SPSS (and PSPP) Guide
Dr. Paul Hensel • University of North Texas
Version 1.30 (updated 21 March 2024)

Table of Contents

Introduction

This document is meant to provide students with a basic understanding of SPSS and PSPP, the statistical software packages that are used for homework assignments in my PSCI 3300 course ("Political Science Research Methods") at the University of North Texas. Some of this material is presented in class lecture, and each homework assignment includes the specific instructions needed to complete the assignment, but this document provides more detail. Hopefully this document will be helpful to students while they are taking that course, as well as to those or other students if they choose to do quantitative research for a senior honors thesis, capstone research project, or other similar purposes.

Most screenshots in this document are from earlier versions of SPSS and PSPP on the Mac -- I do not update all of the images every time a new version of the software comes out -- but other versions and platforms should be very similar in all but cosmetic ways.

Please contact Dr. Hensel (phensel@unt.edu) with any questions or suggestions regarding this document. My goal is to make this as useful as possible, so I appreciate any reactions or suggestions you might have.

Version History

• 1.30: 21 March 2024 (Updated for Spring 2024 semester to reflect improved functionality in recent PSPP versions and to update links to documentation and other online resources for each command. Also added more details and syntax for descriptive analyses, including the procedures used for homework assignments #1 and 2, as well as more detail on alternative procedures for descriptive analyses.)

--Other additions planned for this semester or the future, as time allows:
    --add more details and syntax for the statistical analyses used in homeworks #3 and 4,
     as well as for logistic regression (even though this isn't currently used in any homeworks)
    --add TEMPORARY-SELECT IF-LIST (to display selected data for all or selected cases)
    --add examples of output from each procedure, with guidelines on how to interpret it (this will
     probably have to wait for a future semester because it is a larger change)

• 1.20: April 2022 (Updated for Spring 2022 semester. Added better explanations and clarifications about procedures used in class homework assignments)

• 1.10-1.12: October-November 2019 (Updated for Fall 2019 semester. Added more details and screenshots related to the use of PSPP rather than SPSS, added explanation of how to copy SPSS/PSPP output so you can paste it into your word processing document, and added more links to online resources on topics that are not yet covered in this guide. Added links to the official SPSS and PSPP documentation for each command's syntax and to additional tutorials or references for most commands, improved the discussion of how to open various types of data sets, and added better descriptions of how PSPP is similar/different for commands used in class homework assignments. Also added more details related to crosstabs, scatterplots, and correlations as needed for the next homework assignment, including notes about the differences between SPSS and PSPP in running these types of analyses.)

• 1.05: April 2018 (Added more details about the analyses that have been assigned for class homework assignments, and mentioned the SPSS techniques used for the more advanced techniques that were discussed in class but not used in any homework assignments)

• 1.0-1.04: March 2018 (The first preliminary version of this document was released during the Spring 2018 semester and was updated several times as new assignments were due)

Accessing SPSS or PSPP

SPSS
There are several different ways for you to use SPSS statistical software to complete several of the homework assignments for this course:

• SPSS is installed in many (most? all?) UNT computer labs, so you may use it there for free.

• If you are interested in getting your own copy of SPSS rather than depending on computer labs, you may rent a license for it through UNT at a substantial student discount. You will need the "SPSS Statistics" version of the SPSS Grad Pack, which is available for both Mac and Windows at a cost of approximately $62 (6 month license) or $93 (12 month license) at the following site for UNT students:
    <https://untsystem.onthehub.com/WebStore/ProductsByMajorVersionList.aspx>

• After you have graduated and no longer qualify for the campus discount, you will probably have to pay full price for SPSS, unless you enroll at another college/university for graduate/professional school or you are hired by an employer that provides it for employees. This can be very pricey; they currently charge $99/month for the "Base" package and even more for add-on packages with additional features (for example, such techniques as logistic regression, multinomial logistic regression, event count analysis, and survival analysis are only available in the optional "Custom Tables and Advanced Statistics" package that costs another $79/month):
    <https://www.ibm.com/products/spss-statistics>

PSPP
There is also a free open-source statistical package called PSPP that is claimed to be very similar to SPSS: "it behaves as experienced SPSS users would expect, and their system files and syntax files can be used in PSPP with little or no modification, and will produce similar results (the actual numbers should be identical)". In my testing, this works perfectly well for this course's homework assignments; it reads SPSS data files, offers all of the commands that are used in the homeworks (usually in the same place in the dropdown menus and with the same command syntax), and produces identical results. This may be downloaded freely for Mac, Windows, and Linux platforms:
    < https://www.gnu.org/software/pspp/get.html >
Be sure to download the latest stable release of PSPPP, rather than a "debug" version or a nightly build. Also, if you are using this on a Mac, be sure to follow the instructions in the support note from Apple that PSPP links to (https://support.apple.com/en-us/102445), explaining how to install software that is not downloaded from Apple's official App Store.

Which to Use?
Students are welcome to use PSPP if they would like to avoid paying for their own SPSS license or having to go to a campus computing lab, and it will remain available for free usage after you graduate and no longer qualify for the SPSS student discount or have access to UNT computer labs. There are a few limitations of PSPP, although these will not limit its usability for this course. Some of the more advanced statistical techniques discussed at the end of the class are not currently implemented in PSPP, so it may not be able to do everything you will need to do in your own research. PSPP also feels much clunkier, which isn't surprising -- SPSS has been around since the 1960s and is owned by a major corporation (IBM), so there has been a lot of time for its many well-paid programmers to improve the interface. PSPP is free software that is usually programmed by volunteers and is meant to run on many platforms (Mac, Windows, Linux), trying to look and work the same on all of them. PSPP's dropdown menus are only available at the top of the Data Editor window (they are not included in the Output Viewer or Syntax Editor screens, and unlike most native Mac or Windows software, the menus are not included in the menu bar at the top of the computer screen).

<center>**Using SPSS or PSPP**</center>

Opening the Software
The easiest way to start either SPSS or PSPP is to double-click on a .sav SPSS data file or a .sps SPSS syntax file, which should automatically open the application.

You may also open either SPSS or PSPP manually by double-clicking on the SPSS or PSPP app's icon directly or by clicking on an alias/shortcut from the main screen, the MacOS dock, or the Windows Start menu.

Switching Between Windows
SPSS uses a number of different windows or screens for different purposes. Whenever SPSS is open, you may use the "Window" item in the menu bar at the top of the screen to move between the three main windows that are used.
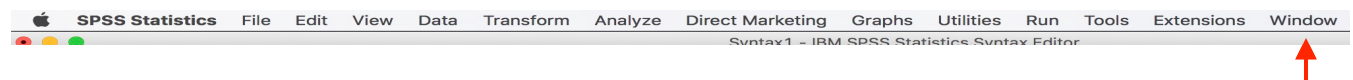
Using PSPP
PSPP uses similar windows, but the graphical user interface menus do not appear at the top of the screen. Most of these menus only appear at the top of the Data Editor window, not the Syntax Editor or Output Viewer windows, although there is always a Windows menu that can be used to move between the various windows..
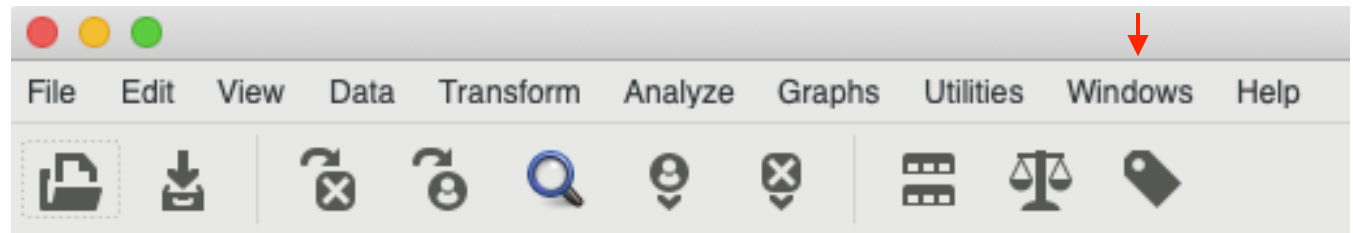
The "Data Editor" Window
Assuming that you opened SPSS by double-clicking on an SPSS ,sav data file, once SPSS has finished opening, you will probably see a spreadsheet with a list of variable names across the top. SPSS calls this spreadsheet the Data Editor window, and these numbers are the data set that you will be using. You may view or edit the data from this screen if desired, although that usually won't be necessary.

<center>3</center>

Using PSPP
This window looks similar in PSPP. As noted above, this window is the only place in PSPP where you can access the dropdown menus to carry out commands.

The "Output Viewer" (or Output) Window (and copying output for homework assignments)
Once you have run at least one operation or analysis (whether using a dropdown menu or a syntax file), this window will display the results, and allow you to copy the tables or graphs so you can paste them into a word processing document. In SPSS, you can usually use your mouse to click on the specific table, graph, or result and copy it (by right-clicking or using the dropdown "Copy" command), then paste this into your word processing document.

Using PSPP
This window looks similar in PSPP. It is a little trickier copying output from PSPP to use in your homework assignments, because PSPPs Output Viewer doesn't allow right-clicking directly on the output to copy it and paste it elsewhere. The solution is to select the specific portion of the output in the list on the left side of the Output Viewer (you may need to click the triangle in the list to show specific portions of the output), then go up to the Edit menu above the top of the list, and click "Copy"; this copied output can then be pasted into your word processing document.
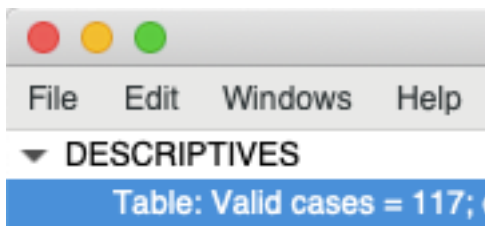
The "Syntax Editor" (or Syntax) Window

This window is used to enter or edit commands manually, without using the graphical user interface. You won't need to use this window for this course, but it will probably be very useful to you if you ever need to use SPSS later for your research or for a job. More details on using the syntax editor are provided in an appendix to this document.
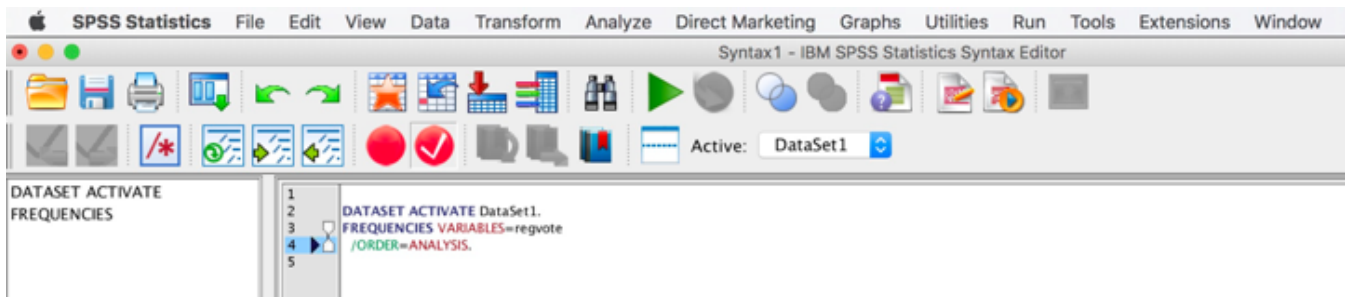
Using PSPP
This window looks similar in PSPP.

Opening SPSS (.sav) Data Sets

All data files needed for homework assignments in my course are provided for you in SPSS .sav format (posted alongside the homework assignments that require them). With either SPSS or PSPP, you can open an SPSS .sav data file by double-clicking directly on the file itself on your computer, whether or not the program is already running. You can also open such a file by dragging it to the application's icon on your computer (or to an alias/shortcut for the application, which you might have on your desktop or in your Mac's dock).

Once SPSS or PSPP is running, you can use the File menu to select and open the .sav data file that you want to work with: **File > Open > Data**, or else you will need to open the data from a syntax file (as described in the Syntax Files appendix to this document):
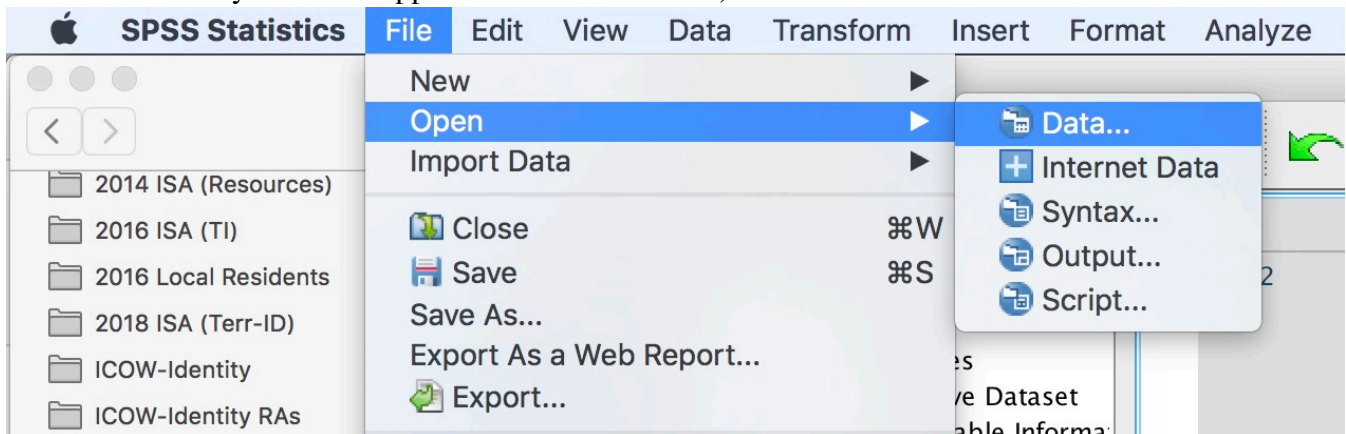
Opening or Importing Other Types of Data

If you go beyond the class homework assignments, you may want to try to open other types of files. SPSS and, to a certain extent, PSPP can open data sets in some alternative formats, as well as importing raw data or spreadsheet files.

Comma-Delimited or Tab-Delimited Files (.csv, .txt ,or .dat)

Many professional data sets are downloadable in this format, because it is easy to open in any spreadsheet or statistics package.
• Official SPSS Syntax Documentation: GET, GET DATA
• Official PSPP Documentation: GET, GET DATA, GET DATA-text files
• Kent State Statistical Consulting: Importing Data into SPSS
• UCLA Statistical Consulting Group: How to Read Raw Data into SPSS

Spreadsheet Files (.xls, .xlsx, or .xlsm / ,gnm or .ods)

Many data sets are originally created and maintained in Excel or another spreadsheet format. SPSS can open Excel (.xls, .xlsx, or .xlsm) spreadsheet files directly without first converting them to .csv or a similar format. PSPP is more limited, only opening spreadsheet files from Gnumeric (.gnm) or OpenDocument (.ods) formats. To use formats that can not be opened directly, you will need to convert them to comma- or tab-delimited text files and follow the instructions above.
• Official SPSS Syntax Documentation: GET, GET DATA
• Official PSPP Documentation: GET, GET DATA, GET DATA-spreadsheet files
• Kent State Statistical Consulting: Importing Data into SPSS
• UCLA Statistical Consulting Group: How to Read Raw Data into SPSS

Other Data Formats

SPSS can also read data saved in other formats, although PSPP does not currently offer this. There are separate commands to read in data files from the SAS and STATA statistical software, as well as a command to read in files from many types of databases (dBase, Lotus 1-2-3, Symphony, Systat).
• Official SPSS Syntax Documentation: GET SAS, GET STATA, GET TRANSLATE

Entering Data in SPSS

Most scholars prefer to enter their data in spreadsheets like Microsoft Excel, and then read the raw data into SPSS using one of the commands described above. If you prefer to enter your data directly in SPSS, here are some resources to help.
• Kent State Statistical Consulting: Data Creation in SPSS
• NC State's SPSS Tutorials: Creating a Dataset
• UCLA Statistical Consulting Group: How to Input Data into the SPSS Data Editor
• UNT's SPSS Short Course: Familiarization with SPSS - Creating a Data File

Summarizing the Data Set

Variable Codebook: "DISPLAY"
You can also get a list of basic variable information with the following command: **File > Display Data File Information > Working File**.
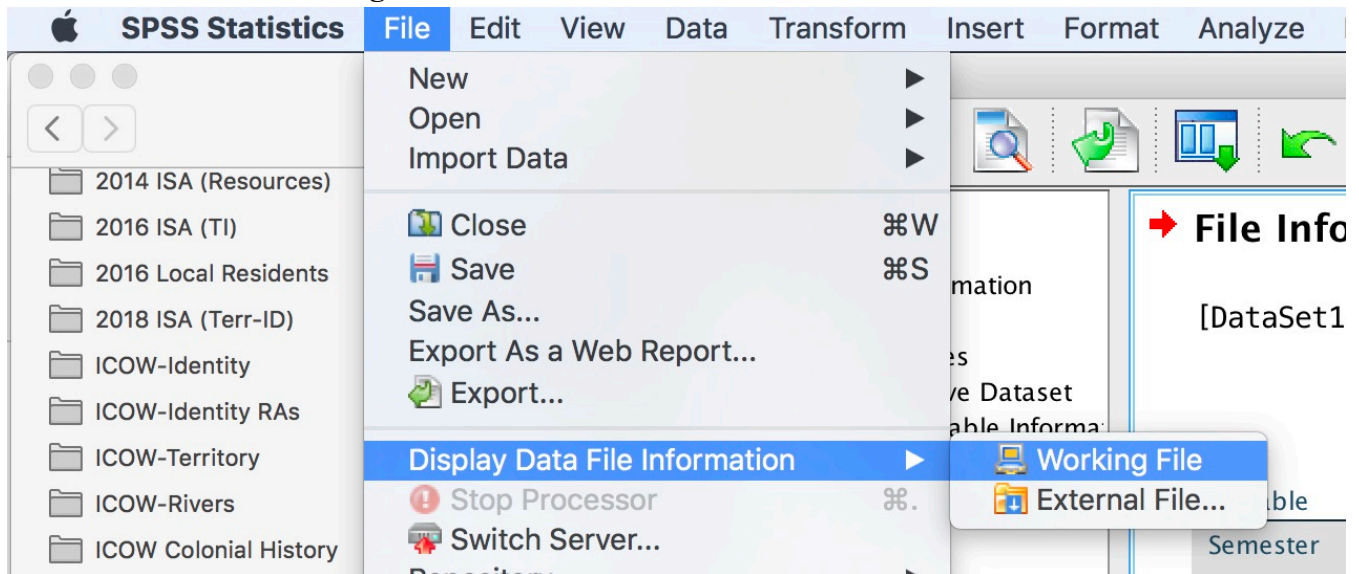
This will provide two tables with information about each variable in the data set. Most important, the first table lists the name of each variable and the full variable label:

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format
Semester	1	Year of course (YYYYxx)	Scale	Input	11	Right	F11	F11
School	2	University (FSU or UNT)	Nominal	Input	3	Left	A3	A3
gender	3	Gender	Nominal	Input	11	Right	F11	F11
age	4	Age	Scale	Input	11	Right	F11	F11
yrscolg	5	Years of college	Scale	Input	11	Right	F11	F11
numchld	6	Number children in family	Scale	Input	11	Right	F11	F11

The second table only applies to variables that have specific value labels, or text attached to one or more of the possible values that the variable can take on (typically used for nominal or ordinal variables, but this can sometimes be used for a few meaningful values of interval/ratio variables):

Variable Values

Value		Label
gender	1	Male
	2	Female
relig	1	Protestant
	2	Catholic
	3	Jewish
	4	Other
	5	None
relguid	1	Some
	2	Quite a bit
	3	A great deal
	4	None/Not important

Using PSPP

The same basic menu command (**File > Display Data File Information > Working File**) and syntax command (DISPLAY DICTIONARY) also work in PSPP. In fact, the output in PSPP seems to be better organized -- whereas SPSS displays separate tables for variable labels and value labels, PSPP displays value labels inside the main variable labels table, so all of the information about each variable is easily available in a single place.

Syntax

To do this in a syntax file:
    DISPLAY DICTIONARY.

To limit the display to selected variables only:
    DISPLAY DICTIONARY
        /VARIABLES = VAR1 VAR2 VAR3.

To sort the variables alphabetically, rather than in the order in which they appear in the data file:
   DISPLAY SORTED DICTIONARY

Other Resources
• Official SPSS Syntax Documentation: <u>DISPLAY</u>
• Official PSPP Documentation: <u>DISPLAY</u>

Alternative Option: "CODEBOOK" (SPSS only)

One of the first things you should do when you start using a new data set is put together a list of all variables (with their full name/label), and for nominal or ordinal variables, a list of all values (with the label for each). Without doing this, you won't know what your data look like, you won't know what the possible values are for each variable, and you may not be able to identify values that are impossible or that represent missing data when you run descriptive analyses.

The easiest way to do this in SPSS is using the **Codebook** command, which you can access from the Analyze menu: **Analyze > Reports > Codebook**. Unfortunately, this option is not available for PSPP.
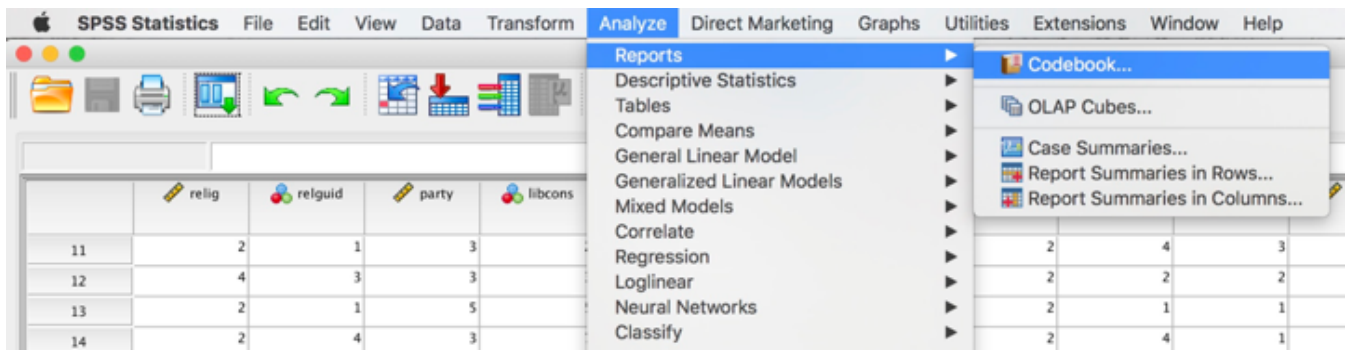
Once you have selected the Codebook command, highlight the variables that you want to include, click the button in the middle of the window to transfer these to the list of Codebook Variables, then click OK:
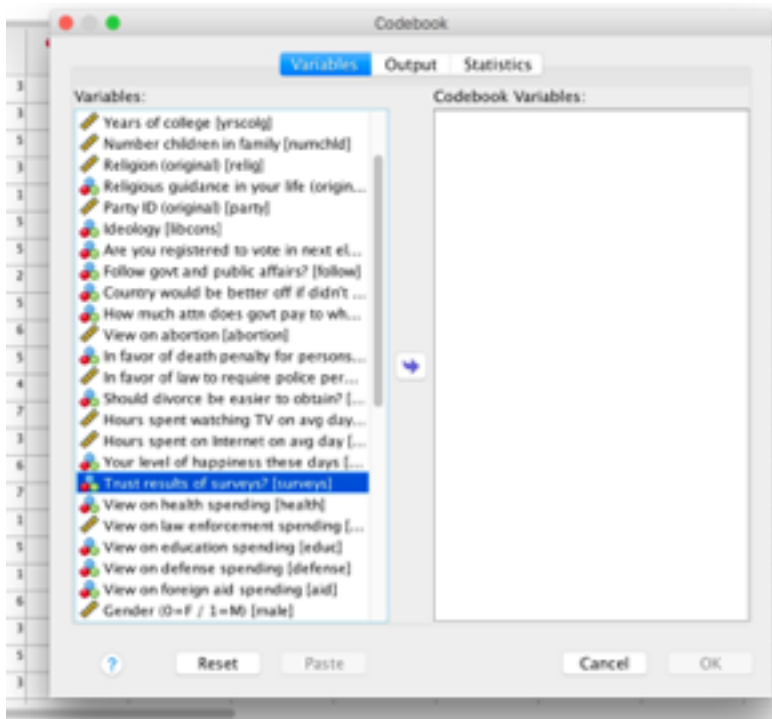
This will produce something like the following for each variable -- note especially the Label cell (which shows the full variable label), the Valid Values (the numeric value for all values of this if it is a nominal or ordinal variable), and the value label for each valid value.

surveys

		Value	Count	Percent
Standard Attributes	Position	22		
	Label	Trust results of surveys?		
	Type	Numeric		
	Format	F11		
	Measurement	Nominal		
	Role	Input		
Valid Values	1	Almost always right	2	2.2%
	2	Right most of the time	31	34.4%
	3	Right some of the time	57	63.3%
	4	Hardly ever right	0	0.0%

Using PSPP
PSPP does not implement the Codebook command in version 2.0.0. It may be added in a future version, though.

Syntax
To do this in a syntax file:
    CODEBOOK.

You may choose to limit this command to certain variables:
    CODEBOOK VAR1 VAR2 VAR3.

You may choose to limit this command to a subset of the possible information about each variable; the following list will limit this to the most useful information:
    CODEBOOK
      /VARINFO LABEL TYPE FORMAT VALUELABELS MISSING.

You may choose to add several basic descriptive statistics to this output, although most of the descriptives you might need are not available here:
    CODEBOOK
   /STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.

Other Resources
• Official SPSS Syntax Documentation: [CODEBOOK](#)

Descriptive Statistics

Frequencies, Central Tendency, and Dispersion: "FREQUENCIES"
The easiest way to display the frequency distribution for any variable in the data set is using the
Frequencies command, which you can access from the Analyze menu: **Analyze > Descriptive
Statistics > Frequencies**:
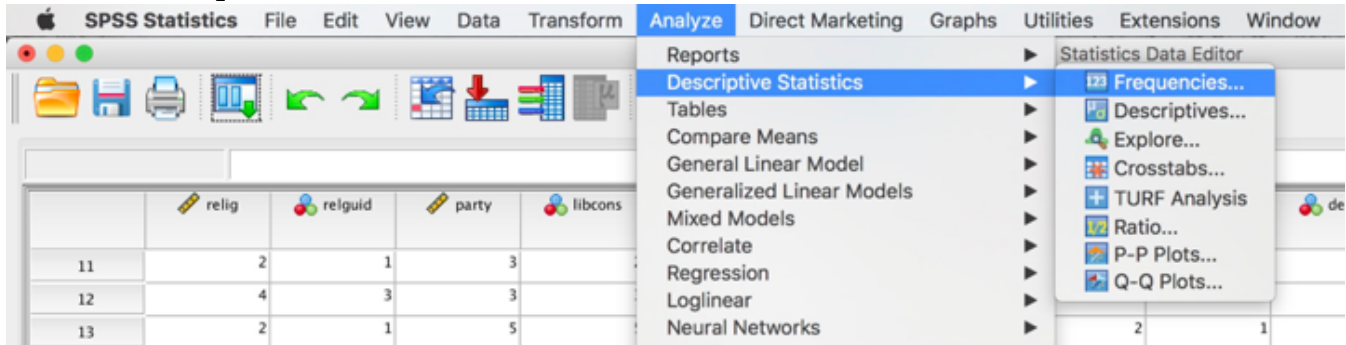
This command will open a dialog box that allows you to choose one or more variables, in order to
display their frequencies. Select the variable(s) you want to examine on the left side of this box, push the
arrow button in the middle of the screen, and the variable(s) should then appear in the list at the right
side of the box; once this happens, click OK.
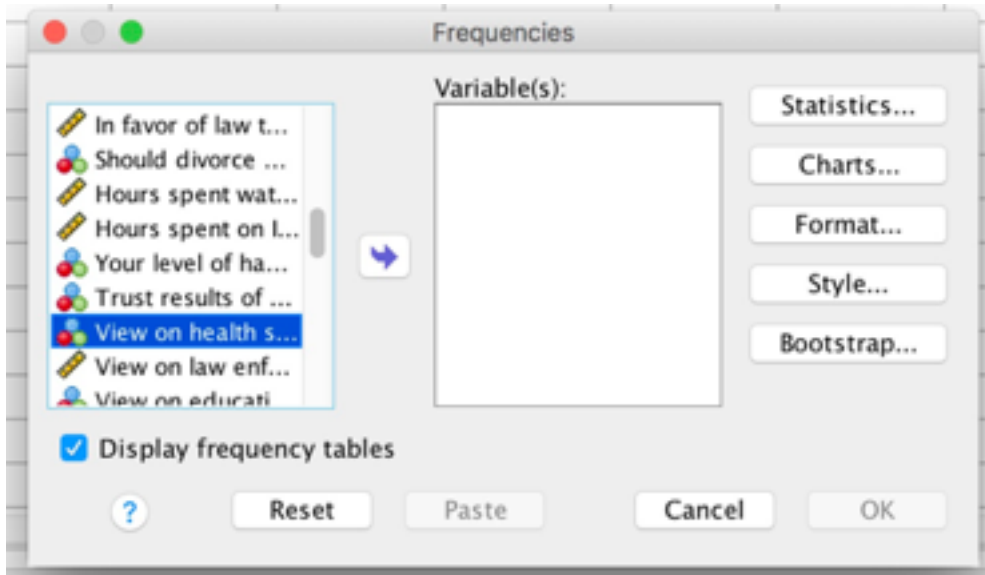
This will give you a frequency table with a variety of frequency-related details for each value of the
variable: the frequency count (the number of cases that have this value), the percentage of cases, the
valid percentage of cases, and the cumulative percentage of cases. (You can modify this list to add or
remove statistics by clicking the "Statistics" button, as will be described below)
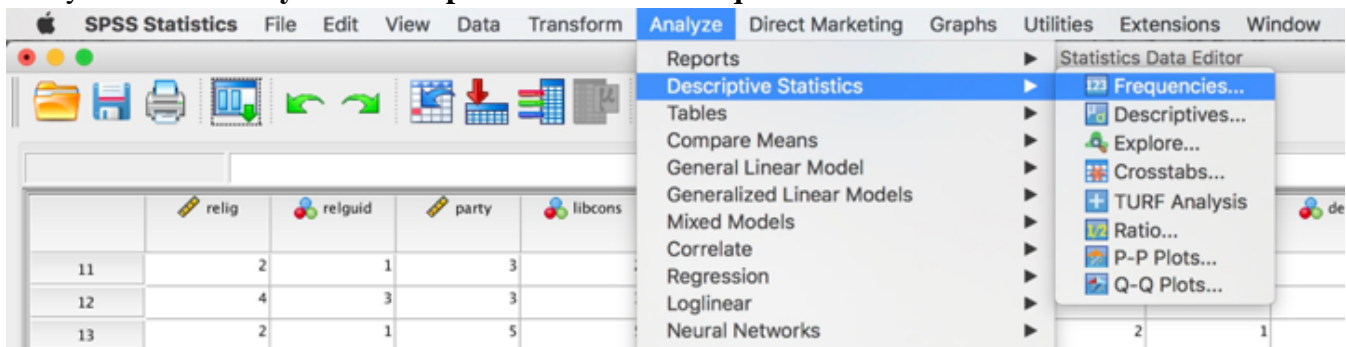
→ Frequencies

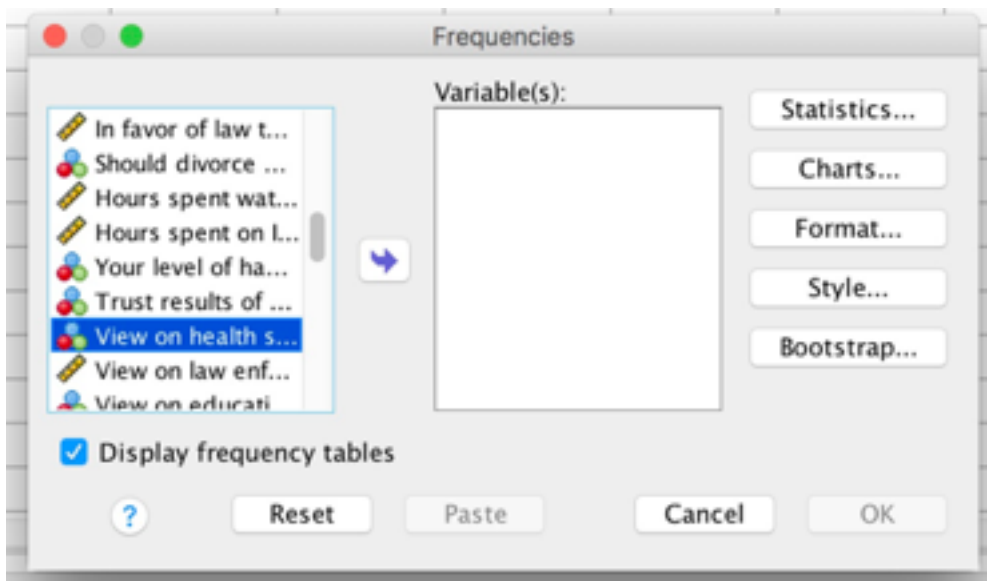**Statistics**

View on health spending

N	Valid	90
	Missing	0

View on health spending

|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Too little	62	68.9	68.9	68.9
	About right	19	21.1	21.1	90.0
	Too much	5	5.6	5.6	95.6
	Don't know/Other	4	4.4	4.4	100.0
	Total	90	100.0	100.0	

The easiest way to display the common measures of central tendency and dispersion for any variable in the data set is (surprisingly) also using the **Frequencies** command, which you can access from the Analyze menu: **Analyze > Descriptive Statistics > Frequencies**:

This command will open a dialog box that allows you to choose one or more variables, in order to display their frequencies. Select the variable(s) you want to examine on the left side of this box, push the arrow button in the middle of the screen, and the variable(s) will appear in the list on the right side.

After the relevant variables are selected, click the "**Charts**" button on the right side of the box, which will allow you to select the specific statistics you want included:
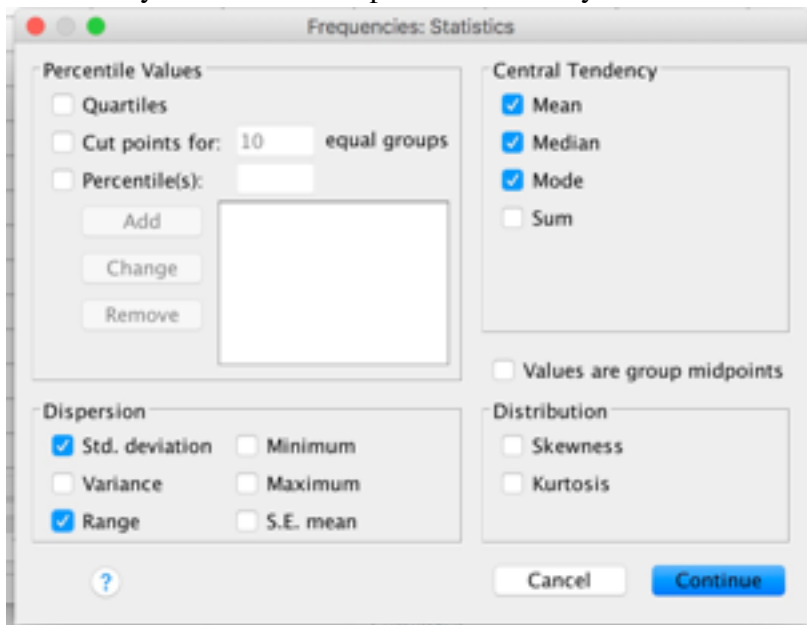
Select whether you want to produce a bar chart, pie chart, or histogram, and then select any of the available options. Once everything is ready, click OK, and the graph will be generated.

Using PSPP
• The same basic menu command (**Analyze > Descriptive Statistics > Frequencies**) also works in PSPP.

Syntax
• The default output just shows the frequency of cases with each value of the selected variable(s), along with the percentage and cumulative percentage of cases with this value.
• Additional statistics can be provided by using the STATISTICS subcommand; some of the more commonly used options are shown below.

13

• If you only want the summary statistics and don't want to see a table with the frequencies of each possible value of a variable (e.g. for continuous variables with dozens or hundreds of values), add "/FORMAT=NOTABLE"

```
FREQUENCIES
  /VARIABLES = xxx yyy zzz.
```

```
FREQUENCIES
  /VARIABLES = xxx yyy zzz
  /FORMAT=NOTABLE.
```

```
FREQUENCIES
  /VARIABLES = xxx yyy zzz
  /STATISTICS=MEAN MEDIAN MODE STDDEV MINIMUM MAXIMUM SUM RANGE.
```

Other Resources
• Official SPSS Syntax Documentation: FREQUENCIES
• Official PSPP Documentation: FREQUENCIES
• Kent State University Libraries Statistical Consulting: Frequency Tables, Descriptive Stats for One Numeric Variable (FREQUENCIES),
• UCLA Statistical Consulting Group: Annotated Output

Alternative Option: "DESCRIPTIVES"
Beyond the **Frequencies** command that I use in my course, there is a more limited option for descriptive statistics using the **Descriptives** command, which you can access from the Analyze menu: **Analyze > Descriptive Statistics > Descriptives**. This does not allow you to see the mode or median for the selected variables, but some of the other commonly used measures are still available:

This is most useful for displaying a single line of statistics for a large number of variables, rather than displaying more information for a single variable at a time:

→ **Descriptives**

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Trust results of surveys?	90	1	3	2.61	.534
Valid N (listwise)	90				

Using PSPP
• The same basic menu command (**Analyze > Descriptive Statistics > Descriptives**) also works in PSPP.

Syntax
• The default statistics (which are automatically used if no statistics are specified, or if "DEFAULT" is included in the requested options) are the mean, standard deviation, minimum, and maximum. To choose one or more of these individually, use MEAN, STDDEV, MINIMUM, and MAXIMUM.
• Other options include the range (RANGE), sum (SUM), skewness (SKEWNESS), and kurtosis (KURTOSIS).
• To delete cases that have missing values on any one or more variables in the list (e.g. if you are trying to produce details on the cases that are actually used in your other analyses), add "/MISSING=listwise".
• This procedure can also be used to save Z-scores using the "SAVE" subcommand; see the documentation for more details on how to do this.

DESCRIPTIVES
   /VARIABLES= all.

DESCRIPTIVES
   /VARIABLES= fuel fwdrive engsize mpg co2emit
   /STATISTICS=DEFAULT RANGE SKEWNESS.

Other Resources
• Official SPSS Syntax Documentation: DESCRIPTIVES
• Official PSPP Documentation: DESCRIPTIVES

Alternative Option: "MEANS"
This command can be used to display the mean and standard deviation (and other optional statistics) of each listed variable, as well as to show this same information for the dependent variable at each value of an independent variable. This is accessed from the Analyze menu: **Analyze > Compare Means > Means:**
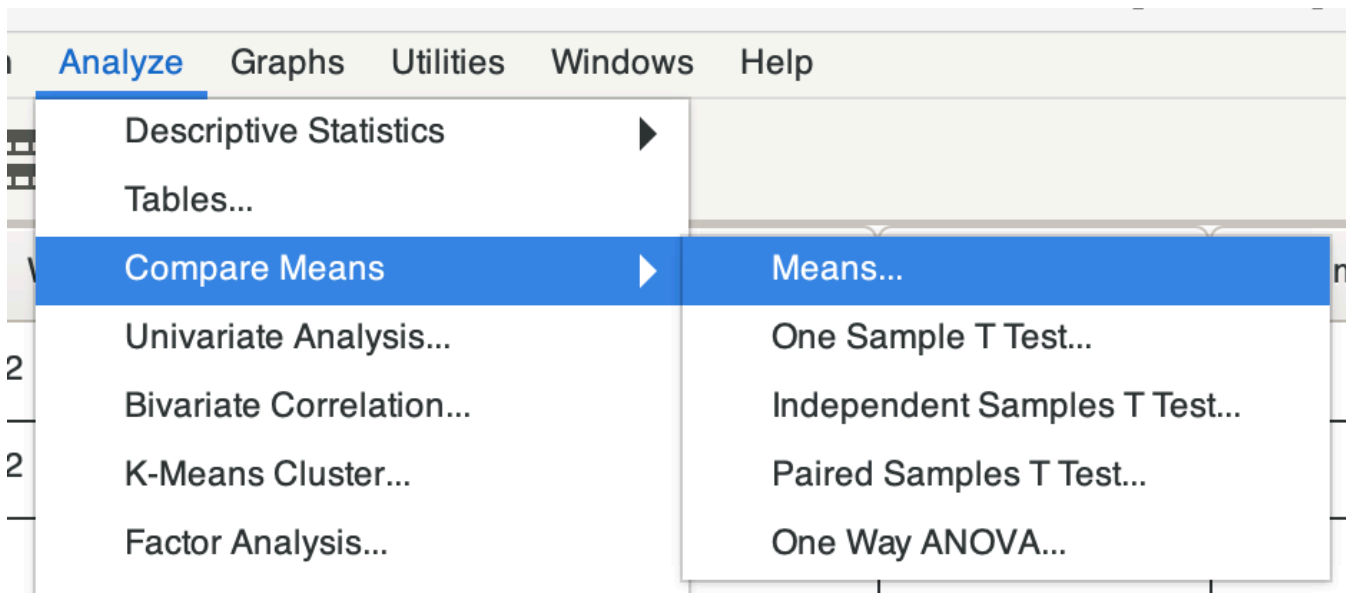
This command will open a dialog box that allows you to choose one or more dependent variables (the continuous variable whose mean you want to examine) and one or more independent variables (the categorical variable whose values you want to use to compare the mean of the dependent variable). Select the variable(s) you want to examine on the left side of this box, push the appropriate arrow button in the middle of the screen, and the variable(s) will appear in the list on the right side:
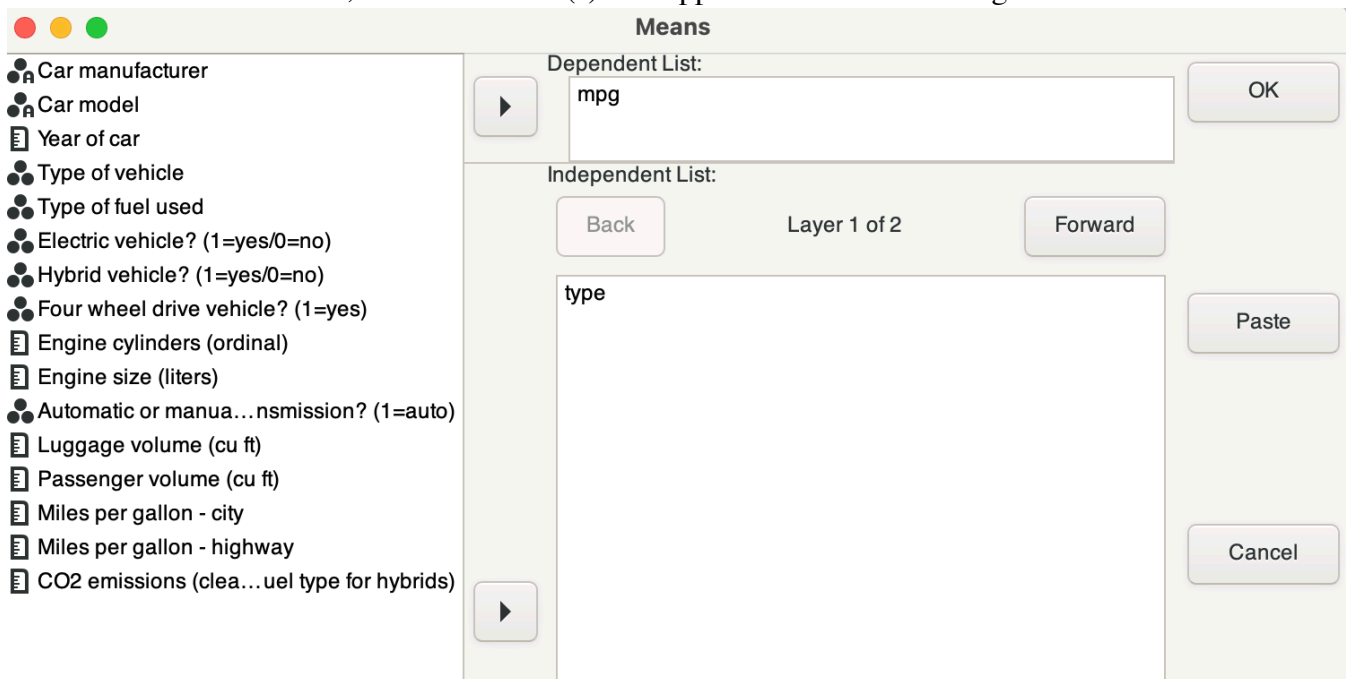
By default, this will produce a table showing the number of cases, mean, and standard deviation of the dependent variable at each value of the independent variable. Additional options are available using syntax files.

Using PSPP
• The same basic menu command (**Analyze > Compare Means > Means**) also works in PSPP.

16

• In PSPP, using the dropdown menu doesn't offer any ability to customize the statistic that are displayed. The only choices are to list the variable(s) being summarized, and optionally the independent variable(s) that will be used to split the dependent variable.

Syntax
• The default output just shows the number of cases, mean, and standard deviation of the dependent variable at each value of the independent variable.
• Additional information can be provided by using the CELLS subcommand; some of the more commonly used options are shown below.
• SPSS can also run a oneway ANOVA by using the STATISTICS subcommand (/ STATISTICS=ANOVA), which is equivalent to using the ONEWAY command. This option is not available in PSPP.
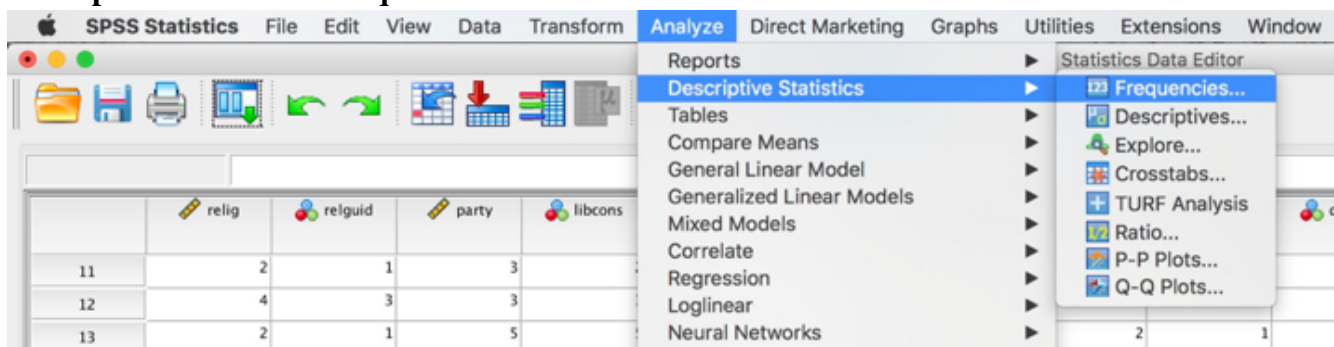
MEANS TABLES = xxx BY yyy.

MEANS TABLES = xxx BY yyy
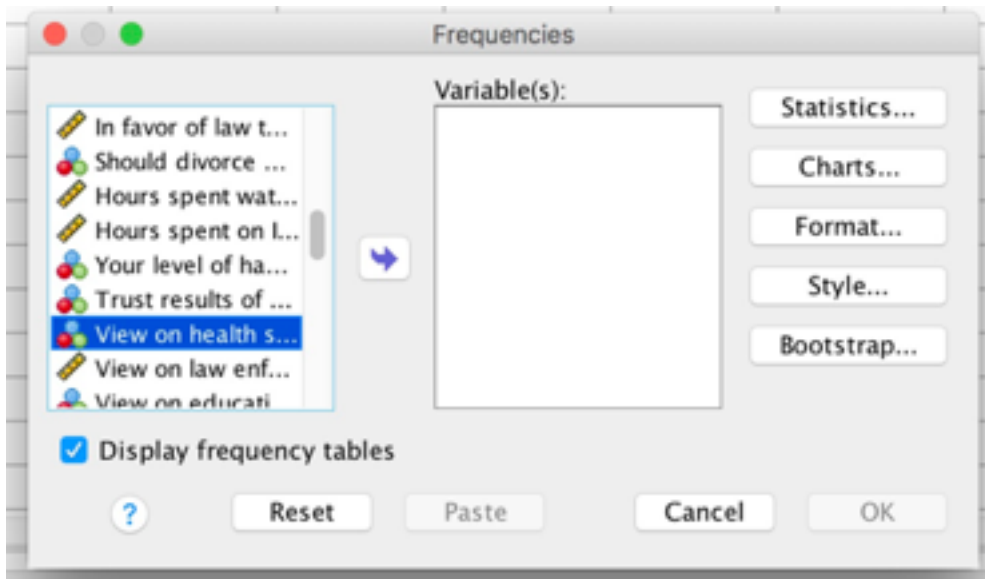 / CELLS=COUNT MEAN STDDEV SEMEAN MIN MAX SUM RANGE.

Other Resources
• Official SPSS Syntax Documentation: MEANS
• Official PSPP Documentation: MEANS
• Kent State Statistical Consulting: Descriptive Stats by Group (Compare Means)

**Basic Bar Graphs, Pie Charts, and Histograms: "FREQUENCIES"**
The easiest way to display graphs, charts, or histograms for any variable in the data set is (surprisingly) using the **Frequencies** command, which you can access from the Analyze menu: **Analyze > Descriptive Statistics > Frequencies**:

This command will open a dialog box that allows you to choose one or more variables, in order to display their frequencies. Select the variable(s) you want to examine on the left side of this box, push the arrow button in the middle of the screen, and the variable(s) will appear in the list on the right side:

After the relevant variables are selected, click the "**Charts**" button on the right side of the box, which will allow you to select the type of chart/graph to be produced:
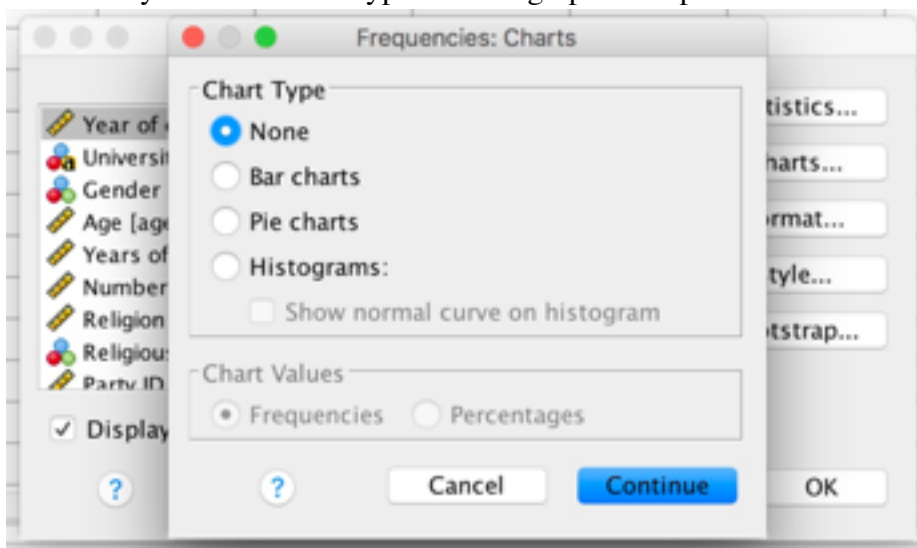
Select whether you want to produce a bar chart, pie chart, or histogram, and then select any of the available options. Once everything is ready, click OK, and the graph will be generated.

Using PSPP
• The same basic menu command (**Analyze > Descriptive Statistics > Frequencies**) also works in PSPP.

Syntax
• The default output includes the frequency of cases with each value of the selected variable(s), along with the percentage and cumulative percentage of cases with this value. If you don't want to include these, and only want to see a histogram or graph/chart, add "/STATISTICS=NONE".
• The HISTOGRAM subcommand can be used to produce a histogram for the selected variable(s). Adding =NORMAL to this command will superimpose a normal curve on the histogram.
• The PIECHART subcommand can be used to produce a pie chart/graph for the selected variable(s).

18

• The BARCHART subcommand can be used to produce a bar chart/graph for the selected variable(s).

```
FREQUENCIES /VARIABLES = xxx yyy zzz
  /HISTOGRAM=NORMAL.
```

```
FREQUENCIES /VARIABLES = xxx yyy zzz
  /PIECHART
  /STATISTICS=NONE.
```

```
FREQUENCIES /VARIABLES = xxx yyy zzz
  /BARCHART.
```

Other Resources
• Official SPSS Syntax Documentation: <u>FREQUENCIES</u>
• Official PSPP Documentation: <u>FREQUENCIES</u>
• Kent State Statistical Consulting: <u>Frequency Tables</u>, <u>Descriptive Stats for One Numeric Variable</u> <u>(FREQUENCIES)</u>,
• UCLA Statistical Consulting Group: <u>Annotated Output</u>

Alternative Option: "GRAPH"
These should produce the same basic output as the Frequencies command, but by giving the user more control. This can be accessed from the GRAPHS menu: **Graphs > Histogram** or **Graphs > Barchart** (in PSPP; other options are available in SPSS):
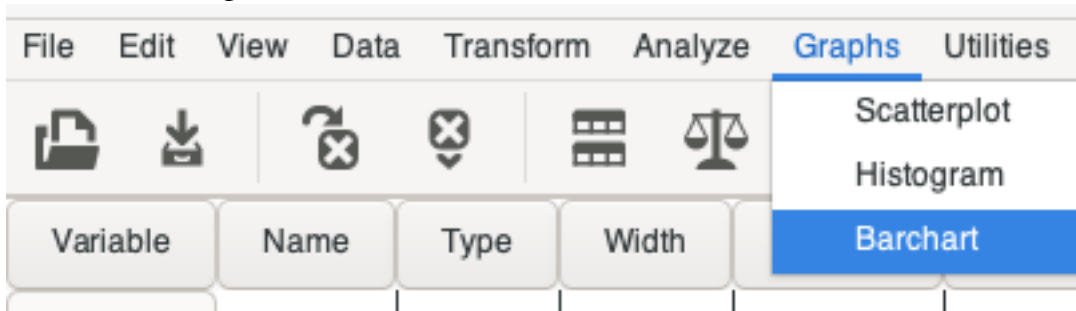
For a bar chart, this command will open a dialog box that allows you to choose the variable to be graphed. Select the variable(s) you want to examine on the left side of this box, and push the arrow button to add the variable to the Category Axis box. You can also choose what the bars represent ("N of cases" is usually the most useful, but you may want the percent of cases or the cumulative N or percentage). Click OK to run the analysis.
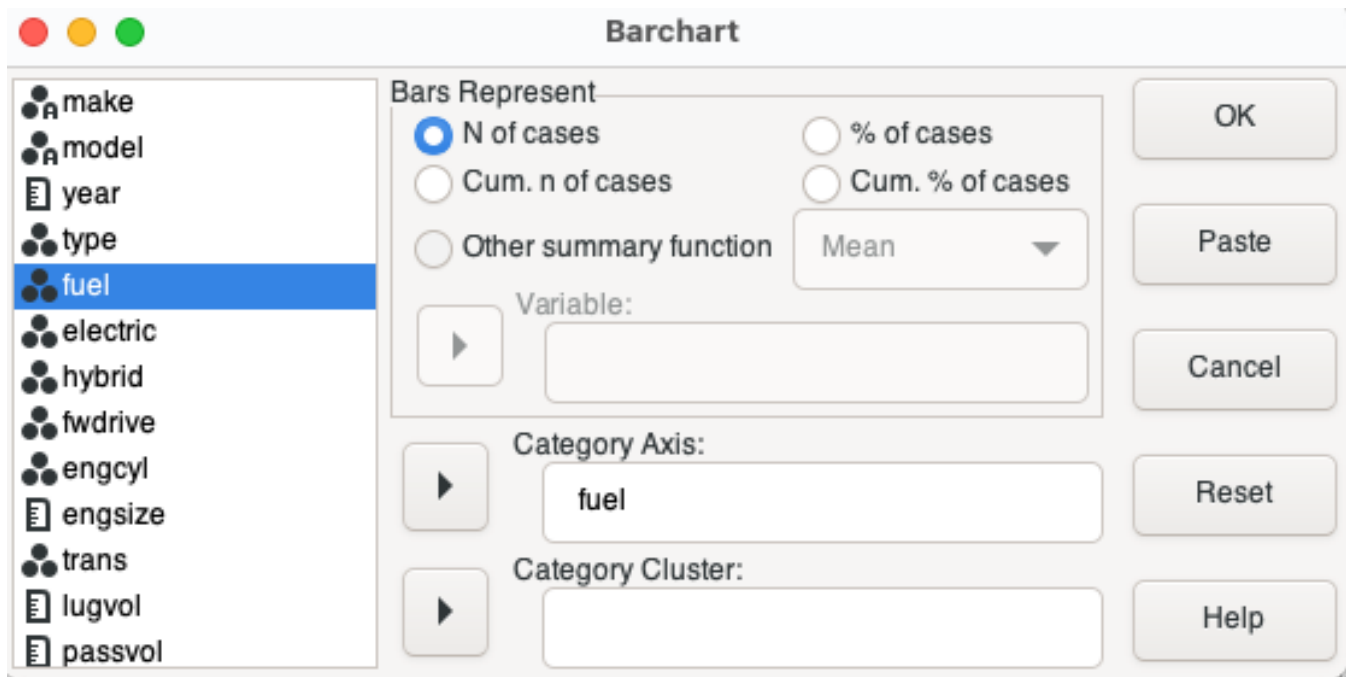
For a histogram, this command will open a dialog box that allows you to choose the variable to be graphed. Select the variable(s) you want to examine on the left side of this box, and push the arrow button to add the variable to the Variable box. You can also use the checkbox to decide whether or not you want to superimpose the normal curve over the histogram. Click OK to run the analysis.
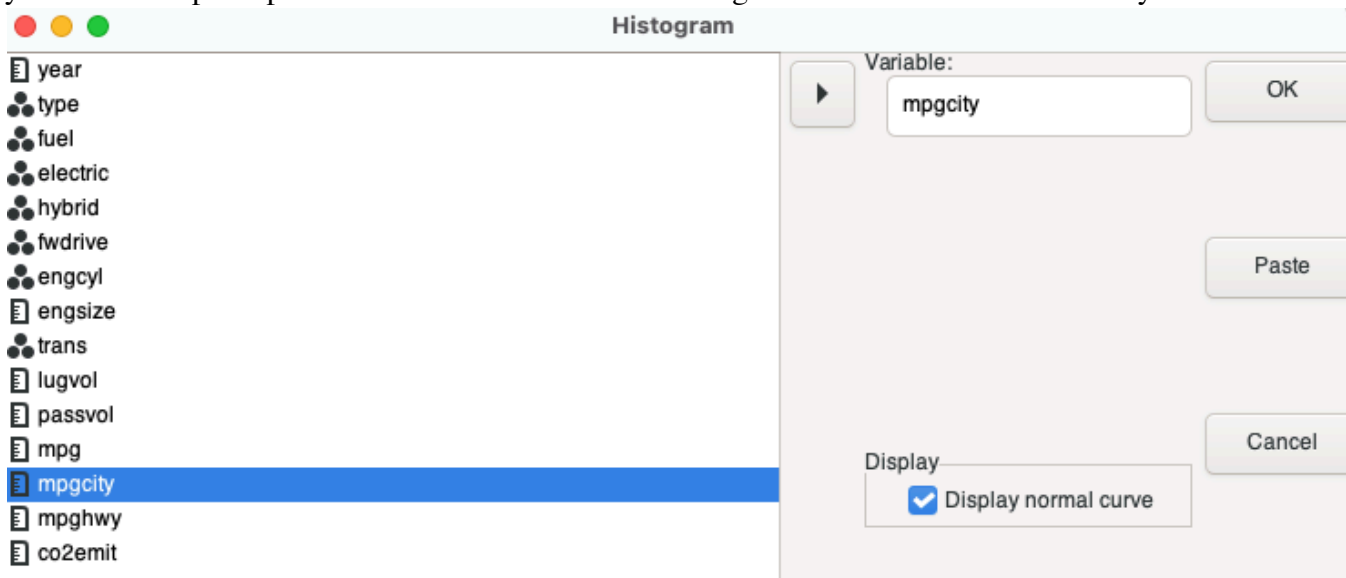
Using PSPP
• PSPP does not produce pie charts/graphs or line charts/graphs with the GRAPH command, although it can produce bar charts and histograms, and pie charts/graphs can still be produced using the FREQUENCIES command.

Syntax

• The HISTOGRAM subcommand will display a histogram for the selected variable(s). Adding (NORMAL) to this command will superimpose a normal curve on the histogram.
• The BAR subcommand will produce a bar chart/graph for the selected variable(s). The default is to have the bar represent the number of cases in each category of the variable. Instead of the number of cases (COUNT), you could specify PCT for the percentage of the total cases in each category. You could also create a bar graph using two variables, such as showing the mean value of one variable at each value of another variable. More options are available in SPSS (see the documentation linked below for more details).
• SPSS only: The default PIECHART subcommand can be used to produce a pie chart/graph for the selected variable(s).

GRAPH
 /HISTOGRAM=xxx.

GRAPH
 /HISTOGRAM (NORMAL)=xxx.

GRAPH
 /BAR=xxx.   [same as /BAR = COUNT by xxx]

GRAPH
 /BAR=MEAN(income) BY gender.

Other Resources
• Official SPSS Syntax Documentation: GRAPH, GRAPH-Bar Graphs , GRAPH-Histograms , GRAPH-Line Graphs , GRAPH-Pie Graphs
• Official PSPP Documentation: GRAPH, GRAPH-Bar, GRAPH-Histogram
• SPSS also has a more advanced graphing feature, which is not included in version 2.0.0 of PSPP: GGRAPH

Statistical Analyses Covered in Class

Crosstabs and Chi-Squares: "CROSSTABS"
To set up a crosstabulation (whether just displaying a crosstab of the values of two variables or also calculating various test statistics), go to **Analyze** > **Descriptive Statistics** > **Crosstabs**. This will open a dialog box that allows you to choose variables for analysis.
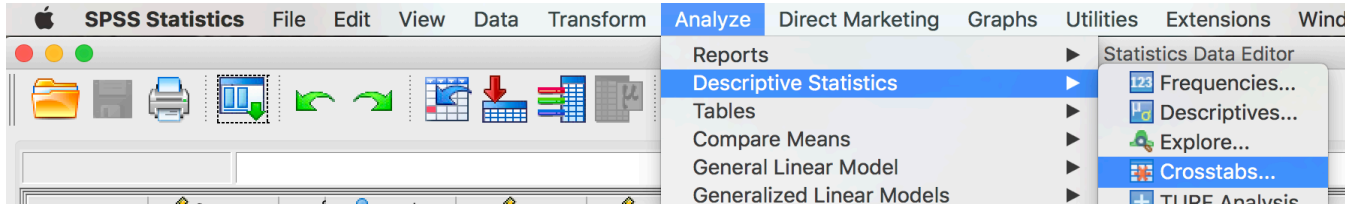
At this screen, select the variable you want to examine (the dependent variable) on the left side of this box, push the arrow button toward the top of the screen, and the variable should then appear in the "Row(s)" window at the right side of the box. Do the same to select the variable that you are using to identify groups for comparison (the independent variable), and push the button in the middle of the screen to select this for the "Column(s)" box.
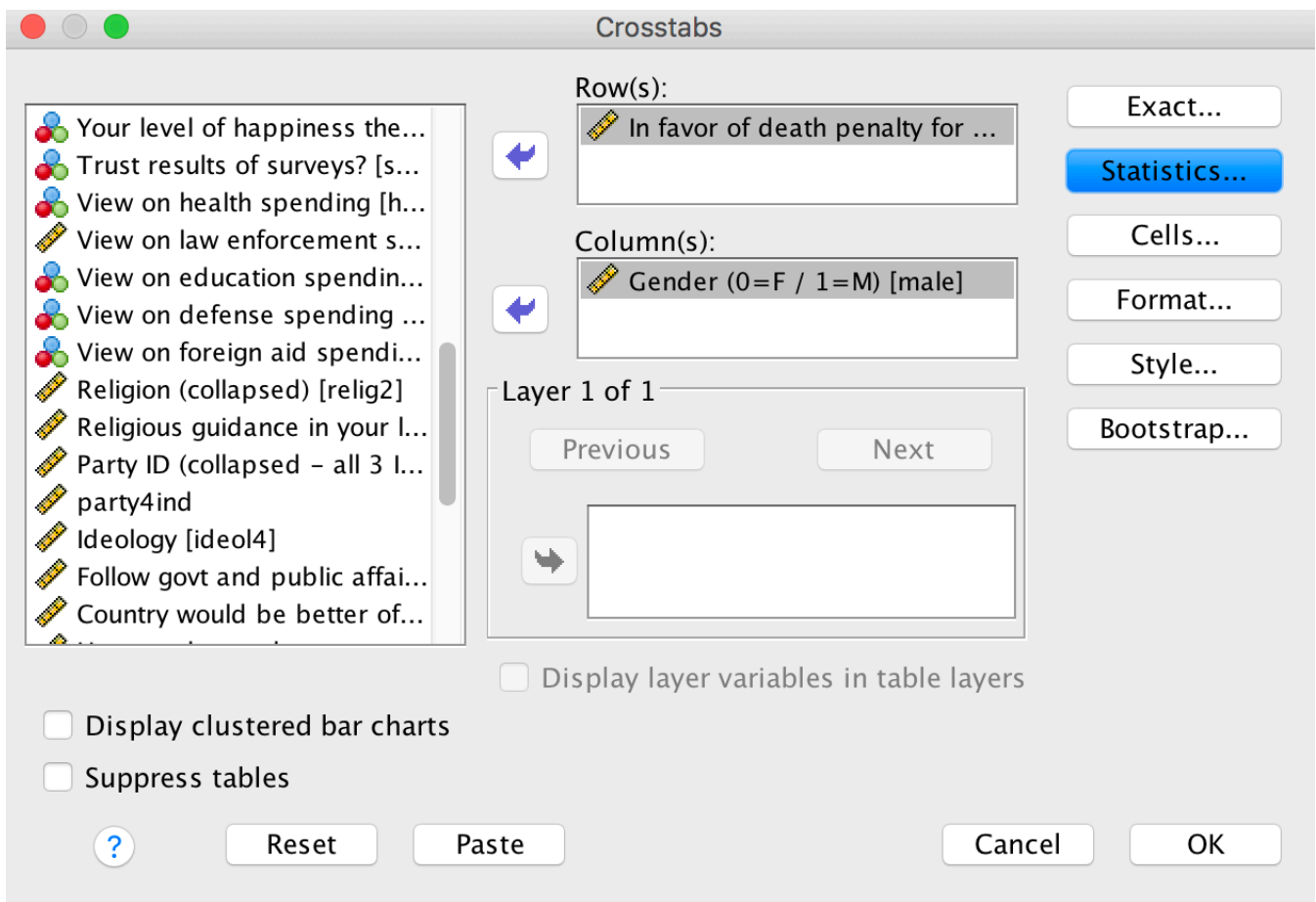
You should then select the "Statistics" box and be sure that "Chi-square," "Phi and Cramer's V," and "Lambda" (or any other desired test statistics) are checked, to make sure that you will get the relevant statistics, then click Continue to return to the main variable selection screen.

From here click the "Cells" box and be sure that the "Observed" counts and "Column" percentages are selected, then click Continue to return to the main variable selection screen. Once everything is ready, click OK in the main screen to run the analysis.
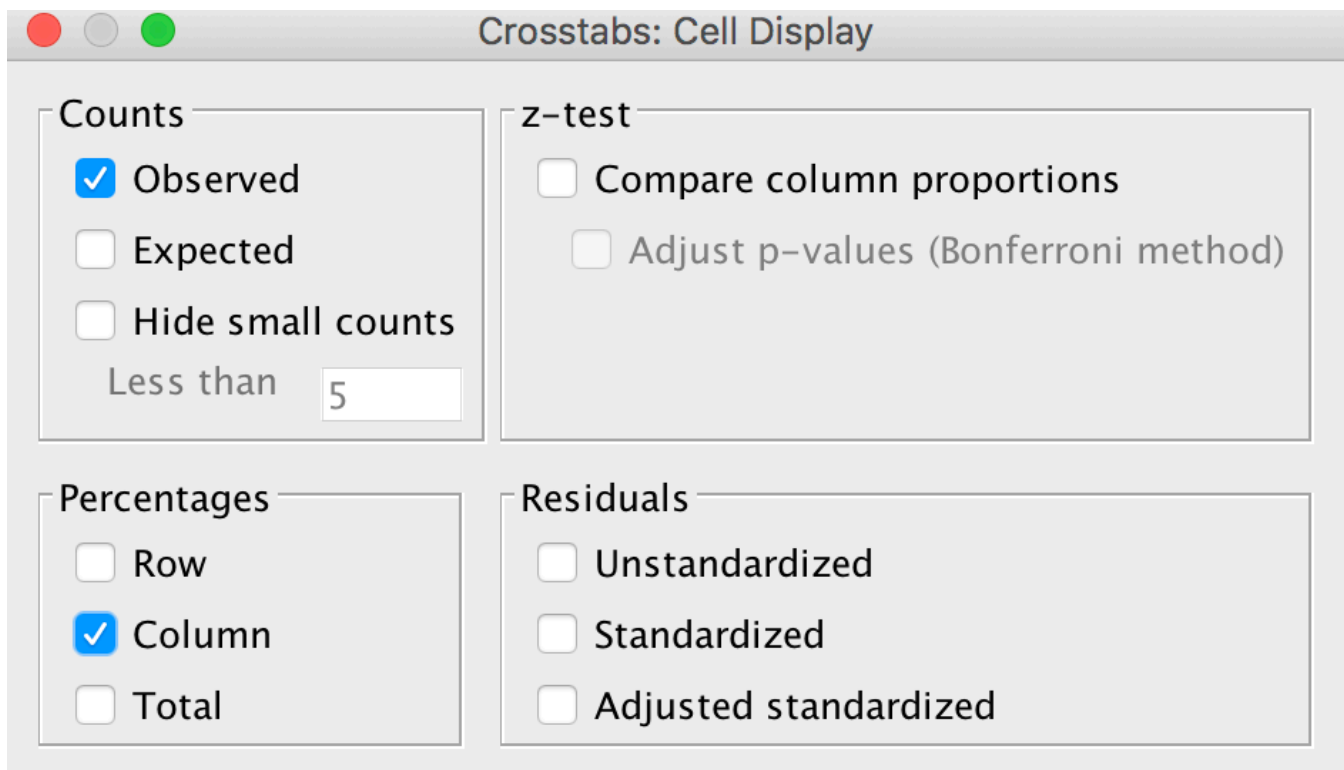
Using PSPP
• The same basic menu command (**Analyze** > **Descriptive Statistics** > **Crosstabs**) also works in PSPP. The options are named somewhat differently, but this usually easy to figure out (e.g. "Chisq" rather than "Chi-square" and "Phi" instead of "Phi and Cramer's V" for test statistics, and "Count" instead of "Observed" for cell contents).

Syntax
• The default setting only displays the count of cases in each cell of the table, along with the totals for each row and column.
• Other cell contents can be specified with the CELLS subcommand. Some of the most useful are count (COUNT), row percentage (ROW), column percentage (COLUMN), percentage of the entire table (TABLE), and expected value (EXPECTED).
• Various statistics can be specified with the STATISTICS subcommand. Some of the most useful are Chi-square (CHISQ), Phi and Cramer's V (PHI), Lambda (LAMBDA), Tau-B (BTAU), Tau-C (CTAU), Gamma (GAMMA), and (for 2x2 tables only) the odds ratio (RISK).

CROSSTABS
 /TABLES=xxx BY yyy.

CROSSTABS
 /TABLES=xxx BY yyy
 /CELLS=COUNT COL.

CROSSTABS
 /TABLES=xxx BY yyy

/STATISTICS=CHISQ PHI RISK.

Other Resources
• Official SPSS Syntax Documentation: CROSSTABS
• Official PSPP Documentation: CROSSTABS
• Kent State Statistical Consulting: Crosstabs, Chi-Square Test of Independence

Difference of Means: "T-TEST"

For a 2-sample difference of means test, SPSS uses a t-test. This is obtained by selecting **Analyze > Compare Means > Independent Samples T test** (because it is assumed that the samples are independent -- one-sample t-tests and paired-sample t-tests are different techniques that we aren't going to use in this course and you probably won't see very often in Political Science research):
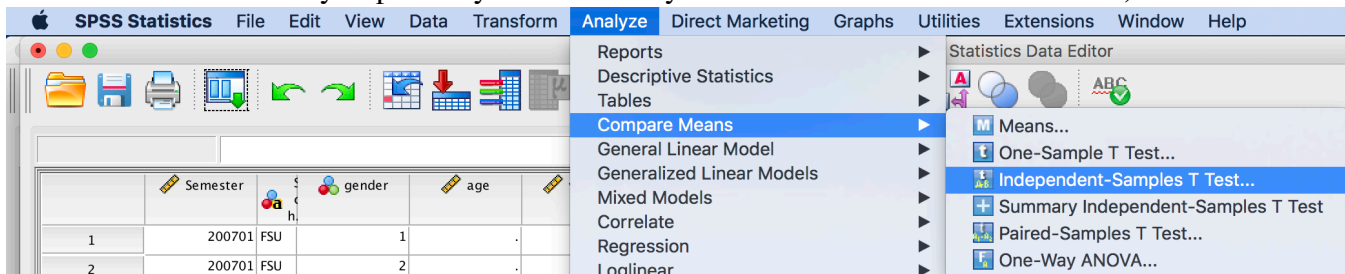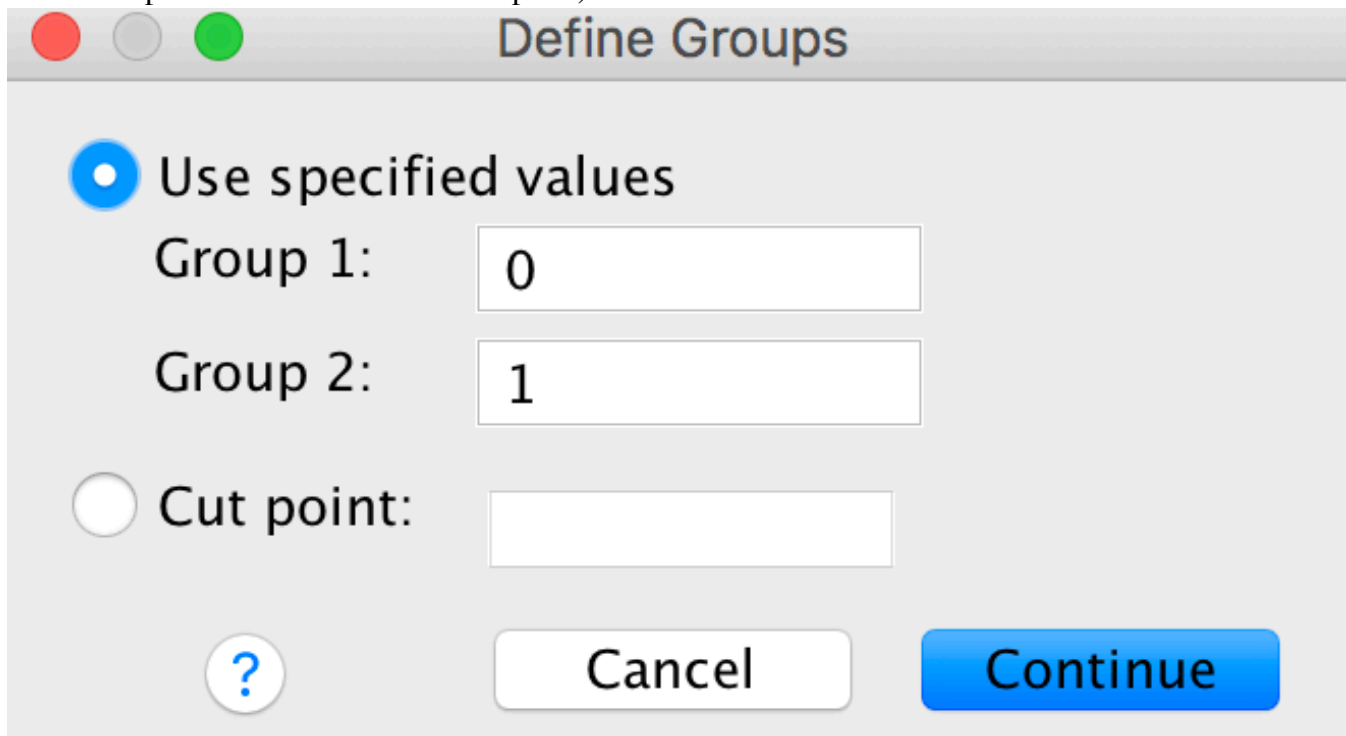
Once the method has been selected, you use the standard interface to select one or more Test Variables (the dependent variables you are trying to explain) as well as the Grouping Variable (the independent variable that includes the categories you are trying to compare):
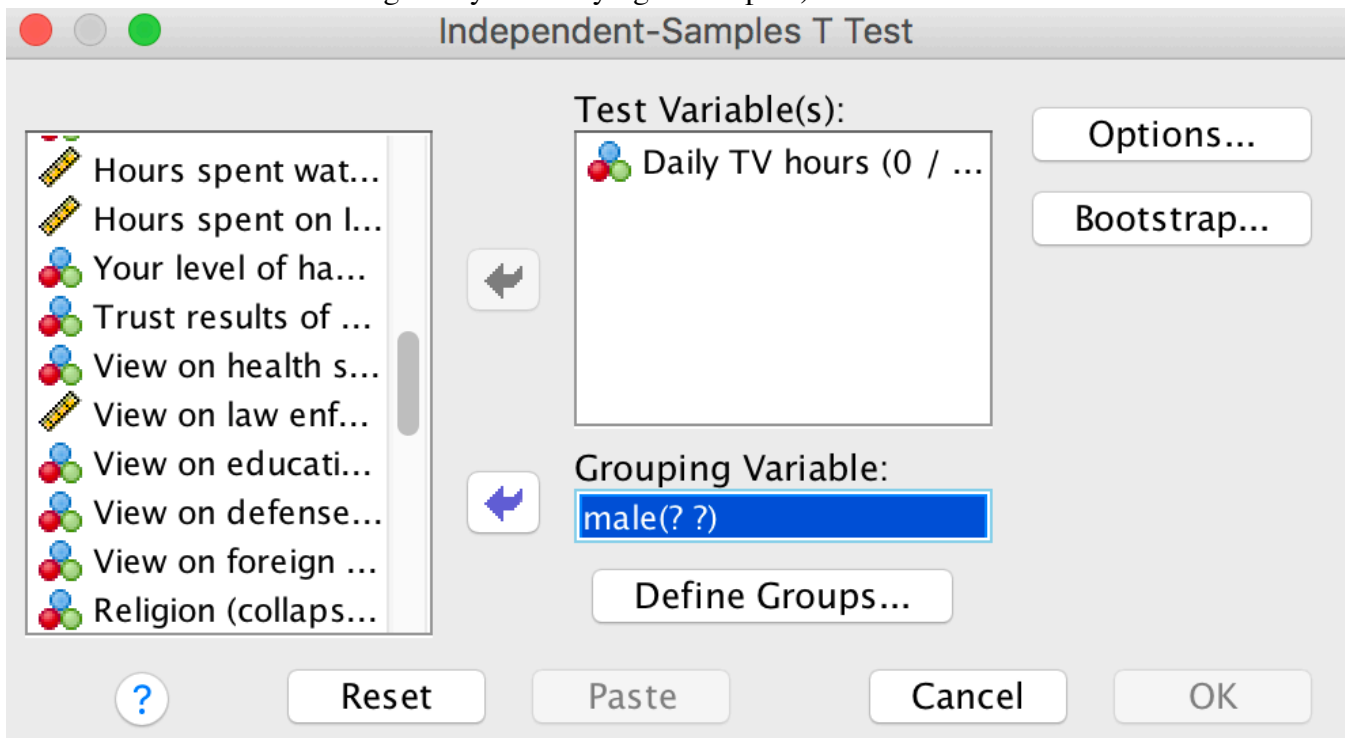
Note that the Grouping Variable is shown by default with two question marks. This is because SPSS needs to be told which two values you want it to compare (regardless of whether the grouping variable only has two values or it has many values and you only want it to compare two of them). This is done

25

by clicking the **Define Groups** button, and then entering the values you want for Group 1 and Group 2 (or for ordinal, interval, or ratio variables, by specifying a cut point, so that SPSS compares values below this point with values above this point):

Using PSPP
• This command works the same way in PSPP, with the added benefit that the Define Groups box gives you the ability to select any value of the variable from a dropdown menu (which can be useful if you don't remember what the values are).

Syntax
• Instructions for doing this in a syntax file will be provided later.

Other Resources
• Official SPSS Syntax Documentation: T-TEST
• Official PSPP Documentation: T-TEST, T-TEST/independent samples
• Kent State Statistical Consulting: Independent Samples T Test,
• UCLA Statistical Consulting Group: Annotated Output
• UNT's SPSS Short Course: *t*-Tests in SPSS

Analysis of Variance: "ONEWAY"
To run an ANOVA, go to **Analyze > Compare Means > One-Way ANOVA**. This will open a dialog box that allows you to choose variables for analysis:
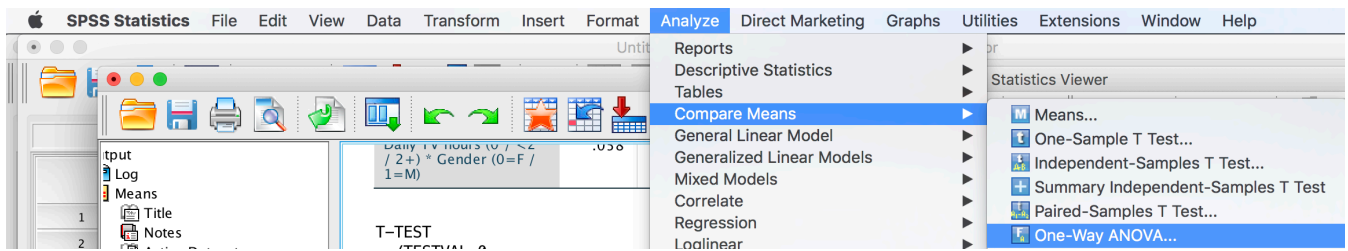
Select the variable you want to examine (the dependent variable) on the left side of this box, push the arrow button toward the top of the screen, and the variable should then appear in the "Dependent List" window at the right side of the box. Do the same to select the variable that you are using to identify groups for comparison (the independent variable), and push the button in the middle of the screen to select this for the "Factor" box.
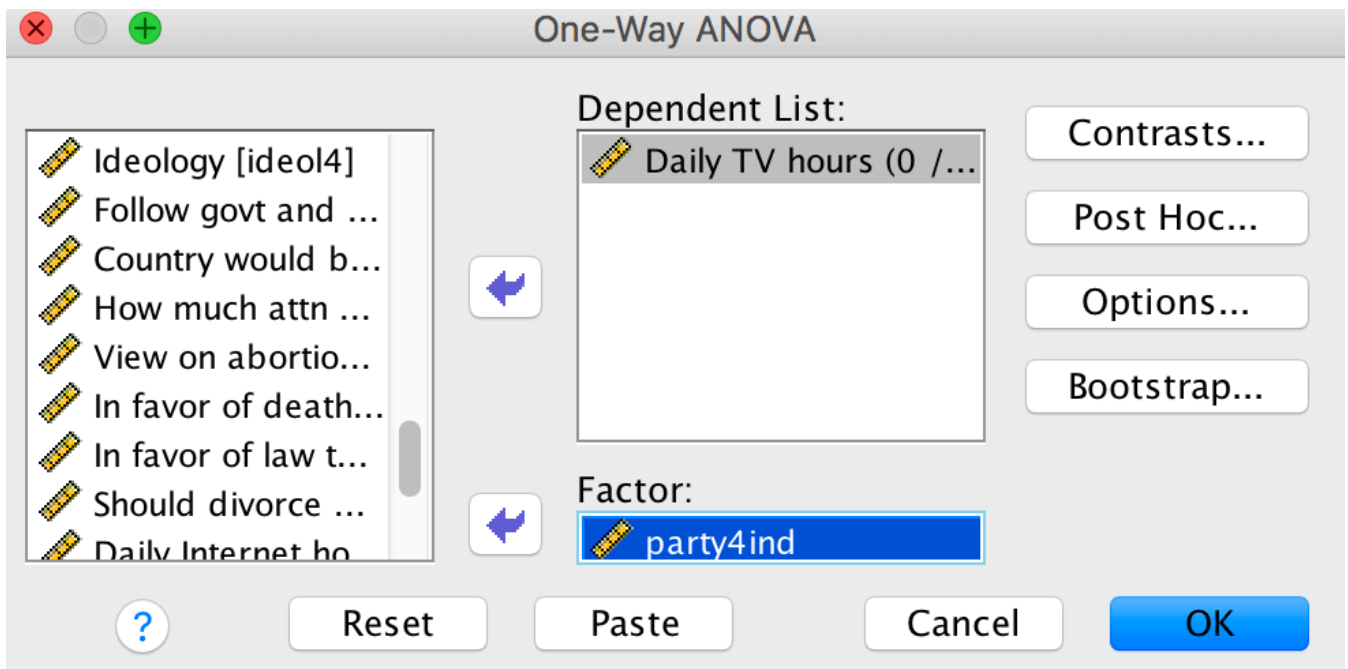
Click the "Options" box and be sure that "Descriptives" is checked, to make sure that you will get the means for each group; once this happens, click OK to run the analysis.
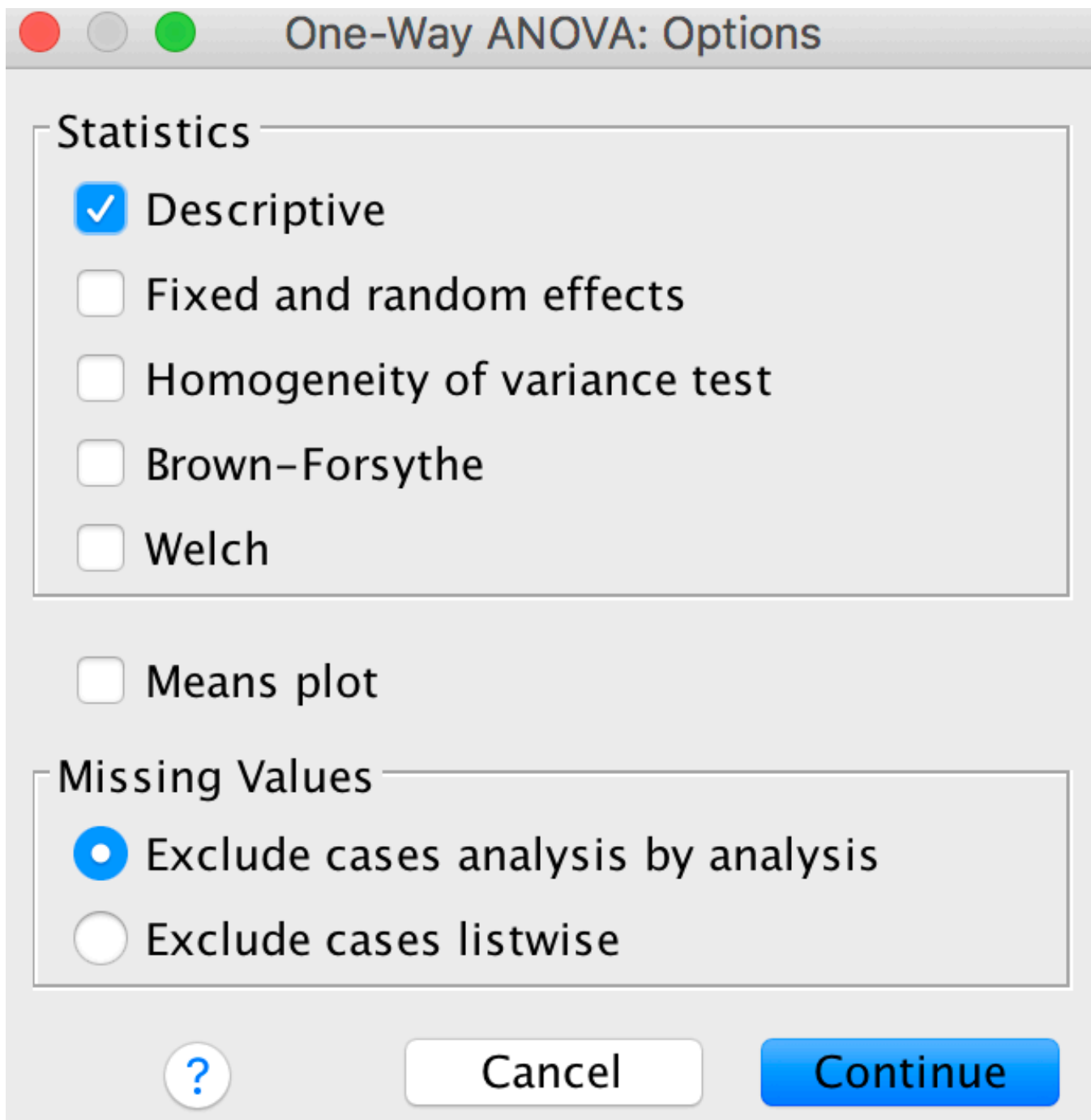
Using PSPP
• PSPP only has the ability to do oneway ANOVAs -- this is all we will be doing in this class, and this is the ANOVA technique that is used by most political scientists, but more advanced ANOVA techniques require SPSS or other software.   For oneway ANOVAs, the same basic menu command (**Analyze** > **Compare Means** > **One-Way ANOVA**) also works in PSPP. „; the main difference is that there are less options, so the option to choose Descriptive statistics is included directly at the bottom of the main ANOVA screen rather than in a separate Options screen.

Syntax
• Instructions for doing this in a syntax file will be provided later.

Other Resources
• Official SPSS Syntax Documentation: <u>ONEWAY</u>, <u>ANOVA</u>
• Official PSPP Documentation: <u>ONEWAY</u>
• Kent State Statistical Consulting: <u>One-Way ANOVA</u>
• UNT's SPSS Short Course: <u>Analysis of Variance (ANOVA) in SPSS</u>

Scatterplots: "GRAPH"

This command can create scatterplots, in addition to some of the types of charts/graphs described earlier. There are several ways to do this from the dropdown menus. Going to **Graphs > Chart Builder** offers more features and more options, but for the purposes of my course, it is better to use the "legacy" graphing options at **Graphs > Legacy Dialogs > Scatter/Dot**.
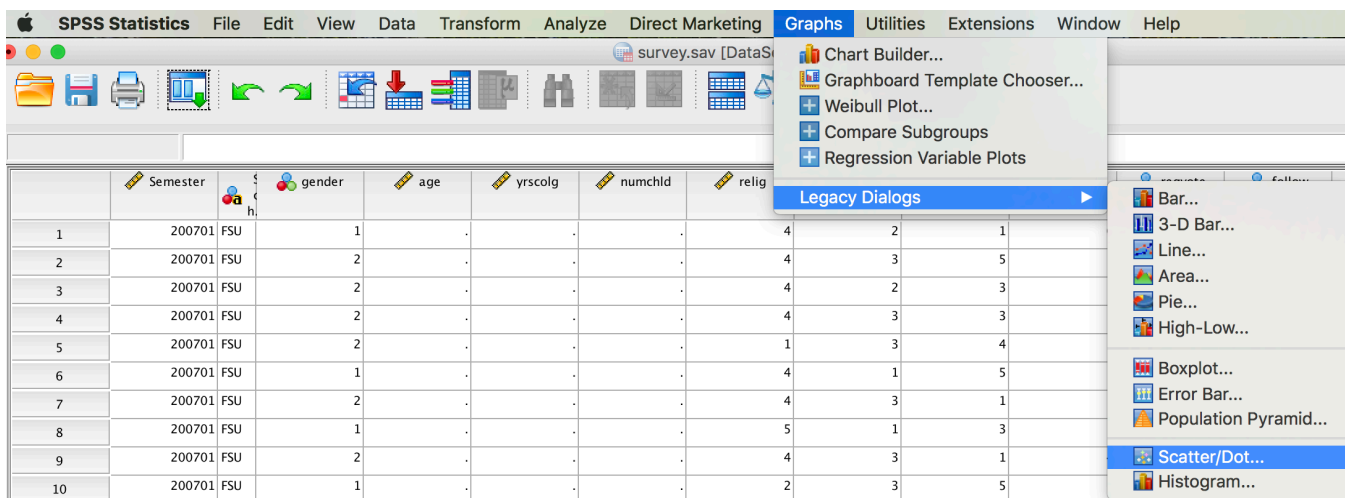
You will then have a choice among five types of graphs; for our purposes in this course, choose **Simple Scatter**, then click the Define button to choose the variables you wish to include in the plot.
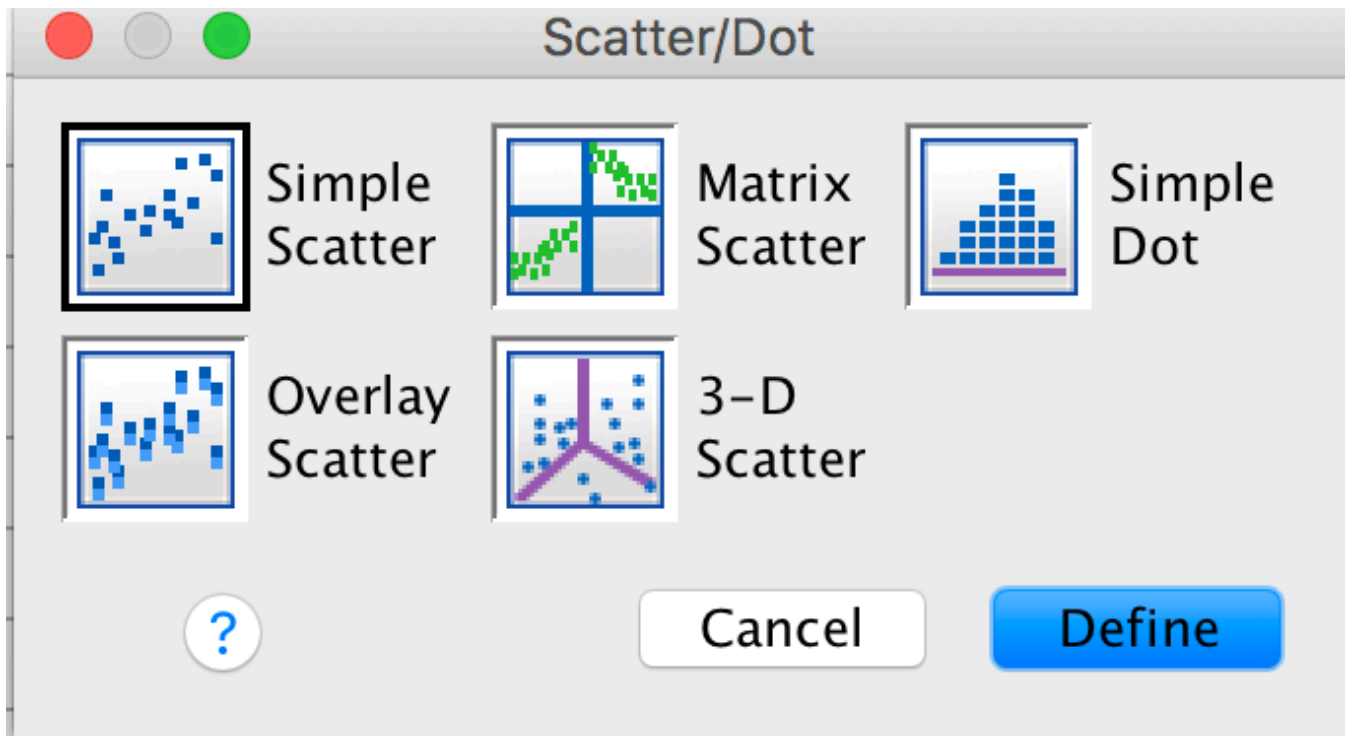
Select the variable you want to examine (the dependent variable) from the list on the left side of this box, push the arrow button next to the Y Axis box, and the variable should then be selected as the dependent variable for your analysis. Do the same to select the independent variable for the X Axis box. (If you want to remove a variable from the list, select the variable in the right side of the box, and click the arrow in the middle of the screen.) Click OK to run the analysis.
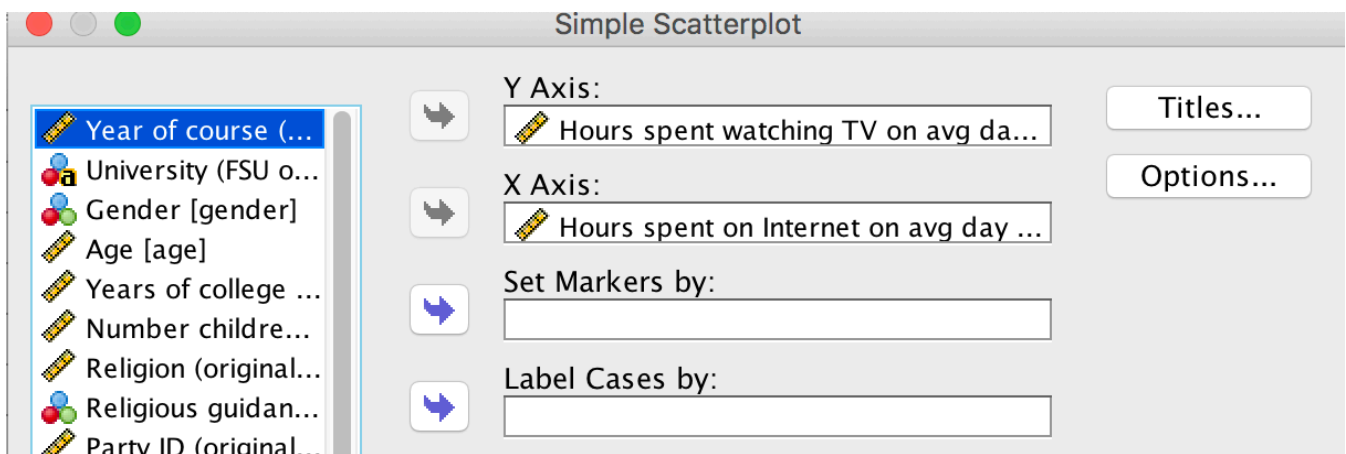
Using PSPP
• PSPP has fewer graphing options available, so setting up a scatterplot is more straightforward. Go to **Graphs > Scatterplot**. Select the variable you want to examine (the dependent variable) from the list on the left side of this box, push the arrow button next to the Y Axis box, and the variable should then be selected as the dependent variable for your analysis. Do the same to select the independent variable for the X Axis box. (If you want to remove a variable from the list, select the variable in the right side of the box, and click the arrow in the middle of the screen.) Click OK to run the analysis.

Syntax
• Instructions for doing this in a syntax file will be provided later.

Other Resources
• Official SPSS Syntax Documentation: GRAPH, GRAPH-Scatterplots
• Official PSPP Documentation: GRAPH, GRAPH-Scatterplot
• NC State SPSS Tutorials: Creating Scatter Plots

Correlation: "CORRELATIONS"

SPSS offers a number of correlation-related analyses, but for our purposes in this course, all you need to run is Pearson's correlation coefficient, which is available at **Analyze > Correlate > Bivariate**.
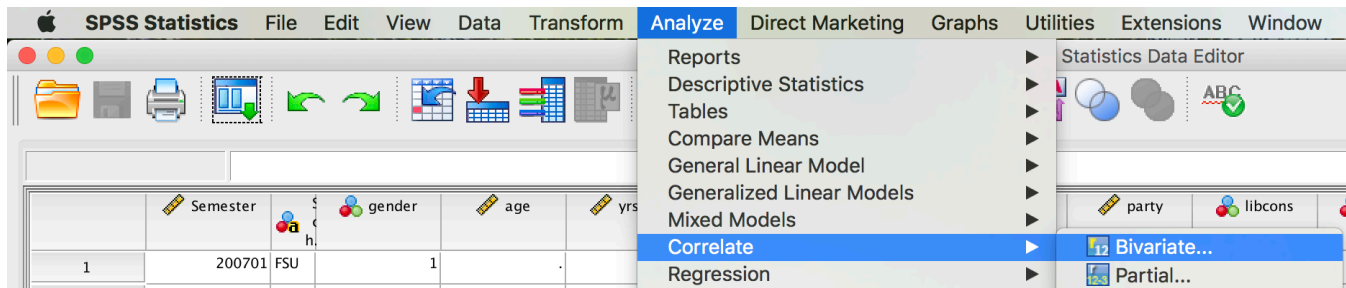
One at a time (don't worry about the order - independent and dependent variables don't matter for correlation), select each variable that you want to examine from the list on the left side of this box, and push the arrow button to add it to the Variables box. (If you want to remove a variable from the list, select the variable in the right side of the box, and click the arrow in the middle of the screen.) Make sure the Pearson box is checked under Correlation Coefficients, and click OK to run the analysis.
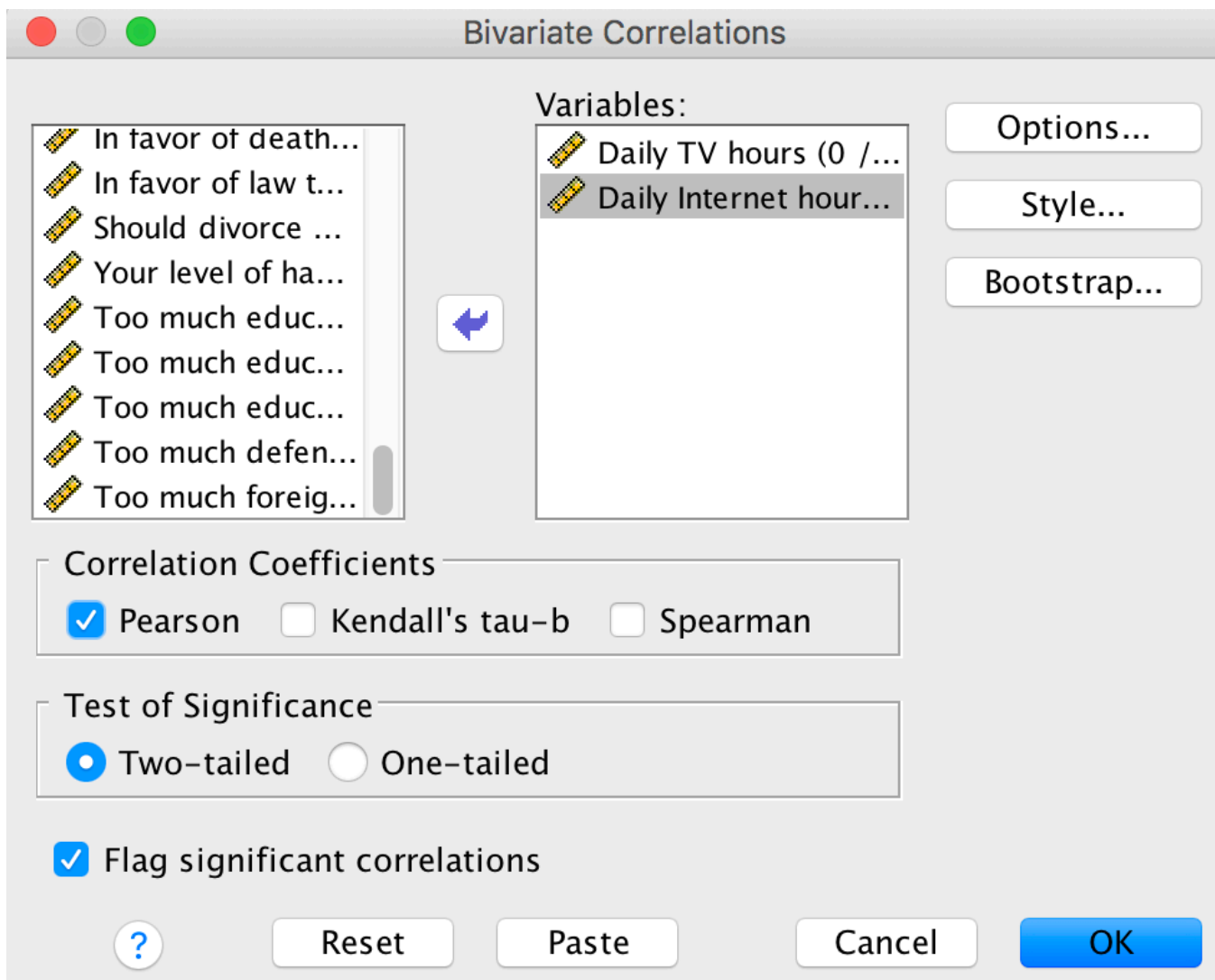
Using PSPP
• PSPP offers fewer options for correlation analysis than SPSS. Go to **Analyze > Bivariate Correlation**. One at a time (don't worry about the order - independent and dependent variables don't matter for correlation), select each variable that you want to examine from the list on the left side of this box, and push the arrow button to add it to the Variables box. (If you want to remove a variable from the list, select the variable in the right side of the box, and click the arrow in the middle of the screen.) Click OK to run the analysis (unlike SPSS, Pearson's correlation coefficient is the only technique offered by PSPP, so you don't need to select it).

Syntax
• Instructions for doing this in a syntax file will be provided later.

Other Resources
• Official SPSS Syntax Documentation: CORRELATIONS
• Official PSPP Documentation: CORRELATIONS
• Kent State Statistical Consulting: Pearson Correlation
• UCLA Statistical Consulting Group: Annotated Output

32

OLS Regression: "REGRESSION"

To run either a bivariate or multiple regression analysis, go to **Analyze** > **Regression** > **Linear**. This will open a dialog box that allows you to choose variables for analysis.

Select the dependent variable on the left side of this box, push the arrow button toward the top of the screen, and the variable should then appear in the "Dependent" window at the right side of the box. Do the same to select the independent variables, and push the button in the middle of the screen to select this for the "Independent(s)" box.
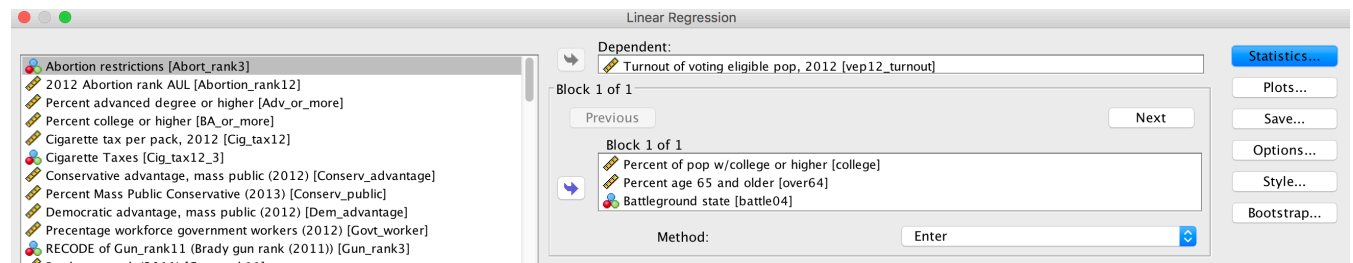
Using the Statistics button, make sure that Estimates, Confidence Intervals, and Model Fit are selected, click Continue to return to the variable selection screen, and click OK to run the analysis
.
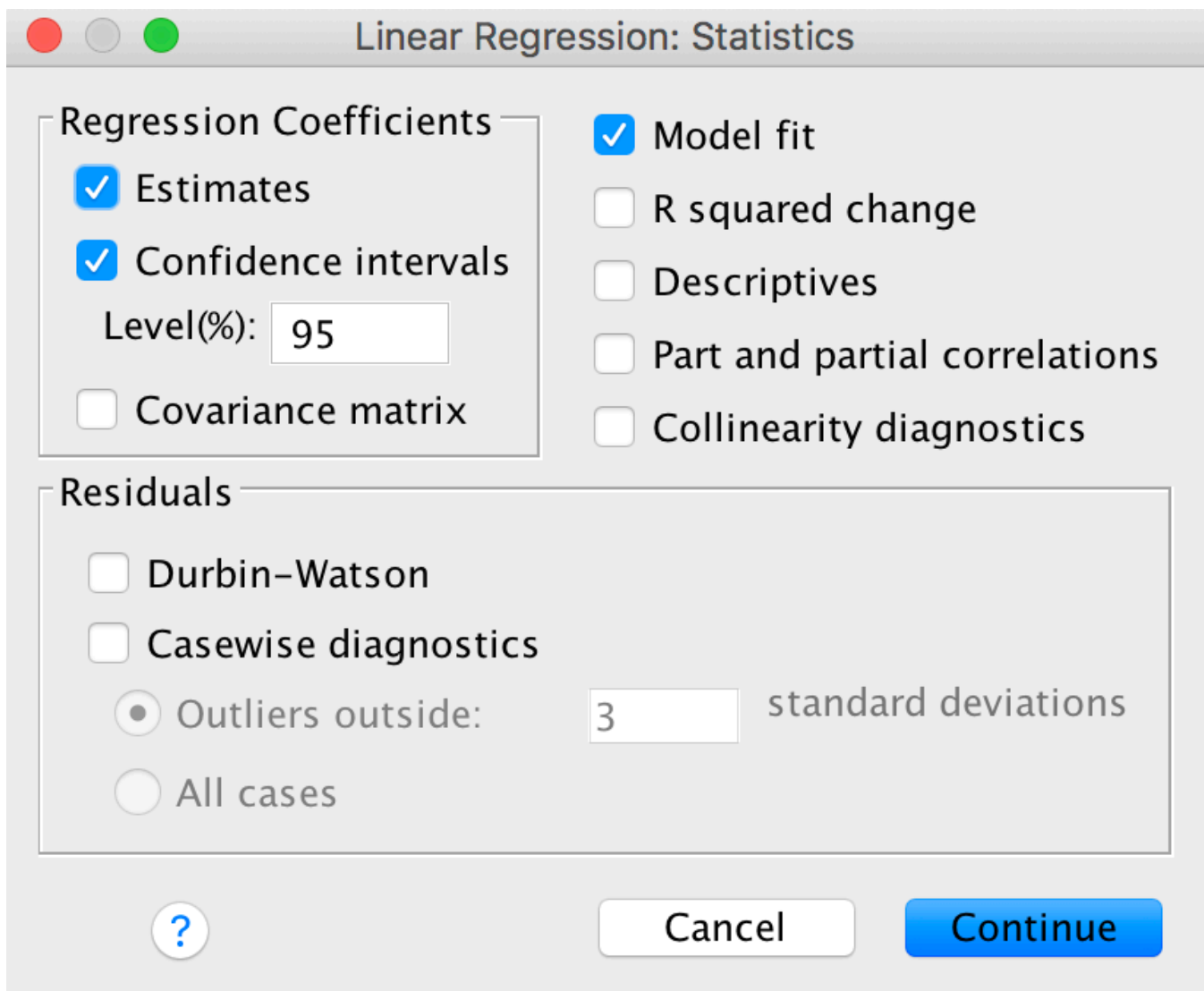
Using PSPP
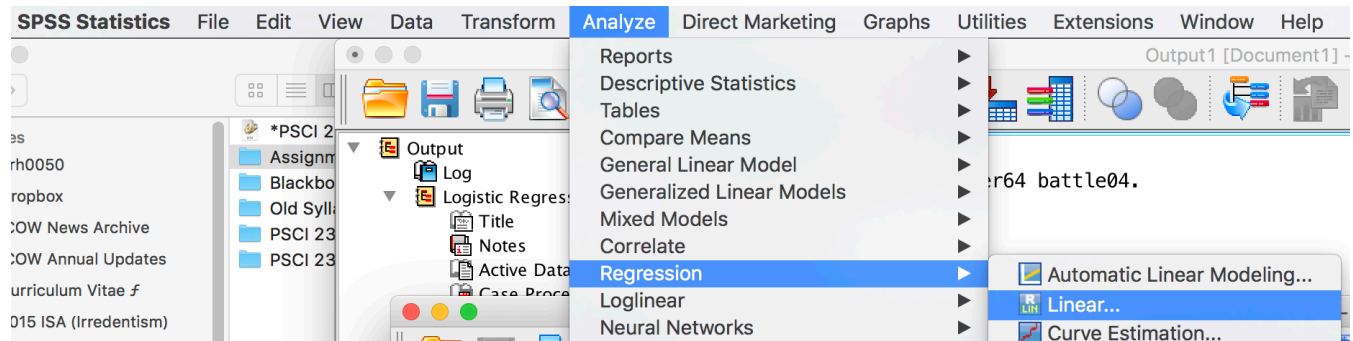• Instructions for doing this in PSPP will be provided later.

Syntax
• Instructions for doing this in a syntax file will be provided later.

Other Resources
• Official SPSS Syntax Documentation: REGRESSION
• Official PSPP Documentation: REGRESSION, REGRESSION-Syntax
• UCLA Statistical Consulting Group: Annotated Output
• UNT's SPSS Short Course: Bivariate Linear Regression in SPSS, Multiple Linear Regression in SPSS

Logistic Regression: "LOGISTIC REGRESSION"
More detail on this method (which was discussed at the end of the semester and is not currently required for any homework assignments in this course) will be provided later.

Using PSPP
• Instructions for doing this in PSPP will be provided later.

Other Resources
• Official SPSS Syntax Documentation: <u>LOGISTIC REGRESSION</u>
• Official PSPP Documentation: <u>LOGISTIC REGRESSION</u>
• UCLA Statistical Consulting Group: <u>Data Analysis Example</u>, <u>Annotated Output</u>
• UNT's SPSS Short Course: <u>Logistic Regression (Binary)</u>

More Advanced Analyses

Multinomial Logit: "NOMREG" (SPSS only)
More detail on these method (which was discussed at the end of the semester and is not required for any homework assignments in this course) will be provided later.

Using PSPP
• This is not available in PSPP version 1.2.0. It may be added in future versions, though.

Other Resources
• Official SPSS Syntax Documentation: NOMREG
• UCLA Statistical Consulting Group: Data Analysis Example, Annotated Output
• UNT's SPSS Short Course: Logistic Regression (Multinomial)

Ordered Logit/Probit: "PLUM" (SPSS only)
More detail on these method (which was discussed at the end of the semester and is not required for any homework assignments in this course) will be provided later.

Using PSPP
• This is not available in PSPP version 1.2.0. It may be added in future versions, though.

Other Resources
• Official SPSS Syntax Documentation: PLUM
• UCLA Statistical Consulting Group: Data Analysis Example, Annotated Output

Event Count Models: "GENLIN" (SPSS only)
More detail on these method (which was discussed at the end of the semester and is not required for any homework assignments in this course) will be provided later.

Using PSPP
• This is not available in PSPP version 1.2.0. It may be added in future versions, though.

Other Resources
• Official SPSS Syntax Documentation: GENLIN
• UCLA Statistical Consulting Group: Data Analysis Examples: Poisson Regression and Negative Binomial Regression, Annotated Output: Poisson Regression

Survival/Duration: "SURVIVAL," "KM," "COXREG," "SURVREG" (SPSS only)
More detail on these method (which was discussed at the end of the semester and is not required for any homework assignments in this course) will be provided later.

Using PSPP
• This is not available in PSPP version 1.2.0. It may be added in future versions, though.

Other Resources
• Official SPSS Syntax Documentation: COXREG, KM, SURVIVAL, SURVREG AFT, SURVREG RECURRENT

**Selection Models: [Not Currently Available]**

As of SPSS version 29 and PSPP version 1.2.0, neither of these applications allows the user to conduct Heckman-style selection models. Most academics who need these procedures use more advanced statistical software like STATA or R.

Appendix: Using Syntax Files

Syntax files are not required for any of the homework assignments in my course. Any student who uses SPSS or PSPP for more advanced courses or for their own research after the class ends, though, will need to learn how to work directly with syntax rather than always doing everything from the dropdown menus at the top of the screen.

Opening a Syntax File

SPSS and PSPP do not automatically create a new syntax file when it is opened, so if you want to work with a syntax file rather than using the graphical user interface, you will need to create or open one manually. To create a new syntax file, use **File > New > Syntax**.
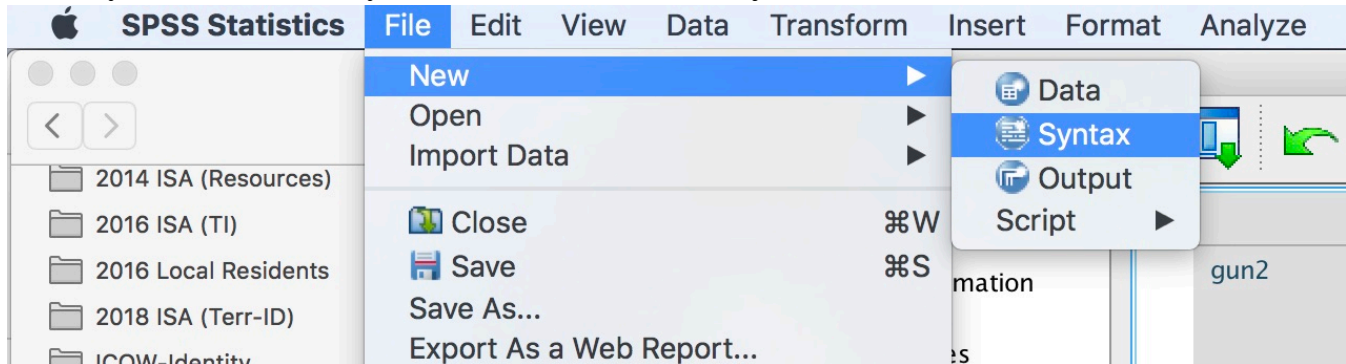
If you want to use a syntax file that you have already saved, you can open SPSS or PSPP by double-clicking on the syntax file (depending on how your computer is set up to handle .sps files). You can also open an existing syntax file from the SPSS/PSPP menu with **File > Open > Syntax**.

Basic Syntax Notes

• Most examples and tutorials use upper-case to indicate commands, but this is not necessary. The following should work equivalently:

    CODEBOOK
      /STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.
    codebook
      /statistics count percent mean stddev quartiles.
    Codebook
      /Statistics Count Percent Mean Stddev Quartiles.

• Each command must start on a new line. Commands that cover multiple lines do not need to start at the very beginning of the line though; it is often helpful to indent the second or later lines in a single command to make it easier for you to follow.

• A single command can be split into multiple lines of syntax to make it easier to follow, but this is not necessary. The following should work equivalently:

    CODEBOOK
      /STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.
    CODEBOOK /STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.

• Each command should end with a period at the end of the final line in that command. If a period is omitted, SPSS should interpret a blank line as ending the previous command, but it is always better to include the period just to be avoid confusion or errors later if you edit the syntax file.

• Subcommands are usually separated from the main command by slashes:

    CODEBOOK

/STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.

• All text included within apostrophes or quotation marks (such as variable labels or value labels) must be contained on a single line of syntax.

Pasting Commands into Syntax Window

Even if you are using the graphical user interface to run your SPSS commands, you may still choose to paste the commands into the syntax window, where they can be saved or edited later. Almost every dropdown menu includes a "Paste" button at the bottom of the menu; clicking this button rather than "OK" will paste the command syntax in the syntax window rather than running it:

You can also see how a command is structured by looking at the output from any command that you have run using the drop-down menus. SPSS automatically places the syntax for each command in the output file before it displays the output of the command.

Running Commands from the Syntax Window

You can run a single command by highlighting the command using your mouse, then using the dropdown menu (**Run > Selected**) in either SPSS or PSPP. SPSS also includes a number of icons at the top of the screen, one of which looks like a "play" button (a sideways triangle); this can be used to run the selected command.

You can run the entire syntax file at once by using the dropdown menu (**Run > All**) in either SPSS or PSPP, by using your mouse (or **Edit > Select All** in SPSS) to select the entire syntax file and then clicking the "play" button as described above.

Appendix: Generating, Recoding, and Labeling Variables

As users get more advanced, they will need to manipulate variables in their data sets, rather than just taking them in their original form. This can be done with dropdown menus, but it is much easier with syntax files.

Generating and Recoding Variables:
• Official SPSS Syntax Documentation: COMPUTE, IF, RECODE, RENAME VARIABLES
• Official PSPP Documentation: COMPUTE, IF, RECODE, RENAME VARIABLES

• Kent State Statistical Consulting: Computing Variables, Recoding Variables
• UCLA Statistical Consulting Group: Creating and Recoding Variables, Using SPSS Functions for Making and Recoding Variables

Formatting and Labeling Variables:
• Official SPSS Syntax Documentation: APPLY DICTIONARY, FORMATS, MISSING VALUES, STRING, VALUE LABELS, VARIABLE LABELS
• Official PSPP Documentation: APPLY DICTIONARY, FORMATS, MISSING VALUES, STRING, VALUE LABELS, VARIABLE LABELS

• Kent State Statistical Consulting: Defining Variables, Variable Types, Date-Time Variables
• UCLA Statistical Consulting Group: Labeling and Documenting Data, Using Dates in SPSS

Appendix: Sorting and Selecting Data for Analysis

Many statistical analyses will require users to analyze only a subset of the data set being used, perhaps to focus only on one election, one Congressional session, or one type of country -- or to run separate analyses for different elections, sessions, or types of countries.

• Official SPSS Syntax Documentation: DO IF, DROP/KEEP subcommands in SAVE, FILTER, IF, SELECT IF, SORT CASES, SPLIT FILE, TEMPORARY
• Official PSPP Documentation: DO IF, DROP/KEEP subcommands in SAVE, FILTER, IF, SELECT IF, SORT CASES, SPLIT FILE, TEMPORARY

• Kent State Statistical Consulting explanation: Sorting Data, Splitting/Grouping Data
• UCLA Statistical Consulting Group explanation: Using the Sort Cases and Split File Commands, How Can I Analyze a Subset of My Data?, Subsetting Data
• UNT's SPSS Short Course: Select Cases / Create a Filter Variable
• UT SPSS Tutorials: SPSS: Data Manipulation and Advanced Topics

Appendix: Merging Two Data Sets

More advanced users may need to combine multiple data sets for analysis, because no single data set is able to meet all of their needs. For example, an international conflict scholar may need to merge together data sets cover the conflict behavior, democracy levels, military power, and other attributes of the countries being studied. There may also be a need to aggregate (or concatenate) several different data sets that contain the same variables for different cases (perhaps different elections or different polls).

• Official SPSS Syntax Documentation: ADD FILES, MATCH FILES
• Official PSPP Documentation: ADD FILES, MATCH FILES

• UCLA Statistical Consulting Group: Concatenating Data Files, Match Merging Data Files
• UNT's SPSS Short Course: Use the Merge Function to Add Cases, Use the Merge Function to Add Variables
• UT SPSS Tutorials: SPSS: Data Manipulation and Advanced Topics

<div align="center">**Other Useful Resources**</div>

My web site for this courses includes an Additional Online Resources page, which lists a number of statistical reference books and textbooks if you want to go further in using statistical methods than we did in this course. That page also includes links to a variety of political polls, survey data sets, and other data sets that are widely used by political scientists, which you might find useful in future research projects:

<div align="center">< **https://www.paulhensel.org/Teaching/psci3300r.html** ></div>

SPSS-Specific Resources
• Official SPSS Guide to Command Syntax:
  <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=reference-introduction-guide-command-syntax>
• Official SPSS Tutorials and Case Studies (step by step guides for common tasks and commands):
  <https:/https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=tutorial-introduction/www.ibm.com/support/knowledgecenter/SSLVMB_22.0.0/com.ibm.spss.statistics.tut/spss/tutorials/introtut2.htm>
  <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=studies-introduction>

• SPSS Tutorials from Kent State University Libraries: <https://libguides.library.kent.edu/SPSS>
• SPSS Tutorials from UCLA: <https://stats.idre.ucla.edu/spss/>
  Learning Modules: <https://stats.idre.ucla.edu/spss/modules/>
  Frequently Asked Questions: <https://stats.idre.ucla.edu/spss/faq/>
• SPSS Tutorials from UNT: <http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/SPSS_SC/>
• SPSS Tutorials from UT: <https://stat.utexas.edu/training/software-tutorials#spss>
• SPSS Tutorials from the University of Wisconsin:
  The Basics: <https://ssc.wisc.edu/sscc/pubs/spss/classintro/spss_students1.html>
  Statistics & Graphs: <https://ssc.wisc.edu/sscc/pubs/spss/classintro/spss_students2.html>

PSPP-Specific Resources
• Official PSPP Manual (more detail on what each command does and what options are available):
  <https://www.gnu.org/software/pspp/manual/>

• Basic Stats in PSPP (from Norm Lewis at the University of Florida):
  <http://www.bikeprof.com/uploads/9/0/6/5/9065192/stats_intro_in_pspp_handout_npl.pdf>
• PSPP for Beginners (from Gary Fisk at Georgia Southwestern State University):
  <https://garyfisk.com/pspp/index.html>
• PSPP Tutorial from NC State University:
  <https://it.chass.ncsu.edu/tutorials/pspp/index.php>

Advice on Choosing Statistical Techniques
• Choosing the Correct Statistical Test (from UCLA):
  <https://stats.idre.ucla.edu/other/mult-pkg/whatstat/>
• SPSS Statistics Coach (from IBM, the makers of SPSS):
  <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=statistics-coach>