Statistical Package for Social Science (SPSS)

Introduction

SPSS 19
Workshop October 9th 2011

The American University in Cairo
Department of University Academic Computing Technologies (UACT)
Table of contents

1- Introducing the interface

2- The variable view

3- The Data View

4- Statistical Example:

5- The Output View

6- Import an excel worksheet

7- The syntax View
SPSS Introduction

SPSS (Statistical Package for the Social Sciences) has now been in development for more than thirty years. Originally developed as a programming language for conducting statistical analysis, it has grown into a complex and powerful application with now uses both a graphical and a syntactical interface and provides dozens of functions for managing, analyzing, and presenting data.

1 Introducing the interface

When you use SPSS, you work in one of several windows: the data view, the variable view, the output view, the draft output view. Eventually you’ll also use the syntax editor (think: code) to save or refine your queries.

The data view

The data view displays your actual data and any new variables you have created.
The variable view

At the bottom of the data window, you’ll notice a tab labeled Variable View. The variable view window contains the definitions of each variable in your data set, including its name, type, label, size, alignment, and other information.
The output view

The output window is where you see the results of your various queries such as frequency distributions, cross-tabs, statistical tests, and charts. If you’ve worked with Excel, you’re probably used to seeing all your work on one page, charts, data, and calculations. In SPSS, each window handles a separate task. The output window is where you see your results.

The syntax view

SPSS has never lost its roots as a programming language. Although most of your daily work will be done using the graphical interface, from time to time you’ll want to make sure that you can exactly reproduce the steps involved in arriving at certain conclusions. In other words, you’ll want to replicate your analysis. The best method of preserving the exact steps of a particular analysis is the syntax view.
2 Creating the data definitions: the variable view

It’s impossible to talk about SPSS (or any analysis program) without talking about data and types of data. Each particular type of information (such as income or gender or temperature or dosage) is called a variable. Thus, Variable View contains descriptions of the attributes of each variable in the data file.

In Variable View:

- Rows are variables.
- Columns are variable attributes.

You can add or delete variables and modify attributes of variables, including the following attributes:

- Variable name
- Data type
- Number of digits or characters
- Number of decimal places
- Descriptive variable and value labels
- User-defined missing values
- Column width
- Measurement level

1) Variable type

Variable Type specifies the data type for each variable. By default, all new variables are assumed to be numeric. You can use Variable Type to change the data type. The contents of the Variable Type dialog box depend on the selected data type. For some data types, there are text boxes for width and number of decimals; for other data types, you can simply select a format from a scrollable list of examples.

The available data types are as follows:

**Numeric**: A variable whose values are numbers. Values are displayed in standard numeric format.
Comma: A numeric variable whose values are displayed with commas delimiting every three places, and with the period as a decimal delimiter.

Dot: A numeric variable whose values are displayed with periods delimiting every three places, and with the comma as a decimal delimiter.

Scientific notation: A numeric variable whose values are displayed with an embedded E and a signed power-of-ten exponent. The exponent can be preceded either by E or D with an optional sign, or by the sign alone—for example, 123, 1.23E2, 1.23D2, 1.23E+2, and even 1.23+2.

Date: A numeric variable whose values are displayed in one of several calendar date or clock-time formats. Select a format from the list. You can enter dates with slashes, hyphens, periods, commas, or blank spaces as delimiters.

Dollar: A numeric variable displayed with a leading dollar sign ($), commas delimiting every three places, and a period as the decimal delimiter. You can enter data values with or without the leading dollar sign.

Custom currency: A numeric variable whose values are displayed in one of the custom currency formats that you have defined in the Currency tab of the Options dialog box.

String: Values of a string variable are not numeric, and hence not used in calculations. They can contain any characters up to the defined length. Uppercase and lowercase letters are considered distinct.

In SPSS Variable names consists of the following rules.

• Names must begin with a letter.
• Names must not end with a period.
• Names must be no longer than eight characters.
• Names cannot contain blanks or special characters.
• Names must be unique.
• Names are not case sensitive. It doesn’t matter if you call your variable CLIENT, client, or CliENt. It’s all client to SPSS.
2) Variable label

Defining a label for a variable makes output easier to read but does not have any effect on the actual analysis. For example, the label "Family Identification Number" is easier to understand than the name of the variable, fam_id.

In order to make SPSS display the labels, go to View -> Value Label or Click on the Value Label icon located in the toolbar.

3) Missing value declaration

Missing Values defines specified data values as user-missing. For example, you might want to distinguish between data that are missing because a respondent refused to answer and data that are missing because the question didn't apply to that respondent. Data values that are specified as user-missing are flagged for special treatment and are excluded from most calculations.

• User-missing value specifications are saved with the data file. You do not need to redefine user-missing values each time you open the data file.

• You can enter up to three discrete (individual) missing values, a range of missing values, or a range plus one discrete value.

• Ranges can be specified only for numeric variables.
• All string values, including null or blank values, are considered to be valid unless you explicitly define them as missing.

• Missing values for string variables cannot exceed eight bytes. (There is no limit on the defined width of the string variable, but defined missing values cannot exceed eight bytes.)

• To define null or blank values as missing for a string variable, enter a single space in one of the fields under the Discrete missing values selection.

Click in the appropriate cell in the Missing column. A button with 3 dots will appear to the right of the word "None". Click on it:

4) Column format

Assist in improving the on-screen viewing of data by using appropriate column sizes (width) and displaying appropriate decimal places. It does not affect or change the actual stored values.

5) Value labels

Similar to variable labels. Whereas "variable" labels define the label to use instead of the name of the variable in output, "value" labels enable the use of labels instead of values for specific values of a variable, thereby improving the quality of output. For example, for the variable gender, the labels "Male" and "Female" are easier to understand than "0" or "1". In effect, using value labels indicates to SPSS that: "When I am using the variable gender, in any and all output tables and charts produced, use the label "Male" instead of the value "0" and the label "Female" instead of the value "1"."
The easiest way to create or modify value labels is under the Variable View tab:

Enter a value and a label.
Click on the Add button.
When you are done, click on OK.
You can return here in the future and change value labels or remove them.
6) Measurement level

You can specify the level of measurement as scale (numeric data on an interval or ratio scale), ordinal, or nominal. Nominal and ordinal data can be either string (alphanumeric) or numeric.

- **Nominal**: A variable can be treated as nominal when its values represent categories with no intrinsic ranking (for example, the department of the company in which an employee works). Examples of nominal variables include region, zip code, and religious affiliation.

- **Ordinal**: A variable can be treated as ordinal when its values represent categories with some intrinsic ranking (for example, levels of service satisfaction from highly dissatisfied to highly satisfied). Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores.

- **Scale**: A variable can be treated as scale (continuous) when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate. Examples of scale variables include age in years and income in thousands of dollars.

<table>
<thead>
<tr>
<th>Label</th>
<th>Values</th>
<th>Missing</th>
<th>Columns</th>
<th>Align</th>
<th>Measure</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Name</td>
<td>None</td>
<td>None</td>
<td>8</td>
<td>Right</td>
<td>Nominal</td>
<td>Input</td>
</tr>
</tbody>
</table>

3  The Data View

The Data Editor provides a convenient, spreadsheet-like method for creating and editing data files. The Data Editor window opens automatically when you start a session. The Data Editor displays the contents of the active data file. The information in the Data Editor consists of variables and cases.

- In Data View, columns represent variables, and rows represent cases (observations).

In Data View, you can enter data directly in the Data Editor. You can enter data in any order. You can enter data by case or by variable, for selected areas or for individual cells.

- The active cell is highlighted.

- The variable name and row number of the active cell are displayed in the top left corner of the Data Editor.
When you select a cell and enter a data value, the value is displayed in the cell editor at the top of the Data Editor.

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

- To enter anything other than simple numeric data, you must define the variable type first.

If you enter a value in an empty column, the Data Editor automatically creates a new variable and assigns a variable name.

4 Statistical Example:

4.1 Crosstab Report

Use this procedure when you want to look at 2-way frequencies of your categorical data. Specify a row variable and a column variable.

Click on Analyze ->
Descriptive Statistics ->
Crosstab
Click on the Statistics button if you want Chi Squares or other statistics computed.

The report will display how many females and males are in each department.

Click on OK when you are ready to generate your results.
Crosstabs

[DataSet4] F:\New folder (3)\Employee data (1).sav

Case Processing Summary

<table>
<thead>
<tr>
<th>Cases</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valid</td>
<td>Missing</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
<td>Percent</td>
</tr>
<tr>
<td>Employment Category * Gender</td>
<td>474</td>
<td>100.0%</td>
<td>0</td>
<td>.0%</td>
</tr>
<tr>
<td></td>
<td>474</td>
<td>100.0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Employment Category * Gender Crosstabulation

Count

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
</tr>
<tr>
<td>Employment Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerical</td>
<td>206</td>
<td>157</td>
<td>363</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Custodial</td>
<td>0</td>
<td>27</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manager</td>
<td>10</td>
<td>74</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>216</td>
<td>258</td>
<td>474</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5 The Output View

When you run procedures from the Analyze or Graphs menu, you will automatically be taken to the SPSS Output Viewer. In the left frame, you will see a list of various procedures with their subordinate objects. More recent results appear at the bottom. You can use this left frame to navigate to previous results: click on the object name on the left and it will appear in the frame on the right, with a thin black indicator box and a red arrow.

![SPSS Output Viewer](image)

**Frequencies**

- **Frequencies VARIABLES=gender jobcat /ORDER=ANALYSIS.**

**Statistics**

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Employment Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Valid</td>
<td>474</td>
</tr>
<tr>
<td></td>
<td>MISSING</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>474</td>
</tr>
</tbody>
</table>

**Frequency Table**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>Female</td>
<td>216</td>
<td>45.6</td>
<td>45.6</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>258</td>
<td>54.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>474</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**Employment Category**

<table>
<thead>
<tr>
<th>Employment Category</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>Clerical</td>
<td>363</td>
<td>78.8</td>
<td>78.8</td>
</tr>
<tr>
<td></td>
<td>Custodial</td>
<td>27</td>
<td>5.7</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>Manager</td>
<td>84</td>
<td>17.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>474</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>
1) To collapse a procedure, click on the - sign to the left of its name.

2) To see the list of its objects under a procedure, click on the + sign to the left of its name.

3) To delete a portion of your results, click on the procedure name on the left and press the Delete key.

4) To see the associated notes for a procedure, double click on the Notes object name listed beneath a procedure.

5) To rename an item on the left, click in the box and type.

6  **Import an excel worksheet**

   Ideally, the worksheet should have the variable names in the first row. You may want to insert or edit them in the Excel file ahead of time.

   1. Go to the File menu and select Open > Data.

   2. Change the location in the "Look in" box to the subdirectory where your file is.

   3. Change the "Files of type" selection to look for Excel (*.xls) files.
4. Select the file.

5. You might get prompted about the variable names:

6. Click on OK. You will see the data appear in the Data Editor window.
7. You may need to modify some of the variable definitions (Variable View)

7- The syntax view

The Syntax Editor provides an environment specifically designed for creating, editing, and running command syntax. The Syntax Editor features:

**Auto-Completion**: As you type, you can select commands, subcommands, keywords, and keyword values from a context-sensitive list. You can choose to be prompted automatically with the list or display the list on demand.

**Color Coding**: Recognized elements of command syntax (commands, subcommands, keywords, and keyword values) are color coded so, at a glance, you can spot unrecognized terms. Also, a number of common syntactical errors—such as unmatched quotes—are color coded for quick identification.

**Breakpoints**: You can stop execution of command syntax at specified points, allowing you to inspect the data or output before proceeding.

**Bookmarks**: You can set bookmarks that allow you to quickly navigate large command syntax files.

**Auto-Indentation**: You can automatically format your syntax with an indentation style similar to syntax pasted from a dialog box.

**Step Through**: You can step through command syntax one command at a time, advancing to the next command with a single click.
The Syntax Editor window is divided into four areas:

• The editor pane is the main part of the Syntax Editor window and is where you enter and edit command syntax.

• The gutter is adjacent to the editor pane and displays information such as line numbers and breakpoint positions.

• The navigation pane is to the left of the gutter and editor pane and displays a list of all commands in the Syntax Editor window and provides single click navigation to any command.

• The error pane is below the editor pane and displays runtime errors.

Set the language of SPSS to include Arabic

Type in the syntax editor (the window that will open) this code `set locale=Arabic` then from the menu bar choose run --> all
## Reference

An icon next to each variable provides information about data type and level of measurement.

<table>
<thead>
<tr>
<th>Measurement Level</th>
<th>Numeric</th>
<th>String</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (Continuous)</td>
<td>![Icon]</td>
<td>n/a</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>Ordinal</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>Nominal</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
</tbody>
</table>

![Diagram of SPSS interface with icons for data types and measurement levels]