https://udrive.oit.umass.edu/statdata/spss.zip

IBM SPSS Statistics Basics

October, 2011

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

1

I. Availability

SPSS Statistics for Windows is available to UMass Amherst campus and Medical School *faculty & staff* through an OIT site license. Licensing and CD distribution are handled in room A119 Lederle Graduate Research Center (LGRC). The license requires an initial fee at the time of purchase, and an annual fee, due each June. Call 545-9730 for pricing and purchasing information.

Students can purchase the SPSS Statistics Grad Pack at the Campus Center Bookstore. It is the same software, similarly priced. It comes with a four year license, but upgrades are not included. Do not confuse this with the Student Edition, which is a much more limited version.

If you have an OIT username and password you can use SPSS Statistics at no cost on OIT's public machines. SPSS Statistics for Windows can be used in the OIT PCCO classrooms and labs, including in LGRC and Library 7th floor.
See http://www.oit.umass.edu/computer-classrooms for other OIT lab locations and hours.

II. SPSS Statistics Versions

Unless otherwise stated, the instructions in this guide can be used with SPSS Statistics version 17 or higher. SPSS datafiles created with version 17 or higher can be used with earlier SPSS versions, as long as they don't use features introduced in version 17. In general, SPSS Viewer files are not backward compatible. SPSS Statistics can open Viewer files created with SPSS 16, but if you modify it using SPSS Statistics, it may not go back to SPSS 16. SPSS Output documents created with version 15 or lower can only be opened using the SPSS Legacy Viewer – not SPSS 16 or higher.

III. Documentation

The manuals listed below are on the distribution CD, or can be downloaded from http://www.umass.edu/statdata/software/news/spss/spss_install.htm.

IBM SPSS Statistics Brief Guide: A getting started manual with tutorial.

IBM SPSS Statistics Core System User's Guide - the basic manual for use with Windows. Working with the user interface, the Data Editor, Transformations, Charts, and many of the statistical procedures.

IBM SPSS Data Preparation

IBM SPSS Statistics Command Syntax Reference Guide describes the syntax for all commands included in your software.

IBM SPSS Statistics Regression – documents Logistic Regression , Probit Analysis, Nonlinear Regression, Two-Stage Least Squares Regression, and others.

IBM SPSS Statistics Advanced Models - documents GLM, Loglinear analysis, Life Tables, Kaplan-Meier Survival Analysis, Cox Regression Analysis, and others.

Additional manuals are also available for more specialized needs.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

2

IV.   Creating SPSS Statistics data files

## Using the SPSS Statistics Data Editor

You can manually enter data directly into SPSS Statistics by using the `Data Editor`. In the example below, we enter four pieces of information (name, birthday, height, gender) for each of four subjects. To open the `Data Editor` window, start up the SPSS Statistics program and Cancel the opening screen, which asks you what file to open.

Click the **`Variable View`** tab and define the variables to be entered.

Variable name	Type:	Width	Decimals	Values	Columns	Align
name	string	8	0		8	Left
birthday	date(mm/dd/yyyy)	10	0		10	Right
height	numeric	8	2		8	Right
gender	numeric	1	0	{1,male…}	8	Right

For Gender, click in the **Values** cell, then click on **…** to open the Value Labels dialog. Enter the codes and labels for gender: **1 for Male, 2 for Female**.

When all the variables are defined, click the **`Data View`** tab and enter the data:
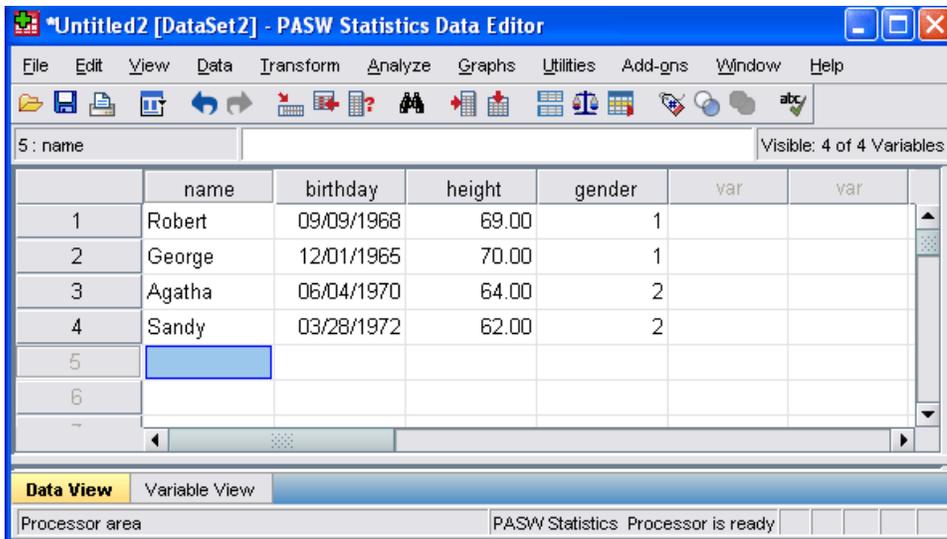
Click the **`Label`** icon on the tool bar (it looks like a luggage tag) to switch the display for Gender from numeric codes to labels.

Importing data from external files

Most of the time you will already have data in a file and just need to import it into SPSS Statistics. From the **File** menu, select **Open → Data.** Look in the **Files of Type** drop down box and select the appropriate file type. As you can see, SPSS Statistics can import data from the most commonly available spreadsheet applications, as well as text data files.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

3

`https://udrive.oit.umass.edu/statdata/spss.zip`

Text files:

We will be using the Minidat data file. If you do not have it, it is described, and can be downloaded from `http://www.umass.edu/statdata/statdata/data`. We will be using the Text version. To import it into SPSS Statistics, go to **File → Open → Data** and select files of type **Text (*.txt,*.dat).** Select the file `minidat.dat` from wherever you saved it. Click **Open**. This starts the Text Import Wizard. The dialog screens and responses are outlined here:

Screen	Question	Response
1	Does your text file match a predefined format?	no
2	How are your variables arranged?	Fixed width
	Are variable names included at top of your file?	no
3	First case of data begins on which line?	1
	How many lines represent a case?	1
	How many cases do you want to import?	All the cases
4	The vertical lines show where SPSS will divide the data into variables	Correct if needed. Otherwise click Next.
5	Specification for variable(s) selected in data preview	See below…

In the "Specification for variable(s) selected in the data preview" dialog, change the variable names from V1 – V8 to something more meaningful…
- Click on V1 column heading in the Data Preview area. The column heading (V1) appears in the "Variable Name" box.
- Change Variable Name V1 to **pulse1**. Do **NOT** click Next!
- Click on V2 column heading in the Data Preview area.
- Change variable name V2 to **pulse2**.
- Repeat for each column, assigning names **gp, smoke, gender, height, weight, activity** to the third through eighth columns.
- Click **Next**.

6	Would you like to save the file format?	No
	Would you like to paste the syntax?	No
		Click Finish.

Other File types:

For spreadsheet files, choose the appropriate file type (Excel, dBase, Lotus). Excel files must be version 5 or later. If your spreadsheet has a single row of column names in the first row, SPSS Statistics will make these your variable names. SPSS Statistics accepts long names provided they do not include blanks and other special characters. However, for ease of use, keep variable names reasonably brief. Variable names that do not meet SPSS Statistics standards are replaced by names of the form V001, V002, etc.

To import data from Access, dBase or Paradox files, choose **File – Open Database – New Query** and follow the Database Import Wizard. MS-Office 95 (or higher) includes ODBC drivers which make these file types available to other software, including SPSS Statistics. If the ODBC drivers were not loaded during Office setup, you will need to load them.

Biostatistics Consulting   4
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

Variable properties

Variable definitions are more meaningful if you assign labels to better describe the variable and how it's coded. If certain codes indicate missing data, you must specify this to SPSS Statistics, otherwise *all non-blank data values are used for computations*. You also need to make sure SPSS Statistics has assigned each variable the correct type and Measurement Level (Nominal, Ordinal or Scale). Correct Measurement Level assignment is important for generating appropriate graphics. Variable Labels and Value Labels are never required, but they make the output easier to understand.

For the `minidat.dat` data, these are the properties we want to assign:

Variable	Type	Measurement Level	Missing	Values
pulse1	Numeric	Scale	0	
pulse2	Numeric	Scale	0	
gp	Numeric	Nominal		1=run, 2=norun
smoke	Numeric	Nominal	9	1=yes, 2=no
gender	Numeric	Nominal		1=male, 2=female
height	Numeric	Scale		
weight	Numeric	Scale	0	
activ	Numeric	Ordinal	9	1=slight,2=moderate,3=very

From the **Data** menu, select **Define Variable Properties**. Select all variables and move them to Variables to Scan box. Click **Continue**. Click on a variable in the **Scanned Variable List**. Enter a Variable Label (optional), Value labels (if appropriate), mark its Missing values, and make sure the Measurement Level is correct. You can also set the number of decimal places to be displayed.

Click on each variable in turn and fill in its properties. When you've finished them all, click **OK**. Here is the completed dialog for variable **smoke**:
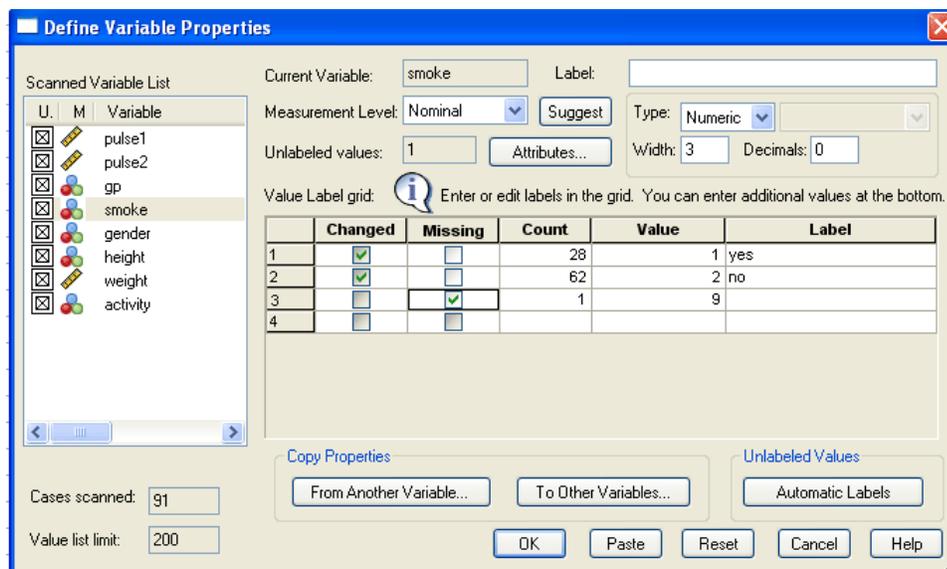
The icons next to each variable in the Scanned Variable list indicate its measurement level – a ruler for Scale, three bubbles for Nominal, and 3 bars for Ordinal. Observe in the picture above that height and activity need their measurement level adjusted.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

5

`https://udrive.oit.umass.edu/statdata/spss.zip`

After completing the Variable Properties dialog, click the **Variable View** tab in the Data Editor. The properties you entered for each variable should be reflected here. Variable properties can also be changed directly in the Variable View window.

V. Transforming Variables

"Transformations" change the values of existing variables, or create new variables based on existing variables.
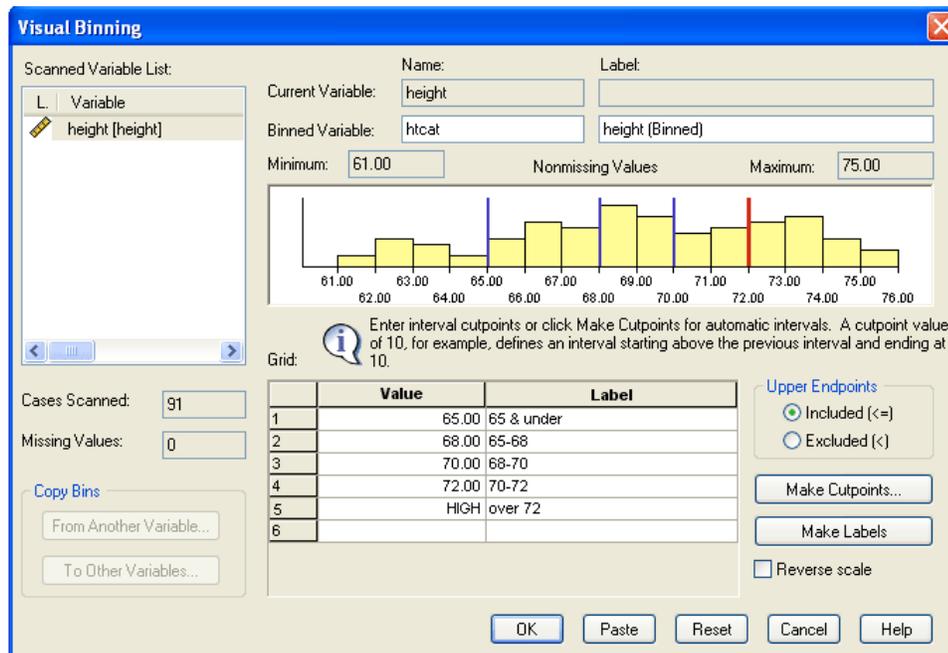
Recode

Recode is used to classify subjects into categories, or change existing categories. For example, let's group the students into five height categories: up to (and including) 65 inches; 65 to 68; 68-70; 70-72; over 72 inches. This can be done using the "Visual Binning" dialog, or using the Recode dialog. Visual Binning is more convenient because it enables you label the categories while you set them up.

Visual Binning:

```
From the Data Editor menu - Transform – Visual Binning
        Select: height
        Click: Continue
        On the Scanned Variable List, Select: Height
        Binned Variable: htcat
        Enter Values (one per row): 65,68,70,72, high
        Enter Labels (one per row): enter suitable value labels
Click OK
```

The completed dialog should look like this:

Click **OK.** The new variable, HTCAT appears on the right of the Data Editor sheet.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

6

Recode Dialog

The Recode dialog is another way to accomplish the same thing.

From the Data Editor menu - **Transform – Recode – Into Different Variables**
In the **Recode** dialog box:
        Select variable: **Height**
        Move it to the **Numeric Variable –> Output Variable** box
        Output variable: **Htcat**
        Label: **Height Category**
        Click: **Change**
NOTE: If you created HTCAT using Visual Binning, you will get a warning that variable HTCAT already exists. Click OK to overwrite HTCAT, or use a different output variable name to avoid the warning.

The dialog now looks like   this:

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

7

https://udrive.oit.umass.edu/statdata/spss.zip
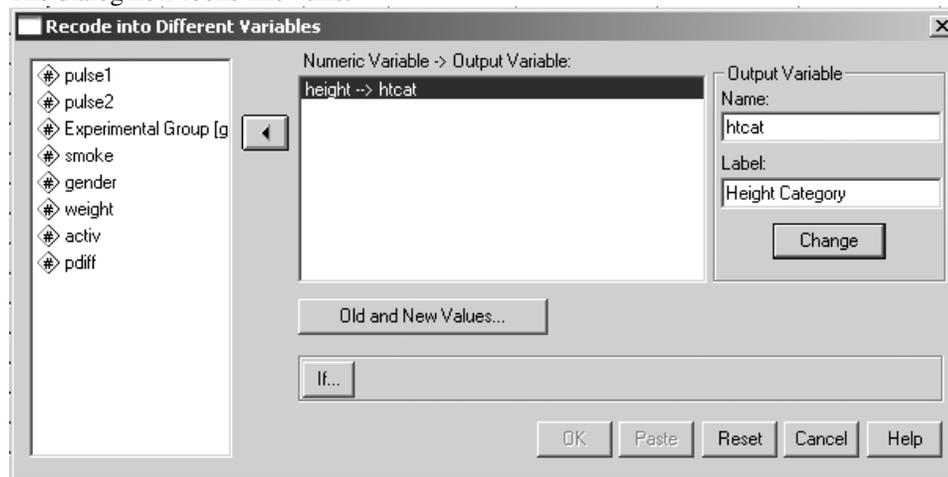
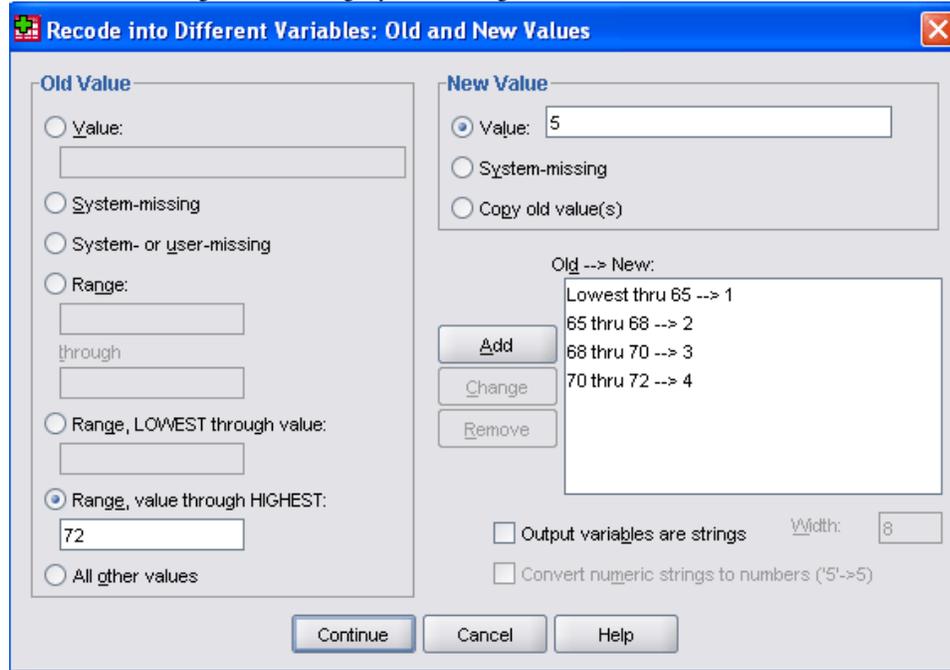
Click **Old and New Values**…
> Under **Old Value**, Range: `Lowest through 65`
> Under **New Value**, Value: `1`
> Click **Add**

Repeat for ranges 65-68→2; 68-70→3; 70-72→4; 72 through Highest→5.

Just before Adding the last category, the dialog should look like this:

Add the final category and click **Continue**, then click **OK**. The new variable, HTCAT appears on the right of the Data Editor sheet. Define Value Labels to make it more meaningful.

NOTE: Values listed in two categories (65, 68, 70 and 72) are classified into the first category in which they are mentioned.

Compute

Use Compute to create a new variable by calculating it from existing variables. For example, let's compute the difference in pulse rates between the first and second measurements for each student. We will call the new variable, *pdiff*. In the Data Editor window:

From the Data Editor menu - **Transform – Compute Variable**
In the **Compute Variable** dialog box:
> Target variable: **pdiff**
> Numerics Expression: **pulse2–pulse1**
Click **OK**

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

8

`https://udrive.oit.umass.edu/statdata/spss.zip`

VI.   Saving your SPSS Statistics data file

Save your completed data file as SPSS Statistics file:

```
From the menu – File – Save As
        Save in:  choose appropriate directory or drive
        Filename: minidat
        Save as type: SPSS Statistics(*.sav)
```

VII.   Quit SPSS Statistics

```
From the menu – File – Exit
```

VIII.   Using a saved SPSS Statistics data file

Double-click the saved SPSS Statistics file, `minidat.sav`. Or, start SPSS Statistics and **Open** the saved SPSS Statistics data file:

```
From the File menu –  Open – Data
        Look in:  choose appropriate directory or drive
        Filename: minidat
        File of type: SPSS Statistics(*.sav)
```

IX.   Descriptive Statistics

Once you have defined your data to SPSS Statistics you are ready to begin data analysis. We will discuss the most common descriptive procedures.

Frequencies

Tabulates the number of cases with each value of a variable. This is suitable for variables that have only a few possible values (categories).

```
From the menu – Analyze – Descriptive Statistics – Frequencies
From the Frequencies  dialog box:
        Select: gp, smoke, gender and activ
        (Optional) Click: charts
                From the charts  dialog box: select bar chart, then click Continue
        (Optional) Click: Statistics
                Select additional statistics you wish to see.
Click OK
```

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

9

https://udrive.oit.umass.edu/statdata/spss.zip

Crosstabulations

Tabulates the number of cases within each combination of two or more categorical variables.

```
From the menu – Analyze – Descriptive Statistics – Crosstabs
From the Crosstabs dialog box:
        Select gp, move to row variables
        Select smoke and active, move to column variables
        Click cells, put a check mark in Row Percents
        Click Continue
Click OK
```

Descriptives

Produces summary statistics (mean, standard deviation, minimum, maximum). This is suitable for variables with many possible values on a numeric scale (continuous/measurement variables).

```
From the menu – Analyze – Descriptive Statistics – Descriptives
From the Descriptives dialog box:
        Select: pulse1 pulse2 height weight
        (optional) Click: Options – select additional statistics to view
Click OK
```

Compare Means

For comparing means of groups defined by a categorical variable. The variables whose means you want to compare are the dependent variables. The variable that defines the groups is the independent.

```
From the menu – Analyze – Compare Means – Means
From the Means dialog box:
        Select pulse1 pulse2 pdiff, move to dependent list
        Select gp, move to independent list
Click OK
```

Two-sample T statistic

The t-test is used to test whether the means of two groups are significantly different. The groups are defined by the values of a categorical variable

```
From the menu – Analyze – Compare Means – Independent Samples T tests
From the Independent Samples T tests dialog box:
        Select pulse1 pulse2 pdiff, move to test variable(s)
        Select gp, move to Grouping variable
        Click: Define groups:
           Group 1: 1
           Group 2: 2
        Click: Continue
Click OK
```

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

10

https://udrive.oit.umass.edu/statdata/spss.zip

X.   Using the Output Viewer

Each time you run something that creates output, an entry is created in the `Output Viewer`'s left side, with subheadings for each section of the output. It is like a Table of Contents for your output.

Opening, Closing and Selecting Sections of Output

To go to any part of the output, click its entry in the `Output Viewer`'s Table of Contents. Your selection is highlighted in the Table of Contents, and the beginning of the selection is displayed in the main (right) section of the viewer.

You can make the output of a procedure invisible by clicking the minus sign to the left of the procedure entry in the `Viewer`. When you do this, the subheadings for the procedure are not listed in the `Viewer`'s Table of Contents, and the corresponding output is no longer visible in the output window. To open that section of output again, click the plus sign to the left of the procedure entry.

Printing Output

File/Print prints only what is visible in the output window. If you've turned off some part of the output (by clicking its minus sign in the `Viewer`), that section will not be printed. You can select a portion of the output to print by clicking on its icon; use **ctrl-click** to select several non-contiguous parts of the output. In the `Print` dialog, click **Selection**.

Saving & Restoring Output

You can save `Viewer Output` using **File/Save As** on the `Output` Window's menu. SPSS Statistics will assign the extension `.spv` to output files. You can open saved output in a future SPSS Statistics session using **File/Open/Output**. Select files of type *.spv. Any new procedure output from the new SPSS Statistics session will be appended to the previous output. Double-clicking an `.spv` file in Windows Explorer will start SPSS Statistics with the saved output file in the Viewer.

SPSS `Viewer` output files (*.spv) cannot be viewed or edited with text editors or word processors. However, you can Export Viewer output directly to Word, Excel, PowerPoint, HTML or Acrobat pdf.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

11

https://udrive.oit.umass.edu/statdata/spss.zip

XI.   Selecting Cases

You can select a subset of the cases to be included in an analysis. Suppose we want to see the means of *pulse1*, *pulse2* and *pdiff* for just the males. From the `Data Editor` window:

```
From the Menu – Data – Select cases
From the select cases dialog box:
        Check: if condition is satistfied
        Click: if
From select cases:IF dialog box:
        Type in: gender=1
        Click: continue
Click OK
```

The `Data View` shows a slash through the row number of the cases that are excluded. Now rerun the means comparison of *pulse1*, *pulse2* and *pdiff* for the two experimental groups. Double click the **Notes** preceding the last analysis (on the left side of the `Output Viewer`) to display information about the selection in effect.

The selection stays in effect until you turn it off. To return to all cases:

```
From the Menu – Data – Select cases
From the select cases dialog box:
        Check: All cases
Click OK
```

XII.   Split File

If you want to run the same analysis on subsets of the cases formed by the value of some variable, the easiest way is to use split file processing. For example, if we wanted to compare the difference in pulse rates of the two experimental groups for males and females separately, we could do it using case selection, or using split file processing. Using case selection, we would have to select the males, and run the means comparison, as above. Then we would have to select the females and run the means comparison again. Using Split File processing, we can get the same results in one step. <u>Make sure all cases are selected, as shown above.</u> From the `Data Editor` Window:

```
From the Menu – Data – Split File
From the Split File dialog box:
        Click: Compare groups
        Select gender, move to Groups based:
Click OK
```

Rerun the means comparison of *pulse1*, *pulse2*, *pdiff* for the experiment groups. The output is now separated according to gender. Split File processing stays in effect until you turn it off.

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

12

```
https://udrive.oit.umass.edu/statdata/spss.zip
```

```
From the Menu – Data – Split File
From the Split file dialog box:
        Click: Analyze all cases, do not create groups
Click OK
```

XIII.   Graphs

Under the Graph menu you will find:

- "Chart Builder", a drag-and-drop interface for defining your chart. In order to use Chart Builder effectively, it is important that the Measurement Levels of all variables be correctly defined.

- "Legacy Dialogs" refer to the old GRAPH procedures, which have been in SPSS since time immemorial, but will likely disappear in some future release.

You may need to experiment to find out which procedure is most useful in any given situation.

Creating Graphs

Scatter plots show the relationship between two variables. In our example, we would like to see the relationship between the two pulse rates. However, since pulse rate should be affected by the experimental condition, we want the plot to distinguish the two groups.

Using Chart Builder

```
From the Graph Menu – Chart Builder
If you have not already set Measurement Levels properly, click Define Variables and make sure
all variables are properly defined.
In the Gallery Tab, Select Scatter/Dot:
        move the pointer over each template to see its identification.
        Drag Grouped Scatter template (1st row, 2nd item) to the Preview area
        Drag Pulse2 to the Y-Axis box
        Drag Pulse1 to the X-Axis box
        Drag Gp to the Set Color box
Click OK
```

Using Legacy Dialogs

```
From the Graph Menu – Legacy Dialogs – Scatter/Dot
From the Scatter plot dialog box:
        Click: Simple
        Click: Define
        Select Y-axis: pulse2
        Select X-axis: pulse1
        Set Markers by: gp
Click OK
```

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

13

`https://udrive.oit.umass.edu/statdata/spss.zip`

Editing Graphs

Double-click on the graph in the `Output Viewer` window to open it for editing. You can change the location, font or text of the axes, legend or title, the frequency of the tick marks, the style or color of the markers, and many other features. You can also add elements, such as regression lines and confidence intervals, where appropriate.

We will add a title, change the range and ticks on the x-axis, and change the marker shape and color for the "No Run" group.

Using Chart Builder or Graph

Double click on the graph. The `Chart Editor` opens in a separate window, with its own menu. The Chart Editor is the same for graphs created with Chart Builder and with Legacy Dialogs

Adjusting the Chart Size and Position:

Before you can add a title you need to make room for it by making the chart smaller.

Click in the chart somewhere between the axis and the axis label (pulse1 or pulse2). A blue border should now surround the chart *excluding* the legend.

Adjusting Chart Size: Move the cursor over the "handles" in the corners of the blue border, where it makes a diagonal double-headed arrow. Drag to make the chart smaller, leaving space for a title to be added above the chart.

Adjusting Chart Position: move the cursor over the blue border around the chart. When it forms a 4-headed arrow, drag it to adjust the chart's position.

Adding a title:

From the `Chart Editor Menu` – **Options – Text Box**
    (You may need to move the Properties Box out of the way to see the new Text box.)
    Click in Textbox, change it to: **Pulse2 versus Pulse1**
    Optionally, use the **blue border** around the title to adjust its size and position.
    Optionally, use the **Properties Box** to change the font style or size.
    If you used the Properties Box to make changes, Click **Apply**

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

14

`https://udrive.oit.umass.edu/statdata/spss.zip`

Changing the Axis Range and Tick Marks:

> From the `Chart Editor Menu` – **Edit – Select X-Axis**
> > `OR`, **double-click on the x-axis.**
> Under the **Scale** tab:
> > Next to `Minimum`: **Uncheck Auto**, enter **40**
> > Next to `Maxiumu`: **Uncheck Auto**, enter **120**
> > Next to Major Increment : **Uncheck Auto**, enter **20**
> Under the **Labels and Ticks** tab:
> > Under Minor Ticks: Check **Display Ticks**
> > Number of Minor Ticks per Increment: **1**
> > Click **Apply**

Changing Marker Colors or Styles:

> In the `Chart Editor`: **Double-Click the 'No Run' marker in the legend**
> > (all green markers are now selected)
> Under the **Marker Tab**:
> > Click Type: **Select a Triangle**
> > In the Color area: **Click fill**
> > Click: **Dark Green swatch**
> > Click: **Apply**
> **Close** the Properties Box
> **Close** the Chart Editor

XIV. Online Help

Under the `Help` menu, you find several choices:

- **Topics** can be used to look up how to do specific tasks, using either the menus or syntax. See example below.

- **Command Syntax Reference –** this is the complete reference manual. It is indispensable for advanced users.

- **Tutorial**- gives helpful hints on using the SPSS Statistics interface.

- **SPSS Inc. Home** – Opens your browser to the SPSS website.

- **Statistics Coach** – asks you a series of questions about what you want to do and chooses the analysis based on your responses. BEWARE – if you don't know enough about statistics to answer the questions correctly, you will not get the right advice!

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

15

https://udrive.oit.umass.edu/statdata/spss.zip

XV.   Help - Topics

Topics is useful for finding instructions for specific tasks. Contents is most useful if you don't know the "magic" word to search the Index.   For example, if you want to find out how to get averages for two groups:

From the `Menu` – **Help – Topics**
      Under the **Contents** tab, Double click **Base System**
      Scrolling through the topics, Analyzing Data seems the most likely suspect.
      Double-click: **Analyzing Data**
      Double-click: **Means**
      If the text describes what you want to do, click: **Show Me**

Organization of Help topics may differ in different releases.

XVI.   Logging your work

The commands associated with everything you do in your SPSS Statistics session are recorded in the output window. These commands, called the log, are a useful document of what you did. You can turn logging on or off:

From the Menu – **Edit – Options**
      Click: **Viewer** tab
      Check or Uncheck: **Display commands in the log**
Click **OK**

The SPSS Statistics journal records all your commands for the current session in an external file called `statistics.jnl`. By default the file is in append mode which means it will quickly become very large and difficult to sort through.

To find out (or change) where the journal is saved, or to change the default to overwrite:

From the Menu – **Edit – Options**
From the **Options** dialog box:
      Click: **File Locations** tab
In the **Session journal** area:
      Select: **Append** or **Overwrite**
      The name and location of the journal is shown. To change it, click: **Browse**
      Click: **apply**
Click **OK**

Biostatistics Consulting
University of Massachusetts School of Public Health
C:\Word\documentation\SPSS\SPSS Basic.doc

16