Working with Survey Data

July 2015

WORKING ONLINE TO FIND AND EXPLORE DATA:

1. FIND data using the <odesi> data catalogue: http://odesi.ca

ODESI lets you search through a collection of 1000's of survey and polling files to find the data that best suits your needs. Search by keyword, or browse the collection by topic.
See: http://library.queensu.ca/data/how_to_guides → **Using <odesi> and Nesstar…**

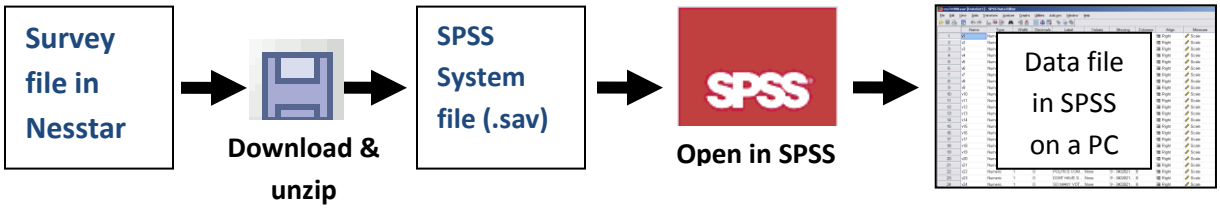
2. EXPLORE data using Nesstar Webview

Once you've found data of potential interest using ODESI, click on '**Explore & Download**' to move into the <odesi> exploration layer, **Nesstar Webview**, where you can work directly with survey data by looking at **FREQUENCIES**, running **CROSSTABULATIONS** (tables of two or more variables), and performing basic **SUBSETTING**, **RECLASSIFICATION OF VARIABLES**, etc. You can also **DOWNLOAD** survey data for use in **SPSS, SAS, STATA**, and other statistical software.

WORKING OFFLINE USING STATISTICAL SOFTWARE:

3. DOWNLOAD data in SPSS, SAS, STATA, or other format

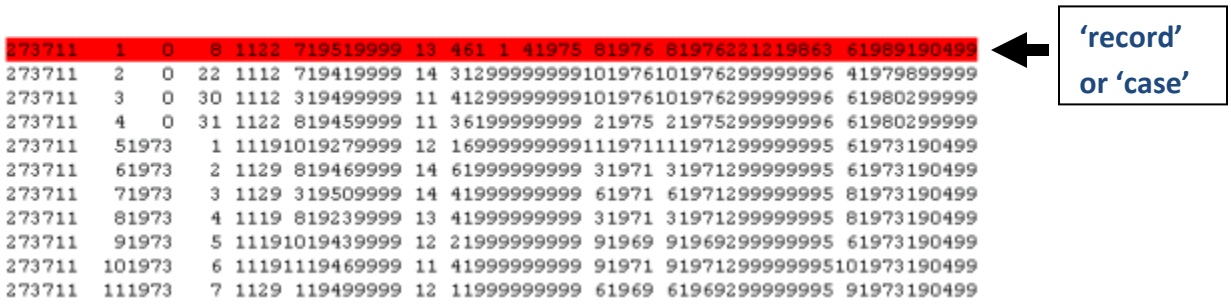
Scenario A: Download data files in your preferred statistical software format.
In most cases, you will be able to obtain survey files in formats that are ready-to-use in statistical software like SPSS and STATA. For example, you could use Nesstar to download a file into an SPSS System file (.sav), ready-to-use in SPSS:



For more on how to do this, see:
http://library.queensu.ca/data/how_to_guides → **Using <odesi> and Nesstar…**

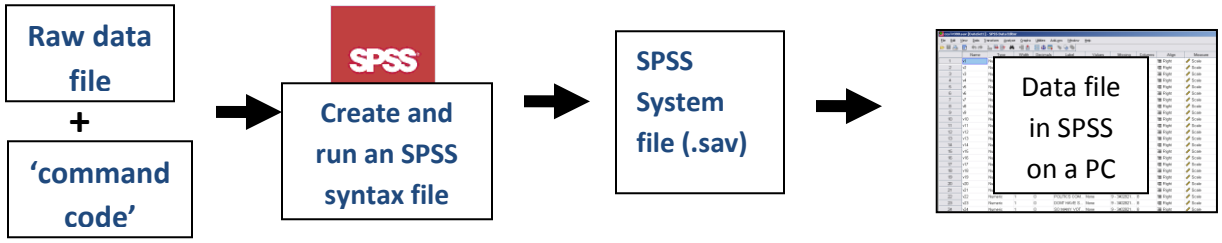
Scenario B: Occasionally, you may encounter data files that are not directly available in statistical software format. Instead, they are provided in 'raw' (i.e. *flat ASCII*) format – which, if opened in a text editor, will look like a large **rectangular array of data (numbers and/or text)**.

Some of the files you find in **ICPSR**[1] will fit this description. If the file you're after *does* come in a ready-to-use statistical software format, by all means use this before tackling a 'raw' version.



Take a look at the ICPSR/NACJD[2] Guide: **How do I interpret a record from an ASCII data file?** for more information.
URL: http://www.icpsr.umich.edu/icpsrweb/NACJD/support/faqs/2007/02/how-do-i-interpret-record-from-ascii

If you are obliged to convert an ASCII dataset into a form readable by statistical software, hopefully, your ASCII data file will come with an *additional* file containing '**command code**' describing what the numbers in **each column** of the data file represent. This file could indicate, for instance, that data in column position 1 contains information about the 'sex' of the respondent, columns 2 to 3 could indicate 'Province', etc. Armed with this information, you can convert the 'raw' data file into a format compatible with your statistical software (e.g. an SPSS System file with a '.sav' extension).



SPSS is available on computers in Stauffer Library. If you have SPSS '**command code**' and a '**raw data file**', you can convert your data in to SPSS format. A sample of SPSS 'command code' is provided below.

[1] ICPSR – Inter-university Consortium for Political and Social Research - http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp

[2] National Archive of Criminal Justice Data

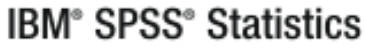
Alternatives to starting with 'Raw' data

1. StatTransfer

The Library has software called **StatTransfer**, which converts statistical files from one format to another. So, for example, if you have an SPSS System file (.sav) and want a STATA file (.dta) you can use **StatTransfer** to do this conversion.

---

2. Use the 'File Save-As' features of SPSS or other software



Using the SPSS 'File → Save As → Save as type' option, you can save an SPSS dataset (downloaded using <odesi>) into any of a number of common formats, including SAS, STATA, Excel, and many more.

Sample SPSS COMMAND CODE

Converting "RAW data" to an "SPSS System File" format (.sav)

```
* Open a file like this in SPSS as a 'syntax file' and run it to convert your ASCII
  data to an SPSS System file.
* Note: all SPSS commands end with a 'period'; comments start with an asterisk.
* These commands provide a title and define the width of the raw data file.

TITLE "ICPSR9897 - Predicting Recidivism in N. Carolina - 1978".
LENGTH=NONE WIDTH=80.

* This command provides a nickname (in this case MYFILE) for the raw data file,
defines its location (path on your PC) and the LRECL or 'logical record length'
(how many columns of numbers are associated with each record).
FILE HANDLE MYFILE/NAME='E:\fullpath_to_data_file'/LRECL=28.


* This command identifies the raw data file using the 'nickname' provided above,
and defines what variables are found in what positions on the data file.
DATA LIST FILE=MYFILE/
 WHITE       1 - 1
 ALCHY       2 - 2
 JUNKY       3 - 3
 STATE       27-28.

* This command provides labels for the variables on the SPSS dataset we are
creating.
```

```
VARIABLE LABELS
 WHITE       "Race - white or black"
 ALCHY       "Alcoholic?"
 JUNKY       "Junky?"
 STATE       "State of residence".
```

*** This command defines what each numeric (or alpha) code means for each variable on the dataset.**

```
VALUE LABELS
 WHITE
           1 "WHITE, and others"
           0 "BLACK"/
 ALCHY
           1 "ALCOHOLIC"
           0 "NOT ALCOHOLIC"/
 JUNKY
           1 "JUNKY"
           0 "NOT A JUNKY"/.
```

*** This command defines missing values that will be recognized as such in SPSS.**

```
Missing values
 ALCHY (-9) WHITE JUNKY (999).
```

*** This command saves the resulting file as an SPSS System file with a dot 'sav' extension.**

```
SAVE OUTFILE='E:\full path\ICPSR8987-1978.sav'.
execute.
```