

EQUINOX AND SPSS

EQUINOX

Equinox is a data extraction service developed and maintained by the Social Science Computing Laboratory at the University of Western Ontario. It provides access to most -- but not all -- of the data available from the DLI (Data Liberation Initiative). The advantage of using an extraction service such as Equinox is that the researcher can readily identify, select and download variables from a datafile that can then be custom-tabulated using statistical software. This guide provides a brief step-by-step introduction to downloading data from the Equinox site, uncompressing it, importing the dataset to SPSS and creating some custom tabulations. It uses two software packages, namely, WinZip (version 7.0) and PASW Statistics 18 (Windows SPSS).

Step 1 - View, select and download variables from a file

At the MUN library homepage (www.library.mun.ca) click on **More Collections, Canadian Statistical Sites, Equinox, Equinox**. At the Equinox homepage, position the cursor on the **Browse** tab, select **Browse by Title** and click on **G** to select the **General Social Survey**; for this exercise, we will work with Cycle 9 of the 1994 General Social Survey (GSS) covering education, work and retirement. Click on the GSS Cycle 9 link; note the **Documentation** link that allows you to view accompanying documentation such as the codebook and questionnaire. Click on **Retrieve Data** which displays a list of all the variables (i.e. questions asked in this survey). When you click on a variable (i.e. not the box in the left margin but the variable itself) a window containing a description of the variable appears, including the choice of answers applicable to that particular variable and the number of responses received.

In this exercise, check the boxes for the following 13 variables, some of which are already tagged by default, such as the weight variable (PERWGHT - third from the top): **CASEID, PERWGHT, A3, DVA15, G9, H2, H4A, H42A, H42B, DVSEX, DVAGEGR, DVPERINC, DVPERNEW**. Click on **Next**; for this exercise, all values will be retrieved, so select the **No** button and click again on **Next**.

In the data submission page, key in your e-mail address, select **SPSS** and submit. Note that the data and code book files are selected by default.

Since this is a small data submission, your request should be processed in a few seconds, presenting you with a link containing an ftp prefix and numerical zip

designation of your data file. You will also receive an e-mail message containing this information along with a PIN (password) for accessing and unzipping the file. (E-mail is a useful feature with large data sets that take much longer to be extracted, thereby enabling users to continue with the task later.)

Step 2 - Download the zipped file to your hard drive and uncompress it

Click on the link mentioned above, containing an ftp prefix and numerical zip designation of your data file. You will then be asked to open or save the file; click on **Save**. In the **Save in** box, indicate where you want to save the file; for the purpose of this exercise, we will designate the C: drive and double click on a folder that will be called **temp** (in which to store the datafile to be uncompressed. If you need to create this folder, right click and select **New, Folder** and name it *temp*.) Double-click on the **temp** folder icon (so that the ftp file name appears in the *file name* bar); in the *Save as type* bar select **WinZip File** and click on **Save**.)

Go to the **My Computer** icon on your desktop and click on the **temp** folder in the C: drive. Double-click on the dataset icon; this automatically switches on the WinZip software so you don't need to activate it from your desktop. A WinZip window then appears, containing a list of three icons comprising: the data file, the SPSS command file and the HTML file that includes the customized codebook for the subset of variables extracted.

Click on the first icon that is the data file (with the **.dat+** suffix), the **Wizard** icon in the tool bar, **Next** and **Unzip Now**. Key in your PIN and click on **OK**. The unzipping/uncompressing of the data file is now complete. An FTP window then appears, containing the 3 icons mentioned above. **Note down the information appearing in the address bar**; it contains the **name and location of the unzipped data file that will be essential later** when the data is imported into SPSS. It should look like this: **c:\unzipped\ftp#####**

Step 3 - Importing the data set into SPSS

In the FTP window containing the 3 icons mentioned above, double-click on the icon containing the SPSS syntax document. (This automatically activates your SPSS Windows software, so there is no need to open SPSS by clicking on the desktop icon.) This will open up an SPSS data editor window and a syntax editor window. It is the latter that contains the SPSS command file.

An important change must be made to this file before the data can be read. As indicated in the 3rd paragraph of the syntax editor page, the location and name of the data file on your workstation must be designated in the **data list** command line below. As per the instructions, delete the characters **[path]** and replace them with the information (i.e., the drive, directory and file name) that you previously noted down from

the FTP window. The entire line should look like this:

Data list list(tab) file='c:\unzipped\ftp#####\ftp#####.dat' skip=1 / 1

Click on **Run/All**. Unless an error was made in specifying the location and name of the data file, the data should be read. The SPSS Viewer window appears, displaying the list of variables that have been input, and the SPSS Data Editor shows the active file whose cells now contain the numerical data for your selected variables.

Step 4 - Creating tabulations using SPSS

Prepare the sampling weight variable for analysis.

The General Social Survey employs a complex sample design. Needless to say, not every respondent in this design had the same probability of being selected in the survey. Consequently, a sampling weight variable must be used to adjust the unequal selection probabilities before generalizations can be made to the population.

Statistics Canada not only provides a variable to correct the unequal selection probabilities of the General Social Survey's sampling method but it also scales the sample size to a population estimate. In other words, using the sampling weight variable changes the N of the file from 11,876 to 21,954,438.

To demonstrate the importance of the sampling weight variable, perform the following analyses. First (in the Data Editor window), generate the frequency distribution of females and males by selecting Analyze / Descriptive Statistics/ Frequencies. Move DVSEX from the list on the left to the Frequencies Variable list and click OK. See the resulting tabulation in the SPSS Viewer window.

A. Record the results of the frequency distribution of females and males:		
Females	(n)	Percent
Males	(n)	Percent
Total	(n)	

Go to the Data Editor by clicking on Window / Untitled – PASW Statistics Data Editor. Then select Data / Weight Cases. Click on the radio button in front of “Weight cases by”. Then move the variable “perwght” from the list on the left to the “Frequency Variable”. Click OK.

Now re-run the frequency distribution for sex by choosing Analyze / Descriptive Statistics / Frequencies / OK.

B. Record the results of the weighted frequency distribution of females and males

Females	(n)	Percent
Males	(n)	Percent
Total	(n)	

Notice that not only do the counts for females and males change between the two runs, but also the percentages of each sex change. The latter phenomenon is due to the adjustment each case receives to equalize the selection probabilities.

The scaling of the sample size to a population estimate is not always desirable when using some statistical techniques. The degrees of freedom in these tests become silly in light of the scaled population size. The scaling factor can be removed while retaining the correction for the sampling weights. The following procedures create a new weight variable to do this:

First, turn off the weight variable from the last frequencies run. In the Data Editor, choose Data / Weight Cases / click on the radio button in front of Do Not Weight Cases / OK.

Second, determine the average of the weight variable. Choose Analyze / Descriptive Statistics / Descriptives. Move "perwght" to the Descriptives Variables list and click OK.

The N of this output is now once again 11,876.

Record the mean of the weight variable perwght:

Third, create a new variable by selecting in the Data Editor: Transform / Compute Variable. Assign as the Target Variable the name: wt. In the numeric expression, enter:
perwght/1848.639
Then click OK.

Fourth, test the new weight variable by selecting in the Data Editor: Data / Weight Cases / Weight Cases By / and move the variable "wt" to the Frequency Variable.

You may need to remove the previous weight variable “perwght” first by clicking on this variable’s name and then clicking on the return arrow. Next, click OK.

Now re-run the frequency distribution for females and males with the weight variable set to “wt”. Analyze / Descriptive Statistics / Frequencies / OK.

C. Record the results of the ‘wt’ weighted frequency distribution of females and males:

Females	(n)	Percent
Males	(n)	Percent
Total	(n)	

D. Do these frequencies match the frequencies in A above?
Do these percentages match the percentages in A above?
Do these percentages match the percentages in B above?

Exit SPSS: File / Exit.

This guide is an adaptation of an exercise prepared by Chuck Humphrey at the Atlantic DLI workshop in March 2000 (<https://ospace.scholarsportal.info/bitstream/1873/205/1/IDLSdal.doc>). The sections covering Equinox extraction, unzipping and importation into SPSS have been altered to allow for different software versions used in this exercise. The SPSS exercises have been left unchanged.