# Introduction to SPSS Version 10.0

Written by Ta Liu

Center for Social Science Computation & Research 145 Savery Hall University of Washington Seattle WA 98195 U.S.A. (206)543-8110

> October 2000 http://julius.csscr.washington.edu/pdf/spss10.pdf

## **INTRODUCTION TO SPSS 10.0**

### **Overview of SPSS.**

SPSS stands for Statistical Package for the Social Sciences. It is general statistical software tailored to the needs of social scientists and the general public. Compared to other software, it is more intuitive and easier to learn; the trade-off is less flexibility and fewer options in advanced statistics than some other statistical software like S-Plus and SAS.

SPSS is good for organizing and analyzing data. You can rearrange data, calculate new data and conduct a variety of statistical analyses. Theoretically, there is no limit to the size of data files, so you can work on large data files in SPSS when you can not do so in Excel. This version also allows easy input/output management, such as exchanging files with other software, changing the appearance of output, or cutting and pasting into different programs.

The best way to learn how to use SPSS is to work with it. A quick way to get familiar with major features is to run through the online tutorial under the Help menu.

SPSS is installed on every computer in CSSCR.

#### Log in to SPSS.

There are two ways to launch the SPSS program. One is to simply click on the SPSS icon shown in red letters on your desktop. If you cannot find the icon, you can click Start, then Program Files, then SPSS. Or if you are not sure whether the computer you are using has SPSS , click Start, then Find, then Files or Folders, then type "SPSS."

When the SPSS window launches, a dialogue box pops up. You have several choices; you can either start a tutorial, type in new data, or open an existing file.

#### Input Data.

If you want to start from scratch and enter data manually in SPSS, select the Type in Data option from the Open dialogue box. A blank window with a spreadsheet appears. You can click on any cell and enter numbers. If you want to enter characters, you need to define the variables as a string first. It is recommended that you define the variables first even if they contain numbers. Now let's assume you wish to open an Excel spreadsheet. You can click File, then Open, then Data (File/Open/Data). A dialogue box should appear. You need to do two things to open your file. First, you need to locate the directory of your file. In this example, it is in C:/Temp. Click the "Look in" directory at the top of the dialogue box, browse to select C:/Temp. Then choose the correct file type, which is Excel. Find the file "World95.xls," then click Open. Another dialogue box pops up. It asks whether you want to read the first line as variable names. Click to check the box and OK. You should have a window filled with data.

Open File				? ×
Look jn:	🔄 temp 🔽	£	<b>r</b>	<b></b>
ing al Dave msoclip1 spss199 tl Vbe	Morld95.xls			
File <u>n</u> ame:	World95.xls			<u>O</u> pen
Files of <u>type</u> :	Excel (*.xls)	-		<u>P</u> aste
				Cancel

For earlier versions of SPSS (9.0 or earlier), it is a little tricky because SPSS then could only read Excel 4.0 or earlier. If you work with Office 97 or 2000, you need to save in Excel 4.0 in order to export to SPSS 9.0. You should also only keep one worksheet in the Excel file when you save it. Version 10.0 can request a specific worksheet from an Excel workbook; so now you can open Excel 2000 workbooks.

Similarly, you can also open a Text file. The Open box for a Text file looks different from the Open box for an Excel file. You need to specify how the data is separated and which part of the data you want to read. A wizard helps you through the process.

To save your file in SPSS format, ending with an extension of .sav. Go to File, then Save As. Choose "SPSS data" and save.

Before we move on to data analysis, let's first look at the basic structure of SPSS.

#### **Basic Structure of SPSS**

Unlike commonly-used Microsoft Office applications, such as Word and Excel, SPSS has many windows. It can be quite confusing in the beginning. You will get used to it as you work along.

The spreadsheet window is called the **Data Editor**. You can also open an output viewer, syntax editor and script editor window from the "File" menu through "New" or "Open." Later we will activate the Chart Editor and Pivot Table Editor.

The most important windows are the Data Editor and the SPSS Viewer. The Data Editor displays data and allows data manipulation and analysis. The **SPSS Viewer** displays output and keeps log of changes in the program. The **Syntax** window displays the command instructions; it helps keep track of analysis and perform automated tasks. The **Chart Editor** and **Pivot Table Editor** are for editing charts and tables. The **Script Editor** is mostly for making specialized formats of table output. The Syntax window and Script Editor are for experienced users.

You can toggle between these windows by clicking on the taskbar at the bottom of your screen or by selecting a window from the Window pull-down menu at the top.

🛗 Untitle	d - SPSS Da	ta Editor						_ 🗆 ×	
<u>F</u> ile <u>E</u> dit	∐iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	nalyze <u>G</u> raph	s <u>U</u> tilities <u>W</u>	indow <u>H</u> elp				
<b>B</b>	🥶 💻 🖻		🗠 💽 🏄	▲ 🗄	1 <b>1 1</b>	0			
1:									
	var	var	var	var	var	var	var	var 🔺	
1									
2									
3									
4									
5								<b></b>	
▲ ▶ \Da	Data View / Variable View /  SPSS Processor is ready								

The pull-down (or drop-down) menus at the top of your screen are similar to Office applications. There are common categories like File, Edit, View, Windows and Help, which you can figure out by looking at the names. Specific to SPSS are four categories, Data, Transform, Analyze (Statistics in earlier versions) and Graphs in the Data Editor Window. It's slightly changed in the Output Viewer window; Data and Transform are replaced by Insert and Format. Some of the commands have icons on the toolbar which provide shortcuts. If you place the cursor on an icon, its name should appear.

📅 Output1 - SPSS Viewer 📃	Ι×
<u>File E</u> dit <u>V</u> iew Insert Format <u>Analyze G</u> raphs <u>U</u> tilities <u>W</u> indow <u>H</u> elp	
····→E Output	
PSSS Processor is ready	_//.

#### Data editor

The Data Editor is a spreadsheet like Excel consisting of columns and rows. Rows are also called cases or records. Columns are also called fields or variables. Looking across a row, a country in this example, you see the values of all variables. Looking down a column, you see values of all cases, in this illustration, all countries.

📷 World	195.sav - SPSS Data	a Editor							_ 🗆 ×	
<u>File E</u> di	rile Edit View Data Transform Analyze Graphs Utilities Window Help									
28										
1 : countr	у	Austria								
	country	popultn	urban	religion	lifeexpf	literacy	pop_incr	babymort	gdp_c:	
1	Austria	8000.00	58.0	Catholic	79.0	99.0	.20	6.70	18396	
2	Bosnia	4600.00	36.0	Muslim	78.0	86.0	.70	12.70	3098	
3	Colombia	35600.00	70.0	Catholic	75.0	87.0	2.00	28.00	1538	
4	Croatia	4900.00	51.0	Catholic	77.0	97.0	10	8.70	5487	
5	i Egypt	60000.00	44.0	Muslim	63.0	48.0	1.95	76.40	748	
E	i Estonia	1600.00	72.0	Protstnt	76.0	99.0	.52	19.00	6000	
7	' Ethiopia	55200.00	12.0	Muslim	54.0	24.0	3.10	110.00	122	
8	) Georgia	5500.00	56.0	Orthodox	76.0	99.0	.80	23.00	4500	
9	Japan	125500.0	77.0	Buddhist	82.0	99.0	.30	4.40	19860	
10	) Lithuania	3800.00	69.0	Catholic	77.0	99.0	.30	17.00	671C 🖵	
	ata View 🖌 Variable \	/iew /			•				<u>▶</u>	
				SPSS Pro	cessor is ready				1.	

If you are familiar with Excel, the first thing you will notice is that variable names in SPSS appear on the top gray rows. They can only have eight characters and the first character must not be a number. You can use underscore or upper case to have abbreviated names. For example, population change as PopChg or pop\_chg. It may be hard to read but you can include variable labels which add meaning to your names.

A very nice improvement in Version 10.0 is the addition of a "variable view" window in the Data Editor. It makes the tasks of managing variables and changing variable characteristics much easier. You can change variable names, define types, control width and decimal points, add or change labels, and change alignment.

🛗 World	95.sav - SPS	S Data Editor						_ 🗆 ×		
<u>F</u> ile <u>E</u> dit	∐view <u>D</u> ata	<u>T</u> ransform <u>A</u> naly	ze <u>G</u> raphs <u>U</u> tilit	ies <u>W</u> indow <u>H</u> e	lp					
28										
	Name	Туре	Width	Decimals	Label	Values	Missing	Colur		
1	country	String	12	0		None	None	12		
2	popultn	Numeric	6	2		None	None	8		
3	urban	Numeric	2	1		None	None	8		
4	religion	String	8	0		None	None	8		
5	lifeexpf	Numeric	2	1		None	None	8		
6	literacy	Numeric	2	1		None	None	8		
7	pop_incr	Numeric	4	2		None	None	8		
8	babymort	Numeric	4	2		None	None	8		
9	gdp_cap	Numeric	5	2		None	None	8		
10	log_gdp	Numeric	16	2		None	None	8		
11										
	ata View <b>)</b> Va	risble Yiew								
		TIDIC VIEW /		SPSS Proces	sor is ready					

In older versions of SPSS, you can work on variables by clicking Define Variables under Data OR double-click on a variable name. Now you can work directly in the variable view window. For example, we can change variable types. There are two common types of variables, numeric and string. Numeric variables have numbers, and string variables have characters. The default type is numeric. You can change variable labels; variable labels make reading output much easier since the variable name is limited to eight characters. String variables cannot be used in calculations and statistical analysis. You must recode them into numeric variables (or categorical variables) with assigned values. For example, religion can be displayed as 1 to 5, each corresponding to a type of religion.

Not only variables have labels, but values of categorical variables can also have labels.

#### Data Manipulation.

Data manipulation changes the layout of data and does not change its values. All data manipulation commands are listed under the Data pull-down menu. Commands like insert data, insert variable and merge file are related to adding new data. Commands like select cases and sort cases change the arrangement of data or sample it. You can choose cases according to certain requirements using select cases.

🛗 World	95. sav	- SPSS Data	Editor								_ 🗆 ×
<u>F</u> ile <u>E</u> dit	<u>File Edit View Data</u> Iransform Analyze <u>G</u> raphs <u>U</u> tilities <u>W</u> indow <u>H</u> elp										
	Insett Variable     Image: Set Variable										
1 : country	l .	Insert Case		а							
	C.	Go to Ca <u>s</u> e.		ltn	urban	religion	lifeexpf	literacy	pop_incr	babymort	gdp_c:
1	Austri	S <u>o</u> rt Cases		0.00	58.0	Catholic	79.0	99.0	.20	6.70	18396
2	Bosni	Tra <u>n</u> spose	•	0.00	36.0	Muslim	78.0	86.0	.70	12.70	3098
3	Colorr	Merge Files	•	0.00	70.0	Catholic	75.0	87.0	2.00	28.00	1538
4	Croati	Aggregate		0.00	51.0	Catholic	77.0	97.0	10	8.70	5487
5	Egypt	Split <u>F</u> ile		0.00	44.0	Muslim	63.0	48.0	1.95	76.40	748
6	Eston	Select <u>C</u> ase	s	0.00	72.0	Protstnt	76.0	99.0	.52	19.00	6000
7	Ethiop	Weight Cas	es	0.00	12.0	Muslim	54.0	24.0	3.10	110.00	122
8	Georgi	ia	550	00.00	56.0	Orthodox	76.0	99.0	.80	23.00	4500
9	Japan		1255	600.0	77.0	Buddhist	82.0	99.0	.30	4.40	19860
10 Lithuania 3800.00			00.00	69.0	Catholic	77.0	99.0	.30	17.00	6710 🚽	
<b>▲ ▶ \ Da</b>	Ita View	🖌 Variable V	iew /				•				
Define Dat	es					SPSS Pro	cessor is ready				

To sort data, click on sort cases and a box will appear. Choose one or more variables in the left blank space. Let us choose "gdp per capita" and leave the Sort Order unchanged, then click OK. All the cases (countries) are rearranged from the lowest to the highest. You can also sort it by string characters from A to Z or reverse.

📲 Sort Cases		×
country     popultn     popultn     w urban     ireligion     ifeexpf     iferacy     pop_incr     w habumort	Sort by: Sort Order Ascending	OK <u>P</u> aste <u>R</u> eset Cancel Help

### **Data Transformation**

SPSS allows you to not only rearrange but also transform data, i.e., to change variable values. You can create new variables or change existing variables. These commands are all under Transform on the pull-down menu.

📷 World	World95.sav - SPSS Data Editor										
<u>F</u> ile <u>E</u> dit	<u>File Edit View Data</u> Iransform Analyze Graphs Utilities Window Help										
	a 🗉 🗠	<u>C</u> orr	pute			kl ⊞ImlF	<u>.</u>				
Random Number Seed											
1 : country	<i>,</i>	C <u>o</u> u	nt								
	country	<u>R</u> ec	ode	_	•	religion	lifeexpf	literacy	pop_incr	babymort	≜ي
1	Austria	Ban	egonze vanable: k Cases	S	B.O	Catholic	79.0	99.0	.20	6.70	
2	Bosnia	Auto	Automatic Recode			Muslim	78.0	86.0	.70	12.70	
3	Colombia	Crea	Create Ti <u>m</u> e Series			Catholic	75.0	87.0	2.00	28.00	
4	Croatia	Rep	eplace Missing <u>V</u> alues			Catholic	77.0	97.0	10	8.70	
5	Egypt	Bur	Pending Trans	forms	<b>1</b> .0	Muslim	63.0	48.0	1.95	76.40	
6	Estonia	1120	1000.00		72.0	Protstnt	76.0	99.0	.52	19.00	
7	Ethiopia		55200.00		12.0	Muslim	54.0	24.0	3.10	110.00	
8	Georgia		5500.00		56.0	Orthodox	76.0	99.0	.80	23.00	
9	Japan		125500.0		77.0	Buddhist	82.0	99.0	.30	4.40	
10	Lithuania	3800.00 6/		69.0	Catholic	77.0	99.0	.30	17.00		
	ita View 🖌 Va	riable \			40.0	K.d	70.0	70 Q	2.40	ZO 00	) )
Compute	Compute SPSS Processor is ready										

#### Compute

Let's first look at Compute. You can create a new variable by calculating values of existing variables. Say we want to create a new variable of total GDP for each country given that we have total population and gdp per capita. Click Compute and you can see a dialogue box. In the top left slot labeled Target Variable, type in "totGDP." In the largest space on the right, enter variables and numeric expressions that can produce the target variable, in this case, totGdp. First select variable gdp\_cap from the variable list on the left and double-click. You should see it appear in the Numeric Expression window. Then choose the multiplication symbol \* from the operator list under the big space. Then choose the variable popultn and enter it into the big space. Click OK. In your Data Editor, you should see the new variable on the far right side.

🗱 Compute Variable		×
Target Variable: totGDP =	Numeric <u>E</u> xpression: gdp_cap * popultn	×
Type&Label		<b>•</b>
▲       country       ▲         ◆       popultn         ◆       urban         ▲       religion         ◆       lifeexpf         ◆       literacy         ◆       pop. incr	+       <>       7       8       9       Eunctions:       ABS(numexpr)         .       <=	4
<ul> <li>➡ pop_inst</li> <li>➡ babymort</li> <li>➡ gdp_cap</li> <li>➡ log_gdp</li> <li>➡</li> </ul>	If OKPasteResetCancelHelp	

#### Recode

Recode is a very useful command. Using it, you can handle missing values or create new categorical variables. For example, we want to divide countries into those with gdp per capita over 2000 US dollars and those less than or equal to 2000 US dollars. Click on Recode Into Different Variables. Chose gdp\_cap from the list of variable at the left. Enter a new Output Variable name as "wealth." On the bottom, find the old and new values button and click it. A new box appears. Choose a range of lowest to 2000 and select a new value as 0 then click add. Similarly, choose a range of 2001 through highest and enter the new value as 1. These equations should be in the right box. Confirm your operation and click Continue. When you come back to the old dialog box, click OK again. Voila, you have another new variable called "wealth" at the far right end of the Data Editor.

<b>***</b> Recode into Different Variables	×
Numeri	c Variable -> Output Variable:
dp ca	ap> wealth
popultn	
(#) urban	
	Label:
A literacu	wealth
pop_ner	
	od New Values
	id New Values
ОК	Paste Reset Cancel Help
Recode into Different Variables: Old a	nd New Values 🛛 🔀
Cld Value	New Value
C ⊻alue:	O Value: 1     O System-missing     O
C System-missing	C Copy old value(s)
C System- or <u>u</u> ser-missing	Old> New:
C Ra <u>n</u> ge:	Add Lowest thru 2000> 0
through	
C Panaci	Lihange
Lowest through	<u>R</u> emove
• Hange:	□ Output varia <u>b</u> les are strings <u>W</u> idth: 8
2001 through highest	Convert numeric strings to numbers ('5'->5)
C All <u>o</u> ther values	Continue Cancel Help

#### Automatic Recode

Though you can have string variables in the data sheet, you cannot do statistical analysis on them. You need to transform them into categorical variables. For example, you can transform gender "male" into 0 and "female" into 1. In SPSS, you can use automatic recode to do it easily. In our example of international cases, say we want to explore any relationship between dominant religion and economic development. We need to change the string variable "religion" into a numeric one. Click automatic recode, type in a new variable name "catRel" and keep the order of values unchanged, i.e., from lowest value. Then click OK. The beauty of this command is that it automatically changes former string values into value labels. It would be more complicated if you did it by using another command, like recode.

<b>*** Automatic Recode</b>		×
<ul> <li>▲ country</li> <li>♦ popultn</li> <li>♦ urban</li> <li>♦ lifeexpf</li> <li>♦ liferacy</li> <li>♦ pop_incr</li> <li>♦ babymort</li> <li>♦ gdp_cap</li> <li>♦ log_gdp</li> </ul>	Variable -> New Name          religion> catRel         New Name       catRel         New Name       catRel         Recode Starting from           © Lowest value       D Highest value	OK <u>P</u> aste <u>R</u> eset Cancel Help

You can view value labels by clicking the value label icon On in the Data Editor window.

```
Statistical analysis
```

Statistical analysis is the core of SPSS. You can use descriptive statistics and graphs to explore statistical relationships between variables. You can test those relationships by using regression analysis.

Say we want to know the mean, median, and standard variations of each variable. You can do so by going to the pull-down menu Analyze/Descriptive Statistics/Descriptives. The results are stored and displayed in the Output Viewer.

📷 World	95.sav - SPSS Da	ata Edil	tor					_	□×
<u>F</u> ile <u>E</u> dit	_ <u>V</u> iew_ <u>D</u> ata_ <u>T</u> rar	nsform	<u>Analyze G</u> raphs <u>U</u> tilitie	es <u>W</u> indow	<u>H</u> elp				
<b>2</b>	🕘 🖳 🗠 🗠	<u>i</u>	Reports D <u>e</u> scriptive Statistics	Erequence	≓l ≂l @				
1 : country	)	A,	Custom <u>T</u> ables	<u>D</u> escripti	ves				
	country	р	Compare <u>M</u> eans General Linear Model	<ul> <li><u>Explore.</u></li> <li>Crosstab</li> </ul>	ç pf	literacy	pop_incr	babymort	▲ي
1	Austria		Correlate		•···· •/9.0	99.0	.20	6.70	
2	Bosnia		<u>Regression</u>	slim	78.0	86.0	.70	12.70	
3	Colombia	3	L <u>og</u> linear	tholic	75.0	87.0	2.00	28.00	
4	Croatia		Classify	tholic	77.0	97.0	10	8.70	
5	Egypt	6	Data Reduction	slim	63.0	48.0	1.95	76.40	
6	Estonia		Sc <u>a</u> le	▶ ptstnt	76.0	99.0	.52	19.00	
7	Ethiopia	5	Nonparametric Tests	slim	54.0	24.0	3.10	110.00	
8	Georgia		Lime Series Survival	hodox	76.0	99.0	.80	23.00	
9	Japan	1	<u>J</u> urviva Multiple Besponse	ddhist	82.0	99.0	.30	4.40	
10	Lithuania			<del>. a</del> tholic	77.0	99.0	.30	17.00	
	ta View 🖌 Variable	e View		A line	70.0	70.0	2 4 2	ZO 00	▶
Descriptive	es			SPSS Pro	cessor is ready				_//

We can also use graphs to visualize the statistical relationships between variables. Say we want to have a sense of any association between gdp per capita and urbanization. We can draw a scatter plot. Go to Graph/Scatter/Simple. Choose two variables and enter them into the Y and X slots. Click on the variable "country," and put it in the last slot, Label Cases by: Click OK and you should see the graph in the Output Viewer. The chart shows a positive relationship between urbanization and economic development.

🗱 Simple Scatterplot			×			
<ul> <li>✤ popultn</li> <li>☑ religion</li> <li>✤ lifeevpf</li> </ul>		Y_Axis: I 余 gdp_cap	OK <u>P</u> aste			
literacy		X Axis:	<u>R</u> eset			
pop_incr     babymort		j ₩ uban	Cancel			
() log_gdp		Set Markers by:	Help			
		Label Cases by:				
Template						
itles						

There are numerous tools of statistical analysis, particularly in the area of regression analysis. Let us look at one of the simplest, linear regression modelling. We want to test whether there is any statistical association between economic development and several social variables, such urbanization, literacy rate and religion.

Go to Analyze/Regression/Linear. Enter "gdp per capita" as the dependent variable and urbanization, literacy rate and categorized religion as independent variables. There are many options to help you interpret statistical relationships. In this example, we will keep it simple. Keep Method "Enter" intact. Click OK. After a few seconds, the results will show in the Output window. There are a number of tables, depending on the options chosen. In our example, there are variables entered, model summary, ANOVA, and Coefficients.

23	Linear Regression			×
	country     popultn     urban	Þ	Dependent:	OK <u>P</u> aste
	▲ religion	Pre <u>v</u> ious	Block 1 of 1 Next	<u>R</u> eset
	<ul> <li>Ifeexpf</li> <li>Iteracy</li> </ul>			Cancel
	pop_incr     babymort     bg_gdp	Þ	<ul> <li>♦ urban</li> <li>♦ literacy</li> </ul>	Help
	(∉) catrei		Method: Enter	
			S <u>e</u> lection Variable:	
			Case Labels:	
	<u>W</u> LS >>	<u>S</u> tatistics	Plots Save Options	

You can also use the Paste button to paste your commands into a syntax window. It not only saves time if you run the regression many times, it also keeps track of what you've done so you can communicate with others or remind yourself of the procedure after a period of time.

#### Output management

You can save your data in formats other than SPSS data. You can change the layout and appearance of tables and graphs to meet your requirements. Data output is easy. When you save, you can select other formats in the drop-down menu, choose a directory and click OK. It is also quite convenient to change the appearance of tables and graphs. If you want to change a table or a chart, just double-click it. A Pivot Table Editor or a Chart Editor will appear.

For example, we want to change the title of a regression coefficient table. Double-click the output table, the Pivot Table Editor pops up. Double-click the title. When the table title is highlighted, you can change it. To place SPSS tables in reports, copy the table using the Copy Object option under Edit and then paste it into Word.

We can also change graphs. Double-click on a scatter plot. The Chart Editor will appear. Let's add a title and the fit line. Go to Chart and click title. Add a title in the small dialogue box. Go to Chart again and click Options and check "fit line" Your chart should be changed and you can close the Chart Editor.

Conclusions

We learned the basic structures of SPSS, how to input data, how to manipulate and transform data, how to do descriptive statistics and regression analysis, and how to draw and change a chart.

The quickest way to get an overview is to run the online tutorial. The best way to learn SPSS is to work on a real project. In the process, you will have a lot of questions. The first thing you can do is check the help topics. Then if you still can not figure them out, ask consultants at CSSCR, we would be happy to help you. Or you can consult SPSS manuals, and the Center has plenty of them. Good luck in your SPSS adventure.