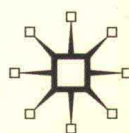


# A TOOLKIT FOR QUANTITATIVE DATA ANALYSIS USING SPSS

SOTIRIOS SARANTAKOS

covers versions  
10 through to 15



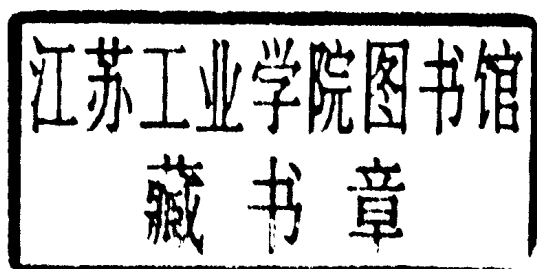
---

# A Toolkit for Quantitative Data Analysis

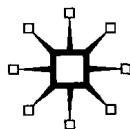
Using SPSS

---

Sotirios Sarantakos



palgrave  
macmillan



© Sotirios Sarantakos 2007

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London W1T 4LP.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2007 by  
PALGRAVE MACMILLAN  
Houndmills, Basingstoke, Hampshire RG21 6XS and  
175 Fifth Avenue, New York, N.Y. 10010  
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN-13: 978-0-230-50045-7

ISBN-10: 0-230-50045-5

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1  
16 15 14 13 12 11 10 09 08 07

Printed in China

---

# About this book

---

This book is about how to conduct quantitative data analysis (QDA), using SPSS (version 15, as well as 10, 11, 12, 13 and 14). It focuses on the essence of data analysis (DA), and avoids unnecessary frills and long-winded jargon-based presentations. More specifically, this book:

- Employs a *simple* approach, making DA easily accessible to the novice.
- Is *free of maths and stats*: doing QDA requires neither mathematical skills nor statistical knowledge. Statistical tests are conducted by SPSS.
- Is *concise*: in about 100 pages it covers the essence of basic DA.
- Is *laconic*: it focuses on the essentials and offers only ‘as much as necessary and as little as possible’.
- Is *systematic*: it follows a step-by-step, recipe-like approach that makes data analysis quick and easy, without sacrificing quality.

The focus of this book is primarily on data analysis and secondarily on statistics. In this sense, it focuses on how to conduct DA, namely on *which* tests to use and *how* to interpret the results. Simply, doing DA means: **focus on an analytic task**, **identify the proper test**, **open the relevant page**, **conduct the test as instructed** and **interpret the results as advised**.

This book is not a text but a guide. It does not introduce you to social research, its epistemological foundations or to the many philosophical or statistical conflicts and debates. It simply tells you how to handle the technical aspects of QDA in a simple but effective way, producing the outcomes required to answer research questions.

In this format, this book is a valuable guide to those who want to learn how to conduct basic QDA without long-winded descriptions and

epistemological puzzles, and certainly a companion to students of sociology, psychology, health, education, journalism and social sciences in general, studying social research. It is also most useful to those who have some insight into the research process, who have access to data and who wish to conduct a quick, direct, easy and reliable QDA. In these cases, this book will be found to be most adequate, most reliable and most useful.

Sotirios Sarantakos  
June 2007

---

# Contents

---

<i>About this book</i>	<i>ix</i>
<b>Introduction</b>	<b>1</b>
<b>1 Getting to know your SPSS: SPSS – its many faces</b>	<b>6</b>
Exploring SPSS for Windows	7
Data Editor	8
SPSS Viewer	9
Getting assistance	10
<b>2 Data analysis: entering and managing data</b>	<b>12</b>
Entering raw data	12
Entering tabular data	16
Entering text data files	20
Entering data with multiple responses	21
Managing SPSS files	24
<b>3 Transforming data</b>	<b>27</b>
Recoding reverse scale items	27
Recoding multiple responses	30
Getting random samples	31
Getting sub-groups	33
Collapsing numeric variables	35
Creating composite variables	37
<b>4 Tables, central tendency and dispersion</b>	<b>39</b>
Frequency tables	40
Getting crosstabs	41
Mean, mode, median and standard deviation	42
Standard scores (Z scores)	44
Statistics for sub-groups	45
<b>5 Data analysis: graphic presentations</b>	<b>47</b>
Histograms	48

Simple bar charts	49
Clustered bar charts	51
Stacked bar charts	52
Pie charts	53
Simple scattergrams	56
Simple box plots	58
Stem-and-leaf plots	60
Transform charts	61
<b>6 Tests of significance: nonparametric tests</b>	<b>63</b>
A note on statistical significance	63
Chi-square test: goodness-of-fit test	65
Chi-square test for independence	69
McNemar test	70
The Wilcoxon signed-ranks test	72
Mann-Whitney U test	74
<b>7 Tests of significance: parametric tests</b>	<b>76</b>
T-test for one sample	77
T-test for unrelated (independent) samples	78
T-test for related (dependent) samples	81
One way analysis of variance (ANOVA)	83
Post hoc comparisons	85
<b>8 Correlation tests: for nominal variables</b>	<b>88</b>
Correlation tests	89
Phi ( $\phi$ ) and Cramer's V	92
Lambda coefficient ( $\lambda$ )	94
<b>9 Correlation tests: for ordinal variables</b>	<b>97</b>
Gamma ( $\gamma$ )	97
Somers' d	100
Spearman's rho	101
<b>10 Correlation – regression: for interval/ratio variables</b>	<b>104</b>
Pearson's correlation ( $r$ )	104
Partial correlation	107
Simple regression	109
Multiple regression	111
<b>Useful concepts</b>	<b>114</b>
<i>Index</i>	<i>117</i>

---

# Introduction

---

Data analysis (DA) is the process of transforming raw data to numbers, applying statistical tools, and aiming to describe, summarize and compare data, and to discover knowledge. This suggests that for an adequate and successful DA the researcher needs firstly data and secondly tools of DA.

## What are data?

Data are basically pieces of information collected by the researcher before DA begins. They are answers to survey questions, responses to experimental stimuli, parts of texts, actions or behaviour options in a context of observation, reactions to situations within a focus group discussion, etc. Data are collected using methods such as experiments, content analysis and most of all through surveys, ie interviewing, and mail questionnaires. Quantitative versions of content analysis, observation and focus groups are also employed to gather data. These raw data are quantized and further prepared for analysis using coding.

## What is coding?

Coding is the procedure of converting raw data into numbers, with each number representing a code and a code standing for a value or category. For instance, the answer 'Yes' becomes '1', 'I don't know' becomes '2', and 'No' becomes '3'. Also, 'Male' is substituted by the number '1'; and 'Female' by the number '2'. Hence, DA deals with the numbers which represent values or categories of variables.

## What are variables?

Variables are empirical constructs that take more than one value. Gender is a variable; it contains the values 'male' and 'female'. Social class is another variable; it contains the values 'upper class', 'middle class' and 'lower class'. Age, finally, is another variable, which contains innumerable



values, because it can be anything from 0 upwards, such as 7, 21, 25.3, 32.5 years, etc. Gender is a *nominal* variable; it names the categories it entails. Social class is an *ordinal* variable; it ranks its categories. Age is a *interval* variable; it contains values with equal intervals. An interval variable that also contains a zero as its starting point is a *ratio* variable.

In data analysis, variables are compared and interrelated with each other, and, depending on their position in the relationship, they can be independent or dependent variables. An *independent* variable is one that is assumed by the analyst to have an impact on another; the variable that is supposed to be affected by another is the *dependent* variable. In a study investigating the effects of religion on scholastic achievement, religion is the independent variable and scholastic achievement the dependent variable.

## What are the types of DA?

There are two types of DA: *manual* and *electronic* (or computer-assisted) DA. DA is conducted predominantly using computers; this is why it is known as *computer assisted data analysis* (CADA). If the data are analysed manually, the analyst will begin the analysis immediately after coding. If, however, the analysis will be conducted electronically, the analyst will enter the data in the computer before analysis begins. In this guide we follow the second option.

## What is the content of DA?

DA is statistical analysis, ie it is a procedure that uses statistical tests. Simply stated, the data analyst focuses on a research question and employs statistical tests to find an answer. The choice of tests is dictated by two factors, namely (a) the nature of the variables (ie whether they are nominal, ordinal, interval or ratio), and (b) the purpose of the analysis, eg whether it aims to describe distributions, to correlate variables, to compare groups, etc. There are different tests for nominal, ordinal and interval/ratio data, and also different tests for answering the various research questions (regarding correlation, inference, etc).

## What are the outcomes of DA?

DA produces descriptions, estimates of central tendency, dispersion, correlation, regression and comparisons of groups. The results of statistical tests appear in the form of graphs, maps and tables, coefficients and

other statistical symbols, all in many formats, as many and as diverse as the tests that produce them. Interpreting the outcomes of DA, ie making sense of the outcomes of the analysis, is the central task of the analyst.

### What does interpretation entail?

During interpretation, analysts complete the path of data analysis, in a way turning back to the start of the research project. Whereas at the beginning of the study analysts converted words and meanings to numbers, at the last step of data analysis they turn numbers to words and meanings. Statistical interpretations are based on mathematical logic, and in this sense they are easy to construct, provided you know the rules of the game, eg the meaning of correlation coefficients and the power of significance tests and values. Nevertheless, making valid statements that are convincing and reflect real situations is another matter. And this is a point that deserves attention, for it is easy for the novice to exaggerate the value of the findings. For instance, there is a great difference between statistical significance and substantive significance – that is the practical importance of the findings for explanation, theory and social policy – and between correlation and causation.

### What is the structure of DA?

In principle, DA is conducted in three steps: data entry, selection of an analytic task and statistical testing. (1) Data entry involves the transfer of data in the computer, providing in this way the material that is required for the computer-assisted data analysis. Although this procedure is indispensable, it usually is not considered a part of DA, and is normally completed before the actual analysis begins. (2) During the second step of data analysis researchers choose a specific task (analytic task) to focus their attention on, and decide *which* tests to employ, *when* and *why*. These tasks are listed in Table 1.1. (3) The third step is about running statistical tests, and focuses mainly on two tasks: (a) the mathematical computations, and (b) the interpretation of the mathematical symbols. The type of statistical tests employed in data analysis are shown in Table 1.1 and are conducted using SPSS.

### What is SPSS?

SPSS is one out of many computer software programs employed by researchers and students when conducting CADA; SPSS stands for

**Table 1.1** Analytic tasks and corresponding statistical tests

Analytic tasks	Corresponding test options
Describe graphically and/or numerically the distribution	Graphs: histograms, bar charts, pie charts Tables: frequency tables, crosstabs
Estimate central tendency	Mean, mode and median
Estimate dispersion	Standard deviation, range
Correlate variables	$\phi$ (phi) coefficient and Cramer's V, Somers' d, Lambda ( $\lambda$ ) coefficient, Spearman's rho ( $\rho$ ), and Pearson's r
Test predictions	Simple/multiple regression
Compare groups	Chi-square ( $\chi^2$ ); McNemar test; Wilcoxon test; t test; and analysis of variance.

Statistical Package for the Social Sciences. SAS and Minitab are another two popular programs but still, SPSS is the most common, particularly in the area of social sciences, and is available for Windows as well as for Mac OS X (SPSS 13).

SPSS is a dynamic, diverse and well integrated computer program, offering a variety of features and modules, tailored to meet the needs of its users. The basic version of SPSS focuses on data analysis but it offers a lot more, for instance it assists the planning of the study and data collection, data preparation and reporting. Put simply, this program contains all that is required for conducting a piece of research including quantitative data analysis, using a simple and easily accessible procedure.

Although all features of SPSS are useful, the central focus of most users is on data analysis, which not only offers a broad spectrum of options but also access to a powerful, fast, valid and reliable statistical analysis. This obviously is most useful to analysts with poor or no mathematical skills, but of equally great importance to those with statistical skills. There is no analyst nowadays who would conduct research and especially data analysis manually. Speed and reliability are two criteria that make CADA the only way even for the experienced statisticians. Access to a variety of forms of tabular and graphical presentation of the data and of reporting methods make SPSS even more practical and more useful for all users.

SPSS is available as a full program, or in a smaller version, the *Student Version*, specially tailored to the needs of the users. This is adequate and also less costly. For those who are interested in more complex procedures and multivariate techniques, *SPSS Advanced Models* is an appropriate extension. Advanced students may opt for *SPSS Graduate Pack* as an option; this includes the full version of SPSS Base, two add-on modules and, for Windows users, software for structural equation modelling (SEM). Also interesting are *SPSS Categories* and *SPSS Classification Trees* which introduce perceptual maps with optimal scaling and dimension reduction techniques as well as visual classification and decision trees.

A closer view of SPSS will be offered in the next section, where its major parts will be introduced.

### Examples of online assistance

- **Social research methods:**  
[www.socialresearchmethods.net/](http://www.socialresearchmethods.net/)
- **Online text:**  
<http://www.isixsigma.com/offsite.asp?A=Fr&Url=http://nilesonline.com/stats/>
- **Glossaries:**  
[http://www.animatedsoftware.com/elearning/Statistics%20Explained/glossary/se\\_glossary.html](http://www.animatedsoftware.com/elearning/Statistics%20Explained/glossary/se_glossary.html)  
<http://stattrek.com/Help/Glossary.aspx>
- **Free statistical software:**  
<http://www.sixsigmalab.com/>

## CHAPTER 1

---

# Getting to know your SPSS

## SPSS – its many faces

---

As noted earlier, over the last ten years, SPSS appeared in a series of versions, of which 10, 11, 12, 13 and 14 are still employed by students; late in 2006, version 15 was published. Although each new version added important innovations and expansions, as far as basic statistics is concerned and with regard to the basic syntax files, versions 10–14 are almost identical. The same is true for version 15. The only area of basic statistics which requires slightly different commands is ‘Graphs’. Improvement and expansion in this area offered additional options, and hence slight adjustments in the approach to charts are required.

In this volume, we shall use all versions and introduce the necessary instructions for each test used, but before we do that, let us get acquainted with the structure of SPSS for Windows, and particularly with the following:

- exploring SPSS for Windows
- Data Editor
- SPSS Viewer
- getting assistance.

After this introduction, we shall look at applying the program in data analysis, beginning with entering the data in the computer.

## Exploring SPSS for Windows

### What is this?






SPSS is a computer program used to carry out statistical tests for data analysis. In this guide we use SPSS versions 10–15.


### The ‘faces’ of SPSS

When we work with SPSS, we are faced with a number of screens, namely the **Data Editor: Data View**, the **Data Editor: Variable View**, the **SPSS Viewer** and the **Statistics Coach**. These are the paths to the program and will be introduced after we explain a few ‘tricks’ that are used regularly when communicating with the computer.

### Communicating with SPSS: abbreviated commands

Computers are instructed by means of commands: we use commands to instruct the computer which test to compute and which aspects of each test to consider. We use also abbreviations of commands, by combining them in a sequence command. For instance, instead of using the commands ‘Click A’, and then ‘Click B’ and finally ‘Click C’, we may use ‘Click A/B/C’, or ‘Go to A/B/C’, or ‘Choose A/B/C’, etc. The following abbreviations are the most common.

- **Go to:** This suggests ‘go to and click on’. Eg ‘Go to Analyze’ means go to Analyze and click on it.
- **Click:** This means press the left-hand side mouse button. For instance, ‘Click on File’ means point to and left-click on File. For pressing the right-hand side mouse button we use the command ‘right-click’; eg ‘right-click the chart’.
- **Activate:** This stands for click on the  or  that stands before or after a specific word. For instance, ‘Activate Pearson’ means ‘click on the sign that stands in front of Pearson’. When activated, these signs change to ,  or .
- **Select A/B/C/X or Choose A/B/C/X or Click A/B/C/X:** These commands suggest to click on A, then on B, on C and on X successively. For instance, ‘Select File/Open/Data’ implies ‘click File, then Open and then Data’.

- **Transfer:** This command is used when a variable is to be moved from one box to another. Example: Transfer Gender to the Rows box. 'Transfer' means (1) highlight the variable in question and (2) click on the  button on the right-hand side of the variables list.

Let us now get acquainted with the many 'faces' of SPSS.

## Data Editor

### What is this?

The Data Editor is the screen where you begin your work with SPSS. This is a spreadsheet-like system and appears in two forms, the **Data View** and the **Variable View**. You can toggle between them by clicking on either of the labels displayed at the bottom of the Data Editor screen. Data View and Variable View will be introduced briefly below.

### Data View

The **Data Editor: Data View** is where the results of a study are entered and viewed. The screen here is arranged so that rows are assigned to cases and columns to variables. Each row contains the responses of one respondent to all questions; and each column contains the responses of all respondents to one variable. Rows are numbered; columns are named after their variable.

### Variable View

The **Data Editor: Variable View** is where variables are defined, edited and viewed. In this editor, rows are for variables and columns are for variable attributes. The titles of the columns indicate the type of variable details that goes in the cells. These titles, their content and meaning, and some helpful hints as to how to fill in these cells are given below.

<b>Name</b>	Variable name; enter a name but do not use blanks, !, ?, ' or *. Also begin with a letter and do not end with a full stop.
<b>Type</b>	Data type; click on the shaded box to choose the right options.
<b>Width</b>	Column width; you may change it but often 8 is sufficient.
<b>Decimals</b>	Number of decimal places; set it according to need – usually 0.
<b>Label</b>	Variable labels; use labels up to 256 characters (eg 'subject's status').
<b>Values</b>	Set values and value labels, eg 1, 2 for men/women respectively.
<b>Missing</b>	Missing values; set it to 9 or 99 (usually 9).
<b>Columns</b>	Set columns to 8.
<b>Align</b>	Set alignment column contents to right, left or centre at will.
<b>Measurement</b>	Set to nominal, ordinal or scale (to be explained later).

As you see, some of the variable attributes are easy to define but others are not. More information on this will be given later in this section.

## SPSS Viewer

### What is this?

The SPSS Viewer (Figure 1.1) is where the output of statistical calculations (graphs, tables, etc) is displayed. It appears automatically after a statistical procedure is completed. Unlike the Data Editor, the SPSS Viewer does not contain rows and columns but two sections, the *outline pane* (narrow part, left) listing the titles of the computed tests, and the *contents pane* (wide part, right) where the actual output (eg tables, charts, etc) is stored and displayed.



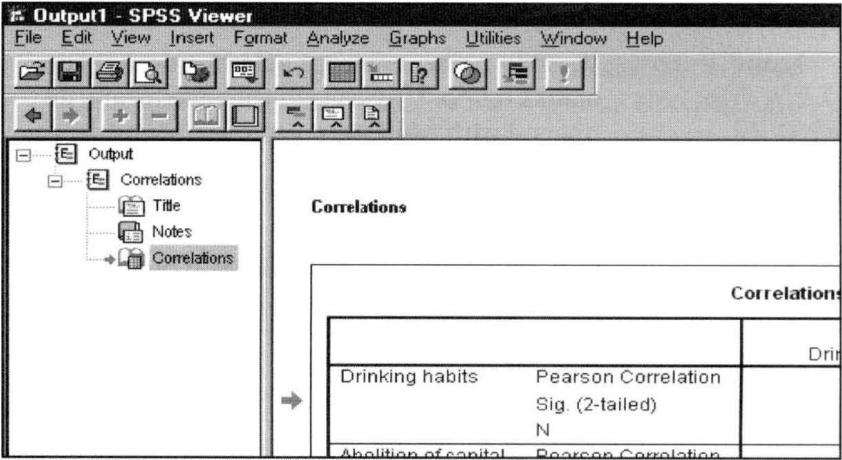


Figure 1.1

Briefly, the output of all tests conducted during one session are listed in the outline pane and are available for further study and comparisons in the content pane. When you click on an item, the corresponding output appears on the contents pane for inspection or further analysis.

The bar that divides the outline pane and the contents pane can be moved sideways at will. Just click on the bar, and while holding the mouse button down, move the pointer horizontally. The bar follows, adjusting the pane size as required. This is particularly necessary when long tables are displayed, too long for the standard-size pane. In such cases, shifting the bar accordingly allows a full display of the table.

### Getting assistance

#### What is this?

SPSS provides a number of services geared towards making the use of the program easy and effective. Most interesting is the ‘Tutorial’ which offers step-by-step instructions on how to use the program. If you are interested in specific aspects of the program, the ‘Statistics Coach’ is very handy (Figure 1.2).

Getting to these sources of assistance is simple: Select **Help** and make your choice. If you choose the ‘Statistics Coach’, you come to the display shown in Figure 1.2 (SPSS 14), or to that shown in Figure 1.3 (SPSS 15).