sociometricsvb

Sociometrics Corporation

Downloading Data Archive Files

This document accompanies datasets downloaded from the following data archives:

American Family Data Archive (AFDA)
AIDS/STD Data Archive (AIDS and PETRA)
Contextual Data Archive (CDA)
Complementary and Alternative Medicine Data Archive (CAMDA)
Child Well-Being and Poverty Data Archive (CWPDA)
Data Archive of Social Research on Aging (DASRA)
Data Archive on Adolescent Pregnancy and Pregnancy Prevention (DAAPPP)
Maternal Drug Abuse Data Archive (MDA)
Research Archive on Disability in the U.S. (RADIUS)

1. Data Set File Compression - How to Uncompress Files

Each data set is comprised of several data and documentation files. To increase download speed and reduce the total number of files that must be downloaded, all data and documentation files for a given data set have been compressed into a single file formatted for either PC or Unix. PC-formatted downloadable files are self-extracting; Unix-formatted downloadable files must be uncompressed on an operating system that uses Unix or Linux.

Extracting PC-Formatted Files

Downloadable dataset files formatted for the PC carry the extension ".exe" (e.g., RAD03.EXE). EXE files have been created using WINZIP.

To extract and uncompress the data and documentation files, move or locate the download file on your computer using Explorer or My Computer and double-click on the EXE filename. The contents of the compressed EXE file will be written into the same folder as the EXE file.

Extracting Unix-Formatted Files

Downloadable dataset files formatted for Unix carry the extension "tar.gz" (e.g., RAD03.TAR.GZ). To uncompress the data and documentation files, move or locate the download file to a directory on your Unix drive and issue the command:

gzip -d fname.tar.gz (e.g., gzip -d rad03.tar.gz)

When the download file has been uncompressed, the "gz" extension will disappear. To extract the data and documentation files, issue the command: tar xvf fname.tar (e.g., tar xvf rad03.tar). Files within the TAR.GZ file are ASCII files.

2. Data & Documentation Files

The file descriptions below apply to all of Sociometrics' archive data sets. In the examples below, "xxxxx" is used to represent the filename prefix for extracted data and documentation files. For example, if you purchased data set number 03 from the Research Archive on Disability in the U.S. (RADIUS), the raw data file will be named rad03.raw and the SPSS/Unix command file will be named rad03.spx.

Raw Data Files (xxxxx.RAW or xxxxx.DAT)

This is a rectangular file of unformatted data values arranged in fixed column positions. Most of our raw data files contain 80 characters records, however several do not. The SPSS or SAS command files for non-standard raw data files indicate the logical record length. Most raw data files have multiple records per case.

SPSS (Batch) Command Files (xxxxx.SPS, xxxxx.SP, xxxxx.SPW, xxxxx.SPC)

SPSS command files contain syntax that reads in and defines the raw data file. The syntax will create an SPSS active file. Datasets from archives DAAPPP, AFDA, MDA, AIDS, RADIUS, CDA, and CWPDA may contain several different versions of SPSS command files in order to accommodate the job control language required by platform-specific versions of SPSS.

All SPSS files contain data list, variable, and value label syntax, and may include recode, missing value, and format declarations where applicable. SPSS command files require additional syntax to create a permanent system file. For example, using SPSS/Unix, the command SAVE OUTFILE = filename.sys would be added at the end of the SPSS syntax to create a permanent system file. SPSS/Windows users simply need to issue the SAVE AS command from the menu to create a permanent SAV file.

A dataset may include versions of SPSS syntax command files named with the following extensions:

SPSS/PC+ command files are named with extension *SPC*.

SPSSx command files are named with extensions *SPX* or *SPS* (datasets created after 2004).

SPSS/WINDOWS command files are named with extensions *SPW* or *SPS* (datasets created after 2004).

SAS (Batch) Command Files (xxxxx.SAS)

SAS command files read the raw data and create a SAS system file and format library. Syntax contents are analogous to the SPSS file described above.

SPSS Data Dictionary (xxxxx.DIC)

The data dictionary consists of output from the SPSS command "display dictionary." This file defines variables in sequential order with names, variable and value labels, missing value designations, and print and write formats.

SPSS Frequencies (xxxxx.FRQ)

The frequency file consists of unweighted frequencies or summary statistics for all variables in a data set. Full frequency listings are provided for variables with fewer than 50 response categories or observed values; descriptive statistics are provided for alphanumeric variables and those with more than 50 observations.

User's Guide, Instruments, Codebooks, Supplementary Documentation (xxxxx.PDF)

User's guides and all other documents are provided in Portable Document Format (PDF). The user's guide includes a study overview, detail on sampling procedures and file processing, variable topic and type reports, key sample characteristics, results of data completeness and consistency checks, and other information of use to the data analyst.

Not all datasets contain instruments, questionnaires or supplementary documentation. Older datasets that did not include digitized documentation will not contain machine-readable or PDF documents, except for the user's guide to the dataset.

3. Technical Assistance

All data archive purchases come with free technical assistance from Sociometrics' Research Support Group (RSG). The User's Guide that was included in your downloaded files contains helpful information that is specific to the data set(s) you purchased. In addition, the answers to several "Frequently Asked Questions" (FAQs) are available online at http://www.socio.com -- from the home page, click on "Help and FAQ".

4. Macintosh Usage Note

Macintosh users should consult the SPSS/Mac manual to determine the most suitable version of SPSS command file for their application.

5. Contacting Sociometrics Corporation

Sociometrics Corporation
170 State Street, Suite 260
Los Altos, CA 94022-2812
1-800-650-949-3282
650-949-3299
socio@socio.com ● http://www.socio.com

FREQUENTLY ASKED QUESTIONS

1. I cannot run all of the SPSS/Windows command file and syntax at one time. Is there a solution?

2. How can I use older SPSS Command Files With SPSS/Windows?

3. How can I use Sociometrics data sets with other statistical packages?

Q1. HOW TO CREATE AN SPSS SYSTEM FILE IN SPSS/WINDOWS

It is recommended that SPSS/Windows users import the SPSS portable system file into the SPSS application, rather than create a new SAV from scratch using the syntax command file. The portable file is named with the extension POR.

Importing an SPSS portable file from the pull-down menus in SPSS/Windows:

> File, Open, Data
> Change file type to SPSS Portable (*.por)
> Locate the portable file on computer
> Click open to launch the import.

Importing an SPSS portable file using syntax and the syntax window in SPSS/Windows (file location must be specified between the single quote marks:

> import file = 'c:\my documents\dapS2.por'.
> save outfile = 'c:\my documents\dapS2.sav'.
> execute.

If you prefer to use the syntax command file to create a new active file, the following steps are recommended:

1. Launch SPSS/Windows

2. Select File, Open, Syntax

3. Find *.spw or *.sps (the dataset name is the filename, e.g. rad0810.spw) and click on the filename to place it in the dialog box. (Unless you have renamed your syntax file with an *.sps extension, you must change the file type by selecting *.* as the type of file that will be viewed in the dialog box.). SPSS should open the syntax command file in the syntax window.

4. Edit the "data list file = *.*" command line so that the location of the raw data file is specified. Example:

    data list file = "c:\my documents\rad0810.raw" records = [number appears in syntax]

    The records subcommand should already have a number specifying the number of data records that will be read from the raw data file. This subcommand does not need to be edited or changed in any way.

5. You may also add a "save outfile = filename.sav". command at the very end of the syntax. This command will save a system file in the directory you are working from.

6. To run the syntax, highlight all of the syntax and select Run from the menu.

Please note that if you have downloaded a very large dataset, you may not be able to process the entire set of commands provided in the syntax file. If you receive the message "***missing command delimiter***" when running the syntax file, the most likely problem is that SPSS cannot process all of the commands in a single run.

The following strategies are suggested:

a) Run pieces of the syntax at a time. Each dataset includes at least the following major syntax portions: DATA LIST FILE, VARIABLE LABELS, VALUE LABELS. A dataset may also include MISSING VALUES and RECODES syntax as well. With the exception of the DATA LIST FILE command, it is possible to run small portions of the other commands using the execute command after each portion. Make sure that the primary command is used at the start of each small portion. For example, if you are running sets of variable labels, the VARIABLE LABELS command must appear at the top of the set of labels starting in the far left column.

You may wish to begin by running the data list command section of the syntax only. To do this, add a period (command delimiter) to the end of the data list file syntax. The end of the data list file command is after the *very last variable* in the data list, NOT the first line of the data list command. An execute command should be placed on the next line after the end of the data list command. New commands always appear in the left most columns in the file (subcommands are indented). Make sure to add a period after the execute command: execute.

After running the data list section, you may wish to save the file as a permanent SAV file.

b) Create syntax extracts with a smaller selection of variables. This step requires major editing of the syntax command file.

Q2. HOW CAN I USE SOCIOMETRICS SPSS/Unix (SPX) COMMAND FILES WITH SPSS FOR WINDOWS?

Older datasets may contain SPSS syntax command files that use outdated job control language or syntax. Most Sociometrics Data sets have SPSS/Windows syntax command files available for download if they have not been included in the download file itself. Please go to ftp://www.socio.com/pub/SPSS_win/ to locate SPSS/Windows files if required.

To modify SPX files:

1. Start SPSS Windows and open the SPSS/Unix syntax file in the syntax window.

2. Some lines in the file must be edited; you may edit the file directly in the syntax window. The following lines should be modified:

   * Delete the line: **SET PRINTBACK / WIDTH = 80**

   * Delete the line: **FILE HANDLE input / FILE=filename.dat**

   * Edit the line: DATA LIST FILE input RECORDS = N as follows:

   DATA LIST FILE = "filename.raw" RECORDS = N

   Add an = (equal) sign, and replace the word "input" with the name of the raw data file, enclosed in quotes. Include the drive designation or path name when the raw data file is located in another directory. Be sure to enclose the filename location in quotes.

   For example, "d:\filename.raw"   or "c:\my documents\filename.raw"

3. Add the command terminator "." (a period) to the end of the last line of each command. A new section of command syntax will always be aligned to the farthest left text column; all indented text that follows is considered a subcommand(s) of the primary command.

4. To create a permanent system file, add the following command line to the bottom of the file:

   save outfile = filename.sys.

5. To run the program, highlight the entire file by clicking on Edit, Select, All. The entire file should be highlighted. Click the RUN button on the upper left part of the syntax window. The progress of the run will be echoed in the output window. The permanent system file will be saved in the c:\spsswin subdirectory unless otherwise specified.

    If you have trouble running a complete set of syntax statements, follow the suggestions described in the answer to question 1 above.

Q3. HOW CAN I USE SOCIOMETRICS DATASETS WITH SPSS FOR MACINTOSH?

Import the SPSS portable system file into SPSS/Mac.

Q4. HOW CAN I USE SOCIOMETRICS DATASETS WITH OTHER STATISTICAL PACKAGES?

SPSS syntax command files may be edited to run with other statistical packages, or the SPSS portable file may be converted to another format using StatTransfer. Characteristics of the raw data and SPSS command files are as follows:

- All raw data files are named by Data Set Number(s) and the file extension RAW or DAT.
- All raw data files contain data values in "fixed-column" positions.
- Data values for a particular variable always appear in the same column positions throughout a raw data file.
- Most raw data files have been formatted with a standard logical record length of 80 characters.
- No "delimiters" (such as commas or blank spaces) are used to separate neighboring data values on records in the raw data files.
- Most raw data files employ multiple records per case, (i.e., a block of several lines is required to hold all the data values for one study respondent, or one case).

To read in and format raw data file with a statistical package other than SPSS, the following information may be helpful. This information is contained in the DATA LIST section of the complete SPSS command file:

- number of records per case
- logical record length of the raw data file;
- record number and column positions in the raw data file that correspond to each variable of interest; and
- variable names.

Other sections of the SPSS command file may be used to further document and format the data values:

- The VARIABLE LABELS section contains a list of all variable names and variable labels.
- The VALUE LABELS section contains a list of variable names and their corresponding response options.
- The RECODE section contains a list of variable names and any data values that need to be changed (not all SPSS command files contain RECODE syntax).

- The MISSING VALUES section contains a list of variable names and data values to be specified as missing data (not all SPSS command files contain MISSING VALUES syntax).

- The FORMATS section contains a list of variable names and required formats. For example, an "income" variable with decimal values such as 38000.55 might be assigned a format of "F8.2", meaning a total format of eight places with two decimal places (not all SPSS command files contain FORMATS syntax).

(updated 10/18/05)