

GETTING YOUR DATA INTO SPSS



UNIVERSITY OF GUELPH

LUCIA COSTANZO

lcostanz@uoguelph.ca

REVISED SEPTEMBER 2011

CONTENTS

Getting your Data into SPSS	0
SPSS availability	3
Data for SPSS Sessions	4
Data entry	6
Opening a data file.....	6
Steps to opening a SPSS data file	6
Steps to opening a text data file	6
Steps to opening an Excel data file	10
Data editor	11
Data View.....	11
Variable View	12
Entering data	13
Inserting new variables.....	14
Deleting variables	14
Inserting new cases	14
Deleting cases	14
Moving to a specific case	14
Data transformations	16
Computing variables.....	18
Steps to compute a variable	18
Recoding values	20
Recode into same variables	20
Steps to recode values of a variable	20

Recode into different variables	22
Steps to recode values of a variable	22
Analyzing Data	24
Descriptive Statistics.....	25
Frequencies	25
Descriptives	27
Journal File.....	29
Working with the Output	30
Viewer.....	30
Saving Output	31
Steps to saving output	31
Export Output	31
Pasting objects from the viewer into other applications	32
Steps to pasting objects in other applications.....	32
Creating charts.....	33
Steps to Creating a Bar Chart.....	33
Adding Error Bars.....	35
Modifying Axis Label	36
Steps to Creating a Scatterplot.....	36
Modifying a Chart	39
Changing Colours	39
Formatting Numbers in Tick Labels	39
Editing Text	39

SPSS AVAILABILITY

Faculty, staff and students at the University of Guelph may access SPSS two different ways:

- 1. Library computers**

On the library computers, SPSS is installed on all machines.

- 2. Acquire a copy for your own computer**

If you are faculty, staff or a student at the University of Guelph, you may obtain the site-licensed standalone copy of SPSS at a cost. A free concurrent copy of SPSS is available to faculty, staff or graduate students at the University of Guelph. However, it may only be used while you are employed or a registered student at the University of Guelph. To obtain a copy, go to the CCS Software Distribution Site (www.uoguelph.ca/ccs/download).

DATA FOR SPSS SESSIONS

DATASET: CANADIAN TOBACCO USE MONITORING SURVEY 2010 – PERSON FILE

This survey tracks changes in smoking status, especially for populations most at risk such as the 15- to 24-year-olds. It allows Health Canada to estimate smoking prevalence for the 15- to 24-year-old and the 25-and-older groups by province and by gender on a semi-annual basis.

The sample data used for this series of SAS workshops only includes respondents from the province of Quebec and only 14 of a possible 202 variables are being used.

To view the data, open the Excel spreadsheet entitled CTUMS_2010.xls

Variable Name	Label for Variable
PUMFID	Individual identification number
PROV	Province of the respondent
DVURBAN	Characteristic of the community
HHSIZE	Number of people in the household
HS_Q20	Number of people that smoke inside the house
DVAGE	Age of respondent
SEX	Respondent's sex
DVMARST	Grouped marital status of respondent
PS_Q30	Age smoked first cigarette
PS_Q40	Age begin smoking cigarettes daily
WP_Q10A	Number of cigarettes smoked – Monday
WP_Q10B	Number of cigarettes smoked – Tuesday
WP_Q10C	Number of cigarettes smoked – Wednesday
WP_Q10D	Number of cigarettes smoked – Thursday
WP_Q10E	Number of cigarettes smoked – Friday
WP_Q10F	Number of cigarettes smoked – Saturday
WP_Q10G	Number of cigarettes smoked – Sunday
SC_Q100	What was the main reason you began to smoke again?
WTPP	Person weight (survey weight variable)

Variable PROV : Province of the respondent

Values	Categories
10	N.L.
11	P.E.I.
12	Nova Scotia
13	N.B.
24	Quebec
35	Ontario
46	Manitoba
47	Saskatchewan
48	Alberta
59	B.C.

Variable HHSIZE : # of people in the household

Values	Categories
1	
2	
3	
4	
5	5 or more

Variable SC_Q100 : What was the main reason you began to smoke again?

Values	Categories
1	To control body weight
2	Stress, need to relax or to calm down
3	Boredom
4	Addiction / habit
5	Lack of support or information
6	Going out more (bars, parties)
7	Increased availability
8	No reason / felt like it
9	Family or friends smoke
10	Other
96	Valid skip
97	Don't know
98	Refusal
99	Not stated

Variable DVURBAN : Characteristic of community

Values	Categories
1	Urban
2	Rural
9	Not stated

Variable DVMARST : Grouped marital status of respondent

Values	Categories
1	Common-law/Married
2	Widow/Divorced/Separated
3	Single
9	Not stated

Variable PS_Q30 : Age smoked first cigarette

Variable PS_Q40 : Age begin smoking cigarettes daily

Variable HS_Q20 : # of people that smoke inside the home

Variable WP_Q10A : # of cigarettes smoked-Monday

Variable WP_Q10B : # of cigarettes smoked-Tuesday

Variable WP_Q10C : # of cigarettes smoked-Wednesday

Variable WP_Q10D : # of cigarettes smoked-Thursday

Variable WP_Q10E : # of cigarettes smoked-Friday

Variable WP_Q10F : # of cigarettes smoked-Saturday

Variable WP_Q10G : # of cigarettes smoked-Sunday

Values	Categories
96	Valid skip
97	Don't know
98	Refusal
99	Not stated

DATA ENTRY

Data analyzed in SPSS comes either from a data file or from manual entry.

OPENING A DATA FILE

In addition to files saved in SPSS format, you can open spreadsheet (Excel, Lotus 1-2-3), database (Access, dBASE), tab-delimited files and other types of ASCII text files without converting the files to an intermediate format or entering data definition information.

STEPS TO OPENING A SPSS DATA FILE

1. Click on File>Open>Data.
2. To view all files, in the Files of Type drop down menu select the All Files (*.*) option.
3. In the 'Open File' dialog box, select the file you want to open.
4. Click Open.

STEPS TO OPENING A TEXT DATA FILE

1. Click on Files>Open>Data.
2. To view all files, in the Files of Type drop down menu select the All Files (*.*) option.
3. In the 'Open File' dialog box, select the file you want to open.
4. Click Open.
5. The Text Import Wizard appears.



TEXT IMPORT WIZARD

In six steps, the text import wizard allows you to define how to read a text data file including tab-delimited, space-delimited, comma-delimited and fixed-field format files. For delimited files, you can also specify other characters as delimiters between values and you can specify multiple delimiters. You will find the detail six-step text import wizard process.

STEP 1 – TEXT IMPORT WIZARD

The text file is displayed in a preview window.

You can apply a predefined format (previously saved from the text wizard) or follow the steps in the Text Wizard to specify how the

Text Import Wizard - Step 1 of 6

Welcome to the text import wizard!
This wizard will help you read data from your text file and specify information about the variables.

Does your text file match a predefined format?
 Yes No

Text file: C:\Temp\Anxiety2.csv

	var1	var2	var3	var4
1				
2				
3				
4				

```
0 ..... 10 ..... 20 ..... 30 ..... 40 ..... 50 ..... 60  
1 subject,anxiety,tension,trial1,trial2,trial3,trial4  
2 1,1,1,18,14,12,6  
3 2,1,1,19,12,8,4  
4 3,1,1,14,10,6,2  
5 4,1,2,16,12,10,4
```

< Back Next > Finish Cancel Help

STEP 2 - TEXT IMPORT WIZARD

The Text Import Wizard needs to determine where the data for one variable ends and the data value for the next variable begins.

If the first row of the data file contains descriptive labels for each variable, you can use these labels as variable names.

Text Import Wizard - Step 2 of 6

How are your variables arranged?
 Delimited - Variables are delimited by a specific character (i.e., comma, tab).
 Fixed width - Variables are aligned in fixed width columns.

Are variable names included at the top of your file?
 Yes
 No

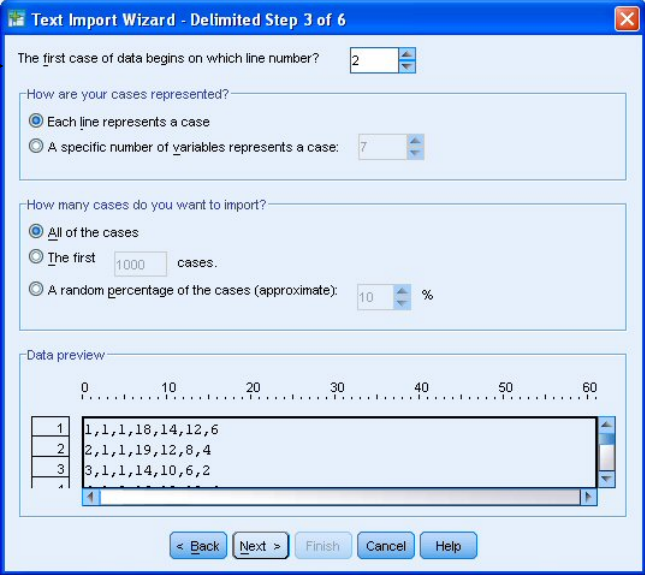
Text file: C:\Temp\Anxiety2.csv

```
0 ..... 10 ..... 20 ..... 30 ..... 40 ..... 50 ..... 60  
1 1,1,1,18,14,12,6  
2 2,1,1,19,12,8,4  
3 3,1,1,14,10,6,2  
4 4,1,2,16,12,10,4  
5 5,1,2,12,8,6,2
```

< Back Next > Finish Cancel Help

STEP 3 – TEXT IMPORT WIZARD (DELIMITED FILES)

Response indicates if the first row of the data file contains descriptive labels for each variable.

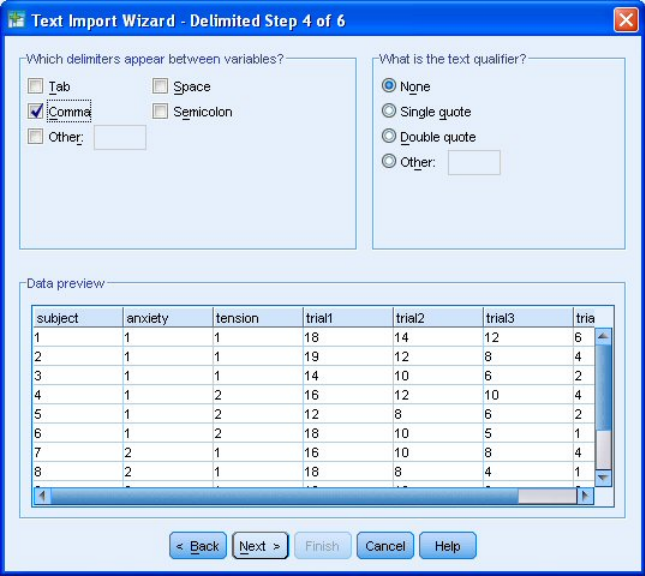


Controls how the Text Wizard determines where each case ends and the next one begins.

You can import all cases in the data file, first n cases, or a random sample of a specified percentage.

STEP 4 – TEXT IMPORT WIZARD (DELIMITED FILES)

Characters or symbols used to separate data values.



Characters used to enclose values that contain delimiter characters.

STEP 5 – TEXT IMPORT WIZARD

By clicking on the variable name in the Data preview window, you can change the variable name and data format.

Text Import Wizard - Step 5 of 6

Specifications for variable(s) selected in the data preview

Variable name: Original Name: subject

Data format:

Data preview

subject	anxiety	tension	trial1	trial2	trial3	trial4
1	1	1	18	14	12	6
2	1	1	19	12	8	4
3	1	1	14	10	6	2
4	1	2	16	12	10	4
5	1	2	12	8	6	2

< Back Next > Finish Cancel Help

STEP 6 – TEXT IMPORT WIZARD

Text Import Wizard - Step 6 of 6

You have successfully defined the format of your text file.

Would you like to save this file format for future use?

Yes No

Would you like to paste the syntax?

Yes No Cache data locally

Press the Finish button to complete the text import wizard.

Data preview

subject	anxiety	tension	trial1	trial2	trial3	trial4
1	1	1	18	14	12	6
2	1	1	19	12	8	4
3	1	1	14	10	6	2
4	1	2	16	12	10	4
5	1	2	12	8	6	2
6	1	2	18	10	5	1

< Back Next > Finish Cancel Help

You can save your specifications in a file for use when importing similar text data files.

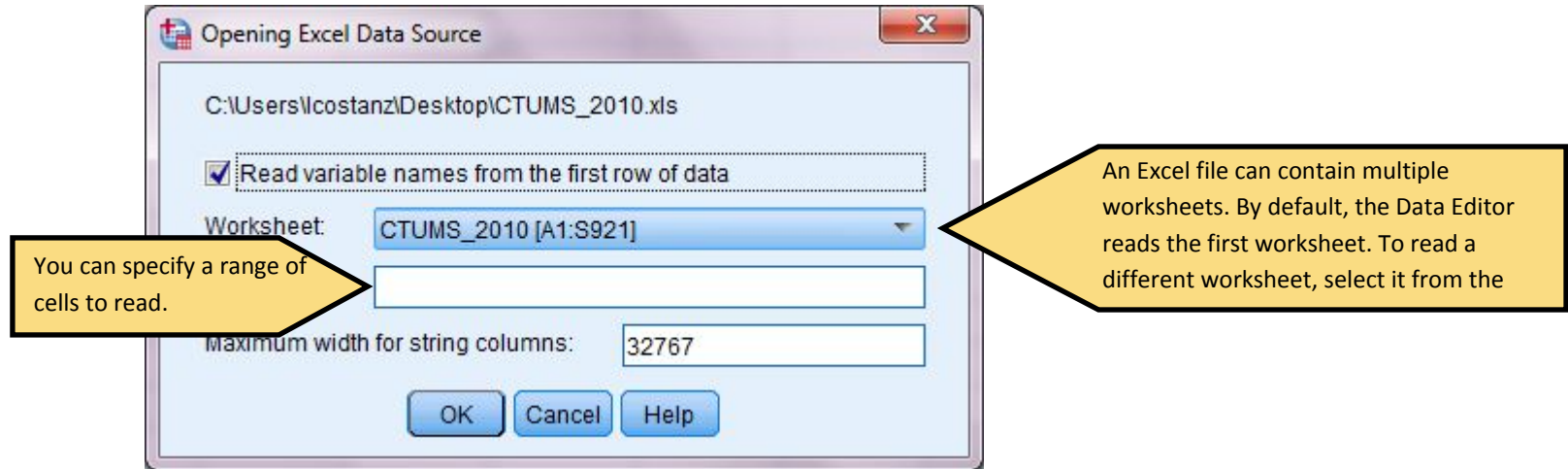
You can also paste the syntax generated by the Text wizard into a syntax window.

Exercise: Open a text data file

The previous SPSS data file was the incorrect file and your boss has sent you text file named 'btempht.dat'. First open the file in Notepad. How is the file organized? How is it delimited ... with spaces, tabs, commas? Open the text file following the steps outlined above.

STEPS TO OPENING AN EXCEL DATA FILE

1. Click on Files>Open>Data.
2. To view all files, in the Files of Type drop down menu select the All Files (*.*) option.
3. In the 'Open File' dialog box, select the file you want to open.
4. Click Open.
5. The following dialog box appears.



Exercise: Open an Excel file

Your supervisor sent you an Excel named 'CTUMS_2010.xls' to analyze. Follow the steps outlined about to open the file.

DATA EDITOR

The data editor provides a spreadsheet like method for creating and editing data files. It automatically opens when you start a session. Within the data editor, there are two views data and variable.

DATA VIEW

Data view displays the actual data values or defined values labels.

The screenshot shows the IBM SPSS Statistics Data Editor window. The main area displays a data grid with 16 rows and 8 columns. The columns are labeled PUMFID, PROV, DVURBAN, HHSIZE, HS_Q20, DVAGE, and SEX. The rows are numbered 1 through 16. A yellow callout box labeled "Column (Variables)" points to the column headers. Another yellow callout box labeled "Cell (Value)" points to the cell containing the value '2' in row 5, column DVAGE. A third yellow callout box labeled "Rows (Cases/Observations)" points to the row numbers. A fourth yellow callout box labeled "These tabs switch between Data and Variable views." points to the "Data View" and "Variable View" tabs at the bottom of the window. The status bar at the bottom indicates "IBM SPSS Statistics Processor is ready".

	PUMFID	PROV	DVURBAN	HHSIZE	HS_Q20	DVAGE	SEX
1	5	24	2	2	0	30	1
2	7	24	1	5	0	17	2
3	17	24	1	2	0	61	2
4	18	24	2	3	0	23	2
5	31	24	9	1	0	22	2
6	36	24	1	3	0	18	1
7	37	24	1	2	2	35	1
8	63	24	2	1	0	70	2
9	66	24	1	4	0	15	2
10	74	24	1	2	0	72	1
11	75	24	2	4	0	16	2
12	91	24	2	3	0	16	2
13	117	24	1	2	0	43	2
14							1
15							1
16	140					23	1



TIPS!

- Many of the features of the Data view are similar to those found in a spreadsheet but it does not have the same functionality. Unlike MS Excel, you can only work on only one data file at a time or an equation does not change the values of a group of cells.
- To see the value labels for variables in Data view either click on the View Labels button or from the menu bar click View>Value Labels.

VARIABLE VIEW

Variable view displays variable definition, information, including defined variable and value labels, data type (i.e. string, data and numeric), measurement level (nominal, ordinal or scale), and user-defined missing values.

The screenshot shows the Variable View window in IBM SPSS. The main table has columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The first row shows a variable named 'PUMFID' with a Numeric type, width of 11, and no decimals. The 'Values' column for this variable is set to 'None'. The 'Measure' column is set to 'Scale'. The 'Role' column is set to 'Input'.

Callouts explain the following fields:

- Variable name:** Variable name
- Number of characters allowed to be entered for the:** Number of characters allowed to be entered for the
- Data types of each variable (Numeric, Comma, Dot, etc.):** Data types of each variable (Numeric, Comma, Dot, etc.)
- Number of decimal places that SPSS will display:** Number of decimal places that SPSS will display
- Number of decimal places that SPSS will display:** Number of decimal places that SPSS will display
- Descriptive variable labels for each value of a variable. For example, codes 0 and 1 for male and female:** Descriptive variable labels for each value of a variable. For example, codes 0 and 1 for male and female
- Display width of column in Data View:** Display width of column in Data View
- Level of measurement (scale, nominal and ordinal):** Level of measurement (scale, nominal and ordinal)
- Specified user defined missing values:** Specified user defined missing values.
- Controls display of data values and/or value labels in the Data View:** Controls display of data values and/or value labels in the Data View
- Defines the role that a variable will play in the analysis:** Defines the role that a variable will play in the analysis

A red pushpin icon is located in the bottom left corner of the slide.

TIPS!

- Variable names must begin with a letter and be 64 digits long. Remaining characters can be any letter, any digit, a period or the symbol @, #, _ or \$ and cannot end with a period. Duplicate names are not allowed.
- To place line breaks in Labels, type \n in the place in the Label where you want the label to wrap.

ENTERING DATA

You can enter data directly into the Data Editor in the Data view. You can enter data in any order.

1. In the data view, select a cell.
2. Enter the data value. The value is displayed in the cell editor at the top of the Data Editor.
3. Click Enter or select another cell to record the value.



TIPS!

Data values are not recorded until you press Enter or select another cell.

Exercise: Entering data

As part of your company's research, a colleague designed and deployed a survey. Your part is to enter the responses from the three surveys below into SPSS.

- First, you will need to ensure the Data editor is clear by clicking on File>New>Data.
- Next, you will need to define the variables in Variable view.
- Then following the steps above, enter the survey responses using the Data view.

Respondent #1

1. What is your gender?
 Male Female
2. What is your age?
 under 18 years
 18 to 30 years
 31 to 40 years
 over 40 years
3. What is your income?
\$12,000
4. Comments?
I work part-time.

Respondent #3

1. What is your gender?
 Male Female
2. What is your age?
 under 18 years
 18 to 30 years
 31 to 40 years
 over 40 years
3. What is your income?
\$50,000
4. Comments?
I work full-time.

Respondent #4

1. What is your gender?
 Male Female
2. What is your age?
 under 18 years
 18 to 30 years
 31 to 40 years
 over 40 years
3. What is your income?
\$72,000
4. Comments?
I work three jobs.

INSERTING NEW VARIABLES

Entering data in an empty column in the Data view or in an empty row in the Variable view automatically creates a new variable with a default variable name (the prefix var and a sequential number) and a default variable name (numeric).

You can also insert new variables between existing variables following these steps:

1. Select any cell in the variable to the right of the (Data view) or below (Variable view) the position where you want to insert the new variable.
2. Click Data>Insert Variable.

DELETING VARIABLES

To delete a variable in the Data view (Variable view), right-click on the variable name in the column (row) header and select Delete.

INSERTING NEW CASES

Enter data in a cell on a blank row automatically creates a new case. The Data Editor inserts the system-missing value for all the other variables for that case.

Here are the steps to insert a new case between existing cases:

1. In the Data view, select any cell in the case (row) below the position where you want to insert the new case.
2. Click on Data>Insert Cases.

DELETING CASES

To delete a case in the Data view, right-click on the case number in the row header and select Delete.

MOVING TO A SPECIFIC CASE

Suppose you have a large data set with about 10,000 cases and wish to see case 3,452. You would either manually scroll down to case 3,452 or use 'Go to Case' option. Here are the steps to using the 'Go to Case' option:

1. Make the Data Editor window active.
2. Click on Edit>Go to Case...
3. Type in the case number or row number you wish to see.

Exercise: Inserting a new case

After you entered all the data, your colleague just gave you the survey for Respondent #2. Using the steps outlined above for inserting new cases, insert the Respondent #2 survey between case 1 and 2.

Respondent #2

1. What is your gender?

Male Female

2. What is your age?

under 18 years

18 to 30 years

31 to 40 years

over 40 years

3. What is your income?

\$22,000

4. Comments?

I work full -time.

SYNTAX FILE

Typically SPSS users read a data file into the Data Editor and then select items from the menus to manipulate the data or perform statistical analyses. This is referred to as *interactive mode* because the program provides a response each time you make a selection. For example if you request a transformation, the data set is immediately updated. If you select an analysis, the results immediately appear in the output window.

Alternatively, it is possible to work with SPSS in *syntax mode*, where the user types code in a syntax window. Once the full program is written, it is then submitted to SPSS to get the results. It is beneficial to have some programming knowledge when using syntax to produce the data transformations and analyses that you want. Generally, the interactive mode is easier and quicker if you only need to perform a few simple transformation or analyses on your data. Here are some situations when it would be better to use syntax over interactive mode:

- Procedures and options are only available in the interactive mode.
- Expect to perform the same procedures on several different data sets and would like to save a copy of the code so that it can easily be re-run.
- Need to perform a large number of similar transformations, such as using vectors and loops.
- Performing a very complicated set of procedures and it would be beneficial to document all of the steps leading to the results.

Each time a selection is made in interactive mode, SPSS records the syntax code in a “journal file”. The location and name of this file can be found/edited by clicking **Edit>Options** then click on the **File Locations** tab. If you ever want to see or use the code in the journal file, it can be edited in the journal file in a syntax window.

SPSS also provides an easy way to see the code corresponding to a particular menu function. Most selections include a **Paste** button that will open a syntax window containing the code for the function, including the details required for any specific options that you have chosen in the menus.

Exercise: Running a syntax file

The excel file containing the data has been imported into SPSS. Now, we will open the syntax file containing the commands which will define the value labels and missing values for each variable. To do this click on **File>Open>Syntax**, then look for the syntax file **Syntax1 . sps** and click on the “Open” button.



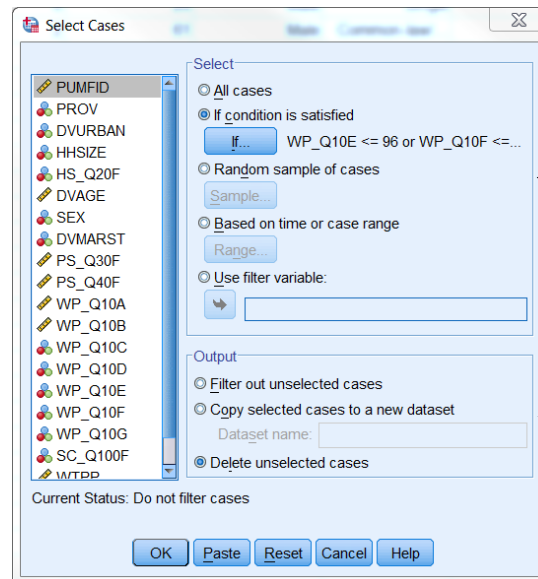
The Syntax Editor will open then select all the commands and click on the button.

SELECT CASES

The Select Cases command excludes from further analysis all those cases that do not meet specified selection criteria.

Steps on how to Select Cases

1. Click on Data > Select Cases.
2. The following dialog box will appear:



Select Cases provides several methods for selecting a subgroup of cases based on criteria that include variables and complex expressions. You can also select a random sample of cases.

Output

DATA TRANSFORMATIONS

COMPUTING VARIABLES

Compute Variable computes values based on numeric transformations of other variables.

- You can compute values for numeric or string variables
- You can create new variables or replace the values of existing variables. For new variables, you can also specify the variable type and label.
- You can compute values selectively for subsets of data based on logical conditions.
- You can use more than 70 built-in functions, including arithmetic functions, statistical functions, distribution functions, and string functions.

STEPS TO COMPUTE A VARIABLE

1. Click on Transform>Compute and the dialog box below appears

The screenshot shows the 'Compute Variable' dialog box in SPSS. It features a 'Target Variable' field, a 'Numeric Expression' field, a list of available variables on the left, a function keypad in the center, and a 'Function group' dropdown on the right. Callout boxes provide instructions: one points to the 'Type & Label...' button, another to the 'Target Variable' field, a third to the 'Numeric Expression' field, and a fourth to the 'Function group' dropdown. A fifth callout box points to the list of variables.

The type and label for the target variable can be modified by clicking this button.

Type the name of a single target variable. It can be an existing variable or a new variable to be added to the

Numeric Expression computes the value of the target variable. Just remember that string constants must be enclosed in quotation marks. Also, numeric constants must be typed in American format with the period (.) as the decimal indicator.

Available variables in working file.

Function List Over 70 built-in functions exist.

2. In the text box labelled 'Target Variable', type in a variable name where the calculation result is to appear. It can be an existing or new variable name. If you choose to type in an existing variable, the values for this variable will be overwritten.
3. To build an expression (calculation), either paste components or type directly into the 'Numeric Expression' field.
4. If you wish the calculation in question be only applied to certain subset of the data, the If function is invoked. For example, you may want to increase the salary for female by \$100. To do this click on the 'If' button located at the bottom of the Compute Variable window shown above. The dialog box below appears and type :

TIP!



Be careful what name you enter in the target variable field. If you enter in an existing variable name and go ahead with the transformation, the transformation cannot be undone, it is permanent.

Exercise: Computing Variable

In this dataset we have a variable which represent the number of cigarettes a respondent smoked each day of a particular week. Let's create a new variable (totcig) that represent the total number of cigarettes smoked by a respondent each week.

RECODING VALUES

You can modify data values by recoding them. This is particularly useful for collapsing or combining categories. You can recode values within existing variables or you can create new variables based on the recoded values of existing variables.

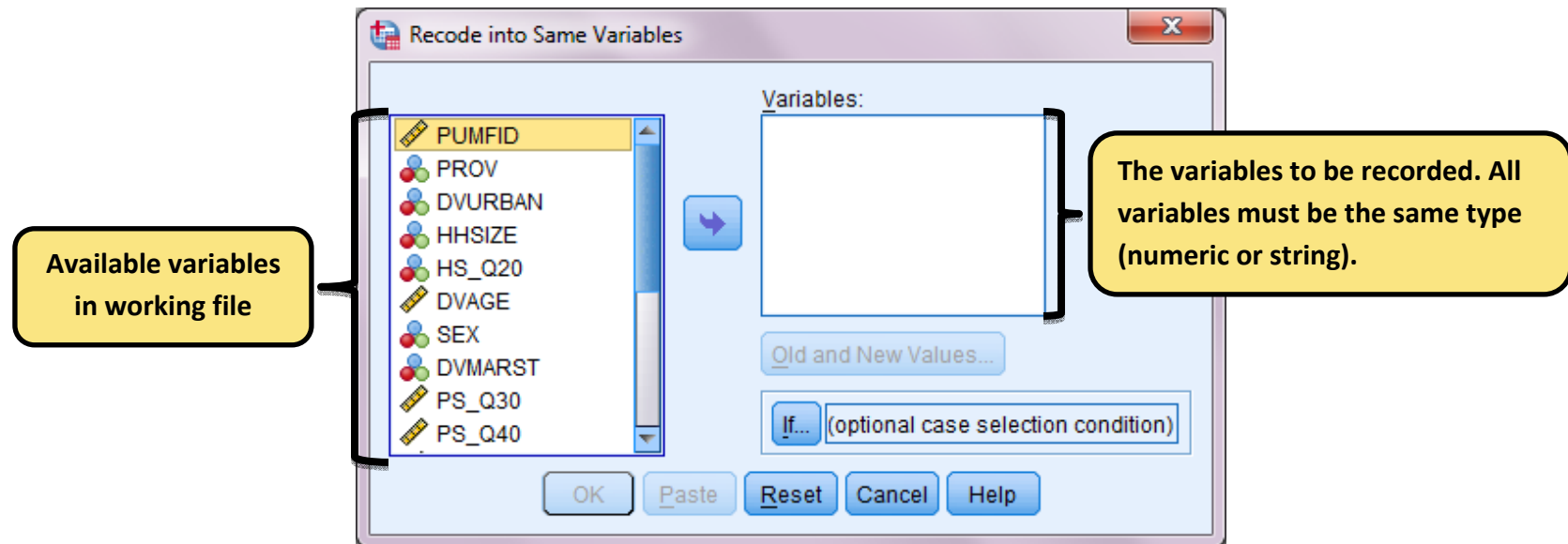
RECODE INTO SAME VARIABLES

Recode into Same Variables reassigns the values of existing variables or collapses ranges of existing values into new values. For example, you could collapse salaries into salary range categories.

You can recode numeric and string variables. If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

STEPS TO RECODE VALUES OF A VARIABLE

1. Click on Transform>Recode>Into Same Variables and the following dialog box appears.



2. Select the variables you want to recode. If you select multiple variables, they must be the same type (numeric or string).
3. Click on Old and New Values and specify how to recode values. The following dialog box appears.

Options to change old values.

Options to change new values.



TIP!

Be careful what name you enter in the target variable field. If you enter in an existing variable name and go ahead with the transformation, the transformation cannot be undone, it is permanent.

RECODE INTO DIFFERENT VARIABLES

Recode into Different Variables reassigns the values of existing variables or collapses ranges of existing values for a new variable. For example, you could collapse salaries into a new variable containing salaryrange categories.

- You can recode numeric and string variables
- You can recode numeric variables into string variables and vice versa.
- If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

STEPS TO RECODE VALUES OF A VARIABLE

1. Click on Transform>Recode>Into Same Variables, the following dialog box appears.

The screenshot shows the 'Recode into Different Variables' dialog box. On the left, a list of variables is shown, including HS_Q20, DVAGE, SEX, DVMARST, PS_Q30, PS_Q40, WP_Q10A, WP_Q10B, WP_Q10C, WP_Q10D, WP_Q10E, WP_Q10F, WP_Q10G, and WTPP. A callout points to this list, stating 'Available variables in working file'. In the center, a list of 'Numeric Variable -> Output Variable:' is shown, with 'SC_Q100 -> ?' selected. A callout points to this list, stating 'The variables to be recorded. All variables must be the same type. (numeric or string)'. On the right, the 'Output Variable' section has 'Name:' set to 'main_reason' and 'Label:' empty. A callout points to the 'Name:' field, stating 'Specifies the output variable that will contain the recoded values from the selected variable. It cannot exceed 8 characters.' Another callout points to the 'Label:' field, stating 'Descriptive name for the output variables.' Below the 'Output Variable' section is a 'Change' button. At the bottom, there is an 'Old and New Values...' button and an 'If...' (optional case selection condition) field. A callout points to the 'Old and New Values...' button, stating 'Specifies old values and the new values they will be recorded into.' At the bottom of the dialog are 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons.

2. Select the variables you want to recode. If you select multiple variables, they must be the same type (numeric or string).
3. Enter an output (new) variable name for each new variable and click Change.

4. Click on Old and New Values and specify how to recode values, the following dialog box appears.

Options to change old values.

Options to change new values.

Exercise: Recoding into Different Variables

Working with the same file, recode the age variable into the following groups:

Group	Ages
1	15-24
2	25-34
3	35-44
4	45-54
5	55-64
6	65-74
7	75-84
8	85 and over

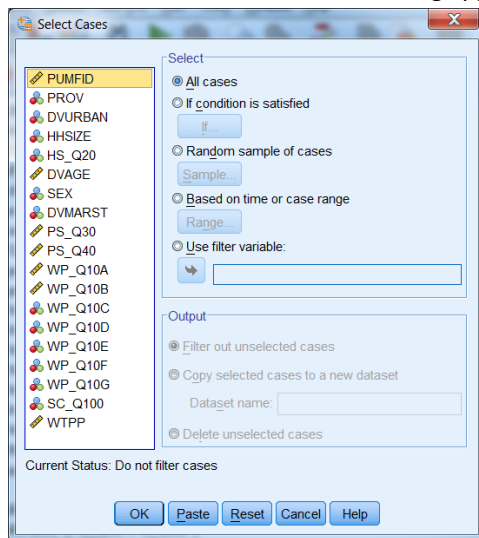
SELECT CASES

There are situations in which a researcher would like to select observations meeting certain conditions. The specified condition is an arithmetic or logical expression that generally consists of a sequence of operands and operators.

SPSS provides a function that allows us to isolate a sub-set of cases in a data file. It is the Select Cases command that excludes from further analysis all those cases that do not meet specified selection criteria. When cases do not meet the selection criteria, they will be turned 'off' or filtered, much like missing values so that they are not counted in further analysis. Alternatively, filtering is for cases to be deleted so that only those cases that meet the selection criteria remain in the data file. Filtering is the default setting, and the one generally used since deleting cases is a fairly drastic step and should only be used when we wish to trim down a very large file.

STEPS TO SELECT CASES

1. Click on Data>Select Cases and following appears.



Exercise: Select a subset of data

Create a subset of the data with only females aged 15-44 yrs of age.

ANALYZING DATA

DESCRIPTIVE STATISTICS

Descriptive statistics describe patterns and general trends in a data set. In most cases, descriptive statistics examine or explore only one variable at a time. However, correlation and regression analysis describe the relationship between two variables. Depending on the type of required descriptive statistics, SPSS offers four types of descriptive statistics procedures including Frequencies, Descriptives, Explore and Crosstabs.

FREQUENCIES

The frequencies procedure provides statistical and graphical displays that are useful for describing categorical variables which take on a value that is one of several possible categories.

STATISTICS AND PLOTS

Frequency counts, percentages, cumulative percentages, mean, median, mode, sum, standard deviation, variance, range, minimum and maximum values, standard error of the mean, skewness and kurtosis (both with standard errors), quartiles, user-specified percentiles, bar charts, and histograms.

DATA

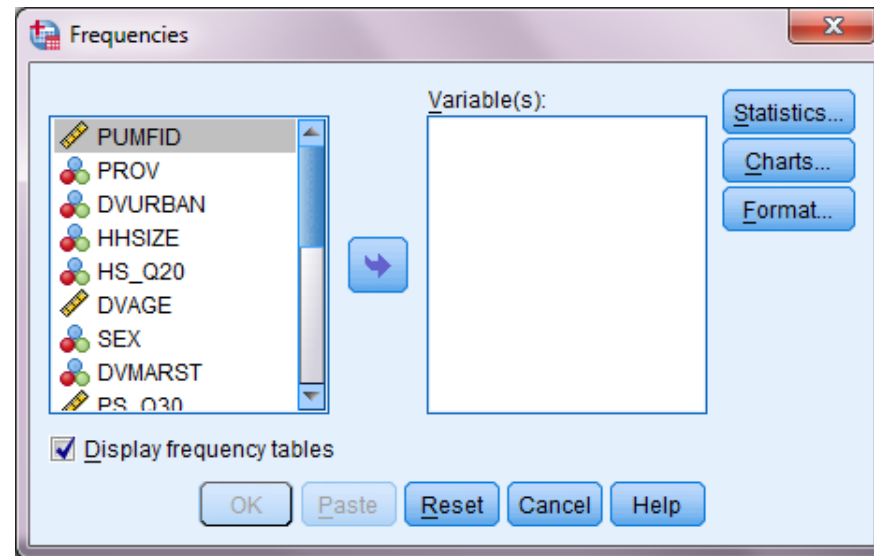
Use numeric codes or short strings to code categorical variables.

ASSUMPTIONS

The tabulations and percentages provide a useful description for data from any distribution, especially for variables with ordered or unordered categories. Most of the optional summary statistics, such as mean and standard deviation, are based on normal theory and are appropriate for quantitative variables with symmetric distributions. Robust statistics, such as the median, quartiles and percentiles are appropriate for quantitative variables that may or may not meet the assumption of normality.

STEPS TO OBTAIN FREQUENCY TABLES

1. Click Analyze>Descriptive Statistics>Frequencies, the below dialog box appears.



2. Select one or more categorical or quantitative variables. Optionally, you can:
 - Click Statistics for descriptive statistics for quantitative variables.
 - Click Charts for bar charts, pie charts, and histograms.
 - Click Format for the order in which results are displayed.

Exercise: Generating summary statistics for a categorical variable

Generate summary statistics for the SEX variable.

DESCRIPTIVES

The Descriptives procedure displays univariate summary statistics for several variables in a single table and calculates standardized values (z scores). It is useful when you want to quickly see summary statistics for interval or ratio variables, but not the frequency tables or graphics output.

STATISTICS

Sample size, mean, minimum, maximum, standard deviation, variance, range, sum, standard error of the mean and kurtosis and skewness with their standard errors.

DATA

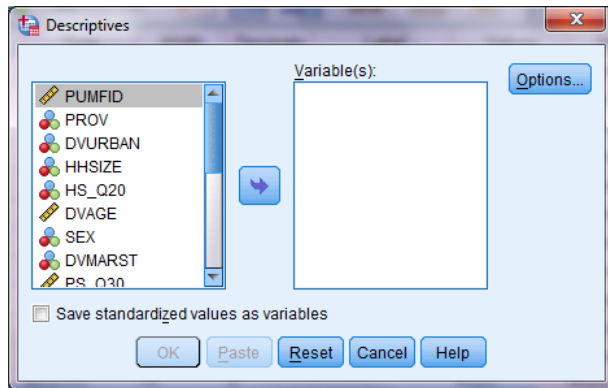
This procedure is best suited to describe continuous variables which can take on any value in a scale. Use numeric variables after you have screened them graphically for recording errors, outliers and distributional anomalies.

ASSUMPTIONS

Most of the available statistics (including z scores) are based on normal theory and are appropriate for quantitative variables (interval or ratio level measurements) with symmetric distributions (avoid variables with unordered or skewed distributions). The distribution of z scores has the same shape as that of the original data; therefore, calculating z scores is not a remedy for problem data.

STEPS TO OBTAIN DESCRIPTIVES

1. Click Analyze>Descriptive Statistics>Descriptives, the dialog box below appears.



2. Optionally, you can:
 - Select Save standardized values as variables to save z scores as new variables.
 - Click Options for optional statistics and display order.

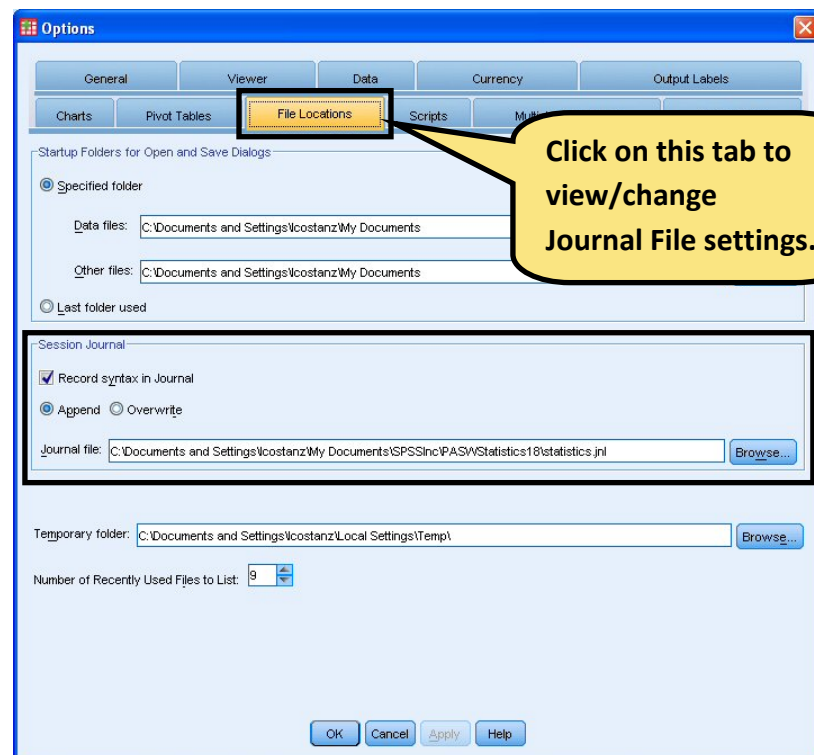
Exercise: Generating summary statistics for a categorical variable

Generate summary statistics for the `totcig` variable.

JOURNAL FILE

All commands issued either from a syntax window or from a dialog box are recorded in a journal file. This file can be reviewed with a text editor. This file also records any error or warning messages generated by commands.

To find the location of the journal file, click on Edit>Options then the following Options dialog box appears. Click on the File Location tab and the location of the Journal File is displayed. As well, the options to turn the journaling feature off or on can be specified. Along with specifying whether the journal file should be overwritten or appended when a new session begins. The default filename is spss.jnl.



This file is particularly useful if you previously ran a statistical analysis and forgot how it was done. You can quickly look at the journal, copy the syntax of interest and run the analysis again.

WORKING WITH THE OUTPUT

When you run a procedure, the results are displayed in a window called the Viewer. In this window, you can easily navigate to whichever part of the output that you wish to see. You can also manipulate the output and create a document that contains precisely the output that you want, arranged and formatted appropriately.

VIEWER

Results are displayed in the Viewer. You can use the viewer to:

- Browse results
- Show or hide selected tables and charts
- Change the display order of results by moving selected items
- Move items between the Viewer and other applications

The view is divided into two panes. The left pane of the Viewer contains an outline view of the contents. The right pane contains statistical tables, charts and text output.

Click to expand or collapse the outline.

Double click a book icon to show or hide an item.

Double click a table to pivot or edit it.

Outline Pane

Contents Pane

*Output1 [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
SEX

Frequencies

[DataSet1]

Statistics

SEX

N	Valid	920
	Missing	0
	Minimum	1
	Maximum	2

SEX

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	432	47.0	47.0	47.0
2	488	53.0	53.0	100.0
Total	920	100.0	100.0	

IBM SPSS Statistics Processor is ready

SAVING OUTPUT

The contents of the Viewer can be saved to a Viewer document. The saved document includes both panes of the Viewer window (the outline and the contents). The file will have the .spo extension and is only viewable using SPSS.

STEPS TO SAVING OUTPUT

1. Make the Viewer window active and click on File>Save
2. Enter the name of the document and click on Save. This file will have the .spn extension and can only be opened by SPSS for windows.

EXPORT OUTPUT

Export output pivot tables and text output in HTML, text, Word/RTF and Excel format. It saves charts in variety of common formats used by other applications.

Steps to Exporting Output

1. Make the Viewer the active window (Click anywhere in the window).
2. Click on File>Export.
3. Enter a filename.
4. Choose the Export Format file type you wish to save the file in.



TIP!

Clicking on File>Save, saves the file as a SPSS .spo which only SPSS can read. To ensure a computer without SPSS installed, export the output using the steps above.

PASTING OBJECTS FROM THE VIEWER INTO OTHER APPLICATIONS

Select an object and paste it into another application such as MS Word.

STEPS TO PASTING OBJECTS IN OTHER APPLICATIONS

1. Select the tables and/or charts to be copied (Shift-click or Ctrl-click to select multiple items).
2. Select Edit>Copy Objects.
3. In the target application, from the menus choose: Edit>Paste.

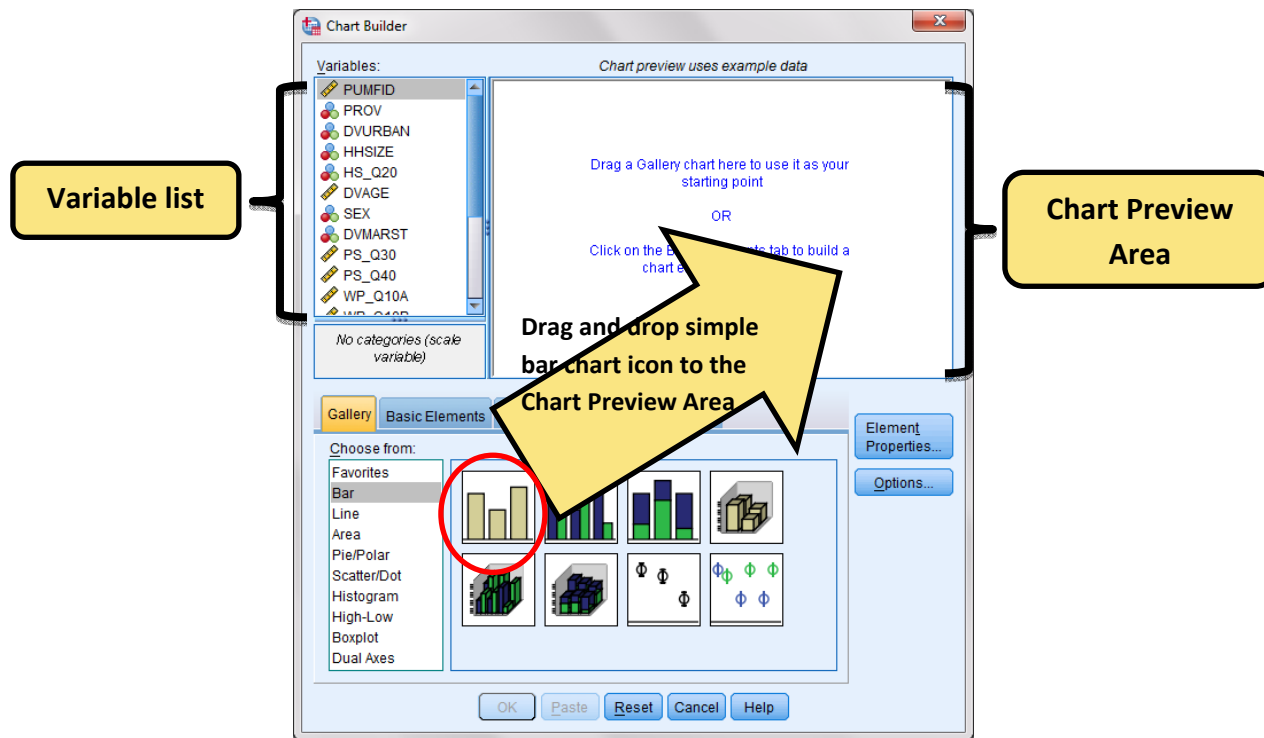
CREATING CHARTS

High resolution charts and plots are created by the procedures on the Graphs menu and by many of the procedures on the Analyze menu. The types of graphs available include Bar, Line, Ribbon, Drop Line, Area, Pie, Scatterplot, Dot Plots, Histograms, Boxplots and Dual Axis. For most of these graphs, it is possible to create a table of charts. This is known as paneling.

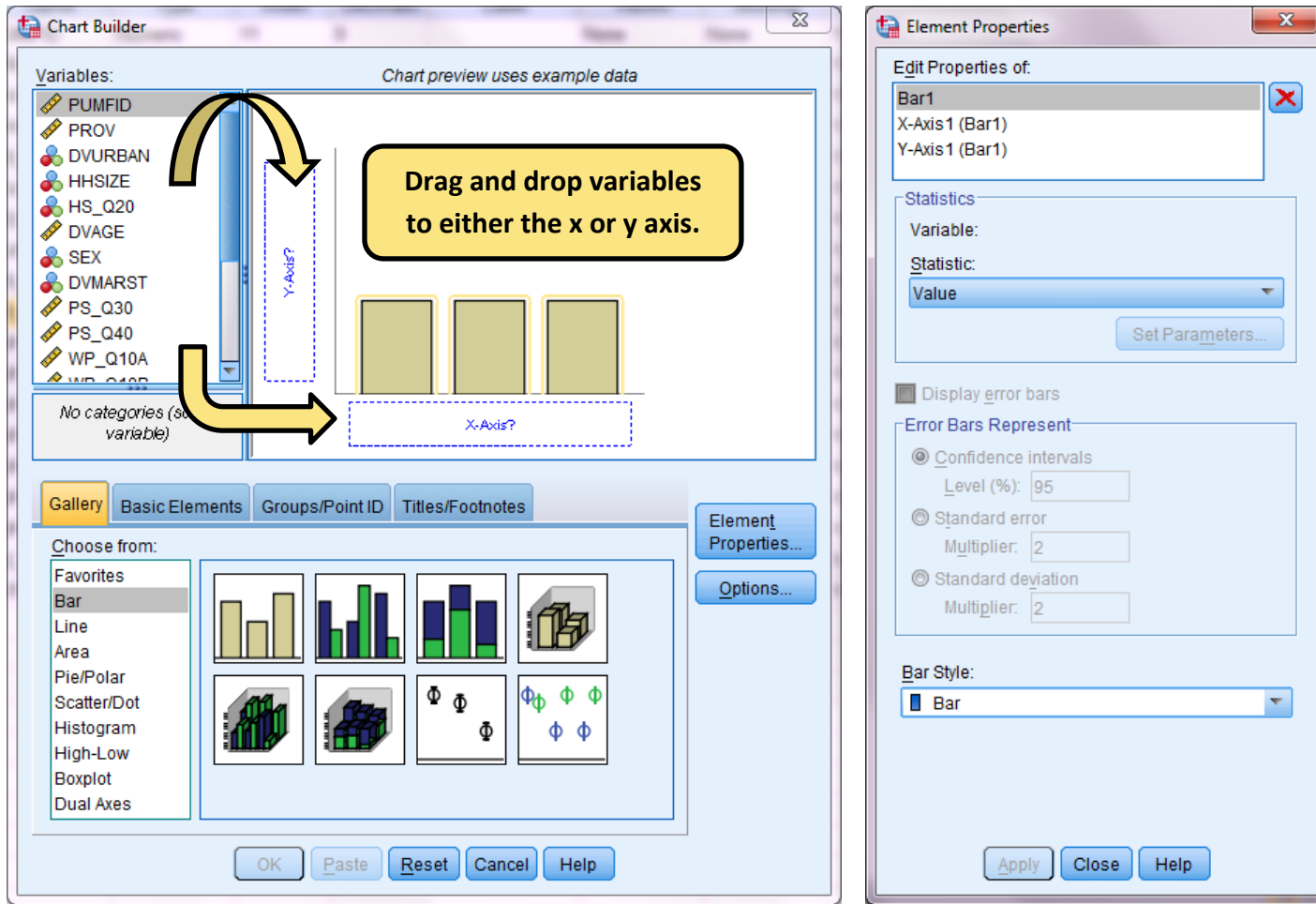
STEPS TO CREATING A BAR CHART

Before you can create a chart, you need to have your data in the Data Editor. You can enter the data directly into a Data Editor, open a previously saved data file, or read a spreadsheet, tab-delimited data file or database file. Follow these steps to create a chart:

1. Click on Graphs and then on Chart Builder. The following interactive dialog box will appear and allows you to preview the chart as you build it. To generate a bar graph, ensure the Gallery Tab is selected then click on Bar to the left. Then drag and drop the simple bar chart icon into the Chart Preview Area.



2. A template of a simple bar chart will be shown in the Chart Preview Area. the Element Properties window appears, it assigns the properties to each of the chart elements. Drag and drop the independent variable (SEX) from the variable list into the “X-Axis?” box in the Chart Preview Area. As well, drag and drop the dependent variable from (PS_Q40) the variable list into the “Y-Axis?” box. Under Statistics select Mean.



ADDING ERROR BARS

Errors bars allow you to graphically see the variability of the data. They are used on graphs to indicate the error or uncertainty in a reported measurement. They give a general idea of how accurate a measurement is, or conversely, how far from the reported value the true (error free) value might be.

In the Elements Properties Dialog box, check off the “Display Error Bars” box and click on the Apply button. The Chart Preview Area will now display the error bars.

The image shows two software dialog boxes side-by-side. The left dialog is titled "Chart Builder" and contains a list of variables on the left, a chart preview area in the center, and a gallery of chart types at the bottom. The right dialog is titled "Element Properties" and shows settings for a selected bar element, including a "Statistics" section with a "Display error bars" checkbox that is circled in red. A yellow callout box with a black border points to this checkbox with the text "Check off to display error bars".

Chart Builder

Variables:

- PS_Q30
- PS_Q40
- WP_Q10A
- WP_Q10B
- WP_Q10C
- WP_Q10D
- WP_Q10E
- WP_Q10F
- WP_Q10G
- SC_Q100
- WTPP

Chart preview uses example data

Mean PS_Q40

Category 1 Category 2 [More...]

PROV

No categories (scale variable)

Gallery Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

- Favorites
- Bar
- Line
- Area
- Pie/Polar
- Scatter/Dot
- Histogram
- High-Low
- Boxplot
- Dual Axes

Element Properties...

Options...

OK Paste Reset Cancel Help

Element Properties

Edit Properties of:

Bar1

X-Axis1 (Bar1)

Y-Axis1 (Bar1)

Statistics

Variable: PS_Q40

Statistic: Mean

Display error bars

Error Bars Represent

- Confidence intervals
- Level (%): 95
- Standard error
- Multiplier: 2
- Standard deviation
- Multiplier: 2

Bar Style:

Bar

Apply Cancel Help

Check off to display error bars

MODIFYING AXIS LABEL

The Axis Label can be edited by selecting the element in the “Edit Properties of” list box and then enter the text in the Axis Label text box.

The Chart Builder window displays a list of variables on the left, including PS_Q30, PS_Q40, WP_Q10A, WP_Q10B, WP_Q10C, WP_Q10D, WP_Q10E, WP_Q10F, WP_Q10G, SC_Q100, and WTPP. The chart preview shows a bar chart with three bars labeled Category 1, Category 2, and [More...]. The Y-axis is labeled Mean PS_Q40. The X-axis is labeled PROV. The chart preview uses example data.

The Element Properties window shows the X-Axis1 (Bar1) element selected. The Axis Label is set to Province. The Scale Range section shows the Variable set to PROV, with checkboxes for Automatic and Custom. The Scale Type section shows the Type set to Linear, with Base set to 10 and Exponent set to 0.5.

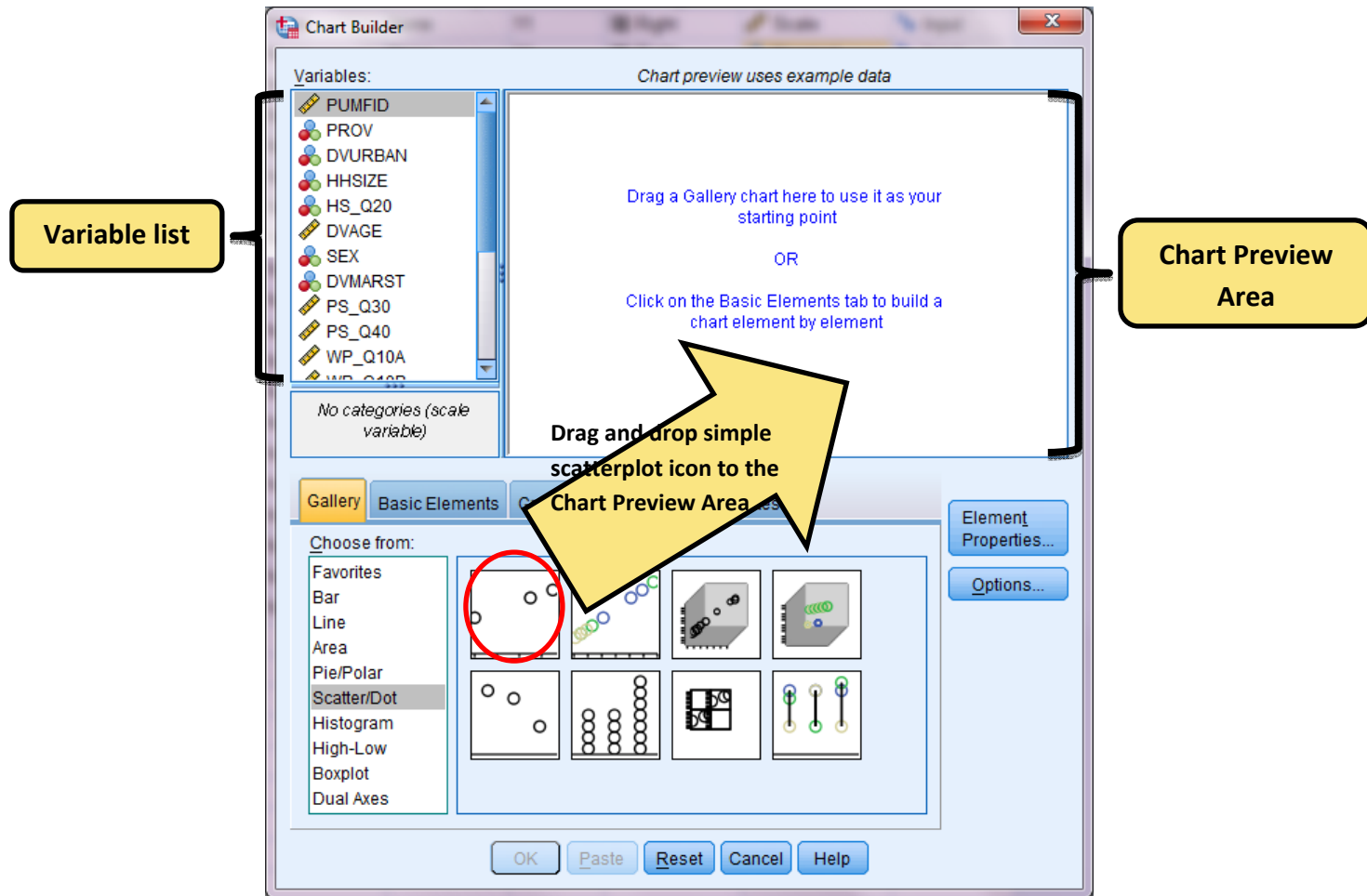
Select element to modify.

Edit text to appear in graph.

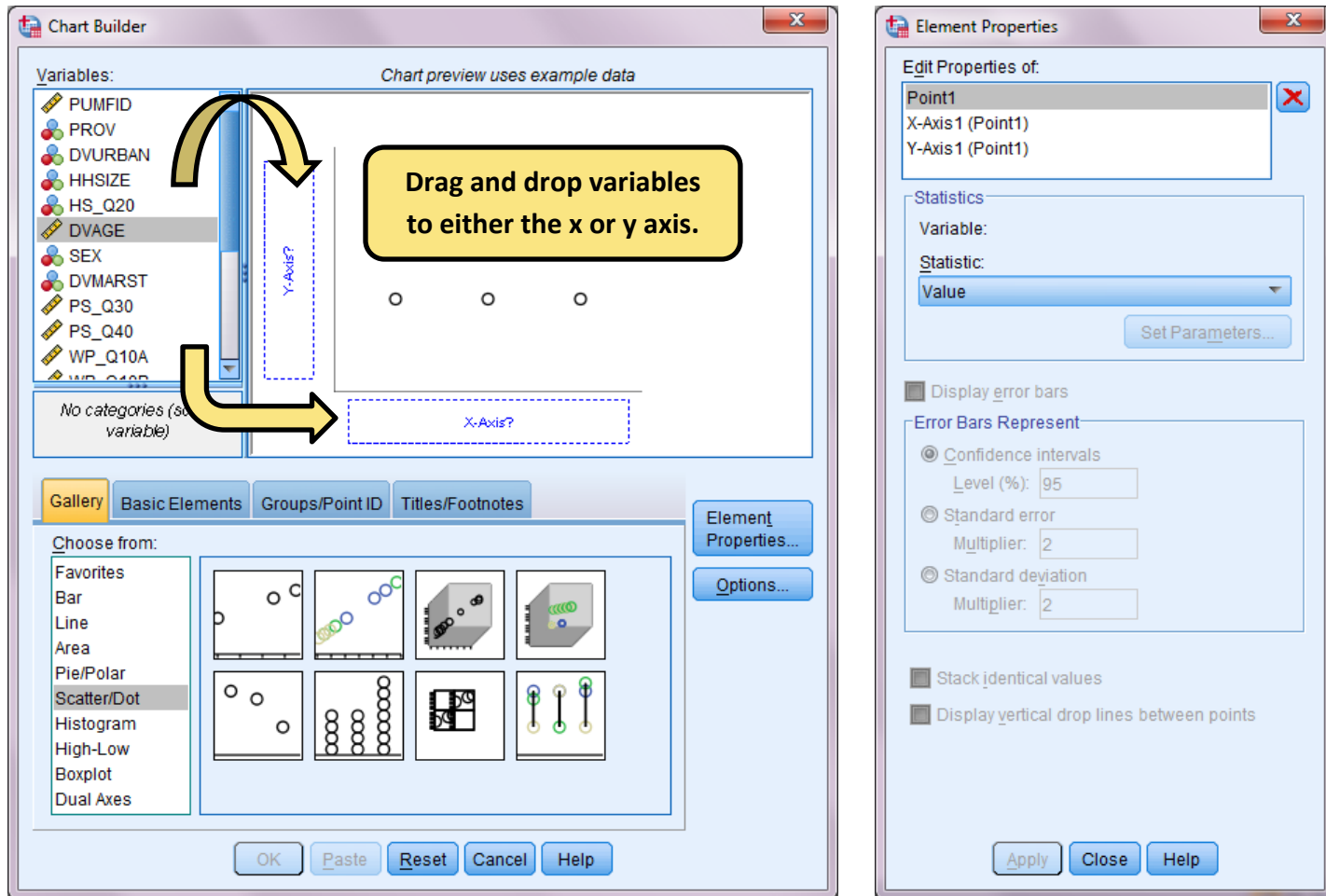
STEPS TO CREATING A SCATTERPLOT

Before you can create a chart, you need to have your data in the Data Editor. You can enter the data directly into a Data Editor, open a previously saved data file, or read a spreadsheet, tab-delimited data file or database file. Follow these steps to create a chart:

1. Click on Graphs and then on Chart Builder. The following interactive dialog box will appear and allows you to preview the chart as you build it. To generate a bar graph, ensure the Gallery Tab is selected then click on Bar to the left. Then drag and drop the simple scatterplot icon into the Chart Preview Area.



2. A template of a simple bar chart will be shown in the Chart Preview Area. the Element Properties window appears, it assigns the properties to each of the chart elements. Drag and drop the independent variable (DVAGE) from the variable list into the “X-Axis?” box in the Chart Preview Area. As well, drag and drop the dependent variable from (totcig) the variable list into the “Y-Axis?” box.



MODIFYING A CHART

To modify a chart, double-click anywhere on the chart that is displayed in the Viewer. This displays the chart in the Chart Editor. Any part of the chart can be modified or changed by clicking on the item of interest.

The most commonly edited part of a chart include the colour scheme, formatting number in the tick labels, editing text and data value labels.

CHANGING COLOURS

1. Double click on the bar chart to open it in the Chart Editor.
2. Click on any of the vars and rectangles will appear around the bars to indicate that they are selected.
3. On the Menu bar select Edit>Properties.
4. Choose the Fill & Border tab from the Properties Dialog Box
5. Click on the colour you wish.
6. Click Apply when the editing is complete.

FORMATTING NUMBERS IN TICK LABELS

1. If you are not in the Chart Editor mode then double click the bar chart.
2. Click on the y-axis tick labels.
3. On the Menu bar select Edit>Properties.
4. Click on the Number Format tab.
5. Make the changes according then click Apply when the editing is complete.

EDITING TEXT

1. To make any text changes, you must be in the Chart Editor mode.
2. Double-click on any of the titles you wish to edit. When in text editing mode, a flashing red bar cursor will appear.
3. Once the changes are complete press Enter to update the text.