly print them up using SPSS in the lab or you can convert SPSS files into WORD or PDF files using SPSS. You convert them, and then can send them home via email to work with (or alternatively, you can print your work up in the lab for a cost).

Should you experience any issues trying to access Scotty, you can email support@kings.uwo.ca.
****

The Census Public use file is a random sample of the 2006 Canadian Census (about 800,000 cases out of Canada’s whole population). This file has remarkably detailed information on the socioeconomic, cultural and demographic characteristics of Canada’s population. The **Canadian Community Health Survey** (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. Data were collected from over 130,000 respondents, aged 12 or older, residing in households across all provinces and territories. The 2009 **General Social Survey** is a survey of over 19,000 Canadians, aged 15 and older, asking them all sorts of questions relating to crime prevention, perceptions of crime, history and risk to crime, abuse, criminal incident reports, internet victimization, etc. The **National Longitudinal Survey of Children and Youth** (NLSCY) is a long-term survey designed to measure child development and well-being. The target population of the NLSCY for Cycle 1 consisted of Canadian children aged newborn to 11 years of age (and their families). The first cycle of the survey was conducted by Statistics Canada in 1994-1995 on behalf of Human Resources Development Canada.

In your final research project for this course you will be asked to do an analysis of one of the variables available in the GSS, CCHS, the Census or the NLSCY. Early on you should select the major “***DEPENDENT” variable*** for your study. In other words, what do you hope to “EXPLAIN” in your research?

Further details will be forthcoming on what sort of analysis is possible for your final paper, but most notably, your analysis will build directly on Assignment # 3 (due in March). This assignment is meant to familiarize you with both datasets as well as with some of the SPSS computing procedures. **I highly recommend that you decide on a topic and dataset ASAP.**

Using SPSS

SPSS is not a difficult software to use. One reason for this is its Windows menu system. However, the ease with which data can be manipulated using the menu system can be its downfall, because it is easy to forget what changes you have made.

One of the most commonly forgotten “rules” of doing social research is “THE NEED FOR PROPER DOCUMENTATION”. This is particularly true if you are a researcher collaborating with many others on the same project. In other words, what did I do with this dataset yesterday (or what did my colleague do with this dataset!!). Therefore, it is always useful to keep your files well organized with proper documentation, so that you can always return to work previously completed without too many difficulties. One way to do so is to use SPSS’s syntax system to keep a record of what you have previously done. While this is a bit more difficult to learn, it is worth doing so as you can save enormous amounts of time over the longer term. It is also useful to title your output, for easy recognition later.

SPSS files

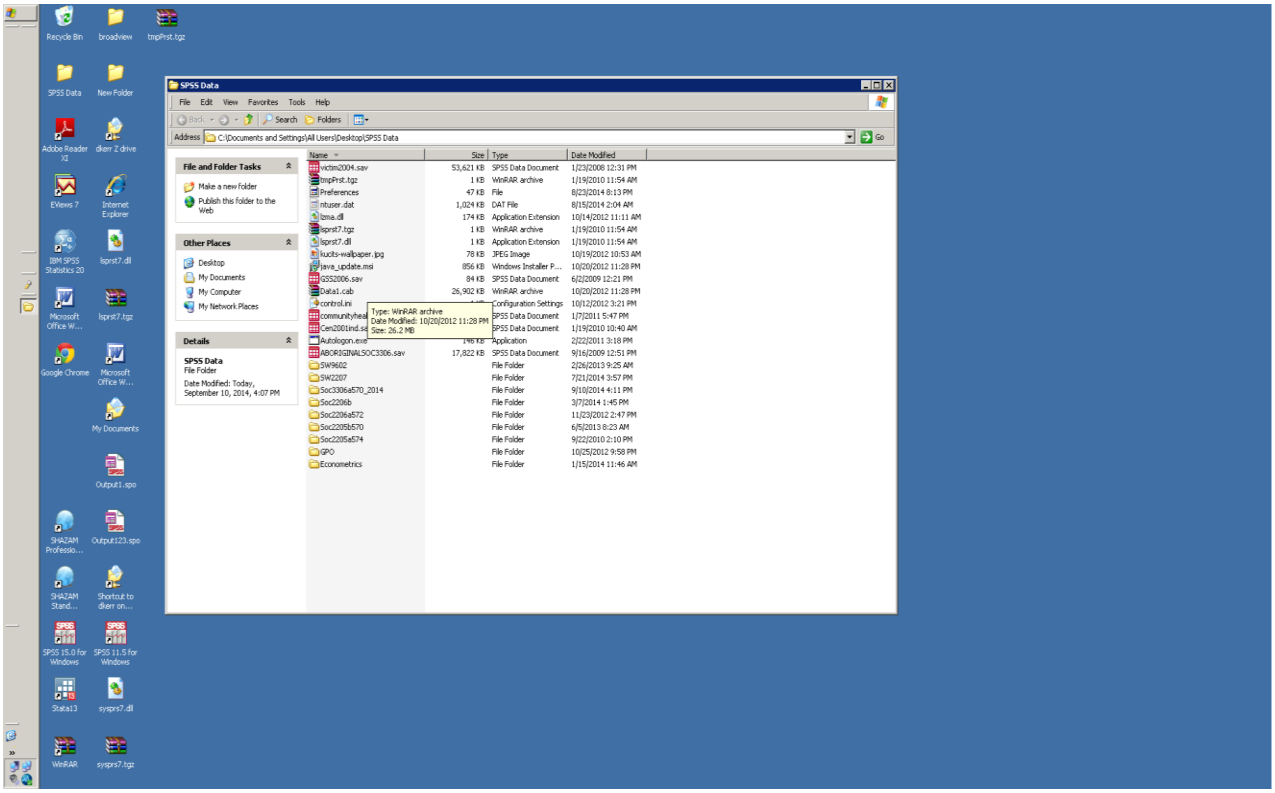
There are 3 types of windows that SPSS handles in order to create 3 different types of files.

1. Data files (*.sav) contain the data that any commands will manipulate and analyze. A data file must be open in order to perform an analysis.
2. Output files (*.spo) contain the output produced by SPSS, including any graphs, tables, or numbers. Results shown in output windows can generally be copied and pasted in to word processing documents.
3. Syntax files (*.sps) contain programs that can be run on SPSS, in its own programming language.

Each type of file opens in its own specific type of window.

Data files:

To demonstrate what a data file looks like, you will need to log into Scotty and gain access to the server at Kings (see above instructions). Once you do so, your screen should look a lot like:



Address: C:\Documents and Settings\All Users\Desktop\SPSS Data

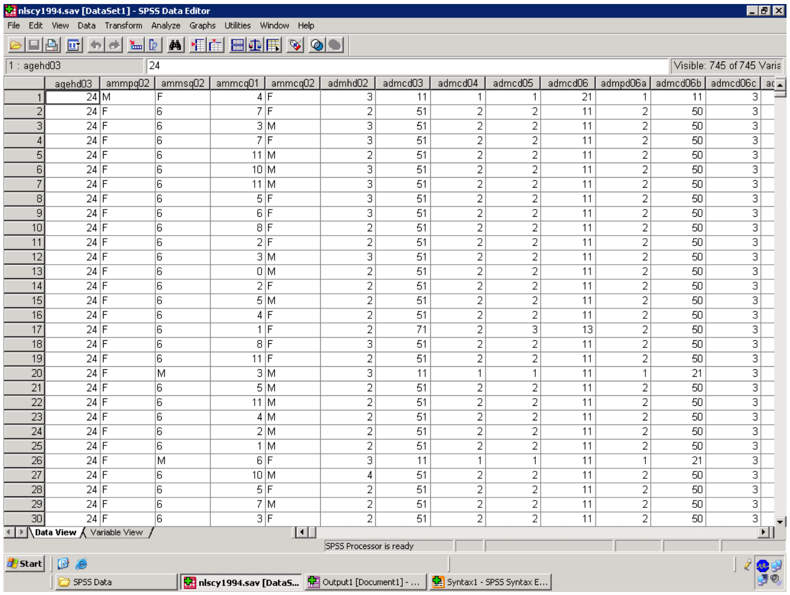
SPSS Data
File Folder
Date Modified: Today, September 10, 2014, 4:07 PM

Name	Size	Type	Date Modified
victim2004.sav	53,621 KB	SPSS Data Document	1/23/2008 12:31 PM
tmpPrst.tgz	1 KB	WinRAR archive	1/19/2010 11:54 AM
Preferences	47 KB	File	8/23/2014 8:13 PM
ntuser.dat	1,024 KB	DAT File	8/15/2014 2:04 AM
lzma.dll	174 KB	Application Extension	10/14/2012 11:11 AM
lsprst7.tgz	1 KB	WinRAR archive	1/19/2010 11:54 AM
lsprst7.dll	1 KB	Application Extension	1/19/2010 11:54 AM
kucits-wallpaper.jpg	78 KB	JPEG Image	10/19/2012 10:53 AM
java_update.msi	856 KB	Windows Installer P...	10/20/2012 11:28 PM
GSS2006.sav	84 KB	SPSS Data Document	6/2/2009 12:21 PM
Data1.cab	26,902 KB	WinRAR archive	10/20/2012 11:28 PM
control.ini	1 KB	Configuration Settings	10/12/2012 3:21 PM
communityheal		SPSS Data Document	1/7/2011 5:47 PM
Cen2001ind.sa		SPSS Data Document	1/19/2010 10:40 AM
Autologon.exe	146 KB	Application	2/22/2011 3:18 PM
ABORIGINALSOC3306.sav	17,822 KB	SPSS Data Document	9/16/2009 12:51 PM
SW9602		File Folder	2/26/2013 9:25 AM
SW2207		File Folder	7/21/2014 3:57 PM
Soc3306a570_2014		File Folder	9/10/2014 4:11 PM
Soc2206b		File Folder	3/7/2014 1:45 PM
Soc2206a572		File Folder	11/23/2012 2:47 PM
Soc2205b570		File Folder	6/5/2013 8:23 AM
Soc2205a574		File Folder	9/22/2010 2:10 PM
GPO		File Folder	10/25/2012 9:58 PM
Econometrics		File Folder	1/15/2014 11:46 AM

Type: WinRAR archive
Date Modified: 10/20/2012 11:28 PM
Size: 26.2 MB

All datasets for this course are available by clicking on the SPSS DATA folder that should appear on your desktop. You can find the four datasets associated with this course in the folder: Soc3306a570_2016_2017 (machine readable only). You can start your SPSS program automatically by clicking on the appropriate data file listed here. If you have trouble opening it for some strange reason, open your SPSS program first.

Once you have opened a dataset, your screen should show the SPSS data editor, with all the appropriate variables and cases, as follows:


nlscy1994.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : agehd03
24
Visible: 745 of 745 Varia
Data View
Variable View
SPSS Processor is ready
Start
SPSS Data
nlscy1994.sav [DataS...
Output1 [Document1] - ...
Syntax1 - SPSS Syntax E...

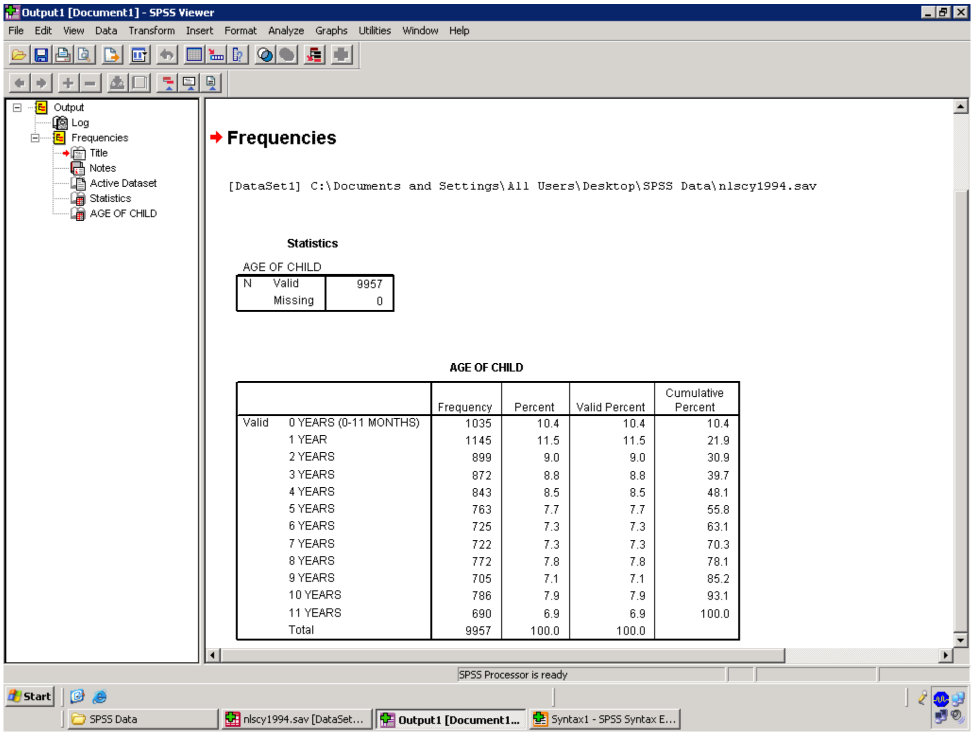
	agehd03	ammpq02	ammsq02	ammcq01	ammcq02	admhd02	admcd03	admcd04	admcd05	admcd06	admpd06a	admcd06b	admcd06c	ac
1	24	M	F	4	F	3	11	1	1	21	1	11	3	
2	24	F	6	7	F	2	51	2	2	11	2	50	3	
3	24	F	6	3	M	3	51	2	2	11	2	50	3	
4	24	F	6	7	F	3	51	2	2	11	2	50	3	
5	24	F	6	11	M	2	51	2	2	11	2	50	3	
6	24	F	6	10	M	3	51	2	2	11	2	50	3	
7	24	F	6	11	M	3	51	2	2	11	2	50	3	
8	24	F	6	5	F	3	51	2	2	11	2	50	3	
9	24	F	6	6	F	3	51	2	2	11	2	50	3	
10	24	F	6	8	F	2	51	2	2	11	2	50	3	
11	24	F	6	2	F	2	51	2	2	11	2	50	3	
12	24	F	6	3	M	3	51	2	2	11	2	50	3	
13	24	F	6	0	M	2	51	2	2	11	2	50	3	
14	24	F	6	2	F	2	51	2	2	11	2	50	3	
15	24	F	6	5	M	2	51	2	2	11	2	50	3	
16	24	F	6	4	F	2	51	2	2	11	2	50	3	
17	24	F	6	1	F	2	71	2	3	13	2	50	3	
18	24	F	6	8	F	3	51	2	2	11	2	50	3	
19	24	F	6	11	F	2	51	2	2	11	2	50	3	
20	24	F	M	3	M	3	11	1	1	11	1	21	3	
21	24	F	6	5	M	2	51	2	2	11	2	50	3	
22	24	F	6	11	M	2	51	2	2	11	2	50	3	
23	24	F	6	4	M	2	51	2	2	11	2	50	3	
24	24	F	6	2	M	2	51	2	2	11	2	50	3	
25	24	F	6	1	M	2	51	2	2	11	2	50	3	
26	24	F	M	6	F	3	11	1	1	11	1	21	3	
27	24	F	6	10	M	4	51	2	2	11	2	50	3	
28	24	F	6	5	F	2	51	2	2	11	2	50	3	
29	24	F	6	7	M	2	51	2	2	11	2	50	3	
30	24	F	6	3	F	2	51	2	2	11	2	50	3	

Here we have the contents from the NLSCY in a spreadsheet format. A lot of work has already gone into setting this dataset up for you. This dataset summarizes the responses of over 20,000 individuals across about 700 variables. Across the top of the dataset you will see the assigned variable names that SPSS uses in reading this data set “agehd03, ammpq02, etc..”. If you move your arrow with your mouse across the variables names, it is possible to see the full name of each. For your information, I have posted the corresponding codebooks and a description of the datasets on my website. Optionally, you can click on UTILITIES then VARIABLES then the variable you are interested in – if you require details on any single variable.

Whereas each column in this dataset represents a variable, each line of this dataset represents a specific “case”. Our “unit of analysis” in working with this dataset is the “individual”, with each row representing the responses across variables for one respondent to the NLSCY. Theoretically, it is possible to make changes on any entry with the SPSS data editor in this spreadsheet (yet obviously, we should not be doing this unless we have a very good reason). Whenever you need to access this dataset, you can follow the above steps. You will know that the file is a data file given that it ends with *.sav .

Output Files

The output window (*.spo) looks like:


Output1 [Document1] - SPSS Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
AGE OF CHILD
SPSS Processor is ready
Start
SPSS Data
nlscy1994.sav [DataSet...
Output1 [Document1...
Syntax1 - SPSS Syntax E...

Frequencies

```
[DataSet1] C:\Documents and Settings\All Users\Desktop\SPSS Data\nlscy1994.sav
```

Statistics

AGE OF CHILD

N	Valid	9957
	Missing	0

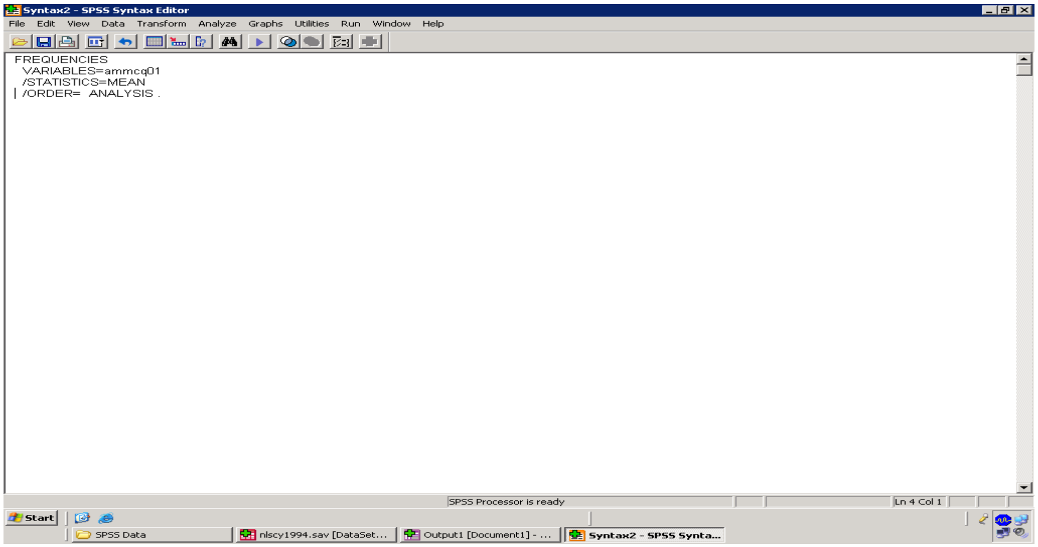
AGE OF CHILD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 YEARS (0-11 MONTHS)	1035	10.4	10.4	10.4
	1 YEAR	1145	11.5	11.5	21.9
	2 YEARS	899	9.0	9.0	30.9
	3 YEARS	872	8.8	8.8	39.7
	4 YEARS	843	8.5	8.5	48.1
	5 YEARS	763	7.7	7.7	55.8
	6 YEARS	725	7.3	7.3	63.1
	7 YEARS	722	7.3	7.3	70.3
	8 YEARS	772	7.8	7.8	78.1
	9 YEARS	705	7.1	7.1	85.2
	10 YEARS	786	7.9	7.9	93.1
	11 YEARS	690	6.9	6.9	100.0
	Total	9957	100.0	100.0	

This output file *.spo gives us a frequency distribution on the fourth variable in our data set "age of child" (ammcq01). In completing your assignments, you will regularly be printing up these output files. I will ask you to regularly provide them when documenting your work.

Syntax Files

A syntax file looks like:


Syntax2 - SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Utilities Run Window Help
SPSS Processor is ready
Ln 4 Col 1
Start
SPSS Data
nlscy1994.sav [DataSet...
Output1 [Document1] - ...
Syntax2 - SPSS Synta...

```
FREQUENCIES
  VARIABLES=ammcq01
  /STATISTICS=MEAN
  /ORDER=  ANALYSIS .
```

This syntax file runs a simple frequency distribution on the variable “age of child” and asks the computer to calculate the “mean” on this variable. At one time, the only way to run SPSS was in creating “syntax” files like this one, whereas one could technically work with SPSS currently without creating syntax files. Each type of file can be saved using the file menu in Windows. Syntax files can potentially run several pages long.

Syntax and the Menu System

There are two ways to execute a command in SPSS. You can use the point-and-click Windows interface, and select the options you desire. This can easily be done from the data window. Unfortunately, using this interface it is possible to alter the data while not realizing what you have done. The other way is by running a program in a syntax window. This method is preferred because a program can be run over and over again, and you have a record of the analysis or data manipulations you have performed. Originally, this was the only way to execute SPSS commands. Unfortunately, this meant understanding the SPSS programming language and its rules, and the specific format required for any commands. However, SPSS now allows you to select commands and options from the menu system and “paste” them into a syntax window, building a program as you go.

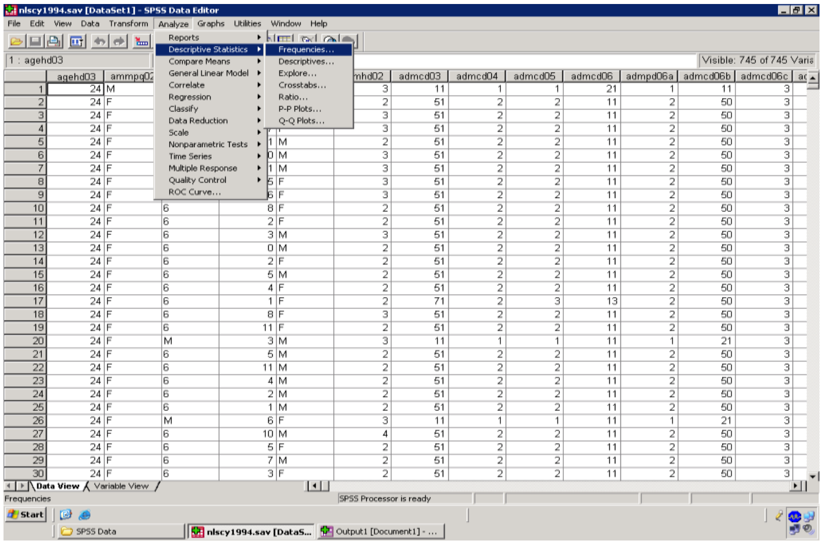
To run a piece of syntax in the syntax window, highlight it and select “run… selection”. For help writing a program in SPSS syntax, you can look at the Syntax Guide under the “help” menu.

Obtaining Descriptive Statistics in SPSS

Frequencies You can obtain a frequency distribution in different ways. In the menu system, you merely follow the hits:

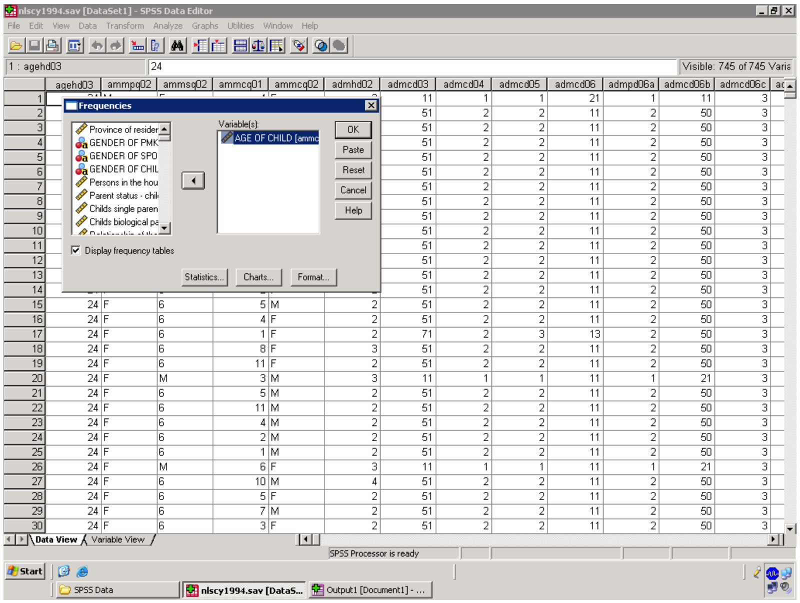
“analyse, descriptive statistics, frequencies”.

For example in creating a frequency distribution and histogram for “ammcq01”:



	agehd03	ammpq02				mhd02	admcd03	admcd04	admcd05	admcd06	admpd06a	admcd06b	admcd06c	ac
1	24	M				3	11	1	1	21	1	11	3	
2	24	F				2	51	2	2	11	2	50	3	
3	24	F				3	51	2	2	11	2	50	3	
4	24	F		7	F	3	51	2	2	11	2	50	3	
5	24	F		1	M	2	51	2	2	11	2	50	3	
6	24	F		0	M	3	51	2	2	11	2	50	3	
7	24	F		1	M	3	51	2	2	11	2	50	3	
8	24	F		5	F	3	51	2	2	11	2	50	3	
9	24	F		6	F	3	51	2	2	11	2	50	3	
10	24	F	6	8	F	2	51	2	2	11	2	50	3	
11	24	F	6	2	F	2	51	2	2	11	2	50	3	
12	24	F	6	3	M	3	51	2	2	11	2	50	3	
13	24	F	6	0	M	2	51	2	2	11	2	50	3	
14	24	F	6	2	F	2	51	2	2	11	2	50	3	
15	24	F	6	5	M	2	51	2	2	11	2	50	3	
16	24	F	6	4	F	2	51	2	2	11	2	50	3	
17	24	F	6	1	F	2	71	2	3	13	2	50	3	
18	24	F	6	8	F	3	51	2	2	11	2	50	3	
19	24	F	6	11	F	2	51	2	2	11	2	50	3	
20	24	F	M	3	M	3	11	1	1	11	1	21	3	
21	24	F	6	5	M	2	51	2	2	11	2	50	3	
22	24	F	6	11	M	2	51	2	2	11	2	50	3	
23	24	F	6	4	M	2	51	2	2	11	2	50	3	
24	24	F	6	2	M	2	51	2	2	11	2	50	3	
25	24	F	6	1	M	2	51	2	2	11	2	50	3	
26	24	F	M	6	F	3	11	1	1	11	1	21	3	
27	24	F	6	10	M	4	51	2	2	11	2	50	3	
28	24	F	6	5	F	2	51	2	2	11	2	50	3	
29	24	F	6	7	M	2	51	2	2	11	2	50	3	
30	24	F	6	3	F	2	51	2	2	11	2	50	3	

Specific variables or sets of variables can be moved over by merely highlighting the variable of interest and clicking on the arrow key. For example, we have moved over the variable of interest "ammcq01".


nlscy1994.sav [DataSet1] - SPSS Data Editor

1 : agehd03 24

Visible: 745 of 745 Varia

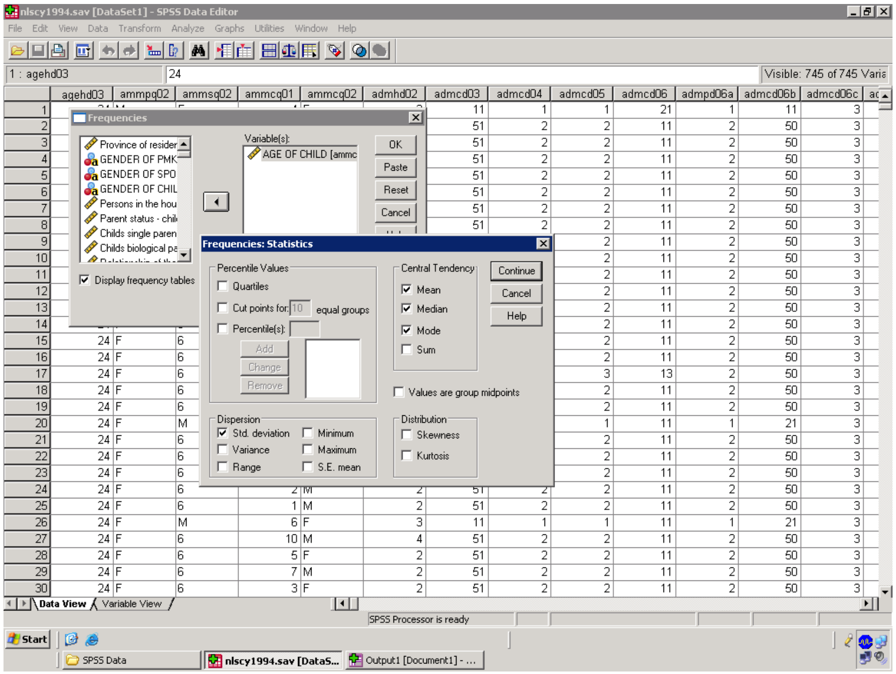
Frequencies

Variable(s): AGE OF CHILD [ammc

☑ Display frequency tables

SPSS Processor is ready

By clicking on "Statistics" you can select whatever descriptive statistics you want (mean, mode, standard deviation, etc). If you click on charts, you can specify that you want a histogram, etc.


nlscy1994.sav [DataSet1] - SPSS Data Editor

1 : agehd03 24

Visible: 745 of 745 Varia

Frequencies

Variable(s): AGE OF CHILD [ammc

☑ Display frequency tables

Frequencies: Statistics

Percentile Values
□ Quartiles
□ Cut points for: 10 equal groups
□ Percentile(s):

Central Tendency
☑ Mean
☑ Median
☑ Mode
□ Sum

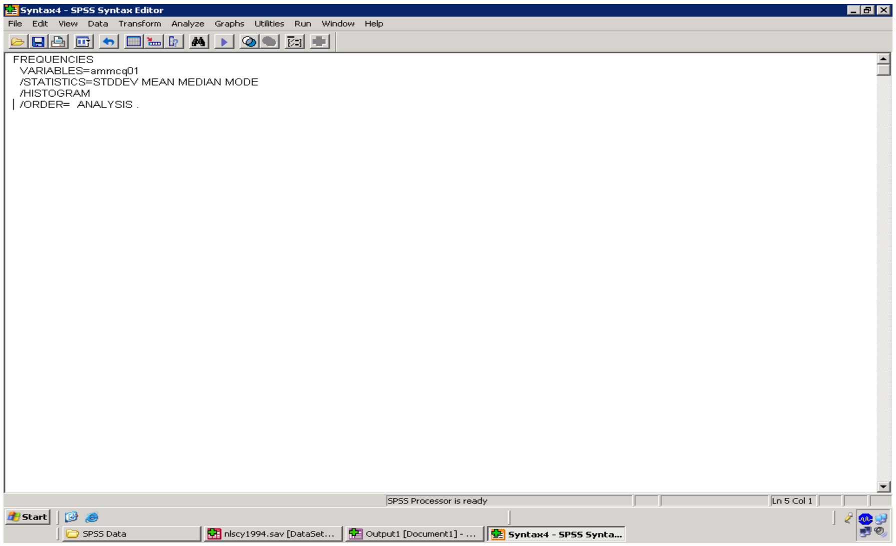
□ Values are group midpoints

Dispersion
☑ Std. deviation
□ Minimum
□ Variance
□ Maximum
□ Range
□ S.E. mean

Distribution
□ Skewness
□ Kurtosis

SPSS Processor is ready

By clicking on "paste" prior to "OK" you can create a SYNTAX file that you can work with:


Syntax4 - SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Utilities Run Window Help
SPSS Processor is ready
Ln 5 Col 1
Start
SPSS Data
nlscy1994.sav [DataSet...
Output1 [Document1] - ...
Syntax4 - SPSS Synta...

```
FREQUENCIES
  VARIABLES=ammcq01
  /STATISTICS=STDDEV MEAN MEDIAN MODE
  /HISTOGRAM
  /ORDER=  ANALYSIS .
```

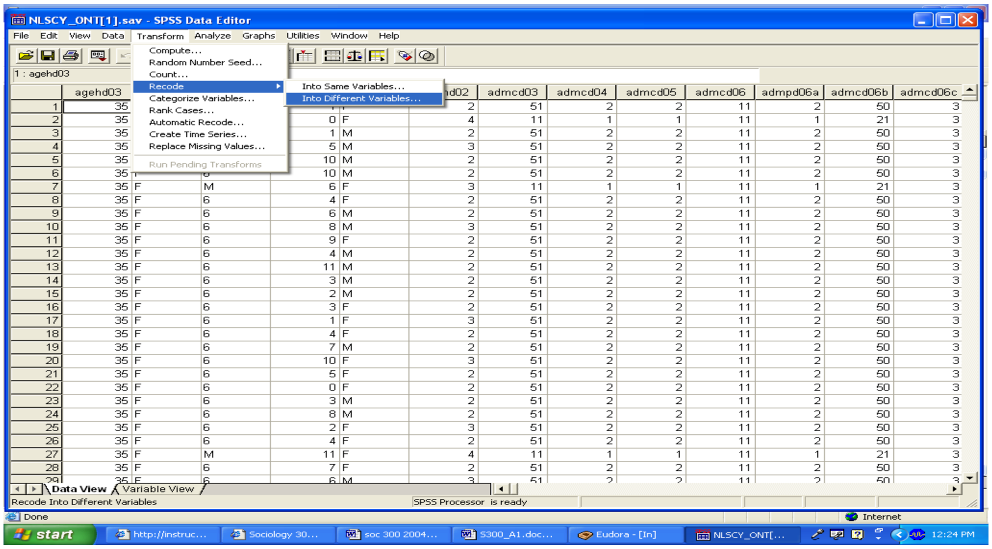
If you highlight and run these commands (this syntax), the software will produce a frequency distribution, standard deviation, median, mode and a histogram with a normal curve superimposed, for the variable of interest “ammcq01”. The latter option of a superimposed normal curve was specified on the “charts” box.

Recoding Variables in SPSS

It is often necessary to create new variables or to re-code existing variables with new values. For example, we may have a variable indicating age in years, but may wish to create a new variable with age in five year intervals. It is generally a good idea to create a new variable, rather than changing the values in existing variables (the new variable that you create would have a new name in your dataset and be placed in a separate column).

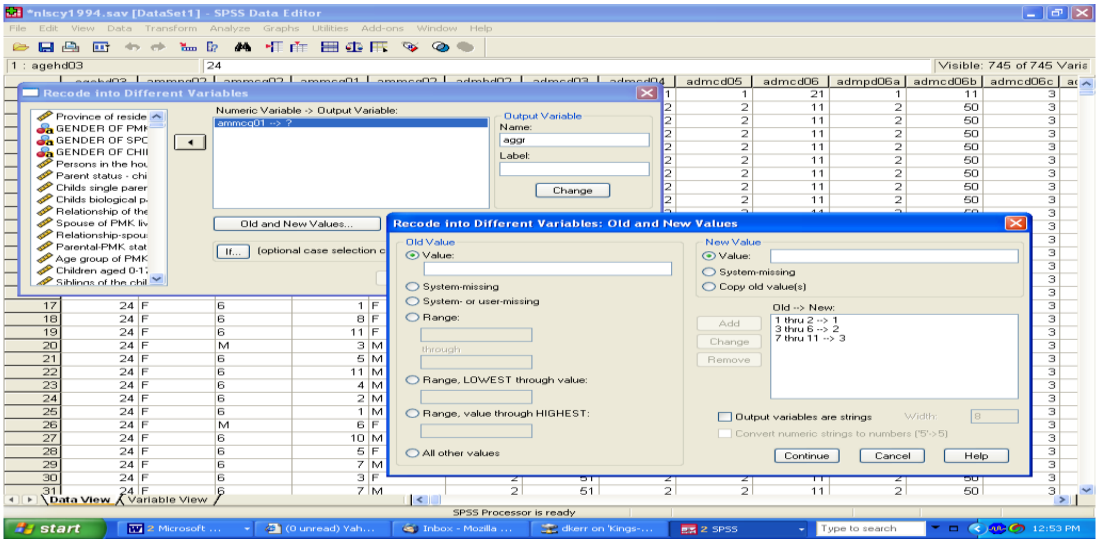
Example: re-categorizing a variable.

To recode a variable using the menu system, choose “transform, recode, into different variables”.



	agehd03					d02	admcd03	admcd04	admcd05	admcd06	admpd06a	admcd06b	admcd06c
1	35					2	51	2	2	11	2	50	3
2	35			0	F	4	11	1	1	11	1	21	3
3	35			1	M	2	51	2	2	11	2	50	3
4	35			5	M	3	51	2	2	11	2	50	3
5	35			10	M	2	51	2	2	11	2	50	3
6	35			10	M	2	51	2	2	11	2	50	3
7	35	F	M	6	F	3	11	1	1	11	1	21	3
8	35	F	6	4	F	2	51	2	2	11	2	50	3
9	35	F	6	6	M	2	51	2	2	11	2	50	3
10	35	F	6	8	M	3	51	2	2	11	2	50	3
11	35	F	6	9	F	2	51	2	2	11	2	50	3
12	35	F	6	4	M	2	51	2	2	11	2	50	3
13	35	F	6	11	M	2	51	2	2	11	2	50	3
14	35	F	6	3	M	2	51	2	2	11	2	50	3
15	35	F	6	2	M	2	51	2	2	11	2	50	3
16	35	F	6	3	F	2	51	2	2	11	2	50	3
17	35	F	6	1	F	3	51	2	2	11	2	50	3
18	35	F	6	4	F	2	51	2	2	11	2	50	3
19	35	F	6	7	M	2	51	2	2	11	2	50	3
20	35	F	6	10	F	3	51	2	2	11	2	50	3
21	35	F	6	5	F	2	51	2	2	11	2	50	3
22	35	F	6	0	F	2	51	2	2	11	2	50	3
23	35	F	6	3	M	2	51	2	2	11	2	50	3
24	35	F	6	8	M	2	51	2	2	11	2	50	3
25	35	F	6	2	F	3	51	2	2	11	2	50	3
26	35	F	6	4	F	2	51	2	2	11	2	50	3
27	35	F	M	11	F	4	11	1	1	11	1	21	3
28	35	F	6	7	F	2	51	2	2	11	2	50	3
29	35	F	6	6	M	3	51	2	2	11	2	50	3

A box will open that will require you to specify a new target variable, and the rules for recoding the variable. For example, suppose that we wanted to recode the variable “ammcq01” (age of child) into a modified variable (agegr), whereby we collapse the original variable into fewer categories.   The first step is to always consider how your variable was originally coded in the dataset.   You can obtain this information from either the “utilities” function in SPSS or via the code book.


*nlscy1994.sav [DataSet1] - SPSS Data Editor
Recode into Different Variables
Recode into Different Variables: Old and New Values
Old and New Values...
If...
Change
Add
Remove
Continue
Cancel
Help
SPSS Processor is ready

Numeric Variable -> Output Variable:
ammcq01 --> ?

Output Variable
Name: aggr
Label:

Old Value
● Value:
○ System-missing
○ System- or user-missing
○ Range:
○ Range, LOWEST through value:
○ Range, value through HIGHEST:
○ All other values

New Value
● Value:
○ System-missing
○ Copy old value(s)

Old --> New:
1 thru 2 --> 1
3 thru 6 --> 2
7 thru 11 --> 3

□ Output variables are strings   Width: 8
□ Convert numeric strings to numbers ('5'->5)

We then type in the name of the new variable “agegr” and give it an optional label (age group of child). Then click on the “old and new values” button in order to specify the rules for creating this new variable. According to the code book, ammcq01 is originally coded such that it ranges from 0 to 11, representing responses from less than one year of age through to 11 years of age.   For the purpose of this exercise, assume that interested in recoding this variable, such that the new variable (agegr) has only 3 categories: (1) aged 0-2 years, (2) aged 3-6, and (3) aged 7-11 years.

This variable can then be recoded using this procedure. In terms of the syntax for an SPSS program, to recode the age of child variable “ammcq01” into the variable “agegr”, you could:

```
compute agegr=0.
IF (ammcq01 ge 0) and (ammcq01 le 2 agegr = 1.
IF  (ammcq01 ge 3) and (ammcq01 le 6) agegr = 2.
IF  (ammcq01 ge 7) and (ammcq01 le 11) agegr = 3.
```

This syntax first creates (initializes) a new variable, “agegr”, with all of the values equal to 0. Then, if the value of the old age variable is greater than or equal to 0 *and* it is less than or equal to 2, the value of the new “agegr” variable is set to 1. Likewise, if the value of the old age variable was greater than or equal to 3 *and* it is less than or equal to 6, the new “agegr” variable is set to 2, and so on. The operators “le”, “gt”, “ge” “lt”, “ne” and “=” can be used to specify “less than or equal to”, “greater than”, “greater than or equal to”, “less than,” “not equal to” and “equal to”, respectively.

Remember that it is always necessary to specify "missing values" in the new variable that you are creating if the old variable has them. In this case, all of the cases which don't fit any of the 3 "IF lines" above will have a "agegr" value of zero. It is important to look carefully at the data to make sure that the transformations have occurred correctly, and to specify the missing values for the variable if they exist.

To include a missing value with the above 3 IF lines, you can use the following syntax:

```
RECODE ammcq01 (99=SYSMIS) INTO agegr.
```

The above syntax which involves creating the new "agegr" variable, could be achieved in a simpler manner through a single recode statement:

```
RECODE
  ammcq01
  (0 thru 2=1)  (3 thru 6=2)  (7 thru 11=3)  (ELSE=SYSMIS)  INTO  agegr .
```

One important thing to note in creating a syntax file (SPSS program) is that your recode or compute procedures must always come before any specific statistical procedures, such as "frequencies", "descriptives" or "explore". For example, the following syntax first creates the new variable "agegr" prior to running the frequencies on this variable (as well as the original variable ammcq01)

```
RECODE
  ammcq01
  (0 thru 2=1)  (3 thru 6=2)  (7 thru 11=3)  (ELSE=SYSMIS)  INTO  agegr .

FREQUENCIES
  VARIABLES= agegr
  /ORDER=  ANALYSIS .
```

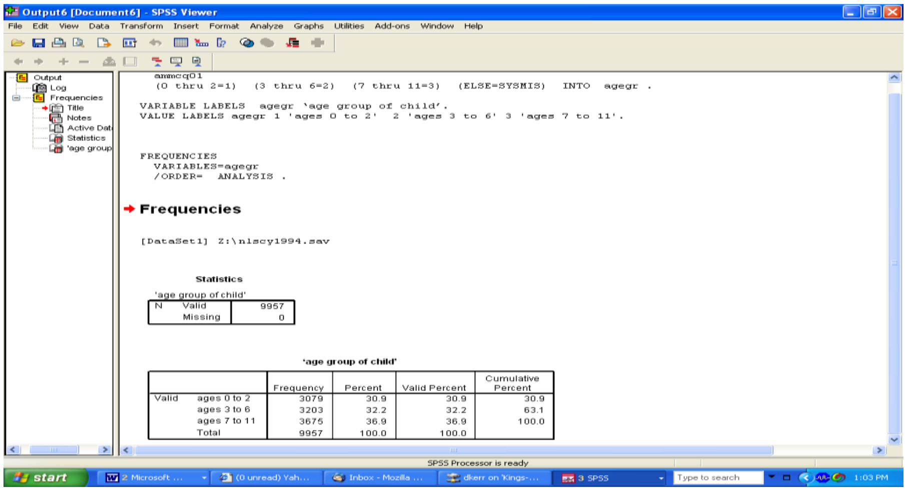
Variable and Value labels

An important part of documenting your work is adding variable and value labels whenever you create new variables.   This can be done relatively easily with SPSS syntax.

Returning to the previous example, after creating the new variable, we can specify the variable label (i.e. what we want to call the new variable "agegr") as well as identify the corresponding value labels (i.e. what we want to call each category of the variable we just created). The variable name must be 7 characters or less, whereas you should also try to keep the value labels relatively short.

```
RECODE
  ammcq01
  (0 thru 2=1)  (3 thru 6=2)  (7 thru 11=3)  (ELSE=SYSMIS)  INTO  agegr .

VARIABLE LABELS  agegr ‘age group of child’.
VALUE LABELS agegr 1 'ages 0 to 2'  2 'ages 3 to 6' 3 'ages 7 to 11'.
```

If we then run a frequency distribution on agegr, we should observe the newly specified labels.



```
   ammcq01
   (0 thru 2=1)   (3 thru 6=2)   (7 thru 11=3)   (ELSE=SYSMIS)   INTO   agegr .

 VARIABLE LABELS  agegr ‘age group of child’.
 VALUE LABELS agegr 1 'ages 0 to 2'  2 'ages 3 to 6' 3 'ages 7 to 11'.


 FREQUENCIES
   VARIABLES=agegr
   /ORDER=  ANALYSIS .
```

Frequencies

```
[DataSet1] Z:\nlscy1994.sav
```

Statistics

'age group of child'

N	Valid	9957
	Missing	0

'age group of child'

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ages 0 to 2	3079	30.9	30.9	30.9
	ages 3 to 6	3203	32.2	32.2	63.1
	ages 7 to 11	3675	36.9	36.9	100.0
	Total	9957	100.0	100.0	

Note: in working with the NLSCY, someone has in fact gone through the trouble of setting up a database that has all the variable names and value labels already allocated. When you create new variables, you should subsequently specify these new names and usually (yet not always) need specify labels for these variables.

NOTE: A few words on creating new variables using SPSS (required for step 4 in the current assignment).

Using SPSS, it is possible to create new variables that combine the scores of variables already on your dataset. For example, it is possible to create what are called “additive scales”. An additive scale can be created by merely adding up scores across variables already in your dataset. For example, consider the following situation (the following variables are not in either of your datasets). Assume that you are interested in studying the victimization of Canadians to crime, and with your dataset you have a series of variables that measure “victimization” (VAR1, VAR2, VAR3 AND VAR4).

Hypothetically, assume you had a series of variables coded in the following manner:

VAR1
Have you ever been robbed by somebody with a weapon?
0. no
1. yes

VAR2
Have you ever been physically assaulted?
0. no
1. yes

VAR3
Has your home ever been burglarized?
0. no
1. yes

VAR4
Have you ever lost money through fraud?
0. yes
1. no

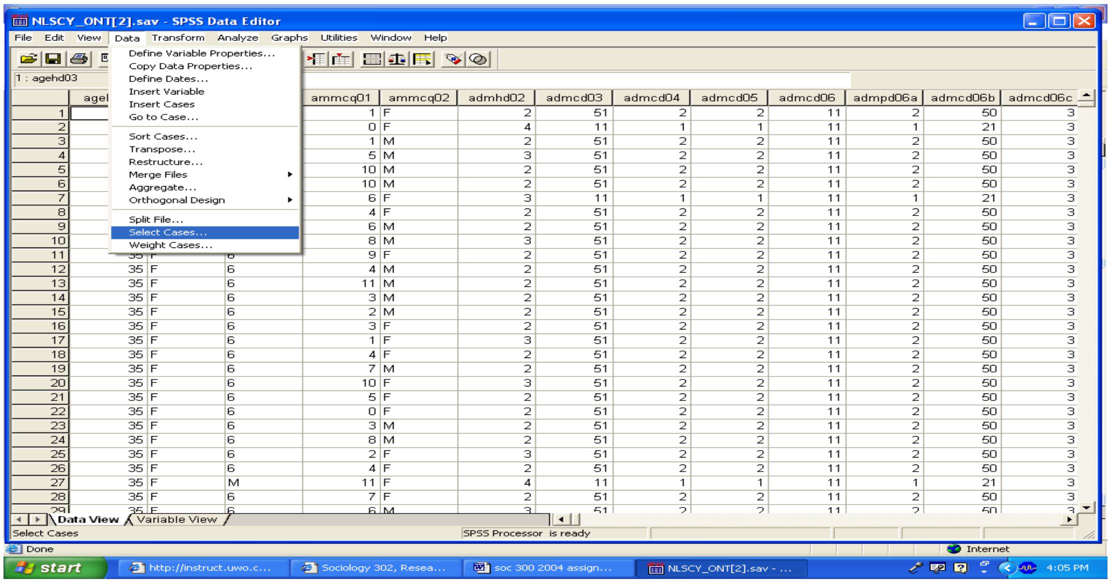
Using SPSS it is possible to create additive scales using these variables, by merely creating a new variable that adds up the scores of these variables, for each and every case. The format of the command for your syntax file is as follows:

```
COMPUTE      CRIME = VAR1 + VAR2 + VAR3 + VAR4.
```

With this command entered into your syntax file, a new variable CRIME could be created, with a specific value for each case in your dataset. Under this hypothetical situation, you would be left with a scale ranging from 0 to 4, depending on how persons responded on VAR1 through VAR4. If an individual in your dataset had never experienced any of the crimes as listed, he/she is left with a score of 0 on the new variable CRIME. If he/she had experienced "a burglary" and "fraud", he/she scores 2., etc. At times, it is necessary to recode variables prior to combining them into an additive scale (which is what you do in Step 3 of the current assignment).

Selecting Cases

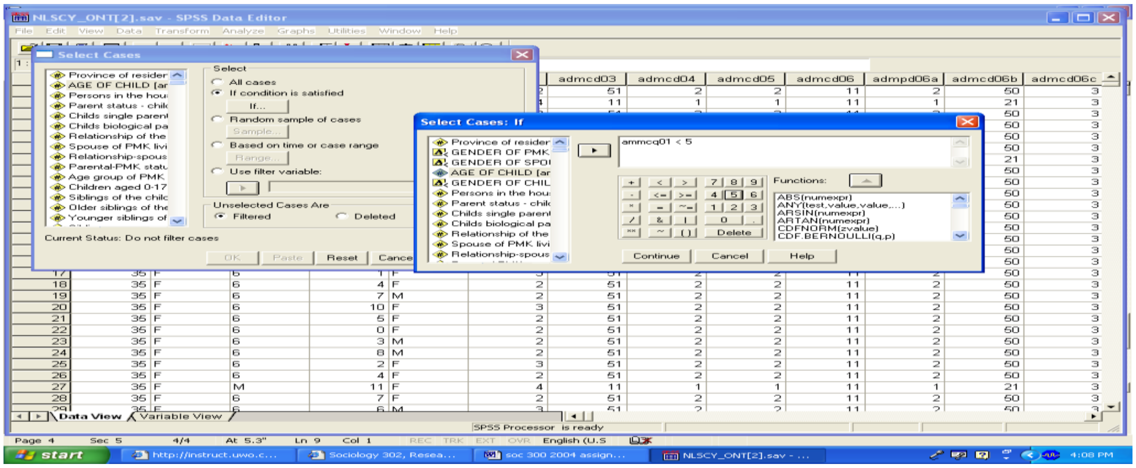
Sometimes you need to perform an analysis on only some of the cases in a dataset. For example, suppose that you wanted to do an analysis involving exclusively children under the age of 5. You can make this selection under the "select cases" option in the data menu.



1 : agehd03

	agel			ammcq01	ammcq02	admhd02	admcd03	admcd04	admcd05	admcd06	admpd06a	admcd06b	admcd06c
1				1	F	2	51	2	2	11	2	50	3
2				0	F	4	11	1	1	11	1	21	3
3				1	M	2	51	2	2	11	2	50	3
4				5	M	3	51	2	2	11	2	50	3
5				10	M	2	51	2	2	11	2	50	3
6				10	M	2	51	2	2	11	2	50	3
7				6	F	3	11	1	1	11	1	21	3
8				4	F	2	51	2	2	11	2	50	3
9				6	M	2	51	2	2	11	2	50	3
10				8	M	3	51	2	2	11	2	50	3
11	35	F	6	9	F	2	51	2	2	11	2	50	3
12	35	F	6	4	M	2	51	2	2	11	2	50	3
13	35	F	6	11	M	2	51	2	2	11	2	50	3
14	35	F	6	3	M	2	51	2	2	11	2	50	3
15	35	F	6	2	M	2	51	2	2	11	2	50	3
16	35	F	6	3	F	2	51	2	2	11	2	50	3
17	35	F	6	1	F	3	51	2	2	11	2	50	3
18	35	F	6	4	F	2	51	2	2	11	2	50	3
19	35	F	6	7	M	2	51	2	2	11	2	50	3
20	35	F	6	10	F	3	51	2	2	11	2	50	3
21	35	F	6	5	F	2	51	2	2	11	2	50	3
22	35	F	6	0	F	2	51	2	2	11	2	50	3
23	35	F	6	3	M	2	51	2	2	11	2	50	3
24	35	F	6	8	M	2	51	2	2	11	2	50	3
25	35	F	6	2	F	3	51	2	2	11	2	50	3
26	35	F	6	4	F	2	51	2	2	11	2	50	3
27	35	F	M	11	F	4	11	1	1	11	1	21	3
28	35	F	6	7	F	2	51	2	2	11	2	50	3
29	35	F	6	6	M	3	51	2	2	11	2	50	3

Once you have opened up the "select cases" box, you can highlight the variable of interest (which in this case is age of child "ammcq01"), and then use the "if" button to open another dialogue box that allows you to specify the rules by which the cases will be selected.



○ All cases
● If condition is satisfied
○ Random sample of cases
○ Based on time or case range
○ Use filter variable:

● Filtered ○ Deleted

Current Status: Do not filter cases

ammcq01 < 5

In this example, only cases meeting the condition that the value on variable "ammcq01" is "less than 5" are selected. It is important to check whether unselected cases will be deleted (permanently) or merely "filtered" temporarily. Cases that are not selected will not be included in any subsequent analysis, until you specifically tell the computer to do so (or if you close your dataset). In order to use all of the cases again, you must select "all cases" in the select cases dialogue box, which overrides the previous command.

The syntax that achieves the above selection on “children under the age of 5” is as follows:

```
USE ALL.
COMPUTE filter_$=(ammcq01 < 5).
VARIABLE LABEL filter_$ 'ammcq01 < 5 (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

Again, the ordering of procedures in your syntax file is crucial. You must select your cases prior to running any statistical procedures. To move back to the full sample, you can specify this again using the “data, select cases” option. The following example is a program file which contains the syntax to:

1) select exclusively children aged under 5
(2) run a frequency distribution on the variable “ammcq01” age of child,
(3) remove this filter and return to the full sample,
(4) run the same frequency.

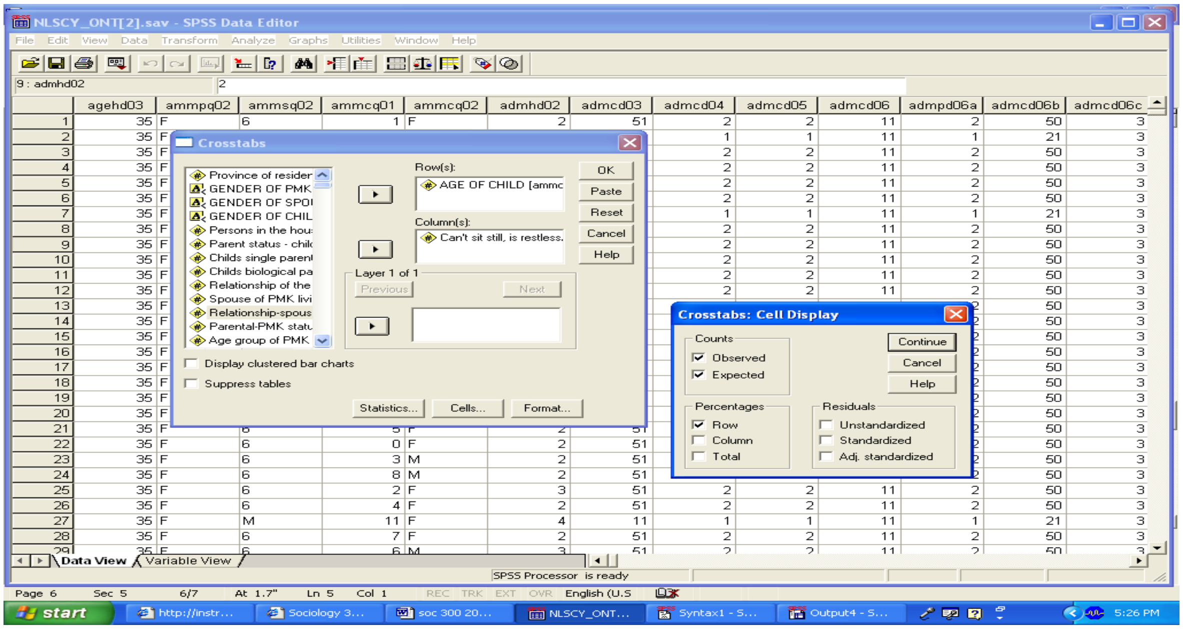
```
USE ALL.
COMPUTE filter_$=(ammcq01 < 5).
VARIABLE LABEL filter_$ 'ammcq01 < 5 (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

FREQUENCIES
  VARIABLES=ammcq01
  /ORDER=  ANALYSIS .

FILTER OFF.
USE ALL.
EXECUTE .

FREQUENCIES
  VARIABLES=ammcq01
  /ORDER=  ANALYSIS .
```

Crosstabs

The crosstabs procedure in SPSS allows you to create a contingency table. Using crosstabs, you can easily create two way or even three way cross tabulations. You can also calculate statistics based on this cross-tabulation, such as Chi squared, Gamma or Lambda (these were introduced in Soc 2205). This procedure can be found in the menu system under “analyze, descriptive statistics, crosstabs”. The dialogue box asks you to specify the row and column variables. “Layer” variables are those that allow you to create three-way cross tabulations. Obviously, when working with this procedure, variables that have a very large number of categories become unmanageable (i.e. it is sometimes necessary to recode variables if you want to work with the crosstabs procedure).

Assume that we wanted to examine the relationship between “ammcq02” (i.e. gender of the child) and “abecq6b” (can’t sit still, is restless?). Note that this latter question was only asked of children aged 2-11 (and not for infants). Using the crosstabs procedure, it is conventional to place the dependent variable in the columns and the explanatory (independent) variable in the rows. We can assume in this case that gender is an independent variable and that our indicator of behavioral problems is our dependent variable. In this case it is useful to examine the row percentages which give us the conditional distributions on the behavioral problem for girls and boys, each separately.

The “cells” box allows you to specify the contents of the cells in your output (observed frequencies, percentages or absolute values, type of percentage: row, column or total). The “statistics” box allows you to produce the chi squared test, as well as measures of association (gamma, lambda, etc.).   Certainly it is useful to at a minimum, to ask for the row percentages as well as the observed counts.



Crosstabs

Row(s): AGE OF CHILD [ammc

Column(s): Can't sit still, is restless.

Layer 1 of 1

□ Display clustered bar charts

□ Suppress tables

Crosstabs: Cell Display

Counts
☑ Observed
☑ Expected

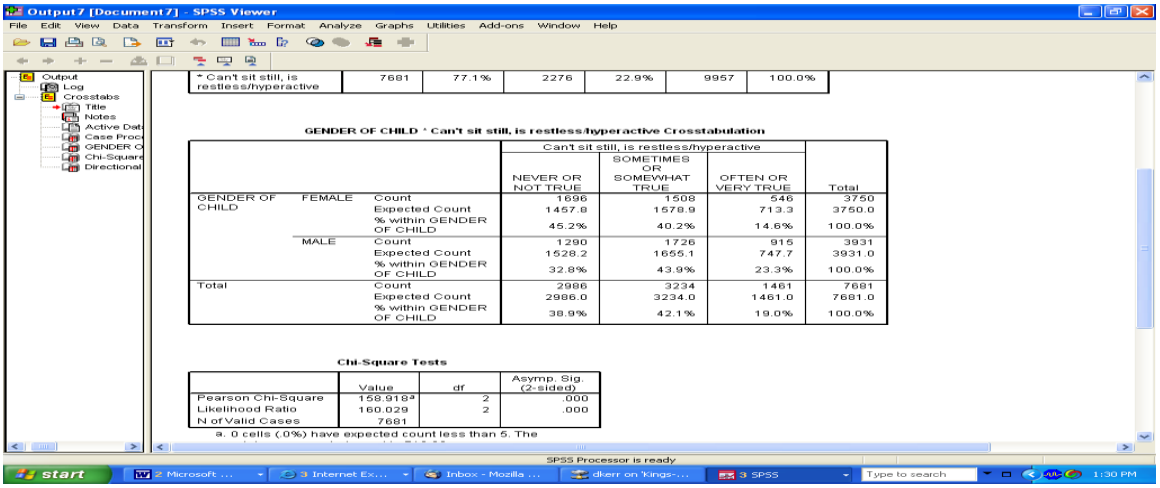
Percentages
☑ Row
□ Column
□ Total

Residuals
□ Unstandardized
□ Standardized
□ Adj. standardized

By pasting your selection, you can also produce the following syntax which does the same thing as the above.

```
CROSSTABS
/TABLES= ammcq02 BY abecq6b
/FORMAT= AVALUE TABLES
/STATISTIC=CHISQ  LAMBDA
/CELLS= COUNT EXPECTED ROW/
```

This procedure produces a cross tabulation of the variable ammcq02 (rows) BY abecq6b (columns), with the observed counts, row percentages, the chi squared statistic and Lambda.   The following output provides you with row %'s that assist in interpretation.



* Can't sit still, is restless/hyperactive	7681	77.1%	2276	22.9%	9957	100.0%

GENDER OF CHILD * Can't sit still, is restless/hyperactive Crosstabulation

			Can't sit still, is restless/hyperactive			
			NEVER OR NOT TRUE	SOMETIMES OR SOMEWHAT TRUE	OFTEN OR VERY TRUE	Total
GENDER OF CHILD	FEMALE	Count	1696	1508	546	3750
		Expected Count	1457.8	1578.9	713.3	3750.0
		% within GENDER OF CHILD	45.2%	40.2%	14.6%	100.0%
	MALE	Count	1290	1726	915	3931
		Expected Count	1528.2	1655.1	747.7	3931.0
		% within GENDER OF CHILD	32.8%	43.9%	23.3%	100.0%
Total		Count	2986	3234	1461	7681
		Expected Count	2986.0	3234.0	1461.0	7681.0
		% within GENDER OF CHILD	38.9%	42.1%	19.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	158.918a	2	.000
Likelihood Ratio	160.029	2	.000
N of Valid Cases	7681		

a. 0 cells (.0%) have expected count less than 5. The

Note that this table does not include the full sample, given that the question on child's behavior was asked only of children aged 2-11. The other cases (i.e. younger children) are treated as missing values and excluded from this crosstab.

Note: In Soc 2205 you learned about the chi square test. The resultant chi squared test in your output (chi squared=158.918 with 2 df) provides us with a P value which is clearly <.001).   If the P-value is less than .05, the relationship is considered to be statistically significant (and not by chance). The chi squared statistic tells us nothing about the "strength of the association", but merely whether or not a "significant" association exists. On the other hand, statistics like lambda (or some other measure of association) provide an indication of the strength of relationships.

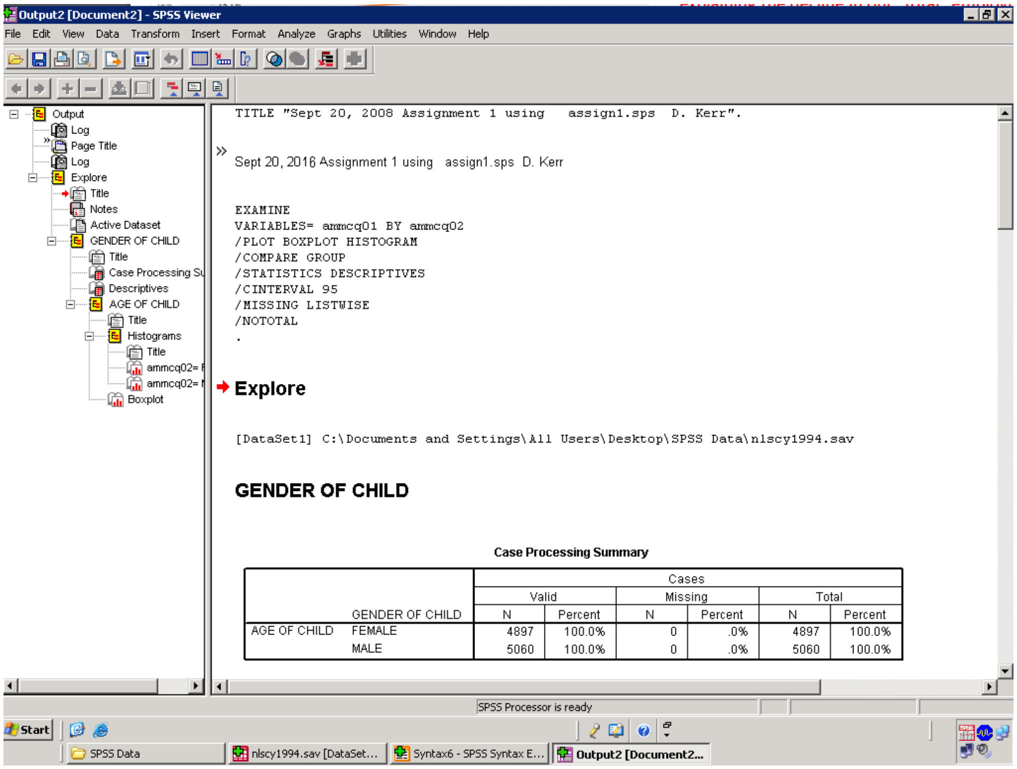
Documenting your work!

It is not a bad idea to get into the habit of using SPSS syntax files rather than merely working with the menu system build into SPSS. This allows you to go back to it at a later point in time, if need be, to make minor modifications to your work. Note: This sort of documentation is crucial in conducting collaborate research.

In the following syntax file, I’ve specified a TITLE for documentation purposes. I’ve specified the date the program was last modified, the name of the program file (assign1.sps) as well as the person who developed the program.

```
TITLE "Sept 20, 2017 Assignment 1 using  assign1.sps  D. Kerr".
EXAMINE
VARIABLES= ammcq01 BY ammcq02
/PLOT BOXPLOT HISTOGRAM
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

As you can see below, this title is found at the top of the resultant output file. You can subsequently save your output file under whatever name you consider appropriate (for example: assign1.spo). You should also save your syntax file which in this case was called “assign1.sps”. By properly documenting your work, you will have a good record of what you have done in the past, just in case you wanted to work with it again.



```
TITLE "Sept 20, 2008 Assignment 1 using   assign1.sps  D. Kerr".
```

Sept 20, 2016 Assignment 1 using   assign1.sps D. Kerr

```
EXAMINE
VARIABLES= ammcq01 BY ammcq02
/PLOT BOXPLOT HISTOGRAM
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL
.
```

Explore

```
[DataSet1] C:\Documents and Settings\All Users\Desktop\SPSS Data\nlscy1994.sav
```

GENDER OF CHILD

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	GENDER OF CHILD	N	Percent	N	Percent	N	Percent
AGE OF CHILD	FEMALE	4897	100.0%	0	.0%	4897	100.0%
	MALE	5060	100.0%	0	.0%	5060	100.0%

Note: If you save stuff on your C drive in the computing lab, it might not be there the next time you check (i.e. these computers are cleaned up nightly). On the other hand, you can save it directly on the network under “my documents” (something which I highly recommend that you start getting into the habit of doing) if you are properly logged into the network. ALWAYS SAVE YOUR WORK UNDER THE “MY DOCUMENTS” SPACE ON THE NETWORK. Also, it is also possible to convert your spss output file into a word document using the SPSS program, which you can then send to your hotmail or uwo email account to be printed up at your convenience. Alternatively, you can also merely select all, copy and paste into a word document.

Note: The order of commands in your syntax file are important. Procedures like “crosstabs” or “frequencies” should always be at the bottom of syntax files.. whereas procedures like “select cases” should be at the top. Commands that recode or create new variables should always fall before the procedures, yet after selections that specify your subsample.

ASSIGNMENT # 1
Working with either the CCHS, the GSS, the Canadian Census or the NLSCY:

I am asking you to produce 4 syntax files and 4 output files. On each syntax file, include a title that specifies the syntax program name, the date and name of programmer (i.e. your name).

1. Create 1 syntax file "Soc1.sps" and an output file "Soc1.spo" that includes:

Working with the dataset of your choice, run a series of frequency distributions (and histograms) on variables that you think might be interesting for your final project. See the corresponding code book on my web page for details on how to interpret these variables. Provide a brief summary of what you see with these variables **(select at least a dozen**). Are there many "missing" values (N.A; Don't Know; Not asked, etc). on these variables? If so, why? (you might need to look at the code book in answering this question: for example, was it asked of everyone?

2. Create a 2nd syntax file "Soc2.sps" and output file "Soc2.spo" that:

Working with either the CCHS (Canadian Community Health Survey), the General Social Survey, the Census or the NLSCY:

From among the variables you selected, recode at least 1 variable into a "dichotomous" variable (2 categories). Be careful with your "missing values" (i.e. the new variable that you create can have "extra" missing value categories and still be dichotomous).

Provide a new variable name and value names for this variable.

Cross tabulate this new variable with at least two variables that you consider as possibly associated or related to this variable. You can choose any relevant variables from the datasets available. Which variable would be the logical "dependent variable" and which would be the logical "independent" variables. Set the crosstab up as previously specified (with the independent variable in the rows and the dependent in the columns). For example, assume you dichotomize smoking into "1" smoker" and "0" non-smoker". Assume you also have education and age as associated with this variable. Logically, education and age are the independent variables and smoking is the dependent. Educational background is an important determinant of whether one starts smoking or not, and of course, age is strongly associated with smoking behavior as well. You could move on to do two separate cross tabs (age and smoking; education and smoking).

Provide a brief interpretation of these cross tabs. Provide a brief interpretation of the conditional distributions in the tables. Does the empirical evidence support the idea that the variables are associated with each other (again, a paragraph is sufficient)?

3. Create a 3rd syntax file "Soc3.sps" and output file "Soc3.spo" that:

Working with either the GSS, the CCHS **OR** the NLSCY file, we will recode a number of variables. **NOTE: one or the other, do not do the recode with all three datasets here.**

The variables listed below all measure a "more general concept". With the GSS we will work with a series of questions that ask persons whether they have been "victimized" in a specific type of crime (assault; property crime, etc). With the CCHS we will consider a series of questions that measure health status (number of chronic conditions). With the NLSCY we will consider a series of questions that entered into developing a scale of parental depression. Working with either the **GSS, CCHS OR NLSCY**, I would like you to begin by recoding the variables as listed.

****** If you use the GSS, you shall work with the following variables:**

OCEQ120B - have you been a victim of "robbery or attempted robbery"
OCEQ120C - have you been a victim of "assault"
OCEQ120E - have you been a victim of "breaking and entry"
OCEQ120F - have you been a victim of "automobile theft"
OCEQ120G - have you been a victim of "property theft"
OCEQ120H - have you been a victim of "fraud"
OCEQ120I - have you been a victim of "theft household propery"

Note: Recode these variables such that they are "1", yes a victim of a crime, or "0" not a victim of a crime. Note: all variables are coded as:

1. Yes
2. No
7. not asked
8. not stated
9. don't know

If persons were "not asked", this is because of a "skip pattern in the questionnaire" i.e. all persons who stated earlier in the questionnaire that they were "not a victim" of a crime over the last year were not asked any of the above questions. For current purposes, we will assume that "not stated" or "don't know" are "not victimized" (merely to simplify things for the current exercise, although of course, one might question this assumption).

When recoding these variables, make sure that you properly deal with categories 7, 8 and 9, which can be treated as "0" in your analysis.

For example, in recoding the above variable OCEQ120B, **you can specify:**

```
RECODE
  OCEQ120B
  (1=1)  (2=0) (7=0) (8=0) (9=0)  INTO  ROBBERY
```

You can place the above command directly into your syntax file, or use the menu system for recoding (and pasting into your syntax file). In this example, the new variable would be called "ROBBERY" (you can specify any name, for that matter). The new variable will be dichotomous, rather than having 5 categories (the reason you are doing this will become more clear in the next step of the assignment.

Provide variable names and value labels for all of the new variables (instructions were given earlier), and run a frequency distribution on all of them as well as with all of the old variables (again, description earlier in this assignment). Compare the new and old variables to make sure that there are no problems (no write up required here).

**** If you use the **CCHS, you shall work with the following variables:**

CCCA_031 – has asthma
CCCA_051 – has arthritis
CCCA_061 – has back problems
CCCA_071 – has high blood pressure
CCCA_081- has migraine headaches
CCCA_091 – has chronic bronchitis, other respiratory
CCCA_101 – has diabetes
CCCA_121 – has heart disease
CCCA_131 - has cancer
CCCA_141 – has stomach or intestinal ulcers

Note: Recode these variables such that they are "1", yes they have the medical problem, or "O" they do not have the difficulty.   Note: variables are coded as:

1. Yes
2. No
6. not applicable
7. don't know
8. refusal
9. not stated

In this case, if "not applicable", this is because of a "skip pattern in the questionnaire" i.e. all persons who stated earlier in the questionnaire that they had no chronic illnesses were not asked the question. For current purposes, we will assume that "not stated" or "don't know" or "refusals" are "not suffering the illness" (again, merely to simplify things for the current exercise).

(see above instructions above relating to the GSS on the specific coding)

****** If you decide to work with the NLSCY**, work with the following four variables that were initially developed in trying to document how depressed the parents who responded to this survey appear to be.

With 3 of the following variables, the higher the score, the lower the likelihood of depression. With 1 of them, it is the opposite situation: the higher the score, the higher the likelihood of depression. I would like all the variables to be coded in the same direction: i.e. higher scores suggest depression. Also, rather than ranging from 1-4, recode them to range from 0-3.

Variable Name: ADPPQ12F
Over the past week, have you felt hopeful about the future:

1. Rarely or none of the time (less than 1 day)
2. Some or little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (all of the days)
7. Don't know
8. Refusal
9. Not stated

Variable Name: ADPPQ12H
Over the past week, have you felt happy:

1. Rarely or none of the time (less than 1 day)
2. Some or little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (all of the days)
7. Don't know
8. Refusal
9. Not stated

Variable Name: ADPPQ12I
Over the past week, have you felt lonely:

1. Rarely or none of the time (less than 1 day)
2. Some or little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (all of the days)
7. Don't know
8. Refusal
9. Not stated

Variable Name: ADPPQ12J
Over the past week, have you felt that you are "enjoying life":

1. Rarely or none of the time (less than 1 day)
2. Some or little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (all of the days)
7. Don't know
8. Refusal
9. Not stated

When recoding these variables, make sure that you properly deal with categories 7, 8 and

9, which can be treated as “missing values” in our analysis (see description above).

For example, in recoding the above variable **ADPPQ12F, you can specify:**

```
RECODE
  ADPPQ12F
  (1=3)  (2=2) (3=1) (4=0) (ELSE=SYSMIS)  INTO  RECQ12F
```

The new variable will be called “RECQ12F” (you can specify any name, for that matter). The new variable will range from 0-3, rather than 1-4 (the reason you are doing this will become more clear in the next step of the assignment.

Provide variable names and value labels for all of the new variables, and run a frequency distribution on all of them as well as with all of the old variables (again, description earlier in this assignment). Compare the new and old variables to make sure that there are no problems (no write up required here).

NOTE: with one of the four above variables, you do not follow the above coding. Alternatively, your coding would be: (1=0) (2=1) (3=2) (4=3) (ELSE=SYSMIS)

4. Create a 4th syntax file “Soc4.sps” and output file “Soc4.spo” that:

Create an additive scale with all of the new variables that you created above. With the GSS the range on this new variable could theoretically be 0-7 (if one was a victim of all types of crimes the score could be 7, although that of course is highly unlikely). With the CCHS which should range from 0 thru 10 (0 being no illnesses listed and 10 suffering from all conditions listed). With the NLSCY, set it up so that the scores range from 0 thru 12 (if a respondent scores 0 on all four items, they will score 0 on this scale; if they score 3 on all 4 items, they will score 12 on this scale).

This merely involves introducing an additional command below the syntax that you created for step 3 immediately following it in your SPSS file (use the COMPUTE command as specified on pages 13 and 14 above). Provide this new scale with a name as well by using the VAR labels command. Note: with scales like this it is not necessary to compute value label.

Run a frequency distribution on this variable, and briefly describe the distribution.

5. Create a 5th syntax file “Soc5.sps” and output file “Soc5.spo” that:

Using the same dataset as in step 4, select exclusively “persons living in British Columbia”. With this new subsample, select two variables that you think might be associated with the scale as created in step 4. Run a crosstab with each variable (considered an independent variable) and the variable (scale) you created in step 4 (as dependent variable). Comment on what you observe in the crosstab. What other variables (available in the respective datasets) might be important in explaining some of the differences observed (list at least a half dozen)?

****

That’s it:
For ease in grading, please submit the following:

1. **the 5 syntax files, stapled together**
2. **the 5 output files, in a separate folder.**
5. **your write up (not more than a couple of pages.. follow above instructions)**

If unclear, consult either me or your SPSS consultant (Donna).