# SPSS 12.0 for Windows

Analysis without Anguish

Sheridan J Coakes



# SPSS 12.0 for Windows

Analysis without Anguish

# SPSS 12.0 for Windows

Analysis without Anguish

Sheridan J Coakes



John Wiley & Sons Australia, Ltd

KaRaNLıK

First published 2005 by John Wiley & Sons Australia, Ltd 33 Park Road, Milton, Qld 4064

Offices also in Sydney and Melbourne

© S. Coakes, L. Steed 2005

National Library of Australia Cataloguing-in-Publication data

Coakes, Sheridan J.

SPSS: analysis without anguish: version 12.0 for Windows

Version 12.0 For tertiary students. ISBN 0 470 80736 9.

 SPSS for Windows. 2. Social sciences — Statistical methods — Computer programs. I. Title.

005.369

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher.

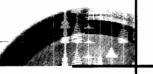
Edited by David Rule

Cover image: © Digital Vision

Printed in Singapore by CMO Image Printing Enterprise

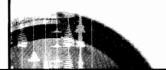
10 9 8 7 6 5 4 3 2 1

SPSS version 12.0 for Windows is a registered trademark of SPSS, Inc. SPSS: Analysis without Anguish is not approved, sponsored or connected with SPSS Inc. Information about SPSS and other software products is available from SPSS, Inc. www.spss.com



# Contents

Prefaceviii	Entering data 24
At-a-glance text examplesx	Inserting and deleting cases and variables 24
	Moving variables 25
CHAPTER 1	Saving data files 25
Introduction to SPSS 1	Opening an existing data file 26
Getting started 1	Practice example 26
The SPSS environment 2	
Data Editor 2	CHAPTER 3
Viewer and Draft Viewer 3	Data screening and
Pivot Table Editor 3	transformation 28
Chart Editor 5	Working example 28
Text Output Editor 6	Errors in data entry 29
Syntax Editor 6	Assessing normality 30
Script Editor 7	Histograms 32
Toolbar 7	Stem-and-leaf plots and boxplots 32
Menus 9	Normal probability plots and detrended
File 9	normal plots 34
Edit 9	Kolmogorov-Smirnov and Shapiro-Wilks
View 10	statistics 35
Data 10	Skewness and kurtosis 35
Transform 10	Assessing normality by group 36
Analyze 11	Variable transformation 36
Graphs 11	Data transformation 42
Utilities 11	Recode 42
Window 12	Compute 45
Help 12	Data selection 47
Dialogue boxes for statistical procedures 12	Practice example 49
Source variable list 13	Solutions 50
Selected variable(s) list 14	Syntax 50
Command pushbuttons 14	Output 51
Accessing sub-dialogue boxes 15	CHAPTER 4
Check boxes, radio buttons and	
drop-down lists 16	Descriptive statistics 56
Check boxes 16	Frequency distributions 56
Radio buttons 16	Measures of central tendency and variability 56
Drop-down lists 17	Working example 56
Saving files and ending a session 17	Descriptives command 60
CHAPTER 2	Practice example 61
	Solutions 62
Preparation of data files 19	Syntax 62
Working example 19	Output 62
Defining variables 19	CHAPTER 5
Variable labels 20	Correlation 65
Value labels 20	
Missing values 20	Assumption testing 65
Variable type 21	Working example 66
Column format 21	Practice example 70
Measurement level 21	Solutions 70
Applying variable definition attributes to	Syntax 70 Output 70
other variables 23	I QUIDUI / U



#### CHAPTER 6

#### T-tests 73

Assumption testing..... 73
Working example..... 73
The one-sample t-test..... 74
T-tests with more than one sample..... 75
Repeated measures t-test..... 75
Independent groups t-test..... 77
Practice example..... 81
Solutions..... 81
Syntax..... 81
Output..... 82

#### CHAPTER 7

#### One-way between-groups ANOVA with post-hoc comparisons 85

Assumption testing..... 85
Working example..... 86
Practice example..... 89
Solutions..... 89
Syntax..... 89
Output..... 89

#### CHAPTER 8

#### One-way between-groups ANOVA with planned comparisons 92

Assumption testing..... 93
Working example..... 93
Practice example..... 96
Solutions..... 96
Syntax..... 96
Output..... 96

#### CHAPTER 9

# Two-way between-groups ANOVA 98

Assumption testing..... 98
Working example..... 98
Practice example..... 104
Solutions..... 104
Syntax..... 104
Output..... 105

#### CHAPTER 10

## One-way repeated measures ANOVA 108

Assumption testing..... 108 Working example..... 108 Practice example..... 112 Solutions..... 112 Syntax..... 112 Output..... 112

#### CHAPTER 11

# Two-way repeated measures ANOVA 114

Assumption testing..... 114
Working example..... 114
Practice example..... 118
Solutions..... 118
Syntax..... 118
Output..... 118

#### CHAPTER 12

#### Trend analysis 120

Assumption testing..... 120
Working example..... 120
Practice example..... 123
Solutions..... 124
Syntax..... 124
Output..... 124

#### CHAPTER 13

# Mixed/split plot design (SPANOVA) 126

Assumption testing..... 126
Working example..... 126
Practice example..... 131
Solutions..... 131
Syntax..... 131
Output..... 131

#### CHAPTER 14

# One-way analysis of covariance (ANCOVA) 135

Assumption testing..... 135
Working example..... 136
Practice example..... 141
Solutions..... 141
Syntax..... 141
Output..... 142

#### CHAPTER 15

#### Reliability analysis 147

Working example..... 147
Practice example..... 150
Solutions..... 151
Syntax..... 151
Output..... 151



#### Spearman's rank order correlation..... 218 CHAPTER 16 Working example..... 218 Factor analysis 154 Practice examples..... 220 Assumption testing..... 154 Practice example 1..... 220 Working example..... 155 Practice example 2.... 220 Practice example..... 164 Practice example 3..... 221 Solutions.... 164 Practice example 4.... 221 Syntax.... 164 Practice example 5..... 221 Output..... 165 Practice example 6.... 221 Practice example 7..... 221 CHAPTER 17 Solutions.... 222 Multiple regression Practice example 1 — chi-square test for goodness of fit .... 222 Assumption testing..... 169 Practice example 2 — chi-square test for Working example.... 169 relatedness or independence..... 222 Practice example.... 179 Practice example 3 — Mann-Whitney U test Solutions.... 179 (Wilcoxon rank sum W test)..... 223 Syntax.... 179 Practice example 4 — Wilcoxon signed-rank Output.... 179 test .... 224 Practice example 5 — Kruskal-Wallis CHAPTER 18 test.... 224 Multivariate analysis of Practice example 6 — Friedman test.... 225 variance (MANOVA) 183 Practice example 7 — Spearman's rank order correlation.... 226 Assumption testing..... 184 Working example.... 184 CHAPTER 20 Data screening..... 185 Multiple response and multiple Practice example..... 192 Solutions.... 192 dichotomy analysis 227 Multiple response analysis..... 227 Syntax.... 192 Output.... 193 Working example.... 228 Hope.... 196 Multiple dichotomy analysis..... 230 Esteem.... 197 Working example..... 231 Optimism.... 199 Practice examples..... 233 Practice example 1..... 233 CHAPTER 19 Practice example 2..... 234 Solutions.... 234 Nonparametric Practice example 1..... 234 techniques 204 Practice example 2..... 236 Chi-square tests..... 204 Assumption testing.... 204 CHAPTER 21 Working example — chi-square test for Multidimensional scaling 237 goodness of fit ..... 205 Working example..... 237 Working example — chi-square test for Practice example..... 245 relatedness or independence..... 208 Mann-Whitney U test CHAPTER 22 (Wilcoxon rank sum W test)..... 211 Working with output 248 Working example.... 211 Editing output in the SPSS Viewer.... 248 Wilcoxon signed-rank test..... 213 Modifying and improving charts for Working example..... 213 presentation.... 254

Practice example..... 260

Appendix 262

Viewer and Draft Viewer output..... 260

Kruskal-Wallis test.... 215

Friedman test.... 217

Working example..... 215

Working example..... 217



## **Preface**

SPSS is a sophisticated piece of software used by social scientists and related professionals for statistical analysis. This workbook is based on version 12.0, which is compatible with the Windows environment.

There are several new features in SPSS 12.0 relating to improved charting, new data management, statistical enhancements and, for those who work with it, a new add-on module for working with sample survey data. Some of these new features are outlined below.

#### **Improved Charting Features**

- Better default chart appearance
- Support for long text strings and automatic text wrapping
- Control of default scale ranges using chart templates
- · 3-D effects for pie and bar charts
- Improved choice of colour patterns.

#### **New Data Management Features**

- Longer variable names now a maximum of 64 bytes, compared with eight in previous versions
- Visual bander assists you in banding scale data into categorical ranges (e.g. age in 10-year ranges)
- Duplicate record finder helps you identify, flag, report and filter duplicate records with the new Duplicate Cases feature
- Output management system allows you to turn output into input with the new OMS command by automatically writing selected categories of output to different output files in different formats HTML, XML, text and SPSS-format for data files
- Command syntax to delete variables the new Delete Variables command makes it
  easy to delete variables you do not need anymore, such as temporary variables used
  in transformations.

#### **Statistical Enhancements**

- · New options of handling weighted data in the Crosstabs procedures
- New stepwise function in Multinomial Logistic Regression.

This edition of SPSS: Analysis without Anguish continues the trend of previous editions in providing a practical text intended as an introduction to SPSS and a guide for users who wish to conduct analytical procedures.

We've also tried to improve the text from the feedback we have obtained from our readers. In this edition, we have included reporting at the end of each relevant chapter to illustrate how significant results from each of the particular tests should be reported.

Many readers also requested that we bring back our chapters on Cluster Analysis, Multidimensional Scaling and Discriminant Analysis. Thanks to the contribution of Dr Mark Fenton from James Cook University in Townsville, this edition includes additional chapters on Mutidimensional Scaling.

Again, we've tried to keep the text clear and simple — no one text will satisfy all our readers. However, in subsequent editions we hope to expand on some of the other techniques offered in SPSS.

Chapters 1 to 3 are generic chapters that introduce the software, detail how to prepare data files and outline data screening methods. Chapters 4 to 20 outline specific procedures within the package. In each of these chapters, a working example is presented.

One major advantage of the workbook is that the text is accompanied by data files, which are available from the website that accompanies this title (www.johnwiley.com.au/highered/spssV12), so that the user can progressively work through each procedure with the text. At the end of each chapter, a practice example is given to help users to consolidate the skills they have learned. Solutions to these practice examples are also provided to clarify the analytical procedure and to facilitate interpretation of the SPSS output.

As highlighted above, chapter 21 considers the application of Multidimensional Scaling and, as with other editions, chapter 22 concludes in dealing specifically with output, outlining ways in which presentation can be improved.

Although the workbook outlines each statistical procedure, this is not a statistical text and a degree of statistical knowledge is assumed. At the beginning of each chapter, assumption testing for each procedure is discussed, and the procedure is approached simply and systematically.

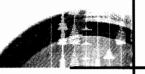
The concept of the workbook arose from our collective experience of teaching and applying research methods. It evolved from a recognised need to make research methodology more accessible and understandable to students who are undertaking research methods courses and to professionals who are taking part in research in an applied context.

We still receive very positive feedback from you, the users of the book, who tell us that we have helped alleviate some of the 'anguish' associated with the analysis of research data. We trust that this edition continues to help with this.

Having worked through the book, you will be well on your way to effective research from coding, entering and exploring data, and undertaking appropriate analysis to creating meaningful data output. We wish you well in your research endeavours.

A very big thank you must also go to Dr Mark Fenton, associate professor, James Cook University, Townsville Queensland, for his chapter contributions and feedback; to Peta Dzidic for her assistance in preparation; and to you, our readers, for your continued support.

Dr Sheridan Coakes, Director, Coakes Consulting, PO Box 1027, Kalamunda WA 6926 Dr Lyndall Steed, Senior Lecturer, School of Psychology, Curtin University of Technology, Perth WA 6001 September 2004



# At-a-glance text examples

Working exa	mples
Work 2	Individual's shopping behaviour
Work 3	Community resident attitudes towards physical exercise
Work 4	Gender differences in tennis performance
Work 5	Relationship between intelligence (IQ) and tertiary entrance examination scores (TEE) among potential university students
Work 6	Comparisons of car engine efficiency
Work 7	Comparisons of utility cost by city
Work 8	Efficacy of weight-reduction programs
Work 9	Toy store sales by store type and location
Work 10	Practice effects on anagram completion times
Work 11	Manager productivity across store type and location
Work 12	Influence of age on reaction time
Work 13	Efficacy of new treatment program for depression
Work 14	Gender differences in sales representatives
Work 15	Internal consistency of a help-seeking instrument
Work 16	Attitudes towards organ donation
Work 17	Effect of shelf space and price on the sale of pet food
Work 18	Influence of attitudes, feelings and exposure on organ donation behaviour
Work 19a	Attitudes towards US military bases in Australia
Work 19b	Reader publication preferences and geographic location
Work 19c	Comparison of productivity across factories
Work 19d	Comparison of factory productivity by time period
Work 19e	Evaluation of different sales training programs
Work 19f	Reaction times across drug conditions
Work 19g	Relationship between sales performance and employee income
Work 20a	Criteria used in selection of rams for breeding
Work 20b	Native trees growing in gardens
Work 21	Tourist perceptions of distance between Australian cities
Practice exam	mples
Practice 2	Attitudes towards new school opening
Practice 3	Adolescent attitudes towards their future
Practice 4	Differences in white goods sales across junior and senior sales staff
Practice 5	Relationship between smoking and lung damage
Practice 6	Hypnosis and memory recall
Practice 7	Comparison of nutrient value of varying food supplements in rats
Practice 8	Comparison of weight gain in rats across varying food supplements

road accidents

Practice 9

Influence of density of traffic and type of intersection on number of

Practice 10	Differences in house sales according to suburb and time of year
Practice 11	Influence of colour and background on visual aesthetics
Practice 12	Effect of caffeine on motor task performance
Practice 13	Influence of auditorium size and sound-proofing on quality of acoustics
Practice 14	Influence of peer tutoring on level of computer anxiety
Practice 15	Reliability of independent personality scales — hope, optimism, locus of control and self-esteem
Practice 16	Factor structure of independent personality scales — hope, optimism, locus of control and self-esteem
Practice 17	Relationship between the volume of wood and forest characteristics (number of trees, average age of trees, average volume of trees)
Practice 18	Gender differences across personality measures
Practice 19a	Customer preference for specialty coffee bags
Practice 19b	Gender differences in drink preference
Practice 19c	Relationship between personality types and religious affiliation
Practice 19d	Effect of temperature on pilot performance
Practice 19e	Number of injuries across different sports
Practice 19f	Companionability of varying dog breeds
Practice 19g	Effect of load on truck fuel consumption
Practice 20a	Reasons for getting married
Practice 20b	Factors influencing purchase of a used car
Practice 21	Tourist perceptions of similarity among distance between nine cities

*Note*: there are no working or practice examples for chapters 1 and 22.



## Introduction to SPSS

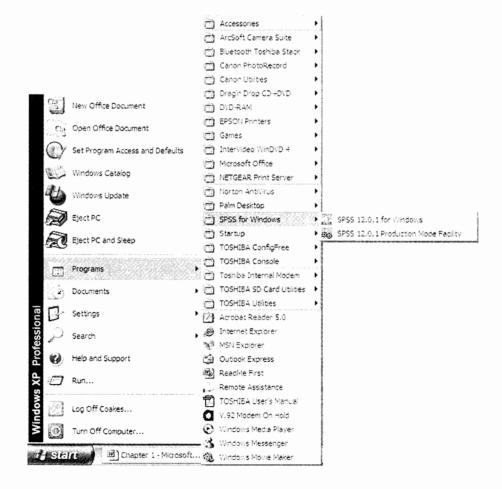
This chapter provides an introduction to using SPSS version 12.0 for Windows. It addresses aspects of the SPSS environment, describes the menu options and toolbars, and provides instructions on how to begin and end an SPSS session.

## Getting started



If this is one of your first experiences with the SPSS package, do not be put off. Before long, you will be able to manoeuvre around the package with ease and carry out all kinds of analytical procedures. Now, it's time to familiarise ourselves with the SPSS program and its attributes.

When SPSS is initially installed, the SPSS program group is created in the **Programs** menu. To start an SPSS session, double-click on the SPSS icon.



## The SPSS environment



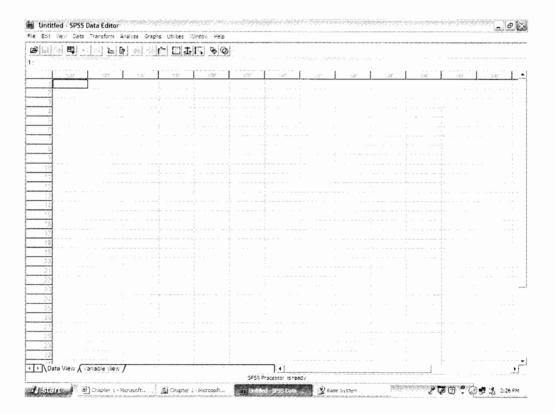
SPSS for Windows provides a powerful statistical analysis and data management system in a graphical environment, using descriptive menus and simple dialogue boxes to do most of the work for you. Most tasks can be accomplished simply by pointing and clicking the mouse.

In addition to the simple point-and-click interface for statistical analysis, SPSS has eight different types of windows:

- · Data Editor
- Viewer
- · Draft Viewer
- · Pivot Table Editor
- · Chart Editor
- Text Output Editor
- · Syntax Editor
- · Script Editor

#### **Data Editor**

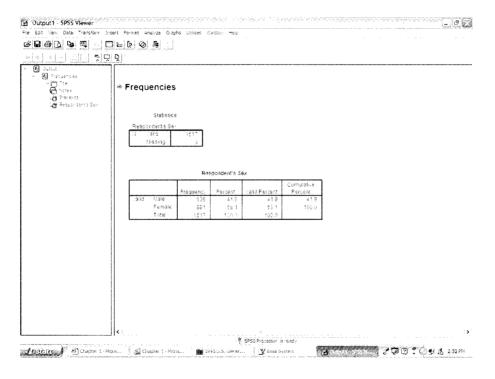
The Data Editor is a versatile spreadsheet-like system for defining, entering, editing and displaying data. This window opens when you start an SPSS session and displays the contents of a data file. In this window you can create new data files or modify existing data files. As outlined, the data editor is like a spreadsheet where cases are represented in rows and variables are represented in columns.



#### Viewer and Draft Viewer

The Viewer makes it easy to browse your results, selectively show and hide output, change the display, order results and move presentation-quality tables and charts between SPSS and other applications. This window opens automatically the first time you run a procedure that generates some output. The window displays all statistical results, tables and charts and allows you to edit the output and save it in an output file for later use. This window also allows you to access the Pivot Table Editor, Text Output Editor and Chart Editor and to move between SPSS and other applications, e.g. Word.

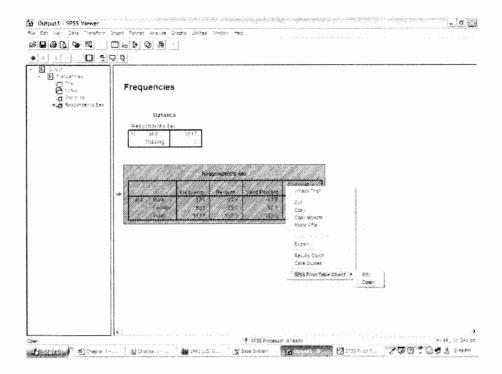
The window is divided into two panes. The left pane, referred to as the outline pane, contains an outline view of the output contents and can be used to navigate through your output and control the output display — very handy if you have a lot of output. The right pane of the window is the contents pane, which contains statistical tables, charts and text output. If you select an item in the outline pane, the corresponding item in the contents pane is highlighted. Moving an item in the outline pane moves the corresponding item in the contents pane. The width of the outline pane can also be changed by clicking on the right-hand border and dragging it to the desired width.



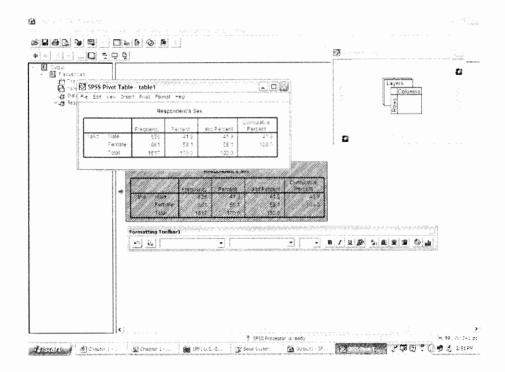
You can also display output as simple text (instead of in interactive pivot tables) in the Draft Viewer.

#### **Pivot Table Editor**

Output displayed in pivot tables can be modified in a number of ways. Using this editor, it is possible to edit text, rearrange rows, columns and layers, add colour, create multidimensional tables and selectively hide and display results. You move into the Pivot Table Editor from the Viewer by selecting the table you want to edit and clicking the right mouse button to open the SPSS Pivot Table Object.

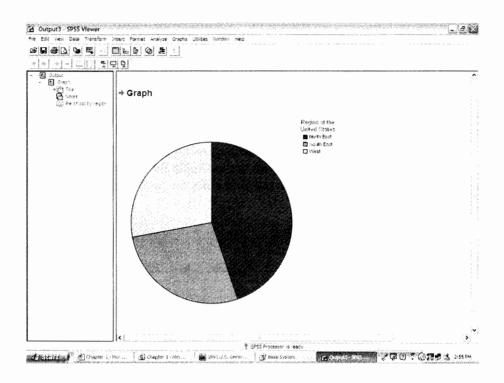


This will reveal the Pivot Table, formatting toolbar and pivot trays and allow editing of tables to begin.

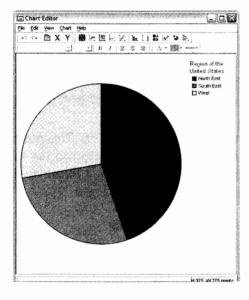


#### **Chart Editor**

High-resolution, full-colour pie charts, bar charts, histograms, scatterplots, 3-D graphics and more are included as standard features in SPSS. These can be edited in the Chart Editor. Changes in colour, font, axes, rotations and chart types can also be made using the Chart Editor.



You move into the Chart Editor from the Viewer by selecting the chart and clicking the right mouse button to open the SPSS Chart Object. This will reveal the Chart Editor and allow editing of charts to begin.



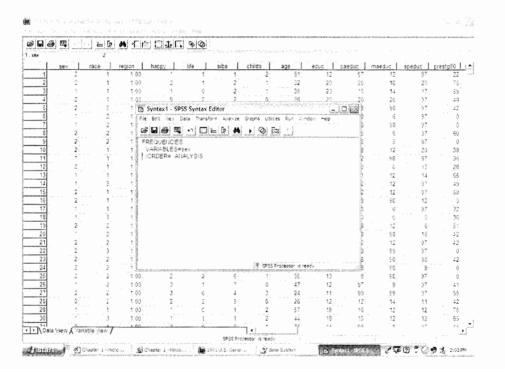
#### **Text Output Editor**

Text in the Viewer that is not displayed in pivot tables or charts can also be modified using the Text Output Editor. Possible modifications include changes to font characteristics such as colour, type, style and size.

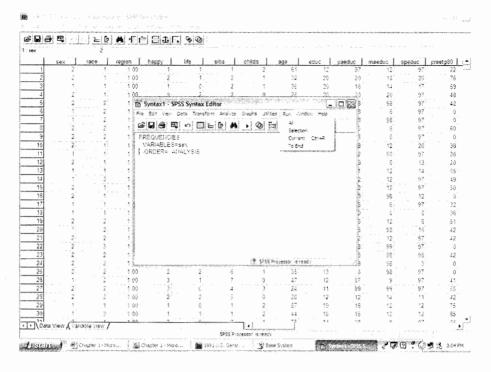
#### **Syntax Editor**

Although most tasks can be accomplished by simply pointing and clicking, SPSS also provides a powerful command language, which allows you to save and automate many common tasks. The command language also provides some functionality not found in the menus and dialogue boxes.

As you undertake particular procedures in SPSS, you can paste your dialogue box choices into a syntax window. This window contains the command syntax that reflects the choices you have made in selecting menu options. These commands can then be edited to include special features not available through the pull-down menus and dialogue boxes, and they can be saved in a file for further use.



At the top of the syntax window there is a menu titled **Run**. This menu allows you to process the commands you have pasted. You choose whether to run the current command, all the commands currently in the syntax window, a selection of commands or just to the end of the command syntax.



If you have decided to paste your syntax in this manner, you will need to use the **Run** menu to obtain your output and move into the Viewer window.

#### **Script Editor**

Scripting and OLE automation allow you to customise and automate many tasks in SPSS. With the Script Editor, you are able to create and modify basic scripts within the program.

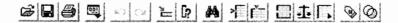
At the top of each window you will have noticed a menu and an icon bar. Let's learn more about these particular features, which provide quick and easy access to the special features available in each of the SPSS windows.

### Toolbar



The toolbar, located just below the menu bar, provides quick, easy access to many frequently used features. The toolbar contains tools that are available when a particular type of window is active. Each window has its own toolbar. When you put the mouse cursor on a tool in an active window, a brief description of that tool is displayed.

Tools in the Data Editor include:



- File Open allows particular data files to be opened for analysis.
- File Save saves the file in the active window.
- File Print prints the file in the active window.
- **Dialogue Recall** displays a list of recently opened dialogue boxes, any of which can be selected by clicking on their name.
- Go to Case allows the typing in of the number of the case you want to go to and it will be found in the data file.
- Variables provides data definition information for all variables in the working data file.

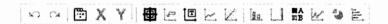
- Find allows data to be found easily within the data editor.
- Insert Cases inserts a case above the case containing the active cell.
- **Insert Variable** inserts a variable to the left of the variable containing the active cell
- **Split File** splits the data file into separate groups for analysis based on the values of one or more grouping variables.
- Weight Cases gives cases different weights for statistical analysis.
- Select Cases provides several methods for selecting a subgroup of cases based on criteria that include variables and complex expressions.
- Value Labels allows toggling between actual values and value labels in the Data Editor.
- Use Sets allows the selection of sets of variables to be displayed in the dialogue boxes

Some of the tools in the Viewer include:



- Print Preview allows what will be printed to be viewed.
- Export Output saves pivot tables and text output in HTML, text, Word/RTF, and
  Excel format, and it saves charts in a variety of common formats used by other
  applications.
- Go to Data moves directly into the data file and makes the Data Editor window active.
- Insert Heading, Title, Text allows headings, titles and text to be added into your output.
- Show/Hide allows output to be shown or hidden.

The Chart Editor has a range of tools that can be used to make your charts more interpretable and attractive, including:



- Show Properties Window shows the properties of the chart.
- Insert a Text Box allows the insertion of a text box.
- Show/Hide Legend allows the legend on a chart to be shown or hidden.
- Show Data Labels allows the data labels to be shown on the chart.

In the Syntax Editor Window the tools are similar to those in the Data Editor, with a few additions:



- **Run Current** runs commands that are selected or, if there is no selection, the command in which the cursor appears.
- Syntax Help assists you with the syntax for the analysis you are undertaking.

### Menus



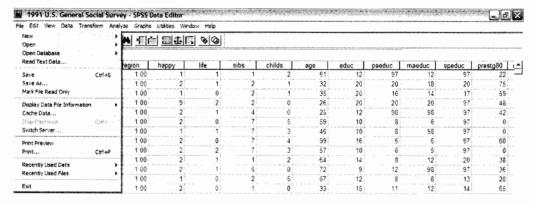
SPSS is menu driven and has a variety of pull-down menus available for the user. The main menu bar in the Data Editor contains ten menus:

- File
- Edit
- View
- Data
- Transform
- Analyze
- Graphs
- Utilities
- Window
- · Help.

The **Analyze** and **Graphs** menus are available in all windows, making it much easier to generate new output without having to switch between windows.

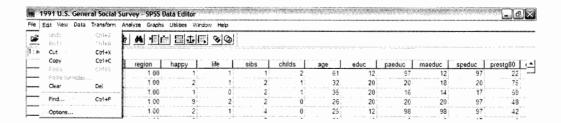
#### **File**

The **File** menu allows you to create new files, open existing files, read in files from other software programs, save files and print.



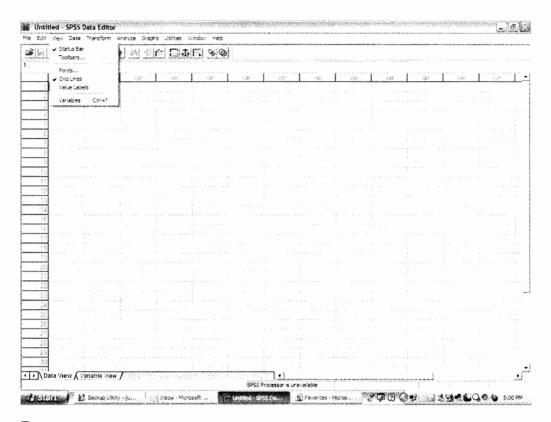
#### **Edit**

The **Edit** menu allows you to modify or copy text from the output or syntax windows, and to search for and replace text or data. It also offers a number of personal preference options.



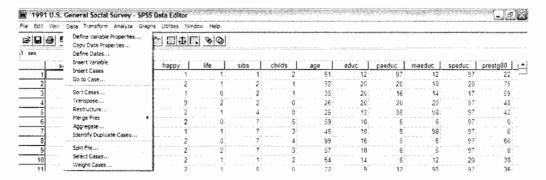
#### **View**

The **View** menu allows you to make the status bar and toolbar active, and to change particular characteristics of the window (for example, by removing grid lines, displaying value labels and changing font style and size).



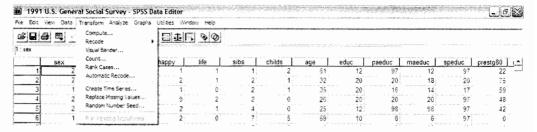
#### Data

The **Data** menu allows you to define variables and create variable templates. In addition, more global changes to SPSS data files are available, such as merging files, inserting, sorting and transposing variables and cases, and selecting and weighting cases.



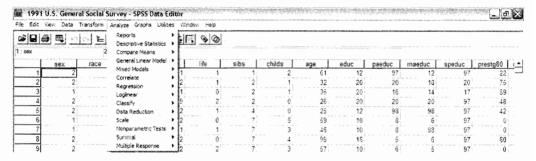
#### **Transform**

The **Transform** menu allows you to change certain variables in the data file using commands such as **Recode** and **Rank Cases**, as well as to create new variables using the **Compute** command.



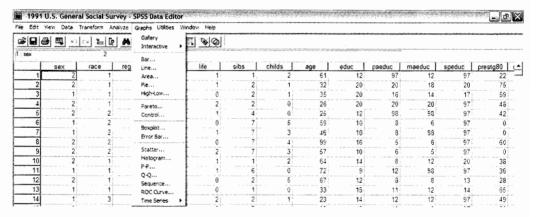
#### **Analyze**

The **Analyze** menu allows you to select the analysis you require. A variety of statistical procedures are available, ranging from summarising data through to more complex designs.



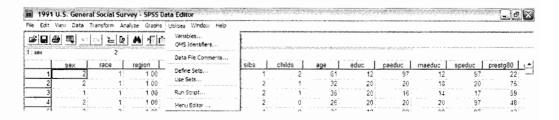
#### Graphs

The **Graphs** menu allows you to create bar, line, area and pie charts, as well as histograms and scatterplots.



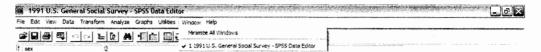
#### **Utilities**

The **Utilities** menu allows you to display file and variable information. In addition, it allows you to define and use different variable sets.



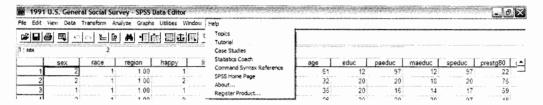
#### Window

The **Window** menu allows you to arrange, select and control the attributes of the various windows. Using this menu, you can move efficiently between data, syntax, output and chart windows.



#### Help

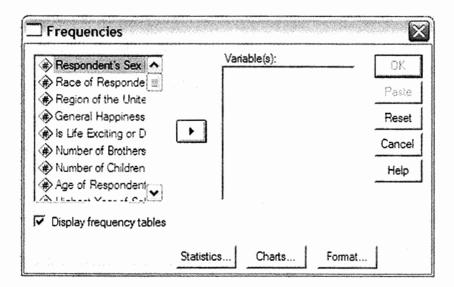
The **Help** menu allows you to access information on how to use the many features of SPSS. The SPSS tutorial can be accessed through the **Help** menu.



## Dialogue boxes for statistical procedures



When you choose a statistical procedure, a dialogue box appears on the screen. Each main dialogue box has four basic components: the source variable list, the target variable(s) list, command pushbuttons and the option to choose sub-dialogue boxes.



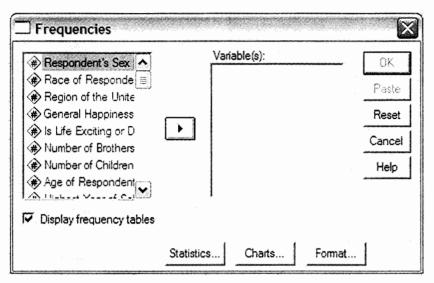
#### Source variable list

The source variable list is a list of all the variables in the data file.



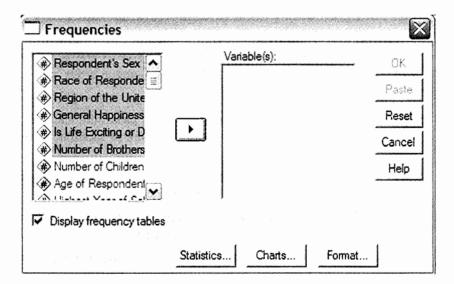
#### To select a single variable

Highlight the variable and click on the ▶ button next to the selected variable list box.



#### To select multiple variables that are grouped together in the variable list

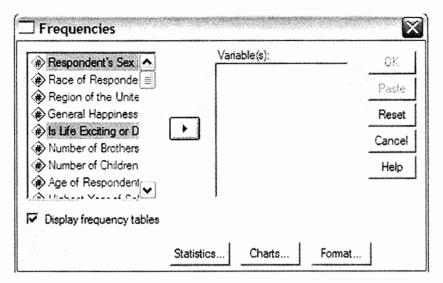
Click on the first variable you wish to select, then hold the Shift key and click on the last variable in the group. Then click on the **>** button.





#### To select multiple variables that are not grouped together

Click on the first variable, then hold the Ctrl key and click on the next variable, and so on. Then click on the ▶ button.



#### Selected variable(s) list

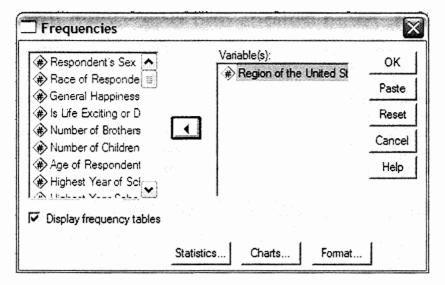
The selected variable(s) list is a list or lists of variables you have chosen for certain analyses. For certain statistical procedures, both dependent and independent variable lists are created.

If you wish to remove variables from this list, highlight the variable and click on the button, which you will notice is now reversed. This process is called deselection of variables.

#### **Command pushbuttons**

The five standard command pushbuttons in most dialogue boxes are:

- **OK** Runs the procedure.
- Paste Pastes the syntax associated with a procedure into the Syntax Editor window. This command syntax can then be modified if necessary or new syntax can be added.
- **Reset** Deselects the variables in the selected variable list(s) and resets all specifications in the dialogue and sub-dialogue boxes.
- Cancel Cancels any changes in the dialogue box settings since the last time it was opened, then closes the dialogue box.
- Help Allows access to a Help window relevant to the current procedure.

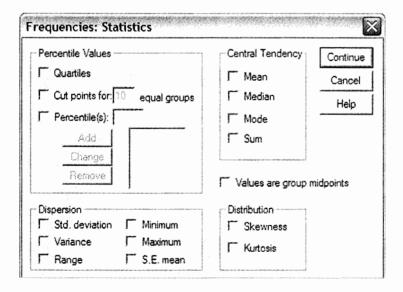


#### Accessing sub-dialogue boxes

When selecting a statistical procedure, you can make additional specifications that are available in the sub-dialogue boxes. These are accessed by selecting the buttons at the bottom or on the side of the main dialogue box. These buttons may include: **Statistics...**, **Charts...**, **Format...**, **Options...**, **Save...**, **Plots...**, **Cells...**, etc.

As outlined in the screen dump below, the three command pushbuttons within subdialogue boxes are:

- Continue Saves any changes to the settings and returns to the main dialogue box.
- Cancel Ignores any changes, restores the previous settings and returns to the main dialogue box.
- Help Allows access to a Help window relevant to the current procedure.



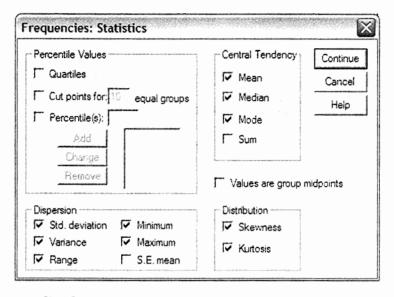
## Check boxes, radio buttons and drop-down lists



Within sub-dialogue boxes, choices can be made using check boxes, radio buttons and drop-down lists.

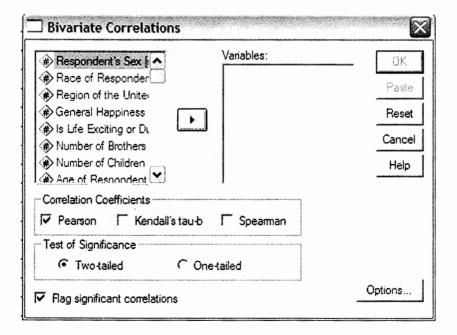
#### Check boxes

Check boxes ( $\square$ ) allow you to select certain options within sub-dialogue boxes. When you click on the box,  $\checkmark$  is displayed in the check box  $\boxdot$ . To deselect this option, click the box again. Multiple check boxes can be selected if required.



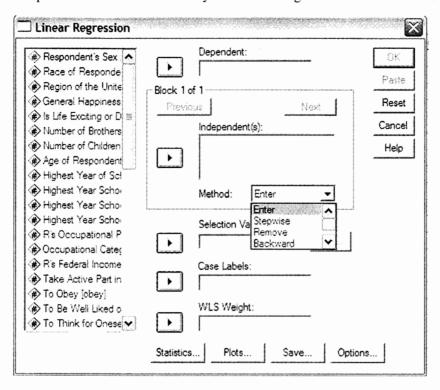
#### Radio buttons

Radio buttons ( $\bigcirc$ ) allow you to make single selections within sub-dialogue boxes. When you click on the radio button, a solid black circle appears in the centre of the button  $\bigcirc$ .



#### **Drop-down lists**

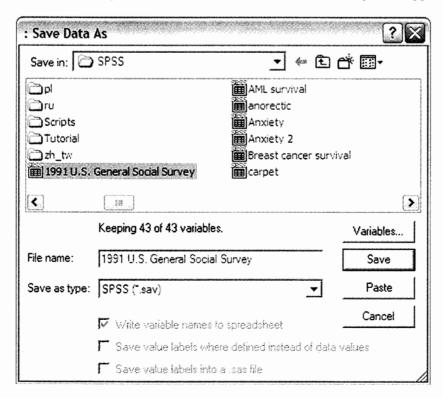
Drop-down menus or lists allow you to make single selections from a list of alternatives.



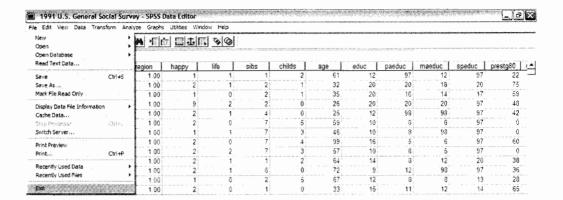
## Saving files and ending a session



To save files in SPSS, the **Save As...** command is selected from the **File** menu. If you are working in the Data Editor window, then the file extension .sav will be displayed in the **Save as type:** box and you will be prompted to type in a filename in the **Filename:** box. If you are in the Viewer window, then the extension .spo will be displayed. And if you are in the Syntax Editor window, then the extension .sps will appear.



To end a session, select the **File** menu and the **Exit** option. If you exit a session without first saving your files, then SPSS will prompt you to save the contents of each window.



# Preparation of data files

This chapter describes the process involved from data source to data file: that is, the conversion of raw source material to a useable data file. It focuses on defining variables, assigning appropriate numeric codes to alphanumeric data and dealing with missing data. These preparatory steps are desirable before data entry can begin. Other procedures such as applying variable definition attributes to other variables, entering data, inserting and deleting cases and variables, saving data and opening existing data files will also be addressed.

## Working example



You have developed a questionnaire that asks a number of questions relating to an individual's shopping behaviour. The variables you have measured include: gender, age, desire for 24-hour shopping, choice of shopping area and amount spent on groceries per week. For each case you have assigned a participant identification number.

## Defining variables



The process of defining variables has seven optional steps. The primary step is naming your variable and the other steps cover labels (variable and value), missing values, variable type, column format and measurement level.



#### Naming a variable

Variable names must comply with certain rules:

- Must begin with a letter. The remaining characters can be a letter, any digit, a full stop or the symbols @, #, \_ or \$.
- · Cannot end with a full stop or underscore.
- Blanks and special characters cannot be used (for example, !, ?, ', and \*).
- Must be unique; duplication is not allowed.
- The length of the name cannot exceed 64 bytes (typically 64 characters in single-byte languages e.g. English, German, French, Spanish etc.).
- Reserved keywords cannot be used as variable names e.g. ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.
- Names are not case sensitive, i.e. can be written in upper or lower case.
- Long variable names need to wrap to multiple lines in the output, so SPSS attempts to break lines at underscores, periods, and a change from lower to upper case.

In the case of the variable *age*, the variable name can be *age* because this name complies with all of the rules listed. The variable *choice of shopping area* can be labelled *area*, and the other variables on the questionnaire could be *id*, *gender*, *allday* and *cost*.

#### Variable labels

The variable label is the full description of the variable name and is an optional means of improving the interpretability of the output. For example, the first variable you will name in the data file is *id* and the label for this variable is 'participant identification number'. Suggested variable labels for the other variables appear in the following table.

Variable name	Label
gender	optional
age	optional
allday	desire for 24-hour shopping facilities
area	choice of shopping area
cost	amount spent on groceries per week

You will notice that the gender, age and choice of shopping area variables do not require variable labels because the variable names are self-explanatory.

#### Value labels

It is possible to use alphanumeric codes for the variables; however, you may also wish to use a numeric code. For example, for gender you could assign a code of 1 for female and 2 for male. This type of variable is categorical because it has discrete categories. The variables *allday* and *area* are also categorical. When variables are measured using interval or ratio scales, then coding is not relevant unless categorisation is required.

Value codes and labels for the above variables are illustrated in the following table.

Variable name	Label
id	not applicable
gender	1 = female 2 = male
age	not applicable
allday	1 = would use 24-hour shopping 2 = would not use 24-hour shopping
area	<ul><li>1 = shop in suburb where living</li><li>2 = travel to next suburb</li><li>3 = travel further to shop</li></ul>
cost	not applicable

#### Missing values

It is rare to obtain complete data sets for all cases. When dealing with missing data you may leave the cell blank or assign missing value codes. If you choose the latter, then a number of rules apply:

- Missing value codes must be of the same data type as the data they represent. For example, for missing numeric data, missing value codes must also be numeric.
- Missing value codes cannot occur as data in the data set.
- By convention, the choice of digit is usually 9.

#### Variable type

By default, SPSS assumes that all new variables are numeric with two decimal places. However, it is possible to select other variable types (such as date, currency, string) and vary the number of decimal places.

#### Column format

It is possible to adjust the width of the Data Editor columns or change the alignment of data in the column (left, centre or right).

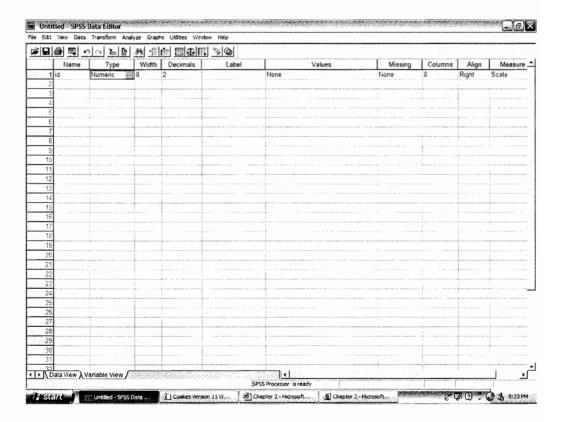
#### Measurement level

You can specify the level of measurement as scale (interval or ratio), ordinal or nominal.



#### To define a variable

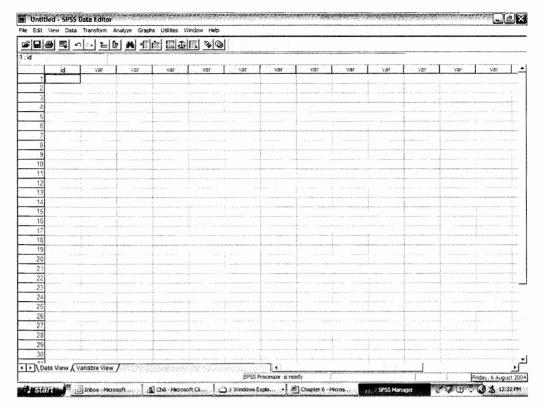
- Working in the Untitled SPSS Data Editor window, double-click a variable name at the top of the column in the Data View or click the **Variable View** tab.
- In the first blank cell of the **Name** column, type the first variable name (i.e. *id*) and press Enter.



3 In the first blank cell of the **Label** column, type the label for the variable, i.e. *Identification Number*. You will notice that the width of the column automatically increases to accommodate the long label.

For the variable *id*, there are no value labels or missing values, and the other properties are appropriate so you can move on to the second variable.

If you return to the Data View, by clicking the **Data View** tab you will notice that the variable name *id* has appeared in the first column as a heading.



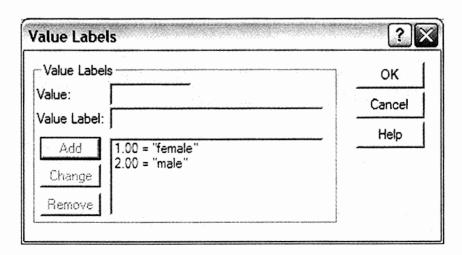


#### To repeat this process for the second variable, gender

Working in the Variable View, in the second blank cell of the **Name** column, type the second variable name, i.e. *gender*, and press Enter.

Again, SPSS automatically supplies other properties such as type, width, values, etc. Since gender requires no further explanation a label will not be typed in. However, values are assigned.

- Click on the second cell of the **Values** column and then on the button on the right to open the **Value Labels** box. In the **Value:** box, type the first value code for the variable (i.e. *I*) then tab. In the **Value Label:** box, type the label for this value, i.e. *female*.
- 3 Click on **Add**. You will notice that the value and its label have moved into the box below.
- 4 Repeat this process for the second value.

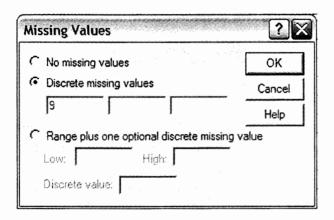


5 Click on **OK**. You will notice that there is now information in the cell.

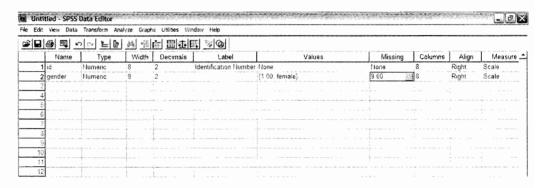


#### To create a missing value

- Click on the second cell of the **Missing** column and then on the button on the right to open the **Missing Values** dialogue box.
- 2 Select the **Discrete missing values** radio button.
- 3 In the first box, type the missing value code, i.e. 9.



4 Click on **OK**. You will notice that the missing value has been recorded.



The previous process is then repeated for each variable you wish to define in your data file. As highlighted earlier in this chapter, other options such as **Type**, **Column Width** and **Measurement Level** are also available if you are dealing with different types of variable that require special conditions. For the variable *gender*, you may wish to select the nominal scale of measurement.

### Applying variable definition attributes to other variables

Once you have nominated variable definition attributes for a variable you can copy one or more attributes and apply them to one or more variables. Basic copy and paste operations are used to apply variable definition attributes. For example, you may have several variables with a Likert scale response format. That is, you may have several variables that use the same response scale where 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree and 5 = strongly agree. Having defined these value labels for one variable you can then copy them to other variables.



#### To apply variable definition attributes to other variables

- In the Variable View, select the attribute cell(s) you want to apply to other variables.
- 2 From the **Edit** menu click on **Copy**.
- 3 Select the attribute cell(s) to which you want to apply the attribute(s). You can select multiple target variables.

4 From the **Edit** menu click on **Paste**.

If you copy attributes to blank rows, then new variables are created with default attributes for all but the selected attributes.

## **Entering data**

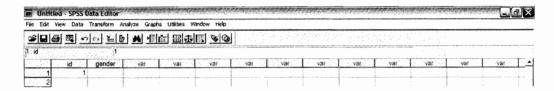




To enter the following two cases

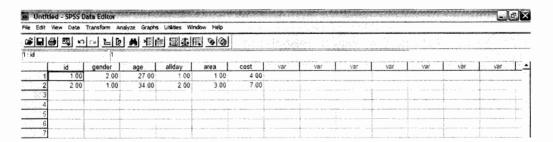
id	gender	age	allday	area	cost
1	male	27	1	1	4
2	female	34	2	3	7

- In the Data View click on the first cell in the **Untitled SPSS Data Editor** window. You will notice that a heavy border appears around the cell, indicating that the cell is now active.
- 2 Type in the first value for *id*, i.e. 1. This value is displayed in the cell editor at the top of the **Data Editor** window and also in the cell itself.



- 3 Press Enter or move to another cell by using the arrow keys or mouse. Data values are not recorded until you press Enter or select another cell. Remember that it is more efficient to code gender numerically.
- 4 To move around your data file quickly, you can hold down the Control key with an arrow key to take you to the limit of the file in that direction.

Having entered data for the first two cases, your **Data View** window will look like this:



## Inserting and deleting cases and variables



Often you may need to insert or delete extra cases (rows) and variables (columns) in the existing data file. You can achieve this by using the menus as described below or by using the appropriate tools from the toolbar.



#### To insert a new case between existing cases

- 1 Select any cell in the case (row) below the position where you want to insert the new case.
- Select the Data menu and click on Insert Case or click on the Insert Case tool. A new case (row) will be inserted.



#### To insert a new variable between existing variables

- 1 Select any cell in the variable (column) to the right of the position where you want to insert a new variable.
- 2 Select the **Data** menu and click on **Insert Variable** or click on the **Insert Variable** tool. A new variable (column) will be inserted.



#### To delete a case (row)

- 1 Click on the case number on the left side of the row if you wish to delete the entire case or select any cell in the row that you wish to delete.
- Select the **Edit** menu and click on **Clear**. Alternatively, you can use the Delete button on the keyboard.



#### To delete a variable (column)

- 1 Click on the variable name at the top of the column if you wish to delete the entire variable, or select any cell within the column that you wish to remove.
- Select the **Edit** menu and click on **Clear**. Again, you can use the Delete button if you wish.

## Moving variables



You may wish to change the sequence of variables in the **Data Editor** window. If you want to position the variable between two existing variables, then insert a new variable in the position where you want to move the variable.



#### To move a variable

- 1 For the variable you want to move, click the variable name at the top of the column in the Data View or the row number in the Variable View. The entire variable is highlighted.
- 2 Select the **Edit** menu and click on **Cut**.
- 3 Click the variable name in the Data View or the row number in the Variable View where you want to move the variable to. The entire variable is highlighted.
- 4 Select the Edit menu and click on Paste.

## Saving data files





#### To save a data file for the first time

- 1 First ensure that you are in the **Data Editor** window.
- 2 Click on the **File Save** tool.
- 3 You will be asked to give the file a name.

or

- 1 Select the **File** menu and click on **Save As...** to open the **Save Data As** dialogue box.
- In the box for **File Name:** type in the file name of your choice. SPSS will append the extension .sav automatically. If you are saving data to a floppy disk, then remember to change to the appropriate drive.
- 3 Click on Save.



To save an existing data file

- First ensure that you are in the **Data Editor** window.
- 2 Click on the File Save tool.

or

1 Select the **File** menu and click on **Save Data**. Your changes will be saved to the existing file.

# Opening an existing data file



Once a data file has been saved in SPSS it can be accessed in subsequent sessions. Furthermore, files that have been created in other software packages can be imported into SPSS.



To open an existing SPSS data file

- 1 Select the **File** menu.
- 2 Click on Open and Data... to open the Open File dialogue box.
- 3 Select the file from the file list.
- 4 Click on Open.



To read a text data file

- 1 Select the **File** menu.
- 2 Click on **Read Text Data** to open the **Open File** dialogue box.
- 3 Select the file from the file list.
- 4 Click on Open.
- 5 Follow the steps in the Text Wizard to define how to read the data file.

# Practice example



You have surveyed individuals within your local community to determine their attitude towards the opening of a new school. You have collected data on the variables in the following table.

id	gender	length of residence	number of children	would you use the school?
1	Female	2	2	Yes
2	Male	3	1	No
3	Male	6	2	Yes
4	Male	5	4	No
5	Female	8	3	Yes
6	Female	9	2	Yes
7	Female	11	2	Yes
8	Male	3	0	Yes
9	Male	5	2	Yes
10	Female	12	3	Yes
11	Male	10	1	No
12	Female	8	1	Yes
13	Female	9	4	Yes
14	Male	9	2	Yes
15	Male	3	0	No
16	Female	1	3	Yes
17	Male	0.5	1	Yes
18	Female	4	2	Yes
19	Male	3	1	ś
20	Female	2	1	Yes

Given the above data, your task is to create a data file in SPSS. Remember, you must define each variable, allocating appropriate variable and value labels. Other definitions are optional. Data must then be entered and saved in a data file.



# Data screening and transformation

Data screening and transformation techniques are useful in making sure that data have been correctly entered and that the distributions of variables that are to be used in analysis are normal. If variable distributions deviate dramatically, then this may affect the validity of the results that are produced. If distributions do vary from normal, then non-normal distributions may be transformed before further analysis. Furthermore, if distributions deviate dramatically, nonparametric techniques may also be used because they are less powerful than their parametric counterparts.

Data may also need to be transformed using **Recode** and **Compute** commands. In addition, if data files have missing values, mean substitution may be an alternative.

It is often useful to be able to conduct analyses on subsets of the data and to make conditional transformations of variables. These can be achieved using the **Select If** and **Compute If** commands.

These procedures will be demonstrated in the context of the following research example.

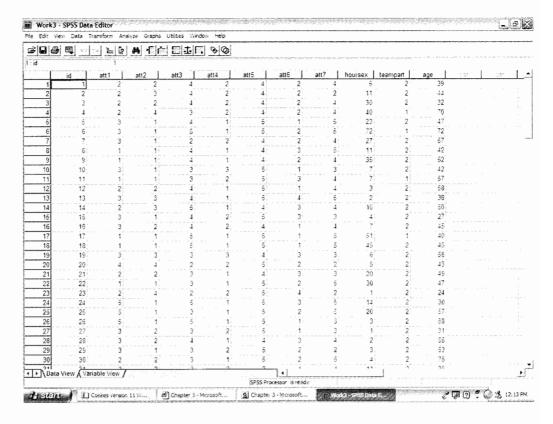
# Working example



Community residents were surveyed to determine their attitudes towards physical exercise. Participants were also asked their age, number of hours per week spent in physical activity and whether they participated in team sports.

Each individual (of the 99 students who participated) was given an identity number. Attitude towards exercise was measured using a seven-item scale comprising statements to which participants agreed or disagreed, using a five-point Likert response format such that: 1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree and 5 = Strongly disagree.

The data file can be found in Work3.sav on the website that accompanies this title and is shown in the following figure. Several items of this scale were negatively worded and thus required recoding.



You wish to determine whether the distributions of the continuous variables (that is, age, hours spent doing exercise and attitude to exercise) are normal. This procedure is not necessary for team participation because this is a categorical variable.

# Errors in data entry

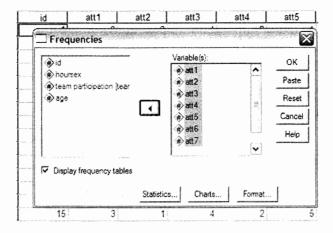


Errors in data entry are common and therefore data files must be carefully screened. For example, while responses to the attitude items are being entered, out-of-range values can be detected easily using the **Frequencies** or **Descriptives** commands and replaced in the data file with the correct value.



#### To obtain frequencies

- 1 Select the **Analyze** menu.
- 2 Click on Descriptive Statistics and then Frequencies... to open the Frequencies dialogue box.
- 3 Select *att1* to *att7* and click on the ▶ button to move these variables into the **Variable(s):** box.



#### 4 Click on OK.

FREQUENCIES
VARIABLES=att1 att2 att3 att4 att5 att6 att7
/ORDER= ANALYSIS.

You will observe that all cases for *att1* are within the expected range of 1–5. You must ensure the other items are also within the expected range.

#### att1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	22	22.2	22.2	22.2
	agree	31	31.3	31.3	53.5
	neither agree nor disagree	38	38.4	38.4	91.9
	disagree	3	3.0	3.0	94.9
	strongly disagree	5	5.1	5.1	100.0
Total		99	100.0	100.0	

# Assessing normality



The assumption of normality is a prerequisite for many inferential statistical techniques. There are a number of different ways to explore this assumption graphically:

- Histogram
- Stem-and-leaf plot
- Boxplot
- · Normal probability plot
- Detrended normal plot.

Furthermore, a number of statistics are available to test normality:

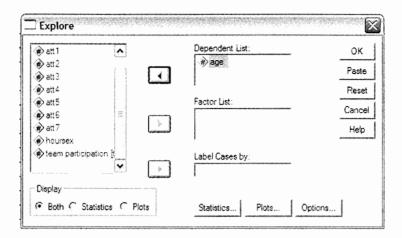
- Kolmogorov-Smirnov statistic, with a Lilliefors significance level and the Shapiro-Wilks statistic
- Skewness
- Kurtosis.

There are several procedures available to obtain these graphs and statistics but the **Explore** procedure is the most convenient when both graphs and statistics are required.

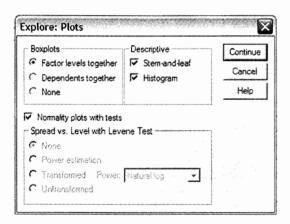


To obtain these graphs and statistics

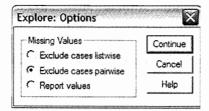
- 1 Select the **Analyze** menu.
- 2 Click on **Descriptive Statistics** and then **Explore...** to open the Explore dialogue box.
- 3 Select the variable you require (i.e. *age*) and click on the ▶ button to move this variable into the **Dependent List:** box.



- 4 Click on the **Plots...** command pushbutton to obtain the **Explore: Plots** subdialogue box.
- 5 Click on the **Histogram** check box and the **Normality plots with tests** check box, and ensure that the **Factor levels together** radio button is selected in the **Boxplots** display.



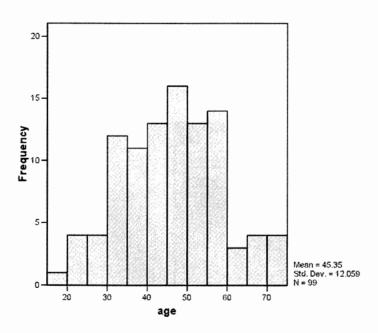
- 6 Click on Continue.
- 7 In the **Display** box, ensure that **Both** is activated.
- 8 Click on the **Options...** command pushbutton to open the **Explore: Options** sub-dialogue box.
- In the Missing Values box, click on the Exclude cases pairwise radio button. If this option is not selected then, by default any variable with missing data will be excluded from the analysis. That is, plots and statistics will be generated only for cases with complete data.



EXPLORE
VARIABLES=age
/PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING PAIRWISE
/NOTOTAL.

### **Histograms**

#### Histogram



Above is a histogram of *age*. The values on the vertical axis indicate the frequency of cases. The values on the horizontal axis are midpoints of value ranges. For example, the midpoint of the first bar is 20 and the midpoint of the second bar is 25, indicating that each bar covers a range of 5. The shape of the distribution is considered normal.

# Stem-and-leaf plots and boxplots

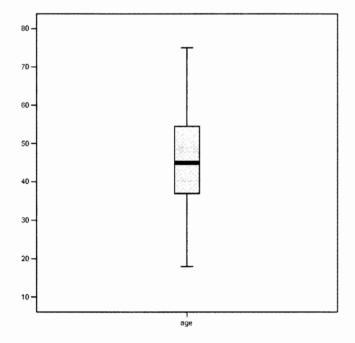
Closely related to the histogram is the stem-and-leaf plot and the boxplot. These plots provide more information about the actual values in the distribution than does the histogram.

The stem-and-leaf plot is very similar to the histogram but is displayed on its side. The length of each row corresponds to the number of cases that fall into a particular interval.

A stem-and-leaf plot represents each case with a numeric value that corresponds to the actual observed value. For example, the stem of the graph corresponds to the first digit of a score (2), while the leaf is the trailing digit (23347889).

```
age Stem-and-Leaf Plot
Frequency
             Stem & leaf
    1.00
               2 . 23347889
    8.00
                3 . 01122233344456777788999
   23.00
                    00001222233445555555556777889
   29.00
                5 . 000000022223455566677778888
    27.00
                6 . 0005566
    7.00
    4.00
                    0025
Stem width: 10
Each leaf:
                 1 case(s)
```

The boxplot, illustrated below, also summarises information about the distribution of scores. Unlike the histogram and the stem-and-leaf plot, which plot actual values, it plots summary statistics such as the median, 25th and 75th percentiles, and extreme scores in the distribution. The lower boundary of the box is the 25th percentile and the upper boundary is the 75th percentile. The median is presented by a horizontal line through the centre of the box. The smallest and largest observed values within the distribution are represented by the horizontal lines at either end of the box, commonly referred to as whiskers.



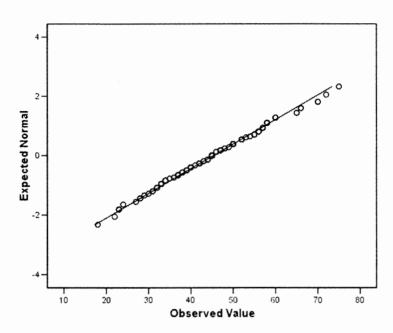
If the distribution has any extreme scores — that is, three or more box lengths from the upper or lower edge of the box — then these will be represented by an asterisk (\*). Cases with values between one-and-a-half and three box lengths from the upper or lower edge of the box are called outliers, and these are designated by a circle.

To determine whether a distribution is normal you look at the median that should be positioned in the centre of the box. If the median is closer to the top of the box, then the distribution is negatively skewed, and if it is closer to the bottom of the box, then it is positively skewed. The spread or variability of the scores can be determined from the length of the box.

## Normal probability plots and detrended normal plots

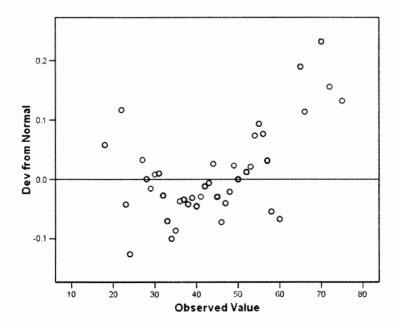
In a normal probability plot, each observed value is paired with its expected value from the normal distribution. If the sample is from a normal distribution, then the cases fall more or less in a straight line.

Normal Q-Q Plot of age



It is also possible to plot the actual deviations of the points from a straight line. If the sample is from a normal distribution, then there is no pattern to the clustering of points; the points should assemble around a horizontal line through zero. This type of plot is referred to as a detrended normal plot and is illustrated in the following figure.

Detrended Normal Q-Q Plot of age



### Kolmogorov-Smirnov and Shapiro-Wilks statistics

The Kolmogorov-Smirnov statistic with a Lilliefors significance level for testing normality is produced with the normal probability and detrended probability plots. If the significance level is greater than .05, then normality is assumed. The Shapiro-Wilks statistic is also calculated if the sample size is less than one hundred.

#### **Tests of Normality**

	Kolmogorov-Smirnov <sup>9</sup>			Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig.
age	.057	99	.200*	.992	99	.825

<sup>\*</sup> This is a lower bound of the true significance.

#### Skewness and kurtosis

Skewness and kurtosis refer to the shape of the distribution, and are used with interval and ratio level data. Values for skewness and kurtosis are zero if the observed distribution is exactly normal. Positive values for skewness indicate a positive skew, while positive values for kurtosis indicate a distribution that is peaked (leptokurtic). Negative values for skewness indicate a negative skew, while negative values for kurtosis indicate a distribution that is flatter (platykurtic). Other descriptive statistics, such as measures of central tendency and variability, can also be used to determine the normality of the distribution.

#### Descriptives

			Statistic	Std. Error
age	Mean		45,35	1.212
	95% Confidence	Lower Bound	42.95	
	Interval for Mean	Upper Bound	47.76	
	5% Trimmed Mean		45.25	
	Median		45.00	
	Variance		145.415	
	Std. Deviation		12.059	
	Minimum		18	
	Maximum		75	
	Range		57	
	Interquartile Range		18	
	Skewness		.126	.243
	Kurtosis		316	.481

You should be familiar with most of the above statistics. However, you may not have encountered the 5 per cent Trim statistic, which is the mean of the distribution with the top 5 per cent and the bottom 5 per cent of scores removed. The purpose of this trimming is to obtain a measure of central tendency that is unaffected by extreme values.

a. Lilliefors Significance Correction

# Assessing normality by group



It is sometimes necessary to assess the normality of a variable across two or more levels of another variable. For example, you may wish to assess the normality of age for team participants and non-team participants separately. This can be achieved using the preceding procedure with one addition. In the **Explore** dialogue box you will notice that there is a **Factor List:** box. By transferring your group variable (that is, team participation) into this box, the chosen statistics and plots will be generated for each group independently.

# Variable transformation



Variables rarely conform to a classic normal distribution. More often, distributions are skewed and display varying degrees of kurtosis. When skewness and kurtosis are extreme, transformation is an option.

The decision to transform variables depends on the severity of the departure from normality. Having decided that transformation is desirable, the researcher must select the most appropriate transformation methods. The options available can be found in any good chapter on data screening.

To illustrate the process of transformation, the hours of exercise variable will be examined using the above steps. Plots and normality statistics were obtained.

#### Descriptives

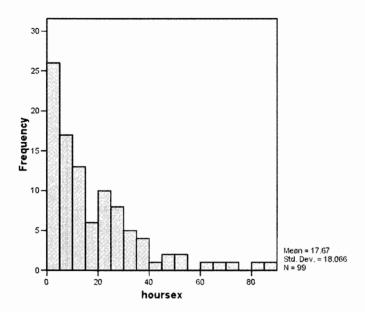
			Statistic	Std. Error
hoursex	Mean		17.67	1.816
	95% Confidence	Lower Bound	14.06	
	Interval for Mean	Upper Bound	21.27	
	5% Trimmed Mean		15.49	
	Median		11.00	
	Variance		326.367	
	Std. Deviation		18.066	
	Minimum		1	
	Maximum		87	
	Range		86	
	Interquartile Range		21	
	Skewness		1.770	.243
	Kurtosis		3.475	.481

#### **Tests of Normality**

	Kolmogorov-Smirnov <sup>s</sup>			Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig.
hoursex	.178	99	.000	.810	99	.000

a. Lilliefors Significance Correction





hoursex Stem-and-leaf Flot

```
Frequency Stem & Leaf

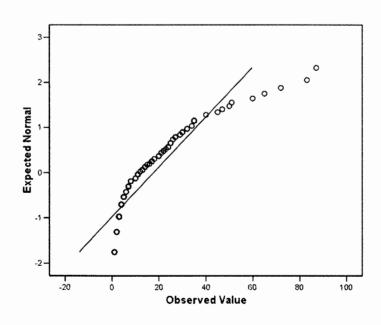
43.00 0 . 11122333334355567773
19.00 1 . 01134734
13.00 2 . 022455744
9.00 3 . 02354
3.00 6 . 4
2.00 5 . 6
5.00 Extremes (>=60)

Stem width: 10

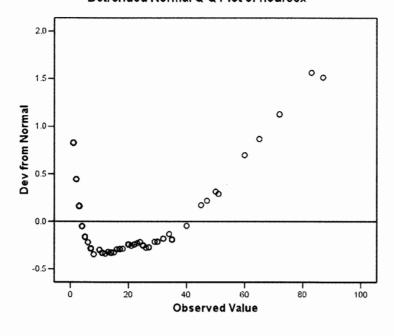
Each leaf: 2 case(s)
```

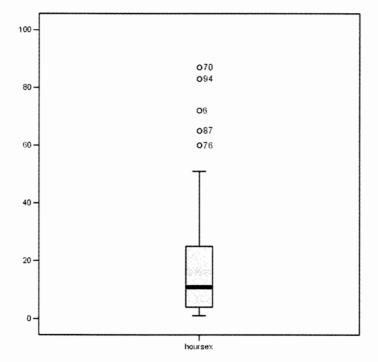
& denotes fractional leaves.

#### Normal Q-Q Plot of hoursex



#### **Detrended Normal Q-Q Plot of hoursex**





All the above charts and statistics suggest that the variable *hoursex* is not normally distributed but is significantly positively skewed. The boxplot indicates that there are five outliers, as illustrated by the circles. Therefore, a natural logarithmic transformation is appropriate.

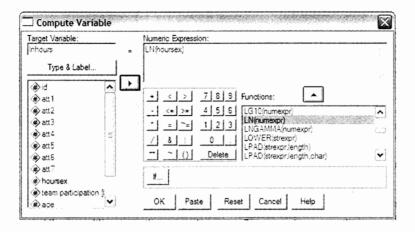
To transform the variable, you will need to use a data transformation command called **Compute**.



To compute values for a variable based on numeric transformations of other variables

- 1 Select the **Transform** menu.
- 2 Click on Compute to open a Compute Variable dialogue box.

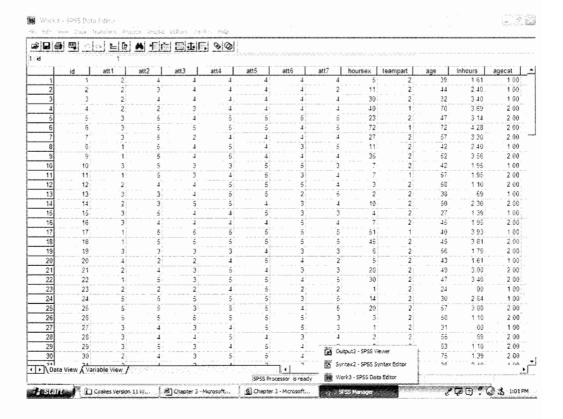
- In the **Target Variable:** box, where the cursor is flashing, type an appropriate variable name (i.e. *Inhours*).
- From the **Functions:** box, select the appropriate transformation (i.e. LN(numexpr)) and press the  $\triangle$  button.
- From the variable list box, select the variable (i.e. *hoursex*) and press the button to insert the variable in the function.



### 6 Click on OK.

COMPUTE Inhours = LN(hoursex) .
EXECUTE .

If you now return to your data file, you will see that a new variable has been created in the data file called *Inhours*. To make things quicker, use the SPSS Manager at the bottom of your screen to quickly move between your data and syntax editor windows and the SPPS Viewer, which displays your output.



Now that you have transformed the variable *hoursex* into *Inhours*, you can obtain normality graphs and statistics for this new transformed variable using the procedures outlined at the beginning of the chapter. This output appears below.

Descriptives

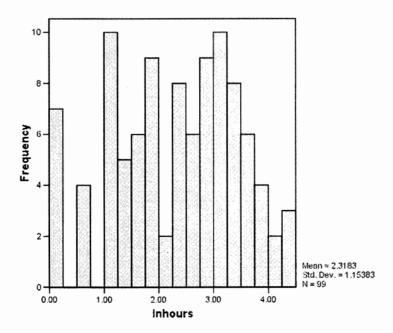
			Statistic	Std. Error
Inhours	Mean		2.3183	.11596
	95% Confidence	Lower Bound	2.0881	
	Interval for Mean	Upper Bound	2.5484	
	5% Trimmed Mean		2.3376	
	Median		2.3979	
	Variance		1.331	
	Std. Deviation		1.15383	
	Minimum		.00	
	Maximum		4.47	
	Range		4.47	
	Interquartile Range		1.83	
	Skewness		-,312	.243
	Kurtosis		706	.481

**Tests of Normality** 

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig.
Inhours	.095	99	.027	.965	99	.010

a. Lilliefors Significance Correction

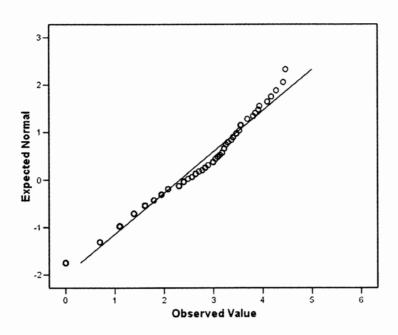
#### Histogram



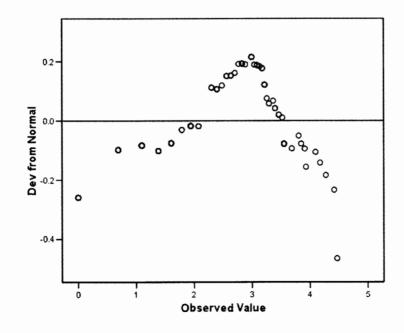
# Inhours Stem-and-Leaf Plot

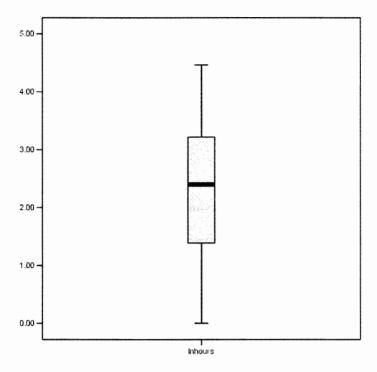
Erequency	Stem &	Leaf
7.00	٥.	0000000
4.00	ο.	6666
15.00	1.	000000000033333
15.00	1.	666666779999999
10.00	2 .	003333334
15.00	2.	556667788889999
18.00	3.	000111222222234444
10.00	з.	5555568899
5.00	4 .	01244
Stem width:	1.0	90
Each leaf:	1 0	ase (s)

### Normal Q-Q Plot of Inhours



### Detrended Normal Q-Q Plot of Inhours





It is apparent from the preceding statistics and graphs that the natural logarithmic transformation was appropriate because the distribution of *hoursex* is now relatively normal. The Lilliefors statistic would suggest that there is still a slight problem, but all the other diagnostic data are satisfactory. It is important to note that when reporting and interpreting results involving transformed variables, you must acknowledge that a transformation has been undertaken on the data.

# Data transformation



### Recode

You can modify data values by recoding. There are three instances where recoding variables is appropriate:

- Collapsing continuous variables into categorical variables (for example, median split).
- Recoding negatively worded scale items.
- Replacing missing values and bringing outlying cases into the distribution.

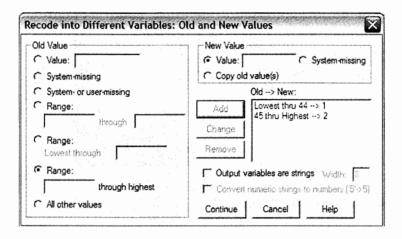


#### To collapse a continuous variable

In relation to the research example, you wish to obtain a median split on the variable age.

- 1 Select the **Transform** menu.
- 2 Click on Recode and Into Different Variables to open the Recode into Different Variables dialogue box. By selecting Into Different Variables, you will retain the original data.
- 3 In the Variable List box, select the variable you wish to recode (i.e. age) and click on the ▶ button to move the variable into the Input Variable -> Output Variable: box.

- 4 In the Name: box of the Output Variable box, type the new variable name (i.e. agecat).
- 5 Click on the **Change** command pushbutton.
- 6 Click on the Old and New Values... command pushbutton to open the Recode into Different Variables: Old and New Values sub-dialogue box.
- 7 Click on the second **Range:** radio button and type the median in the box (i.e. 44).
- 8 In the **New Value** box, type the new value (i.e. 1).
- 9 Click on the **Add** command pushbutton.
- 10 Click on the third **Range:** radio button and type the median plus one (i.e. 45).
- 11 In the **New Value** box, type the new value (i.e. 2).
- 12 Click on the Add command pushbutton.



```
RECODE
age
(Lowest thru 44=1) (45 thru Highest=2) INTO agecat .
EXECUTE .
```

If you now go back into your data file you will notice that the variable *agecat* now has two possible values: 1 = 0–44 years and 2 = 45+ years. This variable can now be used in analysis where categories are required.

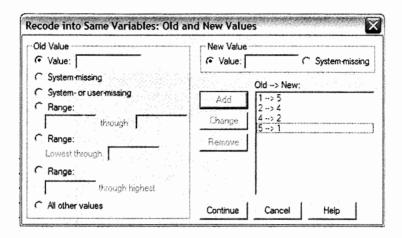


#### To recode negatively worded scale items

Three of the items on the attitude scale required recoding because they were negatively worded. Remember that the response format for the items was 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree and 5 = strongly disagree.

- Select the Transform menu.
- 2 Click on Recode and Into Same Variables to open the Recode into Same Variables dialogue box. By selecting Into Same Variables, you will overwrite the original data.
- 3 In the **Variable List** box, select the variable(s) you wish to recode (i.e. *att2*, *att4* and *att6*) and click on the ▶ button to move the variables into the **Variables**: box.

- 4 Click on the Old and New Values... command pushbutton to open the Recode into Same Variables: Old and New Values sub-dialogue box.
- 5 In the **Old Value** box, type the old value (i.e. 1).
- 6 In the **New Value** box, type the new value (i.e. 5).
- 7 Click on the **Add** command pushbutton.
- 8 In the **Old Value** box, type the second old value (i.e. 2).
- 9 In the New Value box, type the new second value (i.e. 4).
- 10 Click on the Add command pushbutton.
- Repeat steps 8, 9 and 10 for the remaining two values.



```
RECODE
att2 att4 att6 (1=5) (2=5) (4=2) (5=1) .
EXECUTE .
```

Att2, Att4 and Att6 have now been recoded to allow computation of a composite variable.



#### To replace missing values

Missing observations can be problematic. To avoid this problem you can replace missing values with estimates computed with one of several methods. A commonly used method is mean substitution.

- 1 Select the **Transform** menu.
- 2 Click on Recode and Into Same Variables to open the Recode into Same Variables dialogue box.
- In the **Variable List** box, select the variable(s) you wish to recode (i.e. *att1*) and click on the **b** button to move the variable into the **Variables:** box.
- 4 Click on the Old and New Values... command pushbutton to open the Recode into Same Variables: Old and New Values sub-dialogue box.
- 5 In the Old Value box, select the System- or user-missing radio button.
- In the **New Value** box, type the mean of the variable (i.e. 2.37, as obtained by calculating the average score on the *att1* item).
- 7 Click on the Add command pushbutton.

Old Value	New Value			
Value:		C System-missing		
System-missing System- or user-missing	Old> New:			
Range:	Add MIS	SSING> 2.37		
Range: Lowest through	Remove			
Range: through highest				

```
RECODE
att1 (MISSING=2.37) .
EXECUTE .
```

Any missing data for the variable *att1* will be replaced with the mean of 2.37. Recoding missing values using mean substitution allows you to include all cases in your analysis. Alternatively, you may prefer to deal with missing cases in each analysis. Most procedures allow you to exclude missing cases either pairwise or listwise. Exclusion of missing cases pairwise involves the deletion of cases with missing values only on the relevant variables. Listwise deletion involves the elimination of cases with a missing value for any variable on the data list. More advanced procedures, such as factor analysis, give you the opportunity to replace missing data with the mean for particular variables during the procedure itself.

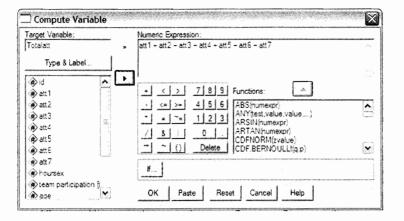
### Compute

Transformation of variables is just one instance where the **Compute** command may be used. The compute command is most commonly used to obtain composite scores for items on a scale. This can be achieved for the whole data set or only a subset if certain conditions apply. In the research example, a total attitude score would be appropriate. This can be obtained by adding the responses to the seven items comprising the scale for each individual case.



#### To compute a new variable

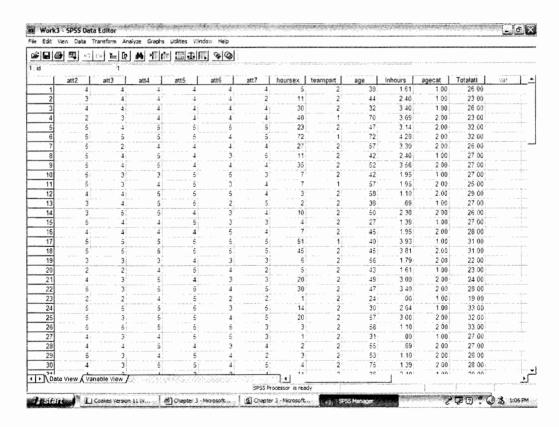
- 1 Select the **Transform** menu.
- 2 Click on **Compute** to open the **Compute Variable** dialogue box. If previous settings remain, these can be cleared with the **Reset** command pushbutton.
- 3 In the **Target Variable:** box, type an appropriate variable name (i.e. *Totalatt*).
- 4 Select the first scale item from the variable list box (i.e. *att1*) and click on the ▶ button to move the variable into the **Numeric Expression:** box.
- 5 Click on the + button.
- 6 Select the second scale item from the variable list box (i.e. *att2*) and repeat steps 4 and 5.



#### 7 Click on **OK**.

```
COMPUTE Totalatt = att1 + att2 + att3 + att4 + att5 + att6 + att7 . EXECUTE .
```

Again if you move back into your data file (remember to use the SPSS Manager) you will notice that items have been added to obtain a new composite variable in your data file — Totalatt.



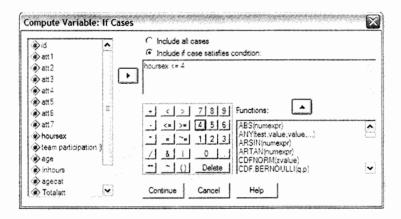
Other transformations are possible using the calculator pad and functions options in the **Compute Variable** dialogue box.

If you wish to compute a new variable based on certain conditions, then a slightly different procedure is required. For example, you may wish to compute a total attitude score for only those people who exercise for four hours per week or less.



To compute new variables based on certain conditions

- 1 Select the **Transform** menu.
- 2 Click on Compute to open the Compute Variable dialogue box.
- 3 In the **Target Variable:** box, type an appropriate variable name (i.e. *Totalatt*).
- Select the first scale item from the variable list box (i.e. *att1*) and click on the button to move the variable into the **Numeric Expression:** box.
- 5 Click on the + button.
- 6 Select the second scale item from the variable list box (i.e. *att2*) and repeat steps 4 and 5 until all items are entered.
- 7 Click on the **If...** command pushbutton to open the **Compute Variable: If Cases** sub-dialogue box.
- 8 Select the **Include if case satisfies condition:** radio button.
- 9 Select the variable on which the condition is based (i.e. *hoursex*) and click on the button to move the variable into the box.
- Select the <= operator button, which moves the symbol into the box above, then select the digit 4.



11 Click on Continue and then OK.

```
IF (hoursex <= 4) Totalatt = att1 + att2 + att3 + att4 + att5 + att6 + att7 . EXECUTE .
```

In this example, a new variable has been computed — *Totalatt* — based on certain conditions, i.e. for those individuals who spent four hours or less (hoursex <= 4) in physical activity.

# Data selection



In the **Select Cases** option in the **Data** menu, there are a number of procedures that can be chosen:

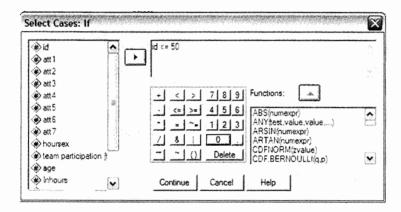
- · Selection of specified cases using the If option
- Selection of a random sample of cases using the Sample option
- Selection of cases based on time or case range using the **Range** option.

Selection of cases using the **If** option is most commonly used. For example, you may wish to examine the descriptive statistics of only males or females, or you may wish to analyse only half your data set.



### To select the first 50 cases in the data file for subsequent analysis

- Select the **Data** menu.
- 2 Click on Select Cases... or click on the Select Cases tool to open the Select Cases dialogue box.
- In the **Select** box, click on the **If condition is satisfied** radio button.
- 4 Click on the **If...** command pushbutton to open the **Select Cases: If** subdialogue box.
- Select the variable you require (i.e. id) and click on the  $\triangleright$  button to move the variable into the box.
- 6 Click on the operator of your choice (i.e. <=), which will then be pasted into the box above.
- 7 Type in the value you require (i.e. 50).



#### 8 Click on Continue and then OK.

USE ALL.

COMPUTE filter\_\$=(id <= 50).

VARIABLE LABEL filter\_\$ 'id <= 50 (FILTER)'.

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.

FORMAT filter\_\$ (f1.0).

FILTER BY filter\_\$.

EXECUTE.

You are now ready to do analysis on the 50 cases selected. If you go back into your data file (using the SPSS Manager) you will see that cases 51 to 99 have been crossed through.

# Practice example



The following scale measured how adolescents feel about their future. Data from 100 year 11 students were collected. The scale comprised six items:

- 1 I feel optimistic about my future.
- 2 I believe that every cloud has a silver lining.
- 3 I doubt that I can achieve what I want to in life.
- 4 I have got what it takes to create a bright future.
- 5 If anything bad can happen, it will.
- **6** I deserve to have good things come my way.

The response format for these items was:

- 1 = strongly agree
- 2 = agree
- 3 = neutral
- 4 = disagree
- 5 = strongly disagree.

Given the data in Prac3.sav, your tasks are to:

- 1 Check for incorrect data entry.
- 2 Recode negatively worded items.
- 3 Recode missing values using mean substitution.
- 4 Compute a total hope score.
- 5 Screen for normality for the composite variable.
- Attempt an appropriate transformation on the composite variable and compare the output of this transformation with the distribution of the original variable.
- 7 Select a random sample of 50 cases from the data set and obtain the mean hope score for this subset.



### Syntax

```
DESCRIPTIVES
VARIABLES=hope1 hope2 hope3 hope4 hope5 hope6
/STATISTICS=MEAN STDDEV MIN MAX .
RECODE
hope3 hope5 (1=5) (2=4) (4=2) (5=1) .
EXECUTE
DESCRIPTIVES
VARIABLES=hope1 hope2 hope3 hope4 hope5 hope6
/STATISTICS=MEAN STDDEV MIN MAX .
RECODE
hope2 (MISSING=4.31) .
EXECUTE .
RECODE
hope4 (MISSING=1.55) .
EXECUTE .
RECODE
hope5 (MISSING=4.34) .
EXECUTE .
RECODE
hope6 (MISSING=3.63) .
EXECUTE .
COMPUTE hopetot = hope1 + hope2 + hope3 + hope4 + hope5 + hope6 .
EXECUTE .
EXAMINE
VARIABLES=hopetot
/PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING PAIRWISE
/NOTOTAL.
COMPUTE refhope = 26 - hopetot .
EXECUTE
COMPUTE sqrefhop = SQRT(refhope) .
VARIABLE LABELS sqrefhop 'Square root of reflected hope' .
EXECUTE .
EXAMINE
VARIABLES=sqrefhop
/PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING PAIRWISE
/NOTOTAL.
USE ALL.
COMPUTE filter_$=(id <= 50).
VARIABLE LABEL filter_$ 'id <= 50 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE
DESCRIPTIVES
VARIABLES=hopetot
/STATISTICS=MEAN STDDEV MIN MAX .
```

### **Output**

The minimum and maximum show that no out-of-range entries have been made. To obtain the means needed for mean substitution of missing values, descriptives must be found for recoded variables. Having recoded the negatively worded items and the missing values, a total hope score can be computed.

### Screening for normality

#### **Descriptive Statistics**

	М	Minimum	Maximum	Mean	Std. Deviation
hope1	100	1	5	1.76	.922
hope2	99	1	5	4.31	.976
hope3	100	1	5	1.26	.705
hope4	98	1	5	1.55	.775
hope5	99	1	5	1.66	1,002
hope6	99	1	5	3.63	1.121
Valid N (listwise)	95				

#### Descriptives

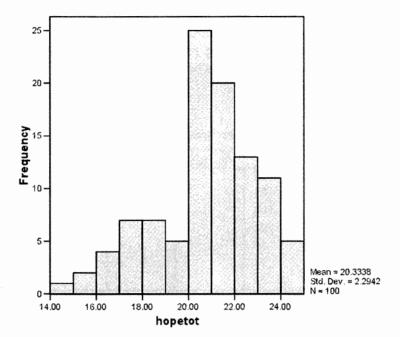
			Statistic	Std. Error
hopetot	Mean		20.3338	.22942
	95% Confidence	Lower Bound	19.8786	
	Interval for Mean	Upper Bound	20.7890	
	5% Trimmed Mean		20.3709	
	Median		20.5900	
	Variance		5.263	
	Std. Deviation		2.29420	
	Minimum		14.00	
	Maximum		25.00	
	Range		11.00	
	Interquartile Range		3.00	
	Skewness		424	.241
	Kurtosis		.162	.478

#### **Tests of Normality**

	Koln	nogorov-Smir	ทอง	Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig.
hopetot	.182	100	.000	.960	100	.004

Lilliefors Significance Correction

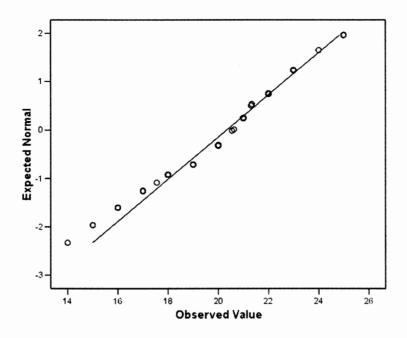
### Histogram

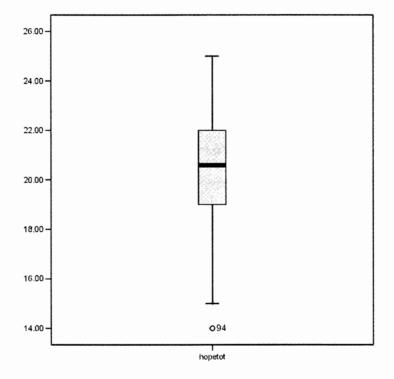


### hopetot Stem-and-Leaf Plot

Frequency	Stem	ě	Leaf
1.00	Extremes		(=<14.0)
2.00	15		00
4.00	16		0000
7.00	27		0000005
7.00	18		0000000
5.00	29		00000
25.00	20		000000000000000000000000000000000000000
20.00	21		0000000000000000033
13.00	22		000000000000
11.00	23		0000000000
1.00	24		0
4.00	25		0000
Stem widt Each leaf		L.0	0 ase(s)

### Normal Q-Q Plot of hopetot





There is no problem with these data but we will do a transformation for practice. We will do a reflection and then a square root.

#### Descriptives

			Statistic	Std. Error
Square root of	Mean		2.3271	.05033
reflected hope	95% Confidence	Lower Bound	2.2273	
	Interval for Mean	Upper Bound	2.4270	
	5% Trimmed Mean		2.3431	
	Median		2.3259	
	Variance		.253	
	Std. Deviation		.50326	
	Minimum		1.00	
	Maximum	•	3,46	
	Range		2.46	
	Interquartile Range		.65	
	Skewness		328	.241
	Kurtosis		.589	.478

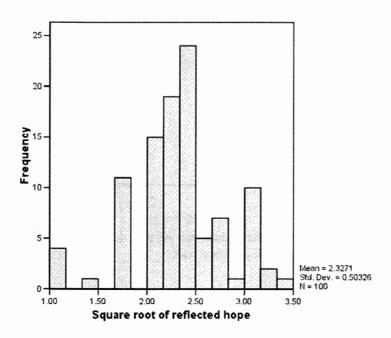
#### **Tests of Normality**

	Koln	nogorov-Smir	ัทอง <sup>ร</sup>	Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig.
Square root of reflected hope	.144	100	.000	.959	100	.003

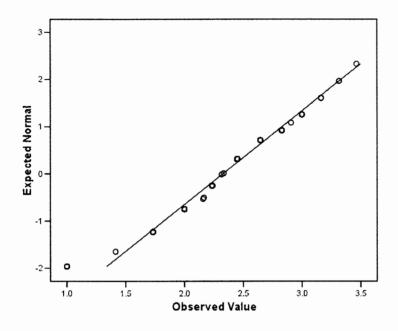
a. Lilliefors Significance Correction

# Square root of reflected hope

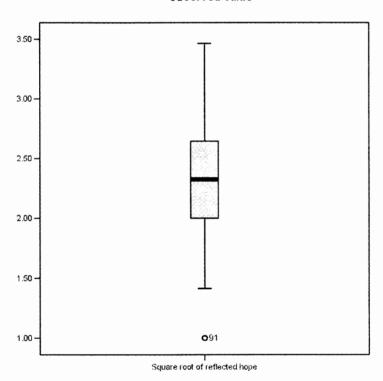
#### Histogram



Normal Q-Q Plot of Square root of reflected hope



#### **Observed Value**



The mean total hope score of a random sample of 50 cases is as follows:

# **Descriptives**

**Descriptive Statistics** 

	N	Minimum	Maximum	Mean	Std. Deviation
hopetot	50	16.00	25.00	20.3366	2.08760
Valid N (listwise)	50				



# **Descriptive statistics**

Descriptive statistics are used to explore the data collected, as shown in chapter 3; and to summarise and describe those data.

Descriptive statistics may be particularly useful if one just wants to make general observations about the data collected, for example number of males and females, the age range and average (mean) age, the average length of residence in a community etc. Other statistics such as standard deviation and variance give more information about the distribution of each variable.

# Frequency distributions



A frequency distribution is a display of the frequency of occurrence of each score value. The frequency distribution can be represented in tabular form or, with more visual clarity, in graphical form. For continuous variables, measured on ratio or interval scales, histograms or frequency polygons are appropriate. For categorical variables, measured on nominal or ordinal scales, bar charts are suitable.

# Measures of central tendency and variability



The three main measures of central tendency are mode, median and mean. The measures of variability include range, interquartile range, standard deviation and variance. All of these measures of variability are more appropriate for interval or ratio data. You can also examine the normality of the distribution through the **Frequencies** procedure.

# Working example



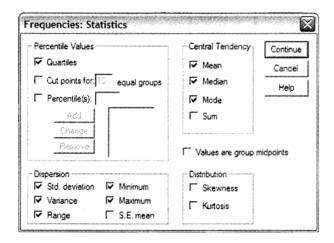
One hundred tennis players participated in a serving competition. Gender and number of aces were recorded for each player. The data file can be found in Work4.sav on the website that accompanies this title and is shown in the following figure.

e Edit	View Data	Transform An	alyze Grapi	is Utilities	Window Help								
		r.∞ <u>≥ </u> [?	M T	d Bi	e 🕟 😵	J							
gender		1											
1	gender	aces	787	V81	481	V38	V31	1-21	Vär	var	SE	931	VB!
1	1 00	5.00											
2	2.00	2 00											
3	1 00	4.00											
4	2.00	8 00											
5	1.00	8.00											
6	2.00	5.00	*****************			·}			gum manonimanimi				
7	1.00	5.00											
8	2 00	9.00	. Th. B. Wohnson to Pro-			±			÷				
9	1.00	9 00	. 1887007 0008870000000	ga na sama nam s s								1	-
10	2 00	6 00							:		m /	†	
11	1.00	5 00											†

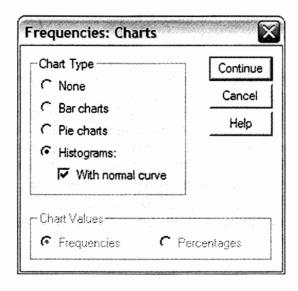


To obtain a frequency table, measures of central tendency and variability

- 1 Select the **Analyze** menu.
- 2 Click on **Descriptive Statistics** and then on **Frequencies...** to open the **Frequencies** dialogue box.
- 3 Select the variable(s) you require (i.e. *aces*) and click on the ▶ button to move the variable into the **Variable(s)**: box.
- 4 Click on the **Statistics...** command pushbutton to open the **Frequencies: Statistics** sub-dialogue box.
- 5 In the **Percentile Values** box, select the **Quartiles** check box.
- 6 In the Central Tendency box, select the Mean, Median and Mode check boxes.
- 7 In the **Dispersion** box, select the **Std. deviation**, **Variance**, **Range**, **Minimum** and **Maximum** check boxes.



- 8 Click on Continue.
- 9 Click on the **Charts...** command pushbutton to open the **Frequencies: Charts** sub-dialogue box.
- Click on the **Histogram(s)** radio button. You will notice that you can also obtain a normal curve overlay, so click on the **With normal curve** check box.



FREQUENCIES

VARIABLES=aces
/NTILES= 4
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
/HISTOGRAMÅ@NORMAL
/ORDER= ANALYSIS.

#### **Statistics**

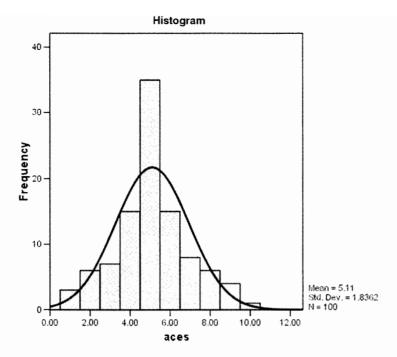
aces		
И	Valid	100
	Missing	0
Mean		5.1100
Median		5,0000
Mode		5.00
Std. Deviation		1.83620
Variance		3.372
Range		9.00
Minimum		1.00
Maximum		10.00
Percentiles	25	4.0000
	50	5.0000
	75	6.0000

aces

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	3	3.0	3.0	3.0
1	2.00	6	6.0	6.0	9.0
	3.00	7	7.0	7.0	16.0
	4.00	15	15.0	15.0	31.0
	5.00	35	35.0	35.0	66.0
1	6.00	15	15.0	15.0	81.0
	7.00	8	8.0	8.0	89.0
	8.00	6	5.0	6.0	95.0
	9.00	4	4.0	4.0	99.0
	10.00	1	1.0	1.0	100.0
	Total	100	100.0	100.0	

In the frequency table, the frequency column summarises the total number of aces served. For example, only one person served ten aces. The percent column displays this frequency in percentage form for *all* cases, including those cases that may be missing. The valid percent column is the proportion of scores only for those cases that are valid. Because you have no missing data in this example, the percent and valid percent columns are identical. The cumulative percent column is the summation of the percentage for that score with the percentage for all lesser scores.

By obtaining the 25th and 75th percentiles for the distribution, the interquartile range can be calculated by subtracting one from the other. Therefore, in this example, the interquartile range is equal to 6 - 4 = 2.





### To obtain the appropriate output for a categorical variable

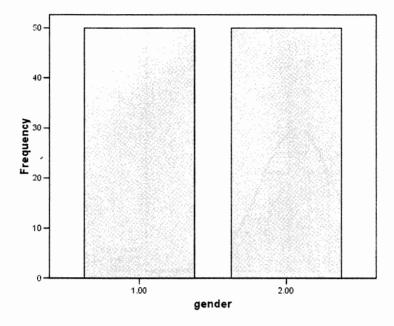
- 1 Select the **Analyze** menu.
- 2 Click on **Descriptive Statistics** and then on **Frequencies...** to open the **Frequencies** dialogue box.
- 3 Select the variable(s) you require (i.e. *gender*) and click on the ▶ button to move the variable into the **Variable(s)**: box.
- 4 Click on the **Statistics...** command pushbutton to open the **Frequencies: Statistics** sub-dialogue box.
- 5 In the **Central Tendency** box, click on the **Mode** check box.
- 6 Click on Continue.
- 7 Click on the **Charts...** command pushbutton to open the **Frequencies: Charts** sub-dialogue box.
- 8 Select the **Bar chart(s)** radio button.
- 9 Click on Continue and then OK.

FREQUENCIES
VARIABLES=gender
/STATISTICS=MODE
/BARCHART FREQ
/ORDER= ANALYSIS.

#### gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	50	50.0	50.0	50.0
	2.00	50	50.0	50.0	100.0
	Total	100	100.0	100.0	

#### gender



# Descriptives command

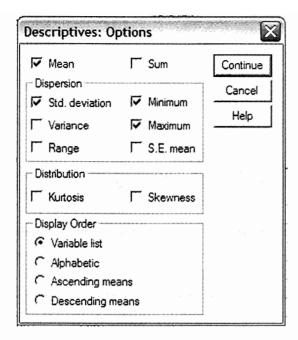


It is also possible to obtain certain measures of central tendency and variability through the **Descriptives** command. This command also allows you to save standardised values as variables. These standardised or Z-scores are useful for further analysis (for example, interaction terms in multiple regression) or in comparing samples from different populations. Furthermore, inspection of Z-scores will allow identification of outlying cases, which is useful in data screening. Z-scores greater than +3 and less than -3 are considered to be outliers.



#### To obtain descriptive statistics and Z-scores

- 1 Select the Analyze menu.
- 2 Click on Descriptive Statistics and then Descriptives... to open the Descriptives dialogue box.
- 3 Select the variable(s) you require (i.e. aces) and click on the ▶ button to move the variable into the **Variable(s)**: box.
- 4 Select the Save standardized values as variables check box.
- 5 Click on the **Options** command pushbutton.



- Note that the **Mean**, **Std. deviation**, **Minimum** and **Maximum** check boxes are automatically selected. If you wish to obtain additional descriptive statistics, select the appropriate check boxes.
- 7 Click on **Continue** and then **OK**.

DESCRIPTIVES
VARIABLES=aces/SAVE
/STATISTICS=MEAN STDDEV MIN MAX.

#### **Descriptive Statistics**

	2	Minimum	Maximum	Mean	Std. Deviation
aces	100	1.00	10.00	5.1100	1.83620
Valid N (listwise)	100				

If you switch back to the Data Editor window, you will notice that the Z-scores have been saved as another variable: Zaces.

# Practice example



Sales in \$1000s by 20 junior and senior salespeople working in a white goods shop were recorded at the end of a week. Given the data in Prac4.sav, your task is to obtain a frequency table and the appropriate chart and descriptive statistics for each variable in the data file.

# **Solutions**



# **Syntax**

FREQUENCIES
VARIABLES=rank
/STATISTICS=MODE
/BARCHART FREQ
/ORDER= ANALYSIS .
FREQUENCIES
VARIABLES=sales
/NTILES= 4
MODE/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN

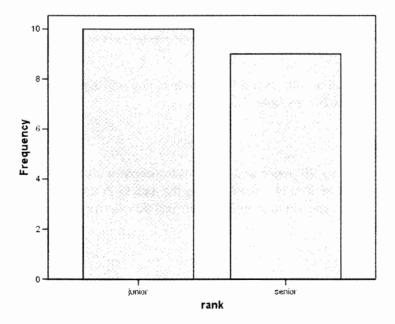
# Output

#### rank

		Frequency	Percent	Valid Percent	Cumulative Percent
∀alid	junior	10	50.0	52.6	52.6
	senior	9	45.0	47,4	100.0
	Total	19	95.0	100.0	
Missing	9.00	1	5.0		
Total		20	100.0		

The appropriate chart for a categorical variable (rank) is a bar chart and the mode is the appropriate measure of central tendency.

#### rank



You will notice that because one case was missing, the percentage and valid percentage columns are not identical.

For continuous data (sales), a histogram is appropriate, and either the mean or median is the chosen measure of central tendency. The interquartile range is 75 - 51 = 24.

Statistics

sales in \$1000

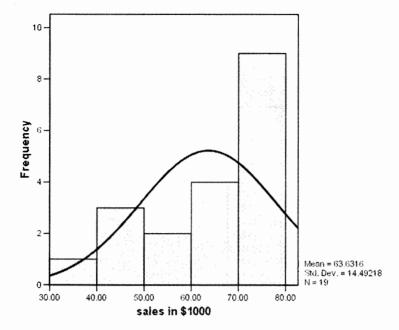
N	∀alid	19
	Missing	1
Mean		63.6316
Median		68.0000
Mode		75.00°
Std. Deviation		14.49218
Variance		210.023
Range		50.00
Minimum		30.00
Maximum		80.00
Percentiles	25	51.0000
	50	68.0000
	75	75,0000

a. Multiple modes exist. The smallest value is shown.

sales in \$1000

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	30,00	1	5.0	5.3	5.3
	43.00	1	5.0	5.3	10.5
	45.00	1	5.0	5.3	15.8
	47.00	1	5.0	5.3	21.1
	51.00	1	5.0	5.3	26.3
	56.00	1	5.0	5.3	31.6
	60.00	1	5.0	5.3	36.8
	64.00	1	5.0	5.3	42.1
	66.00	1	5.0	5.3	47.4
	68.00	1	5.0	5.3	52,6
	70.00	1	5.0	5.3	57.9
	72.00	1	5.0	5.3	63.2
	74.00	1	5.0	5.3	68.4
	75.00	2	10.0	10.5	78.9
	76.00	1	5.0	5.3	84.2
	77.00	1	5.0	5.3	89.5
	80.00	2	10.0	10.5	100.0
	Total	19	95.0	100.0	
Missing	99.00	1	5.0		
Total		20	100.0		

## Histogram



You may also notice that the data for the variable sales appear to be positively skewed.



# **Correlation**

Correlation looks at the relationship between two variables in a linear fashion.

A *Pearson-product-moment correlation* coefficient describes the relationship between two continous variables and is available through the **Analyze** and **Correlate** menus.

A correlation between two dichotomous or categorical variables is called a *phi* coefficient and is available through the **Crosstabs** option from the **Analyze** and **Descriptive Statistics** menus.

A correlation between a continuous and a categorical variable is called a *point-biserial* correlation. This option is not available in the SPSS for Windows package. However, you can use a Pearson product-moment correlation to correlate a dichotomous and a continuous variable but the proportion of each category of the dichotomous variable must be approximately equal, and variables must be coded as 0 and 1.

When the assumptions underlying correlation cannot be met adequately, a nonparametric alternative is *Spearman's rank-order correlation*.

In this chapter you will address bivariate and partial correlations using Pearson's product-moment correlation.

Simple bivariate correlation, also referred to as zero-order correlation, refers to the correlation between two continuous variables and is the most common measure of linear relationship. This coefficient has a range of possible values from -1 to +1. The value indicates the strength of the relationship, while the sign (+ or -) indicates the direction.

Partial correlation provides a single measure of linear association between two variables while adjusting for the effects of one or more additional variables.

# Assumption testing



Correlational analysis has a number of underlying assumptions:

- Related pairs data must be collected from related pairs: that is, if you obtain a score on an X variable, there must also be a score on the Y variable from the same participant.
- 2 Scale of measurement data should be interval or ratio in nature.
- 3 Normality the scores for each variable should be normally distributed.
- 4 Linearity the relationship between the two variables must be linear.
- **Homoscedasticity** the variability in scores for one variable is roughly the same at all values of the other variable. That is, it is concerned with how the scores cluster uniformly about the regression line.

Assumptions 1 and 2 are a matter of research design. Assumption 3 can be tested using the procedures outlined in chapters 3 and 4. Assumptions 4 and 5 can be tested by examining scatterplots of the variables.

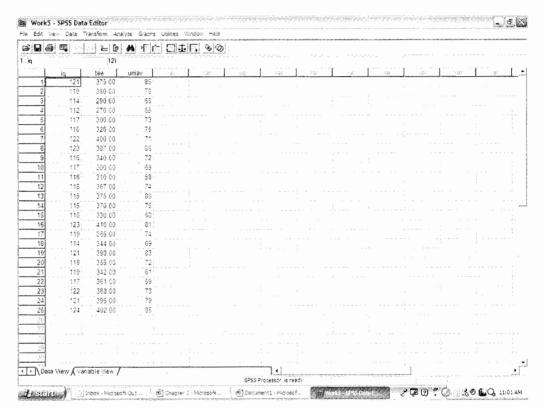
# Working example



Twenty students wishing to enter university were given an intelligence test (IQ) and their tertiary entrance examination scores (TEE) were recorded. You suspect that a positive relationship exists between these two variables and wish to test this directional hypothesis (one-tailed).

At the end of the academic year, the course averages for the same 20 students were obtained. You also wish to determine whether the relationship between TEE scores and course average is significant when IQ is controlled in the analysis.

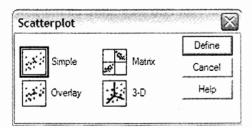
The data file can be found in Work5.sav on the website that accompanies this title and is shown in the following figure.





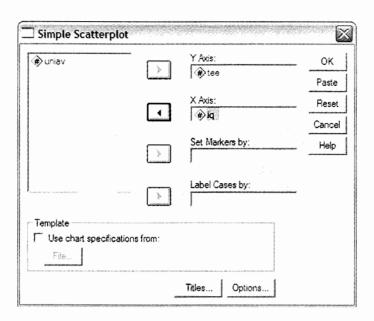
# To obtain a scatterplot

- 1 Select the **Graphs** menu.
- 2 Click on **Scatter...** to open the **Scatterplot** dialogue box.
- 3 Ensure that the **Simple Scatterplot** option is selected.



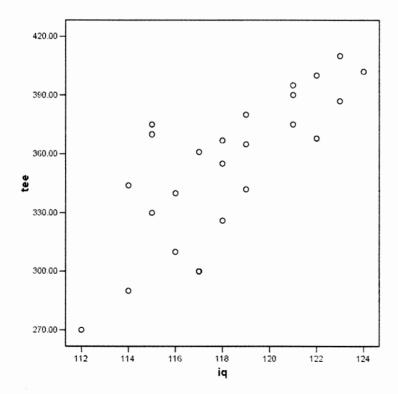
- 4 Click on the **Define** command pushbutton to open the **Simple Scatterplot** sub-dialogue box.
- 5 Select the first variable (i.e. *tee*) and click on the ▶ button to move the variable into the **Y** Axis: box.

Select the second variable (i.e. iq) and click on the  $\triangleright$  button to move the variable into the **X** Axis: box.



## 7 Click on OK.

GRAPH
/SCATTERPLOT(BIVAR)=iq WITH tee
/MISSING=LISTWISE .



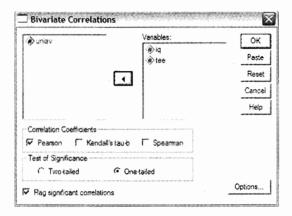
As you can see from the scatterplot, there is a linear relationship between IQ and TEE scores. Given that the scores cluster uniformly around the regression line, the assumption of homoscedasticity has not been violated.

Similarly, the scatterplot was obtained for TEE scores and course average, indicating that assumptions of linearity and homoscedasticity were not violated. The output of this plot is not displayed.



## To obtain a bivariate Pearson product-moment correlation

- 1 Select the **Analyze** menu.
- 2 Click on Correlate and then Bivariate... to open the Bivariate Correlations dialogue box.
- Select the variables you require (i.e. iq and tee) and click on the  $\triangleright$  button to move the variables into the **Variables:** box.
- 4 Ensure that the **Pearson** correlation option has been selected.
- 5 In the **Test of Significance** box, select the **One-tailed** radio button.



#### 6 Click on OK.

CORRELATIONS
/VARIABLES=iq tee
/PRINT=ONETAIL NOSIG
/MISSING=PAIRWISE.

#### Correlations

		iq	tee
iq	Pearson Correlation	1	.767**
	Sig. (1-tailed)		.000
	N	25	25
tee	Pearson Correlation	.767**	1
	Sig. (1-tailed)	.000	
	N	25	25

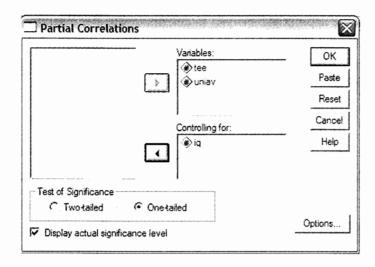
<sup>\*\*.</sup> Correlation is significant at the 0.01 level

To interpret the correlation coefficient, you examine the coefficient and its associated significance value (p). The output confirms the results of the scatterplot in that a significant positive relationship exists between IQ and TEE (r = .767, p < .05). Thus, higher intelligence scores are associated with higher TEE scores.



## To conduct a partial correlation

- 1 Select the **Analyze** menu.
- 2 Click on Correlate and then on Partial... to open the Partial Correlations dialogue box.
- 3 Select the variables you wish to correlate (i.e. *tee* and *uniav*) and click on the button to move the variables into the **Variables:** box.
- Select the variable to be controlled (i.e. iq) and click on the  $\triangleright$  button to move the variable into the **Controlling for:** box.
- 5 In the **Test of Significance** box, select the **One-tailed** radio button.



#### 6 Click on OK.

PARTIAL CORR
/VARIABLES=tee uniav BY iq
/SIGNIFICANCE=ONETAIL
/MISSING=LISTWISE.

#### Correlations

Control Variab	les		tee	นกโลง
įq	tee	Correlation	1.000	.354
		Significance (1-tailed)		.045
		df	0	22
	uniav	Correlation	.354	1.000
		Significance (1-tailed)	.045	
		df	22	0

The output indicates that the relationship between TEE scores and course averages after the first year of university is also significant, having controlled for intelligence scores. Again, the higher the TEE score, the higher the course average for the first year of university (r = .354, p < .05).

# Practice example



Medical researchers believe that there is a relationship between smoking and lung damage. Data were collected from smokers who have had their lung function assessed and their average daily cigarette consumption recorded. Lung function was assessed in such a way that higher scores represent greater health. A negative relationship between the variables was expected.

The researchers also believe that the relationship between these variables may be influenced by the number of years for which the person has been a smoker. Given the data in Prac5.sav, your tasks are to:

- 1 Check your data for violations of assumptions.
- 2 Conduct the appropriate analysis to determine whether cigarette consumption is related to lung capacity.
- 3 Determine whether the above relationship remains significant, having controlled for the length of the smoking habit.

# **Solutions**



## **Syntax**

#### DESCRIPTIVES

VARIABLES=lungfunc cigsday years /STATISTICS=MEAN STDDEV KURTOSIS SKEWNESS .

#### GRAPH

/SCATTERPLOT(BIVAR)=cigsday WITH lungfunc /MISSING=LISTWISE .

**GRAPH** 

/SCATTERPLOT(BIVAR)=years WITH lungfunc /MISSING=LISTWISE .

#### CORRELATIONS

/VARIABLES=lungfunc cigsday /PRINT=ONETAIL NOSIG /MISSING=PAIRWISE .

PARTIAL CORR

/VARIABLES= lungfunc cigsday BY years

/SIGNIFICANCE=ONETAIL

/MISSING=LISTWISE .

#### Output

#### **Assumption testing**

Normality

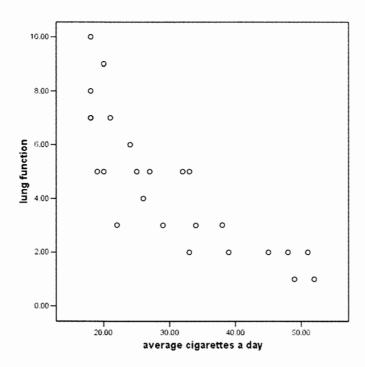
It is evident from the output that all variables are relatively normally distributed.

#### **Descriptive Statistics**

	И	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
lung function	25	4.4800	2.50200	.529	.464	509	.902
average digarettes a day	25	30.3600	11.44509	.665	.464	851	.902
years of smoking	25	10.6800	10.91910	1.002	.464	287	.902
Valid N (listwise)	25						

Linearity and homoscedasticity

The scatterplot of cigarettes per day and lung function suggests a somewhat curvilinear relationship.



However, because this tendency is not marked, the assumptions of linearity and homoscedasticity are not violated. The second scatterplot has a similar shape and indicates no violation of these assumptions.

Correlations

		lung function	average cigarettes a day
lung function	Pearson Correlation	1	825**
	Sig. (1-tailed)		.000
	N	25	25
average cigarettes a day	Pearson Correlation	825**	1
	Sig. (1-tailed)	.000	
	N	25	25

<sup>\*\*.</sup> Correlation is significant at the 0.01 level (1-tailed).

There is a significant negative correlation between average number of cigarettes smoked per day and lung function, indicating that lung function decreases as cigarette consumption increases (r = -.825, p < .001). Given that a directional hypothesis was stated, a one-tailed probability test was appropriate.

Correlations

Control variables			lung function	average cigarettes a da)
years of smoking	lung function	Correlation	1.000	378
		Significance : 1-tailed;		.034
		df	0	22
	average digarettes a da;	Correlation	378	1.000
		Significance (1-tailed)	.034	
		df	22	0

The partial correlation coefficient, whereby years of smoking was controlled, is still significant (r = -.378, p < .05). Thus, it is demonstrated that the relationship between the number of cigarettes smoked per day and lung function is still significant, even controlling for years of smoking.



# **T-tests**

A t-test is used to determine whether there is a significant difference between two sets of scores. Three main types of t-test may be applied:

- · One-sample
- · Independent groups
- · Repeated measures.

# Assumption testing



Each statistical test has certain assumptions that must be met before analysis. These assumptions need to be evaluated because the accuracy of test interpretation depends on whether assumptions are violated. Some of these assumptions are generic to all types of t-test, but others are more specific.

The generic assumptions underlying all types of t-test are:

- **Scale of measurement** the data should be at the interval or ratio level of measurement
- **Random sampling** the scores should be randomly sampled from the population of interest.
- 3 Normality the scores should be normally distributed in the population.

Clearly, assumptions 1 and 2 are a matter of research design and not statistical analysis. Assumption 3 can be tested in a number of different ways, as outlined in chapter 3.

# Working example



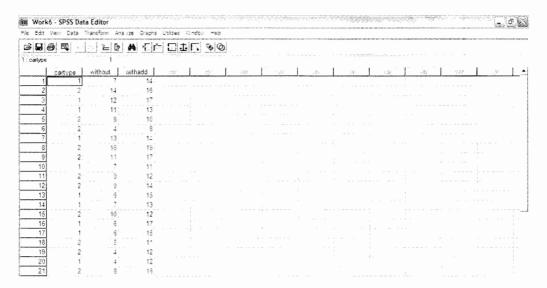
A major oil company developed a petrol additive that was supposed to increase engine efficiency. Twenty-two cars were test driven both with and without the additive, and the number of kilometres per litre was recorded. Whether the car was automatic or manual was also recorded and coded as 1 = manual and 2 = automatic.

During an earlier trial 22 cars were test driven using the additive. The mean number of kilometres per litre was 10.5.

You are interested in addressing the following questions:

- Are the cars in the present trial running more efficiently than those in the earlier trial? The single sample t-test will help answer this question.
- 2 Does engine efficiency improve when the additive is used? This is a repeated measures t-test design.
- 3 Does engine efficiency with and without the additive differ between manual and automatic cars? This is an independent groups t-test.

The data can be found in Work6.sav on the website that accompanies this title and is shown in the following figure.



# The one-sample t-test

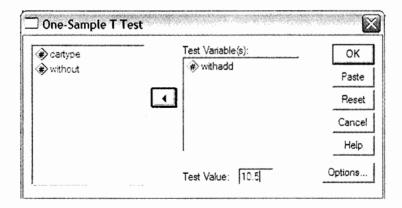


The one-sample t-test is used when you have data from a single sample of participants and you wish to know whether the mean of the population from which the sample is drawn is the same as the hypothesised mean.



## To conduct a one-sample t-test

- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and then One-Sample T Test... to open the One-Sample T Test dialogue box.
- 3 Select the variable you require (i.e. *withadd*) and click on the ▶ button to move the variable into the **Test Variable(s):** box.
- 4 In the **Test Value:** box type the mean score (i.e. 10.5).



5 Click on OK.

T-TEST
/TESTVAL = 10.5
/MISSING = ANALYSIS
/VARIABLES = withadd
/CRITERIA = C1(.95) .

It is possible to determine whether a difference exists between the sample mean and the hypothesised mean by consulting the t-value, degree of freedom (df) and two-tail significance. If the value for two-tail significance is less than .05 (p < .05), then the difference between the means is significant.

The output indicates that there is a significant difference in engine efficiency between the present trial and the earlier trial. That is, the cars in the present trial appear to have greater engine efficiency than that of those in the earlier trial —  $\underline{t}$  (21) = 5.74, p < .05.

#### **One-Sample Statistics**

	Ν	Mean	Std. Deviation	Std. Error Mean
withadd	22	13.86	2.748	.586

#### **One-Sample Test**

		Test Value = 10.5					
				Mean	95% Confidence Interval the Difference		
	t	df	Sig. (2-tailed)	Difference	Lower	Upper	
withadd	5.741	21	.000	3.364	2.15	4.58	

# T-tests with more than one sample



In the last section, a one-sample t-test is used to determine whether a single sample of scores was likely to have been drawn from a hypothesised population. This section extends the understanding of sampling distributions to ask whether two sets of scores are random samples from the same or different populations. If they are random samples from the same population, then any differences across conditions or groups can be attributed to random sampling variability. However, if the two sets of scores are random samples from different populations, then you can attribute any difference between means across conditions to the independent variable or the treatment effect.

# Repeated measures t-test

The repeated measures t-test, also referred to as the dependent-samples or paired t-test, is used when you have data from only one group of participants. In other words, an individual obtains two scores under different levels of the independent variable. Data that are collected from the same group of participants are also referred to as within-subjects, because the same subject performs in both conditions. Studies which employ a pretest–posttest design are commonly analysed using repeated measures t-tests. In this form of design, the same participant obtains a score on the pretest and, after some intervention or manipulation, a score on the posttest. You wish to determine whether the difference between means for the two sets of scores is the same or different.

Before you attempt to answer this question, you must ensure that the assumptions of repeated measures t-test are met. Remember, from the section on assumption testing, that a number of assumptions are generic to all types of t-test. The repeated measures t-test has one additional assumption:

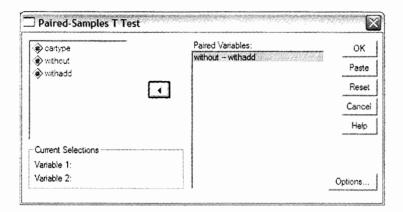
Normality of population difference scores — the difference between the scores for each participant should be normally distributed. Providing the sample size is not too small (30+), violations of this assumption are of little concern.

Testing this assumption involves the same procedures used for the single-sample t-test. Because you have two dependent variables, you need to test the normality of each variable separately, which allows you to assume that the difference scores are normally distributed. Having evaluated the assumption of normality for both pretest and posttest measures, you are ready to conduct a repeated measures t-test.



## To conduct a repeated measures t-test

- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and then Paired-Samples T Test... to open the Paired-Sample T Test dialogue box.
- 3 Select the variables you require (i.e. *without* and *withadd*) and press the ▶ button to move the variables into the **Paired Variables:** box.



#### 4 Click on OK.

T-TEST
PAIRS= without WITH withadd (PAIRED)
/CRITERIA=C1(.95)
/MISSING=ANALYSIS.

#### **Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	without withadd		22 22	3.335 2,748	.711 .586

#### **Paired Samples Correlations**

		N	Correlation	Sig.	
Pair 1	without & withadd	22	.559	.007	

#### **Paired Samples Test**

	***	Pa	aired Differenc	es				
		Std.	Std. Error	95% Confidence Interval of the Difference				Sig.
	Mean	Deviation	Mean	Lower	Upper	t	df	(2-tailed)
Pair 1 without – withadd	-5.364	2.904	.619	-6.651	-4.076	-8.663	21	.000

By looking at the t-value, df and two-tail significance you can determine whether the groups come from the same or different populations. The correct way to determine significance is to consult the critical t-tables that are available at the back of most statistical textbooks, using the degrees of freedom. However, significance can also be determined by looking at the probability level (p) specified under the heading 'two-tail significance'. If the probability value is less than the specified alpha value, then the observed t-value is significant. The 95 per cent confidence interval indicates that 95 per cent of the time the interval specified will contain the true difference between the population means.

As can be seen from the output, a significant difference exists between engine efficiency with and without the additive. The additive significantly improves the number of kilometres to the litre,  $\underline{t}(21) = -8.66$ , p < .05).

# Independent groups t-test

An independent groups t-test is appropriate when different participants have performed in each of the different conditions, in other words, when the participants in one condition are different from the participants in the other condition. This is commonly referred to as a between-subjects design. Again, you wish to determine whether the difference between means for the two sets of scores is significant. The independent groups t-test has two additional assumptions.

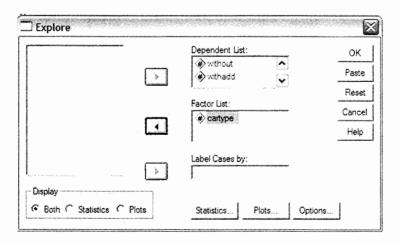
- 1 **Independence of groups** participants should appear in only one group and these groups are unrelated.
- **Homogeneity of variance** the groups should come from populations with equal variances. To test for homogeneity of variance, SPSS uses the Levene test for equality of variances. If this test is significant (p < .05), then you reject the null hypothesis and accept the alternative hypothesis that the variances are unequal. In this instance the unequal variance estimates are consulted. If the test is not significant (p > .05), then you accept the null hypothesis that there are no significant differences between the variances of the groups. In this case you would consult the equal variance estimates. This explanation will make more sense when you consult the output of the independent groups t-test.

Assumption 1 is a matter of research design while assumption 2 is tested in the independent groups analysis. Before proceeding you need to check the normality of the data. Because you have different participants in each condition, you need to check the normality of each set of scores separately. This is achieved through the **Explore** dialogue box using the **Factor List** option.



## To screen for normality

- 1 Select the **Analyze** menu.
- 2 Click on Descriptive Statistics and then Explore... to open the Explore dialogue box.
- 3 Select the dependent variable(s) (i.e. *without* and *withadd*) and click on the ▶ button to move the variables into the **Dependent List:** box.
- 4 Select the grouping variable (i.e. *cartype*) and click on the ▶ button to move this variable into the **Factor List:** box.



5 Click on **OK**.

EXAMINE
VARIABLES=without withadd BY cartype
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.

#### Descriptives

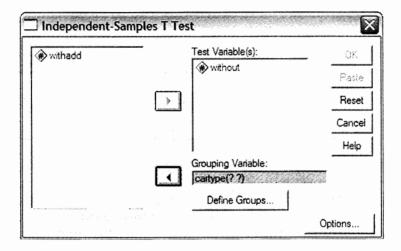
	cartype			Statistic	Std. Error
without	manual	Mean		8.00	.863
		95% Confidence	Lower Bound	6.08	.000
		Interval for Mean	Upper Bound	9.92	
		5% Trimmed Mean	Opper Bound	7.94	
		Median		7.00	
		Variance		8.200	
		Std. Deviation		2.864	
		Minimum		4	
		Maximum		13	
		Range		9	
		Interquartile Range		5	
		Skewness		.656	.661
		Kurtosis		704	1.279
	automatic	Mean		9.00	1.152
	datomatio	95% Confidence	Lower Bound	6.43	1.102
		Interval for Mean	Upper Bound	11.57	
		5% Trimmed Mean	Opper Bourid		
			-	8.89	
		Median		9.00	
		Variance		14.600	
		Std. Deviation		3.821	
		Minimum		4	
		Maximum		16	
		Range		12	
		Interquartile Range		6	
		Skewness		.355	.661
		Kurtosis		253	1.279
withadd	manual	Mean		14.36	.622
		95% Confidence	Lower Bound	12.98	.022
		Interval for Mean	Upper Bound	15.75	
		5% Trimmed Mean	Opper Bound		
				14.40	
		Median		14.00	
		Variance		4.255	
		Std. Deviation		2.063	
		Minimum		11	
		Maximum		17	
		Range		6	
		Interquartile Range		4	
		Skewness		013	.661
		Kurtosis		-1.012	1.279
	automatic	Mean		13.36	1,002
	adiomatio	95% Confidence	Lower Bound	11.13	1.002
		Interval for Mean	Upper Bound	15.60	
			Opper Bourid		
		5% Trimmed Mean		13.35	
		Median		12.00	
		Variance		11.055	
		Std. Deviation		3.325	
		Minimum		8	
		Maximum		19	
		Range		11	
		Interquartile Range		5	
		Skewness		.169	.661
		Kurtosis		749	1.279
		Nul tools		/49	1.219

In reviewing the descriptive statistics and the other output such as stem-and-leaf plots and boxplots (not shown), it is clear that there is minimal violation to the assumption of normality.

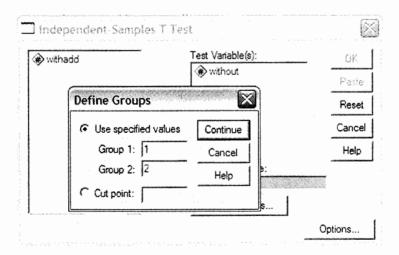


## To conduct an independent groups t-test

- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and then Independent-Samples T Test... to open the Independent Samples T Test dialogue box.
- 3 Select the test variable(s) (i.e. *without*) and then click on the ▶ button to move the variables into the **Test Variable(s)**: box.
- 4 Select the grouping variable (i.e. *cartype*) and click on the ▶ button to move the variable into the **Grouping Variable:** box.



- 5 Click on the **Define Groups...** command pushbutton to open the **Define Groups** sub-dialogue box.
- In the **Group 1:** box, type the lowest value for the variable (i.e. 1), then tab. Enter the second value for the variable (i.e. 2) in the **Group 2:** box.



7 Click on Continue and then OK.

T-TEST
GROUPS=cartype(1 2)
/MISSING=ANALYSIS
/VARIABLES=without
/CRITERIA=CI(.95) .

The syntax for the independent groups t-test is different from that of the repeated measures t-test. In the case of the independent groups t-test you have a grouping variable so you can distinguish between groups 1 and 2 when comparing engine efficiency.

#### **Group Statistics**

	cartype	N	Mean	Std. Deviation	Std. Error Mean
without	manual	11	8.00	2.864	.863
	automatic	11	9.00	3.821	1.152

#### **Independent Samples Test**

			Test for Variances		t-test for Equality of Means					
				,		Sig.	Mean	Std. Error	95% Confide of the D	ence Interval ifference
		F	Sig.	t	df	(2-tailed)		Difference	Lower	Upper
without	Equal variances assumed	.172	.683	695	20	.495	-1.000	1.440	-4.003	2.003
	I variances of assumed			695	18.539	.496	-1.000	1.440	-4.018	2.018

Given that Levene's test has a probability greater than .05, you can assume that the population variances are relatively equal. Therefore, you can use the t-value, df and two-tail significance for the equal variance estimates to determine whether car type differences exist. The two-tail significance for *without* additive indicates that p > .05 and thus is not significant. You therefore accept the null hypothesis and reject the alternative hypothesis. The two groups must come from the same population because no significant differences exist, t = 0.05.

Although it is possible to perform two t-tests with the one command, for the sake of clarity two separate procedures are shown.

T-TEST
GROUPS = cartype(1 2)
/MISSING = ANALYSIS
/VARIABLES = withadd
/CRITERIA = CI(.95)

#### **Grouup Statistics**

	cartype	Ν	Mean	Std. Deviation	Std. Error Mean
withadd	manual	11	14.36	2.063	.622
	automatic	11	13.36	3.325	1.002

#### Independent Samples Test

			Test for Variances		t-test for Equality of Means					
						Sig.	Mean	Std. Error	95% Confidence Interval of the Difference	
		F	Sig.	t	df	(2-tailed)	Difference	Difference	Lower	Upper
withadd	Equal variances assumed	3.390	.080	.848	20	.407	1.000	1.180	-1.461	3.461
	al variances ot assumed			.848	16.704	.409	1.000	1.180	-1.492	3.492

In relation to *withadd*, Levene's test was not significant and thus we interpret the equal variance estimates. Consulting our t-value, df and two-tail significance, again no significant differences are apparent (p > .05). That is, there is no significant difference in engine efficiency between manual and automatic cars either with or without the additive, t (20) = .848, p > .05.

# Practice example



You have been asked to determine whether hypnosis enhances memory. Forty men and women are given five minutes to attempt to memorise a list of unrelated words. They are then asked to recall as many as they can. The next week they are asked to memorise a similar list of words and then to recall as many as possible while under hypnosis. You performed a study last year with another sample, so you have access to descriptive statistics from a similar group of adults. The mean number of words recalled in the earlier study, without hypnosis, was 34.6. Given the data in Prac6.sav, your tasks are to:

- 1 Determine whether the participants in the present study are comparable with those in the earlier study in terms of recall in a normal state.
- 2 Determine whether there was any change in recall as a result of hypnosis for the entire sample.
- 3 Determine whether men and women recall equal numbers of words when under hypnosis.

# **Solutions**



## Syntax

```
EXAMINE
 VARIABLES=natrecal hyprecal BY gender
 /PLOT BOXPLOT STEMLEAF
 /COMPARE GROUP
 /STATISTICS DESCRIPTIVES
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
T-TEST
 /TESTVAL = 34.6
 /MISSING = ANALYSIS
 /VARIABLES = natrecal
 /CRITERIA = CI(.95).
 PAIRS = natrecal WITH hyprecal (PAIRED)
 /CRITERIA = CI(.95)
 /MISSING = ANALYSIS.
T-TEST
 GROUPS = gender(1 2)
 /MISSING = ANALYSIS
 /VARIABLES = hyprecal
 /CRITERIA = CI(.95).
```

## **Assumption testing**

## Descriptives

	gender			Statistic	Std. Error
recall in natural state	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	46.55 40.77 52.33 46.33 45.50 152.261 12.339 21 76 55 12 .357 1.218	2.759 .512 .992
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	39.75 33.67 45.83 39.06 36.50 168.724 12.989 21 71 50 19 .881	2.905 .512 .992
recall under hypnosis	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	69.95 64.11 75.79 70.00 70.50 155.629 12.475 45 94 49 15 218	2.790 .512 .992
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	58.00 52.77 63.23 58.17 58.50 124.737 11.169 35 78 43 14 266 126	.512 .992

Normality was assessed through the **Explore Option**. An examination of the output reveals that the data are normally distributed for each group.

#### **One-Sample Statistics**

				Std. Error
	N	Mean	Std. Deviation	Mean
recall in natural state	40	43.15	12.970	2.051

#### One-Sample Test

		Test Value = 34.6								
				Mean	95% Confidence Interval of the Difference					
	t	df	Sig. (2-tailed)	Difference	Lower	Upper				
recall in natural state	4.169	39	.000	8.550	4.40	12.70				

The t-value is significant (p < .05) and therefore it can be concluded that the present sample is significantly different from the previous sample. Inspection of the means suggests that the present sample has better word recall than that of the earlier sample, t (39) = 4.169, p < .05.

T-tests for paired samples

#### **Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	recall in natural state recall under hypnosis		40 40	12.970 13.161	2.051 2.081

## **Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 recall in natural state & recall under hypnosis		.283	.077

#### **Paired Samples Test**

			P	aired Differe	nces				Sig.
	:		Std.	Std. Error	95% Confide of the Di				
		Mean	Deviation	Mean	Lower	Upper	Т	df	(2-tailed)
Pair 1	recall in natural state – recall								
	under hypnosis	-20.825	15.647	2.474	-25.829	-15.821	-8.418	39	.000

The t-value is significant (p < .05) and therefore it can be stated that there is a significant improvement in recall when the participants are under hypnosis,  $\underline{t}$  (39) = -8.418, p < .05.

T-tests for independent group samples

#### **Group Statistics**

	gender	N	Mean	Std. Deviation	Std. Error Mean
recall under hypnosis	female	20	69.95	12.475	2.790
	male	20	58.00	11.169	2.497

#### Independent Samples Test

			Test for Variances	t-test for Equality of Means						
					Sig. Mean Std. Error		Std. Error		ence Interval ifference	
		F	Sig.	t	df	(2-tailed)	Difference	Difference	Lower	Upper
recall under hypnosis	Equal variances assumed	.036	.851	3.192	38	.003	11.950	3.744	4.370	19.530
	l variances ot assumed			3.192	37.544	.003	11.950	3.744	4.367	19.533

Given that the Levene's test has a probability greater than .05 it can be assumed that there is homogeneity of variance. From the equal estimates t-value it can be seen that there is significance (p < .05) and therefore it can be stated that there was a significant difference between men and women in the recall of words under hypnosis,  $\underline{t}$  (38) = 3.192, p < .05. Inspection of the means in the group statistics table suggests that women (Mean = 69.95) recalled more words under hypnosis than men (Mean = 58).



# One-way between-groups ANOVA with post-hoc comparisons

In the last chapter you tested the null hypothesis that two population means were equal. When you wish to compare the means of more than two groups or levels of an independent variable, a one-way analysis of variance (ANOVA) is appropriate.

At the heart of ANOVA is the notion of variance. The basic procedure is to derive two different estimates of population variance from the data, then calculate a statistic from the ratio of these two estimates. One of these estimates (between-groups variance) is a measure of the effect of the independent variable combined with error variance. The other estimate (within-groups variance) is of error variance by itself. The F-ratio is the ratio of between-groups variance to within-groups variance. A significant F-ratio indicates that the population means are probably not all equal. Because the null hypothesis is rejected if any pair of means is unequal, where the significant differences lie needs to be worked out. This requires post-hoc analysis.

Post-hoc analysis involves hunting through the data for any significance, that is, doing an entire set of comparisons. This type of testing carries risks of type I errors. Unlike planned comparisons, post-hoc tests are designed to protect against type I errors, given that all the possible comparisons are going to be made. These tests are stricter than planned comparisons and so it is harder to obtain significance.

There are a number of post-hoc tests available. The more options a test offers, the stricter its determination of significance. The Scheffe test, for example, allows every possible comparison to be made but is tough on rejecting the null hypothesis. In contrast, Tukey's honestly significant difference (HSD) test is more lenient, but the types of comparison that can be made are restricted. This chapter illustrates the Tukey HSD post-hoc test.

# Assumption testing



Before conducting the ANOVA the necessary assumptions must be met. The assumptions for ANOVA are the same as those for the t-test.

The two assumptions of concern are:

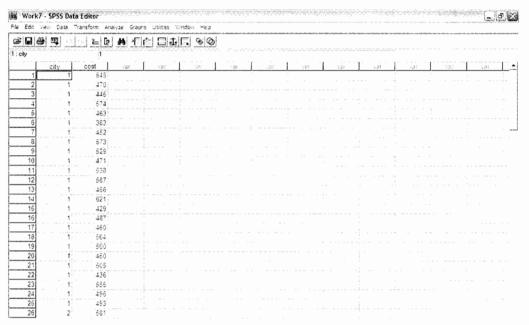
- 1 **Population normality** populations from which the samples have been drawn should be normal. Check this for each group using normality statistics such as skewness and Shapiro-Wilks.
- 2 Homogeneity of variance the scores in each group should have homogeneous variances. As with the t-test, Levene's test determines whether variances are equal or unequal.

# Working example



An economist wished to compare household expenditure on electricity and gas in four major cities in Australia. She obtained random samples of 25 two-person households from each city and asked them to keep records of their energy expenditure over a sixmonth period. Note that this is an independent groups design because different households are in different cities. If the same participants were in all conditions, then it would be a within-subjects or repeated measures design. A repeated measures design requires a different class of procedures, so this design is discussed in chapter 10.

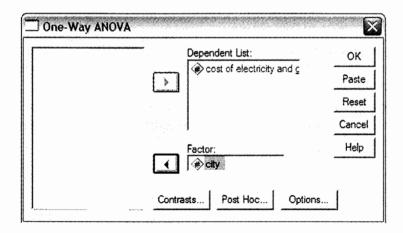
The data file can be found in Work7.sav on the website that accompanies this title and is shown in the following figure.



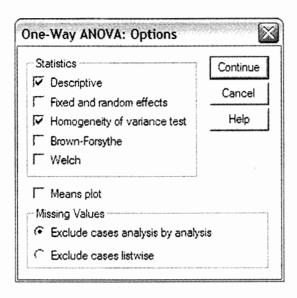


## To conduct a one-way ANOVA with post-hoc analysis

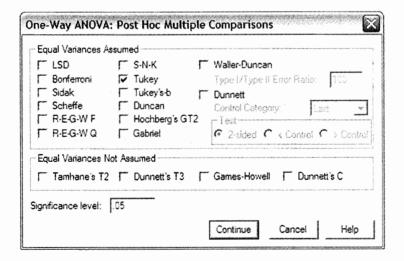
- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and One-Way ANOVA... to open the One-Way ANOVA dialogue box.
- 3 Select the dependent variable (i.e. *cost*) and click on the ▶ button to move the variable into the **Dependent List:** box.
- 4 Select the independent variable (i.e. *city*) and click on the **▶** button to move the variable into the **Factor:** box.



- 5 Click on the **Options...** command pushbutton to open the **One-Way ANOVA: Options** sub-dialogue box.
- 6 Click on the check boxes for Descriptive and Homogeneity-of-variance.



- 7 Click on Continue.
- 8 Click on the **Post Hoc...** command pushbutton to open the **One-Way ANOVA: Post Hoc Multiple Comparisons** sub-dialogue box. You will notice that a number of multiple comparison options are available. In this example you will use the Tukey's HSD multiple comparison test.
- 9 Click on the check box for Tukey.



10 Click on Continue and then OK.

ONEWAY
cost BY city
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS
/POSTHOC = TUKEY ALPHA(.05).

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Adelaide	25	497.28	56.628	11.326	473.91	520.65	383	621
Hobart	25	515.84	56.529	11.306	492.51	539.17	397	647
Melbourne	25	531.20	63.976	12.795	504.79	557.61	397	677
Perth	25	555.12	72.576	14.515	525.16	585.08	429	739
Total	100	524.86	65.385	6.539	511.89	537.83	383	739

In interpreting this output you must first ensure that the homogeneity assumption has not been violated. Levene's test for homogeneity of variances is not significant (p > .05) and therefore you can be confident that the population variances for each group are approximately equal.

Test of Homogeneity of Variances

cost of electricity and gas

Levene Statistic	df1	df2	Sig.
.817	3	96	.488

To determine whether you have a significant F-ratio you use the degrees of freedom (df) (3, 96), the F-ratio and the F-probability. Again, the correct way to determine significance is to use the Critical F tables. Significance can also be determined by looking at the F-probability value. Given that p < .05 you can reject the null hypothesis and accept the alternative hypothesis that states that expenditure on electricity and gas is different across capital cities, F(3,96) = 3.802, p < .05.

**ANOVA** 

cost of electricity and gas

	Sum of squares	df	Mean Square	F	Sig.
Between Groups	44947.000	3	14982.333	3.802	.013
Within Groups	378299.040	96	3940.615		
Total	423246.040	99			

Having obtained a significant result you can also go further and determine, using a Tukey HSD test, where the significance lies; that is, between which cities is there a significant difference in energy costs? You can see by looking at the results of the Tukey test in the table that follows that Adelaide and Perth have significantly different mean energy costs.

**Multiple Comparisons** 

Dependent Variable: cost of electricity and gas

Tukey HSD

		Mean			95% Confide	ence Interval
(I) city	(J) city	Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Adelaide	Hobart	-18.560	17.755	.723	-64.98	27.86
	Melbourne	-33.920	17.755	.231	-80.34	12.50
Hobart	<b>Perth</b>	<b>-57.840*</b>	<b>17.755</b>	. <b>008</b>	-104.26	-11.42
	Adelaide	18.560	17.755	.723	-27.86	64.98
	Melbourne	-15.360	17.755	.823	-61.78	31.06
	Perth	-39.280	17.755	.127	-85.70	7.14
Melbourne	Adelaide	33.920	17.755	.231	-12.50	80.34
	Hobart	15.360	17.755	.823	-31.06	61.78
Perth	Perth	-23.920	17.755	.535	-70.34	22.50
	<b>Adelaide</b>	<b>57.840</b> *	<b>17.755</b>	. <b>008</b>	11.42	104.26
	Hobart	39.280	17.755	.127	-7.14	85.70
	Melbourne	23.920	17.755	.535	-22.50	70.34

<sup>\*</sup>The mean difference is significant at the .05 level.

# Practice example



A biologist wished to examine the nutrient value of six different food supplements. One hundred and fifty-four rats of the same species were randomly assigned to one of six groups. Each group had a different supplement added to its food, and the rats' weight gain over the ensuing six months was recorded in grams. Given the data in Prac7.sav, your tasks are to:

- 1 Test the underlying assumptions of ANOVA.
- 2 Determine whether there are significant differences in weight gain across the food supplements.
- 3 Locate the source of these differences using post-hoc analysis.

# **Solutions**



## Syntax

EXAMINE
VARIABLES=wtgain BY supp
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
ONEWAY
wtgain BY supp
/STATISTICS HOMOGENEITY
/MISSING ANALYSIS
/POSTHOC = TUKEY ALPHA(.05).

## Output

#### **Assumption Testing**

Normality

#### **Descriptives**

	food supplement			Statistic	Std. Error
weight gain	supplement A	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	12.00 9.74 14.26 12.05 11.00 26.000 5.099 1 22 21 7 .062 195	1.087 .491 .953

(continued)

## Descriptives (continued)

food supplement			Statistic	Std. Error
supplement B	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	15.44 13.02 17.86 15.26 16.00 34.257 5.853 7 27 20 9 .486 638	.464 .902
supplement C	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	8.14 6.61 9.66 7.98 6.50 11.838 3.441 4 15 11 5	.734 .491 .953
supplement D	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	8.41 7.58 9.23 8.59 9.00 4.328 2.080 2 11 9 3 -1.367 2.538	.400 .448 .872
supplement E	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	1.13 .86 1.41 1.15 1.00 .533 .730 0 2 2 1 214	.133 .427 .833
supplement F	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range	Lower Bound Upper Bound	6.64 6.22 7.07 6.62 7.00 1.201 1.096 5 9 4	.207
	Skewness Kurtosis		.061 619	.441 .858

The distributions of weight gain scores for each food supplement, except for supplement D, are normal. In the case of weight gain for group D, there is a slight negative skew and kurtosis. However, because only one of six groups is affected, it is appropriate to proceed with the ANOVA.

#### Test of Homogeneity of Variances

weight gain

Levene Statistic	df1	df2	Sig.	
19.822	5	148	.000	

#### **ANOVA**

weight gain

	Sum of squares	df	Mean Square	F	Sig.
Between Groups	3194.264	5	638.853	53.203	.000
Within Groups	1777.165	148	12.008		
Total	4971.429	153			

#### **Multiple Comparisons**

Dependent Variable: weight gain

Tukey HSD

		Mean			95% Confide	ence Interval
		Difference			Lower	Upper
(I) food supplement	(J) food supplement	(I-J)	Std. Error	Sig.	Bound	Bound
supplement A	supplement B	*-3.440	1.013	.011	-6.36	52
	supplement C	*3.864	1.045	.004	.85	6.88
	supplement D	*3.593	.995	.006	.72	6.47
	supplement E	*10.867	.973	.000	8.06	13.68
	supplement F	*5.357	.987	.000	2.51	8.21
supplement B	supplement A	*3.440	1.013	.011	.52	6.36
	supplement C	*7.304	1.013	.000	4.38	10.23
	supplement D	*7.033	.962	.000	4.26	9.81
	supplement E	*14.307	.938	.000	11.60	17.02
	supplement F	*8.797	.954	.000	6.04	11.55
supplement C	supplement A	*-3.864	1.045	.004	-6.88	85
	supplement B	*-7.304	1.013	.000	-10.23	-4.38
	supplement D	271	.995	1.000	-3.14	2.60
	supplement E	*7.003	.973	.000	4.19	9.81
	supplement F	1.494	.987	.657	-1.36	4.34
supplement D	supplement A	*-3.593	.995	.006	-6.47	72
	supplement B	*-7.033	.962	.000	-9.81	-4.26
	supplement C	.271	.995	1.000	-2.60	3.14
	supplement E	*7.274	.919	.000	4.62	9.93
	supplement F	1.765	.935	.414	93	4.46
supplement E	supplement A	*-10.867	.973	.000	-13.68	-8.06
	supplement B	*-14.307	.938	.000	-17.02	-11.60
	supplement C	*-7.003	.973	.000	-9.81	-4.19
	supplement D	*-7.274	.919	.000	-9.93	-4.62
	supplement F	*-5.510	.911	.000	-8.14	-2.88
supplement F	supplement A	*-5.357	.987	.000	-8.21	-2.51
	supplement B	*-8.797	.954	.000	-11.55	-6.04
	supplement C	-1.494	.987	.657	-4.34	1.36
	supplement D	-1.765	.935	.414	-4.46	.93
	supplement E	*5.510	.911	.000	2.88	8.14

<sup>\*</sup> The mean difference is significant at the .05 level.

There is a significant difference in weight gain across the different food supplements, F(5,148) = 53.203, p < .05. However, the homogeneity of variance assumption has been violated (p < .05) and therefore any interpretation of the main effect must be undertaken with caution. The results of the HSD post-hoc test identify where differences in weight gains lie.



# One-way between-groups ANOVA with planned comparisons

Planned (or *a priori*) comparisons are used when you have specific expectations or predictions about some of the results. These comparisons are often of theoretical importance and are planned from the onset of the study. Post-hoc comparisons, on the other hand, such as Tukey's honestly significantly different (HSD) test outlined in chapter 7, are not theoretically driven. Instead, every possible comparison is performed to detect significant effects.

Both planned and post-hoc analyses have problems with type I and type II error properties. For example, in relation to planned comparisons there are problems associated with performing multiple statistical tests on a set of data. Such action increases the vulnerability of the test to Type I errors. The more tests you perform, the greater the chances of Type I error on at least one of those tests. You should be familiar with two types of Type I error rate. These include:

- 1 The per comparison error rate the error rate in a single comparison. For each comparison that you make, the alpha level is set at .05.
- The familywise error rate the cumulative effects of doing many separate tests. It is the probability that you have made at least one type I error in the complete set of comparisons you have undertaken. For example, if you wish to make three comparisons, each at an alpha level of .05, then the familywise error rate would be 0.15, which is equal to  $3 \times .05$ .

There is much debate concerning whether the familywise error rate should be controlled when performing multiple planned comparisons. In most instances you are permitted to perform one less comparison than the number of levels or groups of your independent variable. For example, with three levels, you could perform two comparisons without any difficulty. However, if further comparisons are required using the same data, then a more strict alpha level should be adopted. The Bonferroni test assists you in determining the appropriate alpha level to be used to evaluate the significance of the comparisons. If you wish to conduct four comparisons using a familywise error rate of .15, then each comparison would have to be evaluated against an alpha level of .025 (.15/4 = .025).

In chapter 7 you discovered how post-hoc tests, such as Tukey's HSD test, can be used to locate significant differences among groups on a dependent variable. However, if you have theoretical reasons for hypothesising differences, you can also conduct planned comparisons to investigate the nature of those differences. A comparison between any two of the group means is referred to as a simple pairwise comparison. It is also possible to conduct more complex planned or post-hoc comparisons.

In ANOVA the null hypothesis states that all population means associated with each group or condition under investigation are the same. Furthermore, if the null hypothesis is true, the difference (comparison value) between any simple or complex comparison of population means must be zero. To specify the comparison between any mean (or average of a set of means) with any other mean (or average of a different set of means)

requires that means for the appropriate groups are weighted using a set of coefficients. A description of how these coefficients are established is beyond the scope of this chapter, but information regarding coefficients can be obtained from any good statistical text.

# Assumption testing



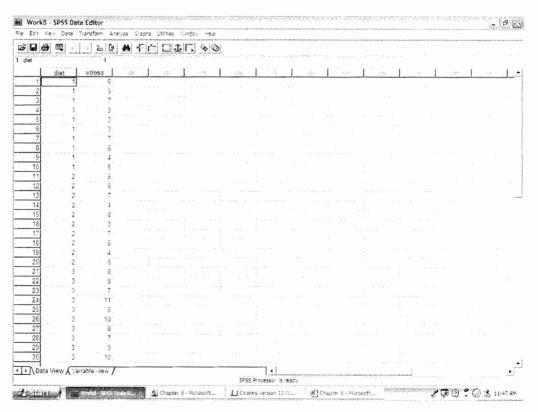
In relation to assumption testing, planned and post-hoc analyses are based on the analytic procedure that precedes them. In this chapter you are addressing the use of planned comparisons with the one-way ANOVA procedure, thus the assumptions underlying this test apply.

# Working example



A dietary consultant has asked you to test the efficacy of three weight-reduction programs. Carbohydrates were restricted in program A, protein was restricted in program B and fats were restricted in program C. Ten overweight volunteers were randomly assigned to each of the programs, and their weight loss after eight weeks was recorded in kilograms. Positive scores signify a weight drop. The dietitian predicted that the diet type would influence the weight loss, and that the loss would be greatest for those restricting fats (program C). Before analysis, checks for normality were conducted and considered to be acceptable.

The data file can be found in Work8.sav on the website that accompanies this title and is shown in the following figure.

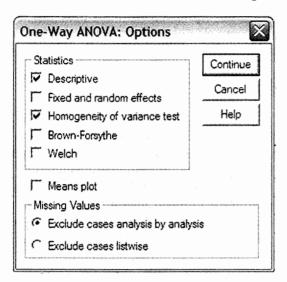




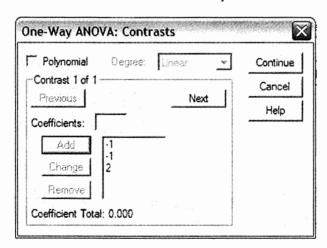
To conduct a one-way ANOVA with a planned comparison

- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and One-Way ANOVA... to open the One-Way ANOVA dialogue box.
- 3 Select the dependent variable (i.e. *wtloss*) and click on the button to move the variable into the **Dependent List:** box.

- 4 Select the independent variable (i.e. *diet*) and click on the ▶ button to move the variable into the **Factor:** box.
- 5 Click on the **Options...** command pushbutton to open the **One-Way ANOVA: Options** sub-dialogue box.
- 6 Select the check boxes for **Descriptive** and **Homogeneity of variance**.



- 7 Click on Continue.
- 8 Click on the Contrasts... command pushbutton to open the One-Way ANOVA: Contrasts sub-dialogue box.
- In the **Coefficients:** box, type the value of the coefficient for the first group (i.e. -1) and click on **Add** to move the coefficient into the box below.
- In the **Coefficients:** box, type the value of the coefficient for the second group (i.e. -1) and click on **Add**.
- In the **Coefficients:** box, type the value of the coefficient for the third group (i.e. 2) and click on **Add**. If all the coefficients have been entered correctly, the **Coefficient Total:** should equal zero.



12 Click on Continue and then OK.

ONEWAY
wtloss BY diet
/CONTRAST= -1 -1 2
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS .

#### Descriptives

#### weight loss

					95% Confidence Interval for Mean			
	Ν	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
restricted carbohydrate	10	4.90	1.595	.504	3.76	6.04	3	7
restricted protein	10	5.50	1.581	.500	4.37	6.63	3	8
restricted fat	10	8.40	1.776	.562	7.13	9.67	5	11
Total	30	6.27	2.227	.407	5.43	7.10	3	11

An examination of the Levene test for homogeneity of variances (below) suggests that this assumption has not been violated (p > .05) and therefore interpretation of the ANOVA can proceed.

#### Test of Homogeneity of Variances

#### weight loss

Levene Statistic	df1	df2	Sig.
.038	2	27	.963

#### ANOVA

#### weight loss

	Sum of squares	df	Mean Square	F	Sig.
Between Groups	70.067	2	35.033	12.817	.000
Within Groups	73.800	27	2.733		
Total	143.867	29			

The F-ratio with an F-probability value less than .05 is significant, suggesting that type of diet does significantly influence weight loss, F(2,27) = 12.817, p < .05.

Means and standard deviations for each group are illustrated. The contrast coefficient matrix gives the coefficients that have been assigned for each group or level of the independent variable. For example, program A, program B and program C have coefficients of 1, 1 and 2 respectively.

## **Contrast Coefficients**

	diet program					
Contrast	restricted carbohydrate	restricted protein	restricted fat			
1	-1	-1	2			

#### **Contrast Tests**

		Contrast	Value of contrast	Std. Error	t	df	Sig. (2-tailed)
weight loss	Assume equal variances	1	6.40	1.281	4.998	27	.000
	Does not assume equal variances	1	6.40	1.329	4.815	16.327	.000

You have not violated the assumption of homogeneity of variance so you can consult the 'Assume equal variances' row for Contrast 1. To interpret this contrast you need to examine both the T-value and the T-probability value. A T-value of 4.988 is highly significant (T-probability < .05). Because you have made only one comparison, it is not necessary to adjust this significance level to evaluate the comparison.

An examination of the means for each of these groups, located in the descriptive statistics section of the output, indicates that a significantly greater reduction in weight was made by those in program C than in either of the other programs. Therefore, you can conclude that individuals restricting their fat intake lose significantly more weight than do those restricting either their protein or carbohydrate intake, F(1,27) = 24.98, p < .05.

As outlined above, to report the planned comparison, the T-value is squared to obtain an F-value. Also note that the df1 for planned comparisons is always 1. The df2 value is that which corresponds to the within-groups estimate in the ANOVA summary table.

# Practice example



You will remember the practice example in chapter 7 that examined the effect of food supplements on weight gain in rats. One hundred and fifty-four rats were randomly assigned to receive one of six food supplements, and their weight gain over six months was recorded. Given the data in Prac8.sav, your tasks are to:

- 1 Test the underlying assumptions of ANOVA.
- 2 Determine whether there are significant differences in weight gain across the six food supplements.
- 3 Test the hypothesis that supplement B is associated with a significantly higher weight gain than that associated with the other five supplements.

# **Solutions**



# **Syntax**

ONEWAY
wtgain BY supp
/CONTRAST= -1 5 -1 -1 -1 -1
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS .

## Output

#### Assumption testing

You will remember that you tested these assumptions as part of the Practice example in chapter 7, so refer to the previous output.

#### Descriptives

weight gain

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
supplement A	22	12.00	5.099	1.087	9.74	14.26	1	22
supplement B	25	15.44	5.853	1.171	13.02	17.86	7	27
supplement C	22	8.14	3.441	.734	6.61	9.66	4	15
supplement D	27	8.41	2.080	.400	7.58	9.23	2	11
supplement E	30	1.13	.730	.133	.86	1.41	0	2
supplement F	28	6.64	1.096	.207	6.22	7.07	5	9
Total	154	8.29	5.700	.459	7.38	9.19	0	27

The ANOVA shows that there is a significant difference in weight gain across the different food supplements (p < .05) — F(5,148) = 53.203, p < .05.

#### Test of Homogeneity of Variances

#### weight gain

Levene Statistic	df1	df2	Sig.
19.822	5	148	.000

#### ANOVA

#### weight gain

	Sum of squares	df	Mean Square	F	Sig.
Between Groups	3194.264	5	638.853	53.203	.000
Within Groups	1777.165	148	12.008		
Total	4971.429	153			

Because the homogeneity of variance assumption has been violated (p < .05), in interpreting the constrast tests, we need to consult the T-probability value in the 'Does not assume equal' row.

#### **Contrast Coefficients**

	food supplement						
Contrast	supplement A	supplement B	supplement E	supplement F			
1	-1	5	-1	-1	-1	-1	

#### **Contrast Tests**

		Contrast	Value of contrast	Std. Error	t	df	Sig. (2-tailed)
weight gain	Assume equal variances	1	40.88	3.791	10.783	148	.000
	Does not assume equal variances	1	40.88	6.016	6.795	26.752	.000

This is significant (p < .05) so we have shown that there is a significantly greater weight gain for supplement B than for any of the other five food supplements, F(1,148) = 46.17, p < .05. Remember, in reporting your result, square the appropriate t-value to obtain an F-value; also remember that df1 for planned comparisons is always 1 and that df2 value corresponds to the within-groups estimate in the ANOVA summary table.



# Two-way between-groups ANOVA

The two-way ANOVA operates in the same manner as the one-way ANOVA except that you are examining an additional independent variable. Each independent variable may possess two or more levels. In a two-factor between-groups design, each participant has been randomly assigned to only one of the different levels of each independent variable. Each of the different cells represents the unique combinations of the levels of the two factors.

## Assumption testing



Before you can conduct the two-way ANOVA, you must ensure that the necessary assumptions are met. The assumptions for two-way ANOVA are the same as those for the one-way ANOVA:

- 1 Population normality the populations from which the samples have been drawn should be normal. Check this for each group using normality statistics such as skewness and Shapiro-Wilks.
- **Homogeneity of variance** the scores in each group should have homogeneous variances.

You are primarily concerned with violations to the second assumption because such violations could mean that your data have been evaluated at a significance level greater than you initially assumed. Rather than being significant at the alpha level of .05, your results may really be significant at only an alpha level of .10. This is because violations of the homogeneity assumption distort the shape of the F-distribution such that the critical F-value no longer corresponds to a cut-off of 5 per cent.

# Working example



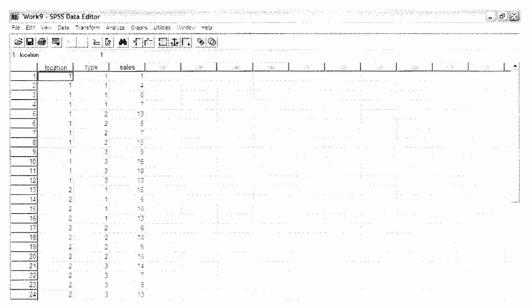
A toy distributor wished to determine which stores were the most successful in selling their stock. He wished to compare the sales in different types of store in different locations: in discount toy stores, department stores and variety stores, and stores in either the central city district or in suburban shopping centres. Therefore, the first independent variable was store type, with three levels; the second independent variable was location, with two levels; and the dependent variable was the amount of toy sales in \$1000s per week. Therefore, you have a  $3 \times 2$  factorial design with six data cells  $(3 \times 2 = 6)$ . Four stores were randomly chosen for each of the six cells (n = 4); sales from a total of 24 stores were recorded (N = 24).

The toy distributor wishes to ask three questions:

- 1 Does the type of store influence the sales of toys?
- 2 Does the location of the store influence the sales of toys?
- 3 Does the influence of the store type on toy sales depend on the location of the store?

Questions 1 and 2 refer to main effects, while question 3 examines the interaction of the two independent variables on the dependent variable.

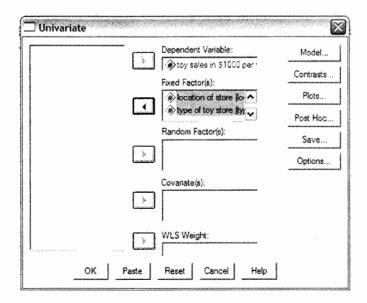
The data file can be found in Work9.sav on the website that accompanies this title and is shown in the following figure.





## To conduct a two-way ANOVA

- Select the Analyze menu.
- 2 Click on General Linear Model and then Univariate... to open the Univariate dialogue box.
- 3 Select the dependent variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Dependent Variable:** box.
- 4 Select the independent variables (i.e. *location* and *type*) and click on the **▶** button to move the variables into the **Fixed Factor(s)**: box.



- 5 Click on the **Options...** command pushbutton to open the **Univariate: Options** sub-dialogue box.
- In the Display box, click on the Descriptive statistics, Estimates of effect size, Observed power and Homogeneity tests check boxes.

Factor(s) and Factor Interactions: (OVERALL) (location type (location type	Display Means for:
	Compare main effects
	Confidence interval adjustment:
	LSD (none)
Display  Descriptive statistics	
Estimates of effect size	Spread vs. level plot
Observed power	Residual plot
Parameter estimates	☐ Lack of fit
Contrast coefficient matrix	General estimable function

## 7 Click on Continue and then OK.

UNIANOVA
sales BY location type
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PRINT = DESCRIPTIVE ETASQ OPOWER HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = location type location\*type .

## **Descriptive Statistics**

Dependent Variable: toy sales in \$1000 per week

location of store	type of toy store	Mean	Std. Deviation	N
city central suburban shopping centre Total	variety store department store discount toy store Total variety store department store discount toy store Total variety store department store department store discount toy store Total variety store department store discount toy store Total	3.00 10.00 14.00 9.00 11.00 12.00 10.00 11.00 7.00 11.00 12.00 10.00	3.162 4.761 3.916 5.970 3.916 5.477 4.082 4.200 5.398 4.870 4.276 5.150	4 4 4 12 4 4 4 12 8 8 8 8

## Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: toy sales in \$1000 per week

F	df1	df2	Sig.
1.000	5	18	.446

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+location+type+location \* type

The Levene's test tells us that the homogeneity of variance assumption has not been violated. The output illustrates that the main effects for location and store type are not significant (p > .05). Therefore, neither store type nor location significantly influences the sales of toys.

### Tests of Between-Subjects Effects

Dependent Variable: toy sales in \$1000 per week

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model Intercept Location Type location * type Error Total Corrected Total	<sup>b</sup> 280.000 2400.000 <b>24.000</b> <b>112.000</b> 144.000 330.000 3010.000 610.000	5 1 2 2 18 24 23	56.000 2400.000 <b>24.000</b> <b>56.000</b> <b>72.000</b> 18.333	3.055 130.909 <b>1.309</b> <b>3.055</b> <b>3.927</b>	.036 .000 .268 .072 .038	.459 .879 .068 .253 .304	15.273 130.909 1.309 6.109 7.855	.742 1.000 .192 .517 .630

- a Computed using alpha = .05
- b R squared = .459 (adjusted R squared = .309)

Because neither main effect is significant, post-hoc analyses are not required. If significant main effects had been obtained, a comparison of marginal means for each effect would have been necessary. You may have noticed that means can be displayed for particular factors in the **Univariate Options** dialogue box.

The output also shows that there is a significant interaction effect (p < .05) for location \* type. That is, the influence of store type on sales does depend on the location of the store, F(2,18) = 3.927, p < .05).

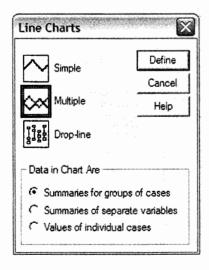
With a significant interaction effect you are interested in conducting a simple effects analysis, and simple comparisons if the simple effects are significant. In addition, interpretation is facilitated by plotting the cell means.

Effect sizes and observed power are also provided. Profile plots could be obtained as part of the analysis but the following procedure gives a clearer picture of the interaction.



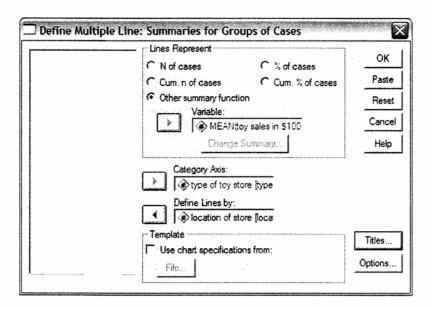
## To plot the cell means

- 1 Select the **Graphs** menu.
- 2 Click on Line... to open the Line Charts dialogue box.
- 3 Click on the **Multiple** box and ensure that the **Summaries for groups of cases** radio button has been selected.



- 4 Click on the **Define** command pushbutton to open the **Define Multiple Line:** Summaries for Groups of Cases sub-dialogue box.
- 5 In the Lines Represent box, select the Other summary function radio button.

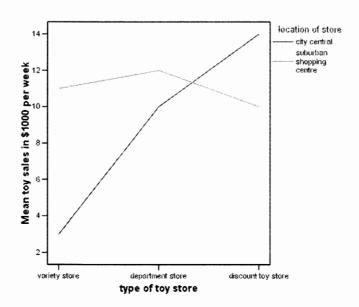
- 6 Select the dependent variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Variable:** box.
- Select the independent variable with the most levels (i.e. *type*) and click on the button to move the variable into the **Category Axis:** box.
- Select the other independent variable (i.e. *location*) and click on the delict button to move the variable into the **Define Lines by:** box. You will notice a **Titles...** command pushbutton in the bottom right-hand side of the window, which allows you to give titles to the lines of the graph. You can add a title for your graph if you want to.



## 9 Click on Continue and then OK.

GRAPH
/LINE(MULTIPLE)MEAN(sales) BY type BY location
/TITLE= 'Comparison of Sales'.

## Comparison of Sales



The graph shows that for suburban shopping centres, the type of store has no influence on the sale of toys. However, in the city centre, the type of store has considerable impact: discount toy stores have the highest sales.

When you have obtained a significant interaction, it is necessary to conduct an analysis of simple effects. That is, you need to look at the effect of one factor at only one level of the other factor. For example, you may be interested in looking at the effect of location on the variety stores, department stores or discount toy stores. Likewise, you could analyse the effect of the type of store just in the city centre or just for suburban shopping centres. In each case you would be analysing simple effects. If you were interested in looking at suburban shopping centres, you would first have to select only those cases in which stores were located in the suburbs. You can achieve this by using the **Select Cases** option in the **Data** menu or clicking the **Select Cases** tool.

USE ALL.

COMPUTE filter\_\$=(location = 2).

VARIABLE LABEL filter\_\$ 'location = 2 (FILTER)'.

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.

FORMAT filter\_\$ (f1.0).

FILTER BY filter\_\$.

EXECUTE .

You would then conduct a one-way ANOVA for store type for only these selected cases.

ONEWAY
sales BY type
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS .

## **Descriptives**

toy sales in \$1000 per week

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
variety store department store discount toy store Total	4 4 4 12	11.00 12.00 10.00 11.00	3.916 5.477 4.082 4.200	1.958 2.739 2.041 1.212	4.77 3.28 3.50 8.33	17.23 20.72 16.50 13.67	6 6 6 6	15 18 14 18

## Test of Homogeneity of Variances

toy sales in \$1000 per week

Levene Statistic	df1	df2	Sig.	
1.050	2	9	.389	

### **ANOVA**

toy sales in \$1000 per week

	Sum of squares	df	Mean Square	F	Sig.
Between Groups Within Groups Total	8.000 186.000 194.000	2 9 11	4.000 20.667	.194	.827

Having conducted this analysis, you could take this simple effect and perform simple comparisons. However, simple comparisons are usually performed only if the relevant simple effect proves to be significant. In this case, the effect is not significant.

If simple comparisons were required, then appropriate coefficients would have to be determined. That is, you would need to establish which means you were interested in comparing. For example, if you have obtained a significant simple effect for store type in suburban shopping centres, then you may be interested in determining whether discount stores had higher sales than department or variety stores. Remember, you are still only dealing with suburban shopping centres.

```
ONEWAY
sales BY type
/CONTRAST= -1 -1 2
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS.
```

Consistent with conducting trend analysis, it is also important to remember that the significance of the contrast F-ratio must be evaluated using the residual mean square from the omnibus test. That is, the residual mean square from the original two-way ANOVA summary table is used. For this example, the mean square residual for the two-way ANOVA is equal to 18.333. Again, the first degree of freedom (df) for all contrasts is 1, while the second df is the value for the residual estimate from the two-way ANOVA summary table (i.e. 18).

SPSS can also be used to analyse data from more complex factorial designs with three or more independent variables. With these designs, the interpretation of interaction effects becomes extremely complex.

# Practice example



A researcher is interested in determining the effect of density of traffic (light, medium or peak) and type of intersection (lights or roundabouts) on the number of accidents. Given the data in Prac9.sav, your tasks are to:

- 1 Check for violations of assumptions.
- 2 Determine whether traffic density influences the number of accidents.
- 3 Determine whether type of intersection influences the number of accidents.
- 4 Determine whether the influence of traffic density on the number of accidents depends on the type of intersection.

## **Solutions**



## Syntax

```
EXAMINE
 VARIABLES=accident BY density intersec
 /PLOT NONE
 /STATISTICS DESCRIPTIVES
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL
UNIANOVA
 accident BY density intersec
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /PRINT = ETASQ OPOWER HOMOGENEITY
 /CRITERIA = ALPHA(.05)
 /DESIGN = density intersec density*intersec .
ONEWAY
 accident BY density
 /STATISTICS DESCRIPTIVES HOMOGENEITY
 /MISSING ANALYSIS
 /POSTHOC = TUKEY ALPHA(.05).
```

## Output

## **Assumption testing**

Normality

## Descriptives

	density of traffic			Statistic	Std. Error
number of accidents	light: 7pm to 7am	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	7.0000 2.4868 11.5132 6.9444 6.5000 29.143 5.39841 .00 15.00 15.00 10.50 .203 -1.165	1.90863 .752 1.481
	medium: 10am to 4pm	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	13.6250 9.5553 17.6947 13.5278 13.5000 23.696 4.86790 8.00 21.00 13.00 9.25 .195 -1.456	1.72106 .752 1.481
	peak hour: 7am to 10am and 4pm to 7pm	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	14,7500 11.1541 18.3459 14,7222 16,0000 18,500 4,30116 9,00 21,00 12,00 7,50 -,117 -1,405	1.52069 .752 1.481

## **Descriptives**

	type of intersection		-	Statistic	Std. Error
number of accidents	lights	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	10.5833 6.2532 14.9135 10.5926 10.5000 46.447 6.81520 .00 21.00 21.00 11.75 151	1.96738 .637 1.232
	roundabouts	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	13,0000 10,0695 15,9305 12,9444 12,5000 21,273 4,61224 6,00 21,00 15,00 7,75 ,207 -1,006	.637 1.232

The distributions of accident scores for each type of intersection and each level of traffic density are normal, as evidenced by the reasonably low skewness statistics.

## Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: number of accidents

F	df1	df2	Sig.
.947	5	18	.475

Tests the null hypothesis that the error variance of the dependent variable is equal across groups. a Design: Intercept+density+intersec+density \* intersec

Levene's test is not significant and therefore homogeneity of variance can be assumed.

## Tests of Between-Subjects Effects

Dependent Variable: number of accidents

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model Intercept Density Intersect density * intersec Error Total Corrected Total	<sup>b</sup> 436.708 3337.042 <b>280.583</b> <i>35.042</i> 121.083 343.250 4117.000 779.958	5 1 2 1 2 18 24 23	87.342 3337.042 <b>140.292</b> 35.042 60.542 19.069	4.580 174.994 <b>7.357</b> 1.838 3.175	.007 .000 <b>.005</b> .192 .066	.560 .907 .450 .093 .261	22.901 174.994 14.714 1.838 6.350	.909 1.000 .893 .250 .534

a Computed using alpha = .05

b R squared = .560 (adjusted R squared = .438)

The output shows that the main effect for traffic density is significant, F(2,18) = 7.357, p < .05, but the type of intersection is not significant (p > .05). In addition, the interaction effect is not significant (p > .05). So the effect of density does not depend on the type of intersection.

To locate the source of difference caused by traffic density, a one-way ANOVA with post-hoc tests can be done.

#### Descriptives

### number of accidents

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
light: 7pm to 7am	8	7.0000	5.39841	1.90863	2.4868	11.5132	.00	15.00
medium: 10am to 4pm	8	13.6250	4.86790	1.72106	9.5553	17.6947	8.00	21.00
peak hour: 7am to 10am and 4pm to 7pm	8	14.7500	4.30116	1.52069	11.1541	18.3459	9.00	21.00
Total	24	11.7917	5.82334	1.18868	9.3327	14.2506		21.00

## Test of Homogeneity of Variances

### number of accidents

Levene Statistic	df1	df2	Sig.
.192	2	21	.827

## **Multiple Comparisons**

Dependent Variable: number of accidents

Tukey HSD

	(J) density of traffic	Mean	Std. Error	Sig.	95% Confide	ence Interval
(I) density of traffic		Difference (I- J)			Lower Bound	Upper Bound
light: 7pm to 7am	medium: 10am to 4pm	*-6.62500	2.43822	.033	-12.7707	4793
medium: 10am to 4pm	peak hour: 7am to 10am and 4pm to 7pm light: 7pm to 7am	*-7.75000 *6.62500	2.43822 2.43822	.012 .033	-13.8957 .4793	-1.6043 12.7707
peak hour: 7am to	peak hour: 7am to 10am and 4pm to 7pm light: 7pm to 7am	-1.12500 *7.75000 1.12500	2.43822 2.43822 2.43822	.890 .012 .890	-7.2707 1.6043 -5.0207	5.0207 13.8957 7.2707
10am and 4pm to 7pm	medium: 10am to 4pm					

<sup>\*</sup> The mean difference is significant at the .05 level.

These results suggest that there are significant differences between light traffic and medium and peak density, but not between medium and peak density. An examination of the means suggests that there are least accidents in light traffic, with no differences between medium and peak traffic.



# One-way repeated measures ANOVA

In chapter 7 you examined the analysis of an independent groups experimental design in which a different group of participants served in each of the conditions. That is, you used an independent groups ANOVA to detect differences in group means.

You are now interested in refining the experimental design to increase its sensitivity to detecting differences in the dependent variable. A major source of experimental error is individual differences, which can be controlled by using a repeated measures or within-subjects design. By having the same participants perform under every condition, you eliminate systematic bias attributable to participants in one group being different from the participants in other groups.

## Assumption testing



Three of the four assumptions underlying the repeated measures ANOVA are similar to those of the independent groups design:

- **Random selection** the sample should be independently and randomly selected from the population of interest.
- 2 Normality each population of scores should have a normal distribution.
- **Homogeneity of variance** the different populations of scores should have homogeneous variances. This is assessed by obtaining variances for each group and dividing the largest variance by the smallest variance to obtain an F-max value. If this ratio is greater than three, then the assumption has been violated and the resulting F-ratio must be evaluated at a more conservative alpha level.
- **Sphericity** the variance of the population difference scores for any two conditions should be the same as the variance of the population difference scores for any other two conditions.

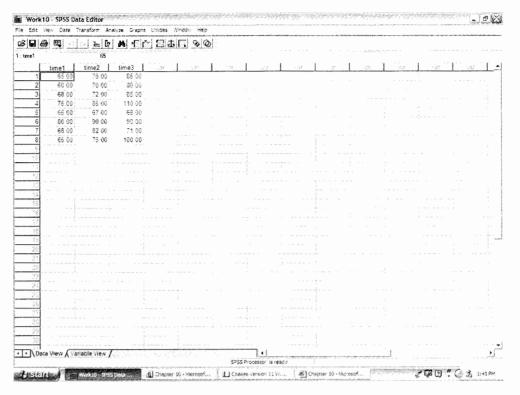
Assumption 1 is a matter of research design while assumption 2 can be tested by using statistics outlined in chapter 3. Assumptions 3 and 4 are assessed as part of the analysis.

# Working example



You wish to determine whether practice increases the ability to solve anagrams. Eight participants were asked to solve as many anagrams as possible in ten minutes. They were then allowed to practise for an hour before being asked to complete another tenminute timed task. Participants were then given another practice session and another timed task. The number of anagrams correctly solved was recorded.

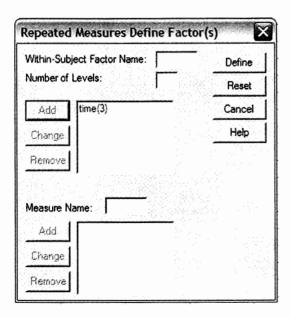
The data file can be found in Work 10.sav on the website that accompanies this title and is shown in the following figure.





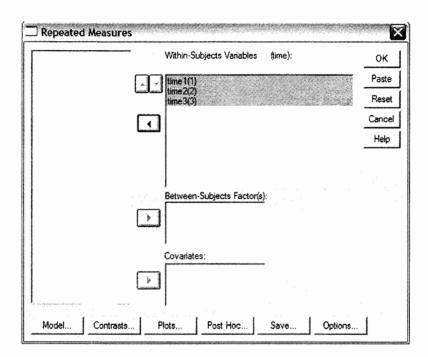
## To conduct a repeated measures ANOVA

- 1 Select the Analyze menu.
- 2 Click on General Linear Model and then Repeated Measures... to open the Repeated Measures Define Factor(s) dialogue box.
- 3 In the **Within-Subject Factor Name:** box, type the name for the within-subjects factor (i.e. *time*).
- 4 In the **Number of Levels:** box, type the number of levels of that particular factor (i.e. 3).
- 5 Click on **Add** to move this information into the box below.

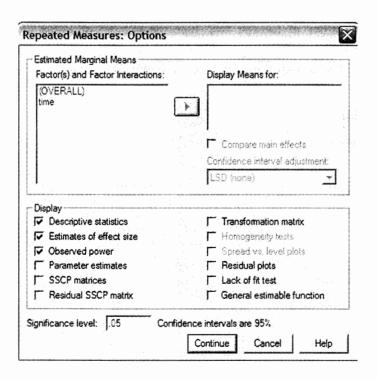


6 Click on the **Define** command pushbutton to open the **Repeated Measures** subdialogue box.

Select the three within-subjects variables (i.e. *time1*, *time2* and *time3*) and click on the ▶ button to move the variables into the Within-Subjects Variables (time): box.



- 8 Click on the **Options...** command pushbutton to open the **Repeated Measures: Options** sub-dialogue box.
- 9 In the **Display** box, click on the **Descriptive Statistics**, **Estimates of effect size** and **Observed power** check boxes.



10 Click on Continue and then OK.

GLM time1 time2 time3 /WSFACTOR = time 3 Polynomial /METHOD = SSTYPE(3) /PRINT = DESCRIPTIVE ETASQ OPOWER /CRITERIA = A'\_PHA(.05) /WSDESIGN = time .

Before interpretation of the output, the assumption of homogeneity of variance was assessed by calculating F-max. No violation of this assumption was found.

For a one-way within-subjects ANOVA, SPSS produces a lot of output: statistics for evaluating the sphericity assumption; traditional univariate ANOVA results; and ANOVA results using multivariate tests. The summary tables for between-subjects effects and within-subjects contrasts are not discussed here.

## **Descriptive Statistics**

	Mean	Std. Deviation	N	
correct anagrams at time 1	68.2500	6.36396	8	
correct anagrams at time 2	77.0000	7.92825	8	
correct anagrams at time 3	86.1250	14.01466	8	

#### Multivariate Tests<sup>a</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>b</sup>
time	Pillai's Trace	.876	°21.234	2.000	6.000	.002	.876	42.469	.994
	Wilks' Lambda	.124	°21.234	2.000	6.000	.002	.876	42.469	.994
	Hotelling's Trace	7.078	°21.234	2.000	6.000	.002	.876	42.469	.994
	Roy's Largest Root	7.078	°21.234	2.000	6.000	.002	.876	42.469	.994

- a Design: Intercept Within Subjects Design: time
- b Computed using alpha = .05
- c Exact statistic

## Mauchly's Test of Sphericity<sup>a</sup>

## Measure: MEASURE\_1

						Epsilon <sup>b</sup>	
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower- bound
Time	.254	8.234	2	.016	.573	.612	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

- a Design: Intercept Within Subjects Design: time
- b May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Before you can interpret the F-ratio of the within-subjects effect (time), you must first ensure that the assumption of sphericity has not been violated. The value for Mauchly is equal to .254 and is significant (p < .05). You must therefore evaluate the obtained F-ratio using new degrees of freedom (df) which are calculated using the Huynh-Feldt Epsilon (.612). The new dfs are 1.22 and 8.56 ( $2 \times .61$ ,  $14 \times .61$ ), compared with 2 and 14. An examination of the critical F-tables (available in any good statistical text) indicates that the F-ratio is still significant when using the new dfs.

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Error (time)	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound	1278.250 1278.250 1278.250 1278.250 759.750 759.750 759.750 759.750	2 1.145 1.223 1.000 14 8.016 8.562 7.000	639.125 1116.226 1045.066 1278.250 54.268 94.778 88.736 108.536	11.777 11.777 11.777 11.777	.001 .008 .006 .011	.627 .627 .627 .627	23.554 13.487 14.405 11.777	.979 .876 .894 .836

a Computed using alpha = .05

You will notice that both the multivariate and univariate summary tables show that the result is significant. There is a large effect size but the study may be overpowered. However, you have demonstrated that practice brings about changes in the ability to solve anagrams, F(1.22, 8.56) = 11.77, p < .05. The exact location of these differences requires that contrasts be performed. Post-hoc analysis of this nature is beyond the scope of this chapter.

# Practice example



A real estate agent wished to determine whether the number of house sales in ten different suburbs changed significantly over the four quarters of the year. She asked her representatives in these ten suburbs to record the number of houses sold in each quarter. Given the data in Prac10.sav, your tasks are to:

- 1 Check your data for violations of assumptions.
- 2 Determine whether differences in the number of house sales exist across the four quarters.

## **Solutions**



## **Syntax**

**DESCRIPTIVES** 

VARIABLES=quart1 quart2 quart3 quart4

/STATISTICS=MEAN STDDEV VARIANCE MIN MAX KURTOSIS SKEWNESS.

GLM

quart1 quart2 quart3 quart4

/WSFACTOR = quarter 4 Polynomial

/METHOD = SSTYPE(3)

/PRINT = ETASQ OPOWER

/CRITERIA = ALPHA(.05)

/WSDESIGN = quarter .

## Output

## **Assumption testing**

Normality

Using descriptive statistics it can be seen that all variables are normally distributed.

### **Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std.	Variance	Skev	ness	Kur	tosis
	Statistic	Std. Error	Statistic	Std. Error						
quart1	10	3.00	13.00	8.6000	3.23866	10.489	556	.687	405	1.334
quart2	10	2.00	8.00	5.2000	1.87380	3.511	2233	.687	564	1.334
quart3	10	1.00	5.00	2.8000	1.31656	1.733	.088	.687	751	1.334
quart4	10	.00	3.00	1.3000	.94868	.900	.234	.687	347	1.334
Valid N (listwise)	10									

## Homogeneity of variance

Examination of the variances for each of the variables shows that F-max is greater than three (10.489/.900), and therefore results must be interpreted with caution because the assumption of homogeneity of variance has been violated.

The value of Mauchly is equal to .275 and is not significant, so the F-ratio can be interpreted.

## Mauchly's Test of Sphericity<sup>a</sup>

Measure: MEASURE\_1

						Epsilon <sup>b</sup>	
Within Subjects Effect	Mauchły's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower- bound
Quarter	.275	9.958	5	.079	.559	.672	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

a Design: Intercept Within Subjects Design: quarter

An examination of the significance of the F-value for the variable quarter indicates that the main effect is significant.

## **Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Quarter	Sphericity Assumed Greenhouse-Geisser	304.275 304.275	3 1.676	101.425 181.572	27.668 27.668	.000	.755 .755	83.005 46.366	1.000 1.000
	Huynh-Feldt Lower-bound	304.275 304.275	2.015 1.000	151.010 304.275	27.668 27.668	.000 .001	.755 .755	55.750 27.668	1.000 .996
Error (quarter)	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt	98.975 98.975 98.975	27 15.082 18.134	3.666 6.562 5.458					
	Lower-bound	98.975	9.000	10.997					

a Computed using alpha = .05

Remembering that the homogeneity of variance assumption was violated, this could be adjusted for by setting a more conservative alpha value. Because the F-ratio has such a low probability, even with this new alpha level it is still significant. Therefore, it can confidently be said that there are significant differences in the number of house sales across the four quarters, F(3,27) = 27.668, p < .001.

b May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.



# Two-way repeated measures ANOVA

# Assumption testing



In the two-way repeated measures design you have two independent variables, with two or more levels, which are within-subject in nature. That is, each participant performs in all conditions.

The assumptions underlying the two-way repeated measures ANOVA are identical to those outlined in the one-way repeated measures ANOVA.

# Working example



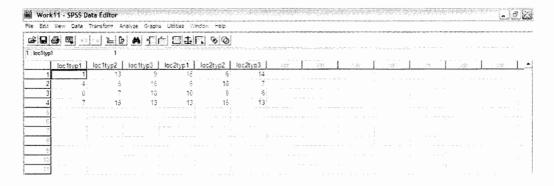
To illustrate the two-way repeated measures ANOVA, you can use an example similar to the one you used for the two-way independent groups design. Here, manager productivity is considered, which is reflected in sales across various store types and locations. Each manager oversees six stores: one of each type in each location. Remember, you wish to determine the effect of store type (variety, department and discount) and location (city centre or suburbs) on productivity, measured as sales of toys. The first independent variable is type of store with three levels, the second independent variable is location with two levels, and the dependent variable is the amount of toy sales in \$1000 per week. Therefore, you have a  $3 \times 2$  factorial design with six data cells ( $3 \times 2 = 6$ ). Given that four managers participated in each of the conditions, you have a total of 24 observations.

You wish to ask three questions:

- 1 Does the store type influence the amount of toy sales?
- 2 Does the store location influence the amount of toy sales?
- 3 Does the influence of store type on the amount of toy sales depend on the location of the store?

Questions 1 and 2 refer to main effects, while question 3 examines the effect of the interaction of the two independent variables on the dependent variable.

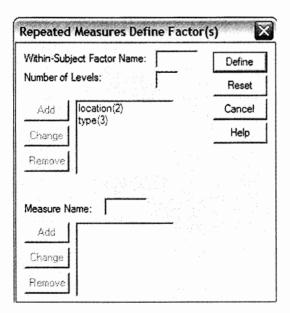
The data file can be found in Work 11.sav on the website that accompanies this title and is shown in the following figure.



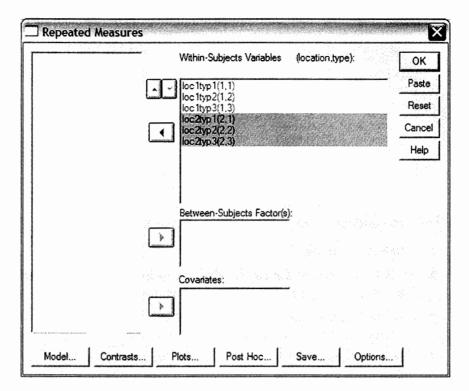


## To conduct a two-way repeated measures ANOVA

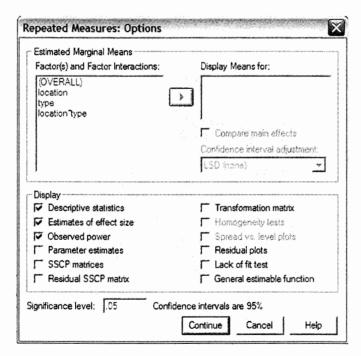
- 1 Select the **Analyze** menu.
- 2 Click on General Linear Model and then on Repeated Measures... to open the Repeated Measures Define Factor(s) dialogue box.
- 3 In the **Within-Subject Factor Name:** box, type the name for the first within-subjects factor (i.e. *location*).
- 4 In the **Number of Levels:** box, type the number of levels of that particular factor (i.e. 2).
- 5 Click on **Add** to move this information into the box below.
- 6 In the **Within-Subject Factor Name:** box, type the name for the second within-subjects factor (i.e. *type*).
- 7 In the **Number of Levels:** box, type the number of levels of that particular factor (i.e. 3).
- 8 Click on **Add** to move this information into the box below.



- 9 Click on the Define command pushbutton to open the Repeated Measures subdialogue box.
- Select the six within-subjects variables (i.e. loc1typ1, loc1typ2, loc1typ3, loc2typ1, loc2typ2, loc2typ3) and click on the ▶ button to move the variables into their appropriate places within the Within-Subjects Variables (location, type): box (i.e. loc1typ1 -> {1,1}, loc1typ2 -> {1,2}, etc.).



- Click on the **Options...** command pushbutton to open the **Repeated Measures: Options** sub-dialogue box.
- In the **Display** box, click on the **Descriptives**, **Estimates of effect size** and **Observed power** check boxes.



13 Click on Continue and then OK.

GLM
loc1typ1 loc1typ2 loc1typ3 loc2typ1 loc2typ2 loc2typ3
/WSFACTOR = location 2 Polynomial type 3 Polynomial
/METHOD = SSTYPE(3)
/PRINT = DESCRIPTIVE ETASQ OPOWER
/CRITERIA = ALPHA(.05)
/WSDESIGN = location type location\*type .

## **Descriptive Statistics**

	Mean	St. Deviation	N
city, variety store	3.00	3.162	4
city department store	10.00	4.761	4
city discount store	14.00	3.916	4
suburb variety store	11.00	3.916	4
suburb department store	12.00	5.477	4
suburb discount store	10.00	4.082	4

## Mauchly's Test of Sphericity<sup>a</sup>

Measure: MEASURE\_1

						Epsilon <sup>b</sup>	
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower-bound
location	1.000	.000	0		1.000	1.000	1.000
type	.918	.171	2	.918	.924	1.000	.500
location * type	.502	1.378	2	.502	.668	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

- a Design: Intercept Within Subjects Design: location+type+location\*type
- b May be used to adjust the degrees of freedom (df) for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
location	Sphericity Assumed	24.000	1	24.000	6.000	.092	.667	6.000	.395
	Greenhouse-Geisser	24.000	1.000	24.000	6.000	.092	.667	6.000	.395
	Huynh-Feldt	24.000	1.000	24.000	6.000	.092	.667	6.000	.395
	Lower-bound	24.000	1.000	24.000	6.000	.092	.667	6.000	.395
Error (location)	Sphericity Assumed	12.000	3	4.000					
	Greenhouse-Geisser	12.000	3.000	4.000					
	Huynh-Feldt	12.000	3.000	4.000					
	Lower-bound	12.000	3.000	4.000					
type	Sphericity Assumed	112.000	2	56.000	10.723	.010	.781	21.447	.891
**	Greenhouse-Geisser	112.000	1.849	60.589	10.723	.013	.781	19.823	.865
	Huynh-Feldt	112.000	2.000	56.000	10.723	.010	.781	21.447	.891
	Lower-bound	112.000	1.000	112.000	10.723	.047	.781	10.723	.600
Error(type)	Sphericity Assumed	31.333	6	5.222					
,	Greenhouse-Geisser	31.333	5.546	5.650					
	Huynh-Feldt	31.333	6.000	5.222					
	Lower-bound	31,333	3.000	10.444					
location * type	Sphericity Assumed	144.000	2	72.000	1.929	.226	.391	3.857	.260
**	Greenhouse-Geisser	144.000	1.335	107.857	1.929	.248	.391	2.575	.200
	Huynh-Feldt	144.000	2.000	72.000	1.929	.226	.391	3.857	.260
	Lower-bound	144.000	1.000	144.000	1.929	.259	.391	1.929	.168
Error (location*type)	Sphericity Assumed	224.000	6	37.333					
, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Greenhouse-Geisser	224.000	4.005	55.926					
	Huynh-Feldt	224,000	6.000	37.333					
	Lower-bound	224.000	3.000	74.667					

a Computed using alpha = .05

The main effect for location is not significant and thus you can conclude that the location of the store does not influence the amount of toy sales.

Because the Mauchly test of sphericity for store type is not significant, you have not violated this assumption. The main effect for type of store is significant (p < .05) and so you can conclude that toy sales are influenced by the type of store in which they are sold, F(2,6) = 10.723, p < .05. There is a large effect size and the power of the study is good.

Because the Mauchly test of sphericity for the interaction is not significant, you have not violated this assumption. The location by type of store interaction effect is not significant and therefore you can conclude that the differences in sales across type of store do not depend on the location of the store.

If post-hoc analysis were to follow, you would conduct a comparison of marginal means for the significant main effect, which can be calculated from the cell means in the output.

# Practice example



A graphic designer wished to determine which combination of colours and backgrounds produce the most aesthetically pleasing display. Five participants were exposed to two different types of background (hatched and spotted) and lettering of four different colours (red, blue, green and yellow). Participants were requested to rate the pleasing nature of these displays on a 20-point scale (1 = least pleasing to 20 = most pleasing). Given the data in Prac11.sav, your tasks are to:

- 1 Check the data for violations of assumptions.
- 2 Determine whether background influences the participants' ratings.
- 3 Determine whether colour of lettering influences the participants' ratings.
- 4 Determine whether the influence of background on rating depends on letter colouring.

## **Solutions**



## **Syntax**

DESCRIPTIVES

VARIABLES=b1c1 b1c2 b1c3 b1c4 b2c1 b2c2 b2c3 b2c4 /STATISTICS=MEAN STDDEV MIN MAX KURTOSIS SKEWNESS.

GLM

b1c1 b1c2 b1c3 b1c4 b2c1 b2c2 b2c3 b2c4

/WSFACTOR = bground 2 Polynomial colour 4 Polynomial

/METHOD = SSTYPE(3)

/PRINT = ETASQ OPOWER

/CRITERIA = ALPHA(.05)

/WSDESIGN = bground colour bground\*colour .

## Output

## **Assumption testing**

Normality

### **Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std.	Variance	Skev	ness	Kur	tosis
	Statistic	Std. Error	Statistic	Std. Error						
hatch, red	5	1.00	9.00	4.8000	3.19374	10.200	.301	.913	-1.344	2.000
hatch, blue	5	3.00	13.00	8.0000	4.12311	17.000	.000	.913	-1.893	2.000
hatch, green	5	5.00	14.00	10.2000	3.42053	11.700	845	.913	.699	2.000
hatch, yellow	5	4.00	12.00	8.0000	3.39116	11.500	.192	.913	-2.234	2.000
spotted, red	5	3.00	10.00	6.8000	3.11448	9.700	.060	.913	-2.299	2.000
spotted, blue	5	5.00	15.00	9.4000	4.15933	17.300	.397	.913	-1.578	2.000
spotted, green	5	9.00	18.00	13.2000	3.49285	12.200	.310	.913	644	2.000
spotted, yellow	5	7.00	15.00	10.4000	3.84708	14.800	.458	.913	-2.939	2.000
Valid N (listwise)	5									

From the descriptive statistics it is necessary to examine the skewness and kurtosis statistics. Although the kurtosis is marked for several variables, the skewness is minimal and thus we can say that we have not violated the normality assumption.

Examination of the variances shows that our F-max is not greater than three (17.3/9.7 = 1.78), so homogeneity of variance is assumed.

### Mauchly's Test of Sphericity<sup>a</sup>

Measure: MEASURE\_1

·					Epsilon <sup>b</sup>			
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower-bound	
bground	1.000	.000	0		1.000	1.000	1.000	
colour	.510	1.833	5	.879	.729	1.000	.333	
bground * colour	.103	6.192	5	.315	.470	.650	.333	

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

- a Design: Intercept Within Subjects Design: bground+colour+bground\*colour
- b May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

### Tests of Within-Subjects Effects

Measure: MEASURE 1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
bground	Sphericity Assumed	48.400	1	48.400	30.488	.005	.884	30.488	.980
	Greenhouse-Geisser	48.400	1.000	48.400	30.488	.005	.884	30.488	.980
	Huynh-Feldt	48.400	1.000	48.400	30.488	.005	.884	30.488	.980
	Lower-bound	48.400	1.000	48.400	30.488	.005	.884	30.488	.980
Error(bground)	Sphericity Assumed	6.350	4	1.588					
	Greenhouse-Geisser	6.350	4.000	1.588					
	Huynh-Feldt	6.350	4.000	1.588					
	Lower-bound	6.350	4.000	1.588					
colour	Sphericity Assumed	175.700	3	58.567	45.196	.000	.919	135.588	1.000
	Greenhouse-Geisser	175.700	2.187	80.330	45.196	.000	.919	98.854	1.000
	Huynh-Feldt	175.700	3.000	58.567	45.196	.000	.919	135.588	1.000
	Lower-bound	175.700	1.000	175.700	45.196	.003	.919	45.196	.998
Error(colour)	Sphericity Assumed	15.550	12	1.296					
	Greenhouse-Geisser	15.550	8.749	1.777					ļ
	Huynh-Feldt	15.550	12.000	1.296					
	Lower-bound	15.550	4.000	3.888					
bground * colour	Sphericity Assumed	3.400	3	1.133	2.325	.127	.368	6.974	.446
	Greenhouse-Geisser	3.400	1.410	2.411	2.325	.184	.368	3.278	.271
	Huynh-Feldt	3.400	1.950	1.744	2.325	.162	.368	4.532	.335
	Lower-bound	3.400	1.000	3.400	2.325	.202	.368	2.325	.219
Error(bground*colour)	Sphericity Assumed	5.850	12	.488	1		}		
	Greenhouse-Geisser	5.850	5.640	1.037					
	Huynh-Feldt	5.850	7.798	.750	1				
	Lower-bound	5.850	4.000	1.463					

a Computed using alpha = .05

The main effect for background is significant (p < .05) and so it can be concluded that type of background does affect the participants' ratings, F(1,4) = 30.488, p < .05.

Because the Mauchly test of sphericity for colour is not significant, this assumption has not been violated.

The main effect for colour is significant (p < .05) and therefore it can be said that the colour of lettering does affect the participants' ratings, F(3,12) = 45.196, p < .05.

Because the Mauchly test of sphericity for the interaction effect (background by colour) is not significant, this assumption has not been violated.

The background by colour interaction effect is not significant and so it can be concluded that although main effects for both background and colour independently were significant, the effect of one independent variable does not depend on the effect of the other in influencing the participants' ratings of pleasing nature. In other words, background does not depend on colour in influencing participants' ratings, F(3, 12) = 2.325, p > .05.

# Trend analysis

Sometimes, having determined that a significant effect exists, you may wish to determine the nature of this relationship by examining trends in the data. Post-hoc and planned comparisons could be performed between the treatment means, as discussed in chapters 7 and 8. This chapter discusses how you can test for different trends in the data — for example, a trend may be linear, quadratic or cubic. However, trend analysis is only appropriate when:

- 1 The levels of the independent variable can be considered to form an increasing or decreasing continuum.
- 2 The intervals between adjacent levels of the independent variable are equal or proportional.

When you conduct planned comparisons you use coefficients to pick out and represent a particular combination of means you wish to compare. When you conduct trend analysis you also use coefficients, but the coefficients you select, when plotted, define the particular shape of the curve that you are testing. These coefficients can be found in the tables of any good statistical text, but when conducting trend analysis within SPSS for Windows, linear and quadratic trend options already exist. In essence, trend analysis is a special case of planned comparisons.

# Assumption testing



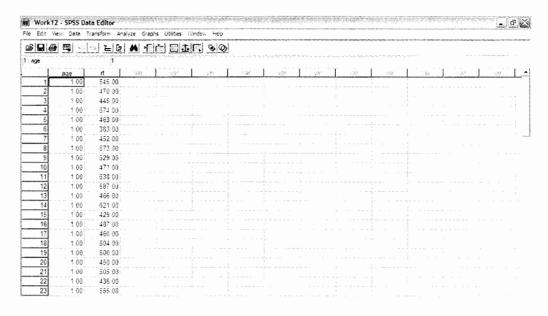
As outlined in chapter 11, trend analysis follows ANOVA and therefore the assumptions of ANOVA must be met.

# Working example



A cognitive psychologist was interested in whether reaction time (RT) declined as a function of age. She obtained random samples of 25 people from each of four age groups and measured the time it took each participant to press a button in response to a visual cue. The age groups she chose were 15–20 years, 35–40 years, 55–60 years and 75–80 years. The researcher anticipated that reaction time would increase consistently with age.

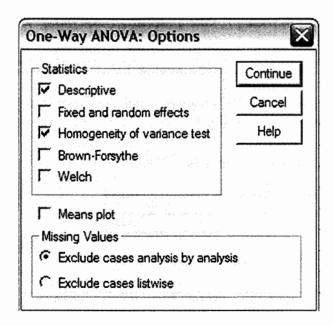
The data file can be found in Work12.sav on the website that accompanies this title and is shown in the following figure.





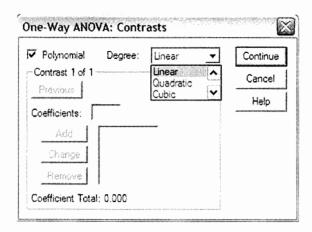
## To conduct a linear trend analysis

- 1 Select the **Analyze** menu.
- 2 Click on Compare Means and One-Way ANOVA... to open the One-Way ANOVA dialogue box.
- 3 Select the dependent variable (i.e. *rt*) and click on the ▶ button to move the variable into the **Dependent List:** box.
- 4 Select the independent variable (i.e. *age*) and click on the ▶ button to move the variable into the **Factor:** box.
- 5 Click on the **Options...** command pushbutton to open the **One-Way ANOVA: Options** sub-dialogue box.
- Select the check boxes for **Descriptive** and **Homogeneity-of-variance**.



- 7 Click on Continue.
- 8 Click on the **Contrasts...** command pushbutton to open the **One-Way ANOVA: Contrasts** sub-dialogue box.

- 9 Select the **Polynomial** check box.
- 10 In the **Degree:** drop-down list, ensure that the **Linear** trend is selected.



## 11 Click on Continue and then OK.

ONEWAY

rt BY age

/POLYNOMIAL= 1

/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS.

## Descriptives

rt

					95% Confidence Interval for Mean			
	Ν	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
15-20	25	497.2800	56.62782	11.32556	473.9052	520.6548	383.00	621.00
35-40	25	515.8400	56.52925	11.30585	492.5059	539.1741	397.00	647.00
55-60	25	531.2000	63.97591	12.79518	504.7920	557.6080	397.00	677.00
75-80	25	555.1200	72.57601	14.51520	525.1621	585.0779	429.00	739.00
Total	100	524.8600	65.38511	6.53851	511.8862	537.8338	383.00	739.00

The Levene test for homogeneity of variances is not significant (p > .05), indicating that the assumption has not been violated.

## Test of Homogeneity of Variances

rt

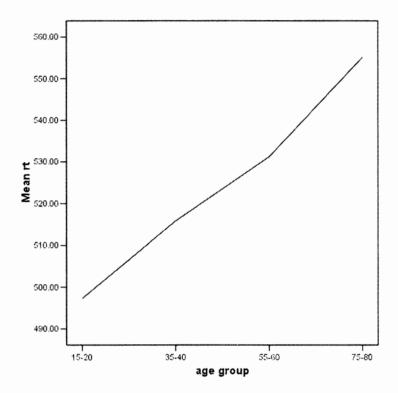
Levene Statistic	df1	df2	Sig.
.817	3	96	.488

## **ANOVA**

rt

			Sum of squares	df	Mean Square	F	Sig.
Between Groups	(Combined) Linear Term	Contrast Deviation	44947.000 44594.568 352.432	3 1 2	14982.333 44594.568 176.216	<b>3.802</b> 11.317 .045	. <b>013</b> .001 .956
Within Groups Total			378299.040 423246.040	96 99	3940.615		

The ANOVA summary table shows a significant between-groups F-ratio (p < .05) and therefore you can conclude that reaction time differs significantly across age groups, F(3,96) = 3.802, p < .05. Furthermore, by examining the linear term, which is also significant (p < .05), you can also conclude that reaction time increases consistently across age groups. Plotting these means illustrates a linear trend, as evident in the following figure. Remember, you can plot means using the **Graphs** menu and **Line Charts** sub-menu.



Trend analyses can also be conducted in the case of a one-way repeated measures design or two-way independent and repeated measures designs. First, the relevant simple effect must be extracted using the **Select Cases...** option in the **Data** menu or toolbar, then analysis for trend components is conducted using the **Contrasts...** procedure available in the **One-Way ANOVA** sub-menu. In the case of two-way designs, it is important to remember that the error term for the different trend F-ratios will be the within-groups or residual mean square from the overall ANOVA. That is, the mean square from the two-way ANOVA summary table is used.

# Practice example



Researchers wished to examine the effect of caffeine on the performance of a simple motor task. Sixty participants performed under one of three caffeine consumption conditions (0.00, 150 and 300 millilitres) and were required to perform a sequential finger-tapping task. The number of errors made was recorded. Given the data in Prac12.sav, your tasks are to:

- 1 Check for violations of assumptions.
- 2 Determine whether caffeine consumption significantly influences ability to perform the task accurately (that is, the number of errors made).
- 3 Determine whether this trend is linear.

# **Solutions**



## **Syntax**

EXAMINE
VARIABLES=errors BY caffeine
/PLOT NONE
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
ONEWAY
errors BY caffeine
/POLYNOMIAL= 1
/STATISTICS HOMOGENEITY
/MISSING ANALYSIS .
GRAPH
/LINE(SIMPLE)=MEAN(errors) BY caffeine
/MISSING=REPORT.

## Output

## **Assumption testing**

Normality

## **Descriptives**

	caffeine consumption	Descriptives		Statistic	Std. Error
number of errors	0.00 caffeine	Mean		1.25	.315
		95% Confidence	Lower Bound	.59	
		Interval for Mean	Upper Bound	1.91	
		5% Trimmed Mean		1.11	
		Median		1.00	
		Variance		1.987	
		Std. Deviation		1.410	
		Minimum		0	
ĺ		Maximum		5	
		Range		5	
		Interquartile Range		2	
		Skewness		1.386	.512
		Kurtosis		1.628	.992
	150 ml caffeine	Mean		5.15	.499
		95% Confidence	Lower Bound	4.11	
		Interval for Mean	Upper Bound	6.19	
		5% Trimmed Mean		5.06	
		Median		4.50	
		Variance		4.976	
		Std. Deviation		2.231	
		Minimum		2	
		Maximum		10	
		Range		8	
		Interquartile Range		_ 4	
		Skewness		.771	.512
		Kurtosis		300	.992
	300 ml caffeine	Mean		10.40	.825
		95% Confidence	Lower Bound	8.67	
		Interval for Mean	Upper Bound	12.13	
		5% Trimmed Mean		10.50	
		Median		11.00	
		Variance		13.621	
		Std. Deviation		3.691	
		Minimum		3	
		Maximum		16	
		Range		13	
		Interquartile Range		5	
		Skewness		640	.512
		Kurtosis		206	.992

Skewness and kurtosis statistics indicate that the number of errors is relatively normally distributed across caffeine groups.

Test of Homogeneity of Variances

number of errors

Levene Statistic	df1	df2	Sig.
6.690	2	57	.002

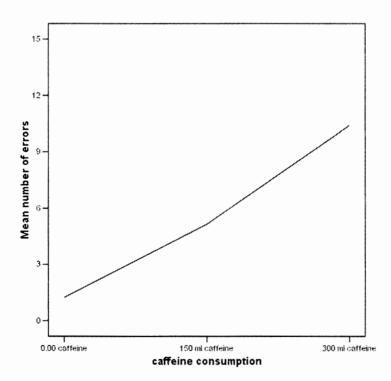
The Levene test for homogeneity of variance is significant (p < .05) and so the linear trend must be interpreted with caution. It may be preferable to assess the F-value at a more conservative alpha level of .001.

### **ANOVA**

number of errors

			Sum of squares	df	Mean Square	F	Sig.
Between Groups	(Combined) Linear Term	Contrast Deviation	843.300 837.225 6.075	2 1 1	421.650 837.225 6.075	<b>61.452</b> 122.019 .885	.000 .000 .351
Within Groups Total			391.100 1234.400	57 59	6.861		

The linear trend is still significant at this more conservative alpha level and so it can be said that there is a significant linear trend in the data. That is, as caffeine consumption increases so does the number of errors made in the motor task, F(2,57) = 61.452, p < .001. This trend is depicted in the following figure.





# Mixed/split plot design (SPANOVA)

In the two-way mixed designs sometimes called a split plot ANOVA or SPANOVA, there are repeated measures on one independent variable and independent groups on the other independent variable.

# Assumption testing



The assumptions underlying the SPANOVA are the same as those for the independent groups and repeated measures ANOVAs. A further assumption specifically applies to this analysis:

Homogeneity of intercorrelations — the pattern of intercorrelations among the various levels of the repeated measures factor(s) should be consistent from level to level of the between-subjects factor(s). This assumption is tested using Box's M statistic. This statistic is very sensitive so it is recommended that an alpha level of .001 be used in its interpretation. Homogeneity is present when the statistic is not significant (p > .001).

# Working example

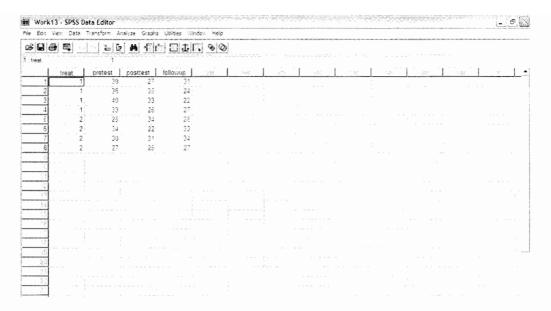


A therapist wishes to determine the efficacy of a new treatment program for depression. She randomly assigns eight participants to either the treatment group or the control group. She asks each participant to complete a depression inventory before the treatment starts, immediately the treatment finishes and three months after completion. The first independent variable, time of test, is within-subjects in nature and has three levels. The second independent variable, treatment status, is a between-subjects factor with two levels. The dependent variable is the score on the depression inventory. Therefore, you have a  $3 \times 2$  factorial design with six data cells ( $3 \times 2 = 6$ ) and 48 observations.

The therapist wishes to ask three questions:

- 1 Do depression scores change with time?
- 2 Is the treatment more effective than the control?
- 3 Are there differential changes in depression scores for the treatment and control groups?

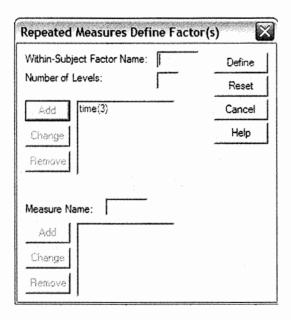
Questions 1 and 2 refer to main effects, while question 3 examines the effects of the interaction of the two independent variables on the dependent variable. The data file can be found in Work13.sav on the website that accompanies this title and is shown in the following figure.



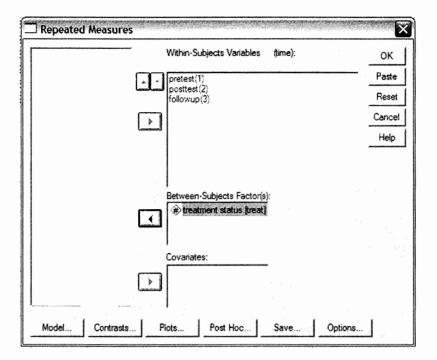


## To conduct a SPANOVA

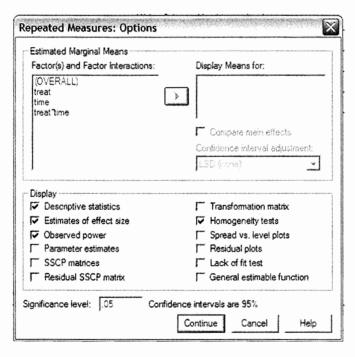
- 1 Select the **Analyze** menu.
- 2 Click on General Linear Model and then on Repeated Measures... to open the Repeated Measures Define Factor(s) dialogue box.
- In the Within-Subject Factor Name: box, type a name for the within-subjects factor (i.e. *time*), then tab. In the Number of Levels: box, type the number of levels of this factor (i.e. 3).
- 4 Click on **Add** to move this information into the box below.



- 5 Click on the **Define** command pushbutton to open the **Repeated Measures** subdialogue box.
- Select the variables representing the within-subjects factor (i.e. *pretest*, *posttest* and *followup*) and click on the button to move these variables into the Within-Subjects Variables: (time) box.
- 7 Select the between-subjects variable (i.e. *treat*) and click on the ▶ button to move this variable into the **Between-Subjects Factor(s):** box.



- 8 Click on the **Options...** command pushbutton to open the **Repeated Measures: Options** sub-dialogue box.
- 9 In the Display box, click on the Descriptives, Estimates of effect size, Observed power and Homogeneity tests check boxes.



10 Click on Continue and then OK.

```
GLM
pretest posttest followup BY treat
/WSFACTOR = time 3 Polynomial
/METHOD = SSTYPE(3)
/PRINT = DESCRIPTIVE ETASQ OPOWER HOMOGENEITY
/CRITERIA = ALPHA(.05)
/WSDESIGN = time
/DESIGN = treat .
```

## **Descriptive Statistics**

	treatment status	Mean	Std. Deviation	N
pre treatment depression score	treatment group	37.00	3.162	4
	control group	29.00	3.916	4
	Total	33.00	5.398	8
post treatment depression score	treatment group	30.00	4.761	4
	control group	28.00	5.477	4
	Total	29.00	4.870	8
follow up depression score	treatment group	26.00	3.916	4
	control group	30.00	4.082	4
	Total	28.00	4.276	8

## Box's Test of Equality of Covariance Matrices<sup>a</sup>

Box's M	9.081
F	.659
df1	6
df2	260.830
Sig.	.683

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+treat
 Within-Subjects Design: time

Because Box's M statistic is not significant (p > .001), you have not violated the assumption of homogeneity of variance-covariance matrices.

The Mauchly test of sphericity for time below is also not significant (p > .05) and so you have not violated the sphericity assumption.

## Mauchly's Test of Sphericity<sup>b</sup>

## Measure: MEASURE\_1

					Epsilon <sup>a</sup>		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower- bound
time	.577	2.749	2	.253	.703	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

- a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
- b. Design: Intercept+treatWithin Subjects Design: time

## Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
time	Sphericity Assumed	112.000	2	56.000	2.632	.113	.305	5.264	.424
	Greenhouse-Geisser	112.000	1.406	79.686	2.632	.137	.305	3.699	.340
	Huynh-Feldt	112.000	2.000	56.000	2.632	.113	.305	5.264	.424
	Lower-bound	112.000	1.000	112.000	2.632	.156	.305	2.632	.278
time * treat	Sphericity Assumed	144.000	2	72.000	3.384	.068	.361	6.768	.525
	Greenhouse-Geisser	144.000	1.406	102.453	3.384	.093	.361	4.756	.421
	Huynh-Feldt	144.000	2.000	72.000	3.384	.068	.361	6.768	.525
	Lower-bound	144.000	1.000	144.000	3.384	.115	.361	3.384	.341
Error(time)	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound	255.333 255.333 255.333 255.333	12 8.433 12.00 0 6.000	21.278 30.277 21.278 42.556					

a. Computed using alpha = .05

### Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
pre treatment depression score	.200	1	6	.670
post treatment depression score	.231	1	6	.648
follow up depression score	.273	1	6	.620

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Levene's test for homogeneity of variance is not significant, so this assumption is not violated.

The main effect for time is not significant (p > .05) and therefore you can conclude that depression scores did not change significantly from the pretest to the followup.

The main effect for treatment is also not significant (p > .05) and so those participants in the treatment group did not fare any better than those in the control group. Consequently, the therapist must be very disappointed!

Tests of Between-Subjects Effects

Measure: MEASURE\_1 Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Intercept	21600.000	1	21600.000	1735.714	.000	.997	1735.714	1.000
treat	24.000	1	24.000	1.929	.214	.243	1.929	.217
Error	74.667	6	12.444					

a. Computed using alpha = .05

Lastly, the interaction effect is not significant (p > .05). Had the main effect for time or the interaction been significant, you would have needed to locate the source of these differences using post-hoc analysis.

a. design: Intercept+treat

Within Subjects Design: time

# Practice example



In designing a new concert hall an architect wished to determine whether the size of the auditorium and the choice of sound-proofing material influenced the quality of the acoustics. Two sizes of auditorium were investigated (small and large), with four different types of sound-proofing material. Ten different conductors were randomly assigned to the different-sized auditoriums. These conductors listened to four orchestral performances under different sound-proofing conditions. Conductors were asked to rate the quality of sound using a 20-point rating scale (1 = poor sound to 20 = excellent sound). Given the data in Prac13.sav, your tasks are to:

- 1 Check the data for violation of assumptions.
- 2 Determine whether auditorium size influences sound quality.
- 3 Determine whether sound-proofing material influences sound quality.
- 4 Determine whether the influence of auditorium size on sound quality depends on sound-proofing material.

## **Solutions**



## **Syntax**

**EXAMINE** VARIABLES=sp1 sp2 sp3 sp4 BY audit /PLOT BOXPLOT /COMPARE GROUP /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL. GLM sp1 sp2 sp3 sp4 BY audit /WSFACTOR = sound 4 Polynomial /METHOD = SSTYPE(3) /PRINT = ETASQ OPOWER HOMOGENEITY /CRITERIA = ALPHA(.05) /WSDESIGN = sound /DESIGN = audit

## Output

## Assumption testing

## **Descriptives**

	size of auditorium			Statistic	Std. Error
sp1	small	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation	Lower Bound Upper Bound	4.8000 .8344 8.7656 4.7778 4.0000 10.200 3.19374	1.42829
		Minimum Maximum Range Interquartile Range Skewness Kurtosis		1.00 9.00 8.00 6.00 .301 -1.344	.913 2.000

(continued)

## Descriptives (continued)

	size of auditorium			Statistic	Std. Error
	size of auditorium	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness	Lower Bound Upper Bound	6.8000 2.9329 10.6671 6.8333 6.0000 9.700 3.11448 3.00 10.00 7.00 6.00	.913
sp2	small	Kurtosis  Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	-2.299  8.0000 2.8805 13.1195 8.0000 8.0000 17.000 4.12311 3.00 13.00 10.00 8.00 .000 -1.893	2.000 1.84391 .913 2.000
	large	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	9.4000 4.2355 14.5645 9.3333 9.0000 17.300 4.15933 5.00 15.00 10.00 8.00 .397 -1.578	.913 2.000
sp3	small	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	10.2000 5.9529 14.4471 10.2778 11.0000 11.700 3.42053 5.00 14.00 9.00 6.00 845	.913 2.000
	large	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	13.2000 8.8631 17.5369 13.1667 13.0000 12.200 3.49285 9.00 18.00 9.00 6.50 .310 644	.913 2.000

(continued)

## **Descriptives** (continued)

	size of auditorium			Statistic	Std. Error
sp4	small	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	8.0000 3.7893 12.2107 8.0000 7.0000 11.500 3.39116 4.00 12.00 8.00 6.50 .192 -2.234	.913 2.000
	large	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	10.4000 5.6232 15.1768 10.3333 9.0000 14.800 3.84708 7.00 15.00 8.00 7.50 .458 -2.939	1.72047 .913 2.000

## Normality

Skewness and kurtosis statistics indicate that data are relatively normally distributed across groups.

## Box's Test of Equality of Covariance Matrices<sup>a</sup>

9.339
.401
10
305.976
.946

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+auditWithin-Subjects Design: sound

## Mauchly's Test of Sphericity<sup>b</sup>

## Measure: MEASURE\_1

					Epsilon <sup>a</sup>		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh- Feldt	Lower- bound
sound	.498	4.680	5	.460	.683	1.000	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

- a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
- b. Design: Intercept+audit
  Within Subjects Design: sound

#### **Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
sound	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound	175.700 175.700 175.700 175.700	3 2.049 3.000 1.000	58.567 85.755 58.567 175.700	65.682 65.682 65.682 65.682	.000 .000 .000	.891 .891 .891 .891	197.047 134.573 197.047 65.682	1.000 1.000 1.000 1.000
sound * audit	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound	3.400 3.400 3.400 3.400	3 2.049 3.000 1.000	1.133 1.659 1.133 3.400	1.271 1.271 1.271 1.271	.307 .308 .307 .292	.137 .137 .137 .137	3.813 2.604 3.813 1.271	.296 .239 .296 .169
Error(sound)	Sphericity Assumed Greenhouse-Geisser Huynh-Feldt Lower-bound	21.400 21.400 21.400 21.400	24 16.391 24.000 8.000	.892 1.306 .892 2.675					

a. Computed using alpha = .05

#### Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
sp1	.000	1	8	1.000
sp2	.004	1	8	.951
sp3	.005	1	8	.948
sp4	.377	1	8	.556

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+audit
 Within Subjects Design: sound

#### Tests of Between-Subjects Effects

Measure: MEASURE\_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Intercept	3132.900	1	3132.900	63.259	.000	.888	63.259	1.000
audit	48.400	1	48.400	.977	.352	.109	.977	.141
Error	396.200	8	49.525					

a. Computed using alpha = .05

In evaluating the necessary assumptions it is found that Box's M statistic is not significant (p > .001) and so the assumption of homogeneity of variance–covariance matrices is not violated. Mauchly's statistic is also not significant (p > .05) and so the sphericity assumption has not been violated; and the Levene's test indicates homogeneity of variances.

The results of the SPANOVA indicate that there is a main effect for sound proofing and that the quality of sound ratings differ with the choice of sound-proofing material, F(3,24) = 65.682, p < .05. However, no main effect is obtained for auditorium size.

The interaction is also not significant, meaning that the effect of choice of sound-proofing material does not depend on auditorium size.

Post-hoc tests could now be performed to determine where the differences in sound proofing lie.



# One-way analysis of covariance (ANCOVA)

Analysis of covariance (ANCOVA) provides an elegant means of reducing systematic bias, as well as within-groups error, in the analysis. To determine whether the independent variable is indeed having an effect, the influence of an extraneous variable (covariate) on the dependent variable is statistically controlled during the analysis. That is, you attempt to reduce error variance due to individual differences. ANCOVA requires that different participants perform in each condition, and because of this is suitable for only between or independent groups designs.

## Assumption testing



There are six assumptions that you need to address before conducting an ANCOVA:

- 1 **Independence** the individual's scores on both the dependent variable and the covariate should be independent of those scores for all the other participants.
- 2 Normality the dependent variable should have a normal distribution for participants with the same score on the covariate and in the same group. You want to obtain normality at each score on the covariate. If the scores for the covariate alone are normally distributed, then ANCOVA is robust to this assumption.
- 3 Linearity a linear relationship should exist between the dependent variable and the covariate for each group. This can be verified by inspecting scatterplots for each group.
- 4 Homogeneity of regression slopes the relationship of the dependent variable to the covariate in each group should be the same.
- 5 Independence of covariate and treatments when you remove the proportion of shared variability between the dependent variable and the covariate you must be careful that you do not also remove some of the effect of the independent variable. You can avoid this, to some extent, by measuring the covariate before the beginning of the experiment and also by randomly allocating participants to the different levels of the independent variable.
- **Reliability of the covariate** the instrument used to measure the covariate should be reliable.

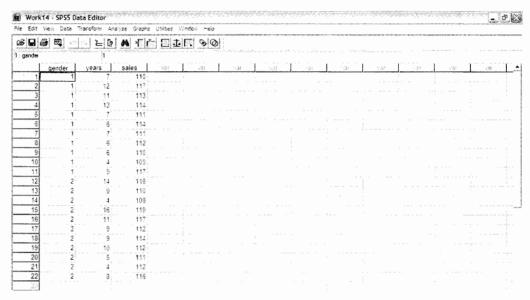
Assumptions 1, 5 and 6 relate to experimental design, while assumption 2 has been explored in chapter 3. To test assumptions 3 and 4, the procedures outlined in the research example can be used.

## Working example



A sales manager wished to determine whether women or men were more successful salespeople. She recorded the sales made by 22 sales representatives over a 12-month period. However, she was aware that years of selling experience would also contribute to the representatives' success and she thought it would be fairer to make the comparison while taking into account this added factor. Thus, the dependent variable is sales in \$1000s per year, the independent variable is gender (coded as 1 = male and 2 = female) and the covariate is years of experience in selling.

The data file can be found in Work14.sav on the website that accompanies this title and is shown in the following figure.

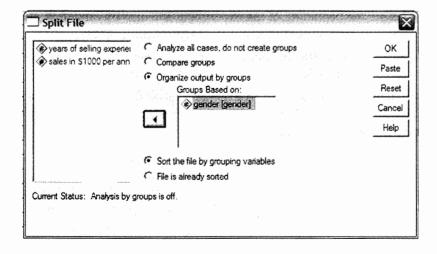




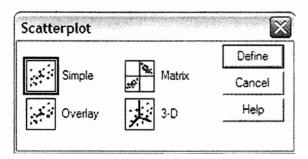
#### To test for linearity

Assumptions of linearity need to be assessed for each level or group of the independent variable separately. The **Split File...** option can be used to select specific groups from the data file.

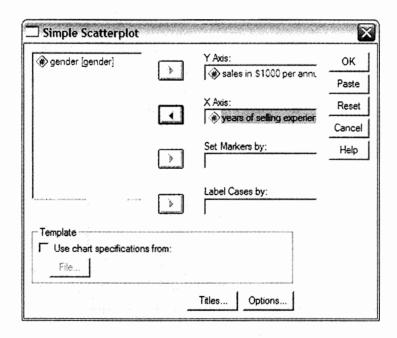
- 1 Select the **Data** menu and click on **Split File...** or click on the **Split File** tool to open the **Split File** dialogue box.
- Select the **Organize output by groups** radio button, then select the group variable (i.e. *gender*) and click on the button to move the variable into the **Groups Based on:** box.



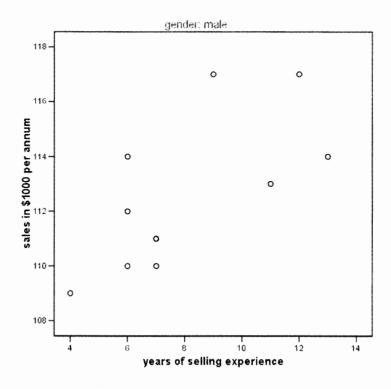
- 3 Click on OK.
- 4 Select the **Graphs** menu.
- 5 Click on **Scatter...** to open the **Scatterplot** dialogue box. Notice that the **Simple scatterplot** option is outlined by default.

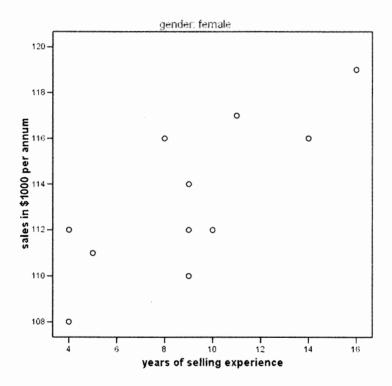


- 6 Click on the **Define** command pushbutton to open the **Simple Scatterplot** sub-dialogue box.
- 7 Select the first variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Y** Axis: box.
- 8 Select the second variable (i.e. *years*) and click on the ▶ button to move the variable into the **X** Axis: box.
- 9 Click on OK.



SORT CASES BY gender .
SPLIT FILE
SEPARATE BY gender .
GRAPH
/SCATTERPLOT(BIVAR)=years WITH sales
/MISSING=LISTWISE .



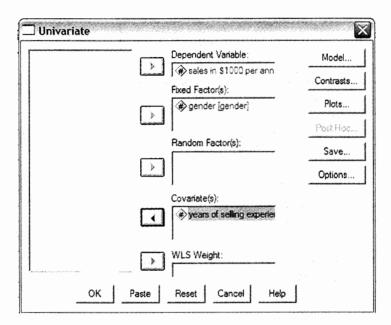


Both of these scatterplots indicate a linear relationship between the dependent variable (sales) and the covariate (years) for each group (male and female). In addition, it appears that the slope of the regression line is similar across groups. Therefore, you can confidently proceed with an analysis of covariance. Remember to turn the **Split File...** option off before conducting the ANCOVA. You do this by selecting the **Data** menu, clicking on **Split File...** to open the Split File dialogue box and selecting the **Analyse all cases, do not create groups** radio button. Click on **OK**. The split file option is now off and you're ready to continue.

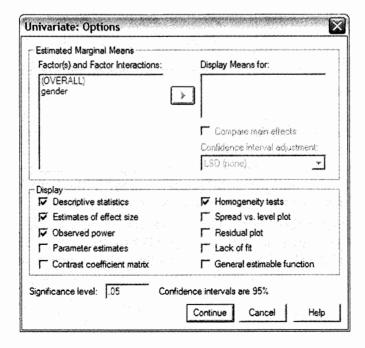


#### To conduct an analysis of covariance

- 1 Select the **Analyze** menu.
- 2 Click on General Linear Model and then Univariate... to open the Univariate dialogue box.
- 3 Select the dependent variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Dependent Variable:** box.
- 4 Select the independent variable (i.e. *gender*), and click on the ▶ button to move the variable into the **Fixed Factor(s)**: box.
- 5 Select the covariate (i.e. *years*) and click on the ▶ button to move the variable into the **Covariate**(s): box.



- 6 Click on the **Options...** command pushbutton to open the **Univariate: Options** sub-dialogue box.
- 7 In the **Display** box click on the **Descriptive** statistics, **Estimates of effect size**, **Observed power** and **Homogeneity tests** check boxes.



#### 8 Click on Continue and then OK.

UNIANOVA
salesBY genderWITH years
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PRINT = DESCRIPTIVE ETASQ OPOWER HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = years gender .

#### **Descriptive Statistics**

Dependent Variable: sales in \$1000 per annum

gender	Mean	Std. Deviation	N		
male	112.55	2.734	11		
female	113.36	3.325	11		
Total	112.95	3.000	22		

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: sales in \$1000 per annum

F	F df1		Sig.	
.022	1	20	.884	

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+years+gender

Levene's test is not significant (p > .5), indicating that the homogeneity of variance assumption has not been violated.

#### Tests of Between-Subjects Effects

Dependent Variable: sales in \$1000 per annum

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	106.353 <sup>b</sup>	2	53.176	12.232	.000	.563	24.463	.988
Intercept	31745.528	1	31745.528	7302.092	.000	.997	7302.092	1.000
years	102.671	1	102.671	23.616	.000	.554	23.616	.996
gender	.116	1	.116	.027	.872	.001	.027	.053
Error	82.602	19	4.347					
Total	280881.000	22						
Corrected Total	188.955	21						

a. Computed using alpha = .05

The output indicates no main effect (p > .05) for gender; but a significant relationship exists between sales and years of experience, F(1,19) = 23.616, p < .05. So it can be said that when you statistically control for years of experience, gender has no influence on the amount of sales made per year. The sales manager's hunch that sales success is related to years of selling experience is also confirmed.

If you had obtained a significant F-ratio for gender, and if you had more than two levels of the independent variable, then you would have had to tease out the source of the significance. Planned and post-hoc comparisons could be used to achieve this.

b. R squared = .560 (adjusted R squared = .517)

## Practice example



Ninety-six male and female second-year psychology students participated in a peer-tutoring program designed to reduce computer anxiety. Of these students, 47 volunteered to act as peer tutors in first-year research methods computing workshops, while 49 preferred not to be tutors. Computer anxiety for all potential tutors was measured before the beginning of the course and then on completion. Given the data in Prac14.sav, your tasks are to:

- 1 Determine whether the reduction in computer anxiety is the same for males and females.
- 2 Determine whether the reduction in computer anxiety is the same for those who acted as tutors and those who did not.

## **Solutions**



#### **Syntax**

**EXAMINE** VARIABLES=anxiety1 anxiety2 BY gender tutors /PLOT BOXPLOT /COMPARE GROUP /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL SORT CASES BY gender tutors . SPLIT FILE SEPARATE BY gender tutors . /SCATTERPLOT(BIVAR)=anxiety2 WITH anxiety1 /MISSING=LISTWISE . SPLIT FILE OFF. UNIANOVA anxiety2 BY gender WITH anxiety1 /METHOD = SSTYPE(3) /INTERCEPT = INCLUDE /PRINT = DESCRIPTIVE ETASQ OPOWER HOMOGENEITY /CRITERIA = ALPHA(.05) /DESIGN = anxiety1 gender . UNIANOVA anxiety2 BY tutors WITH anxiety1 /METHOD = SSTYPE(3) /INTERCEPT = INCLUDE /PRINT = DESCRIPTIVE ETASQ OPOWER HOMOGENEITY /CRITERIA = ALPHA(.05) /DESIGN = anxiety1 tutors .

#### **Assumption testing**

Normality

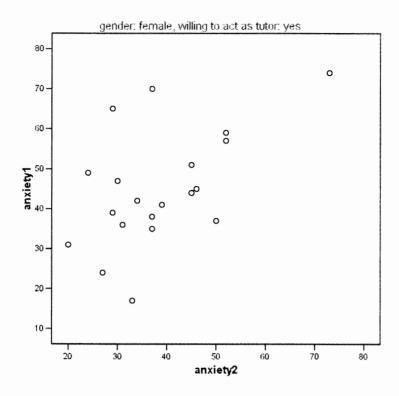
#### Descriptives

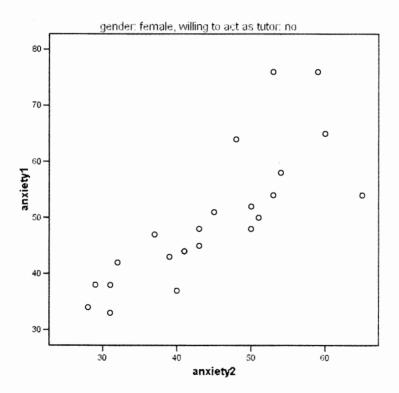
	GENDER			Statistic	Std. Error
ANXIETY1	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	47.49 43.41 51.56 47.37 45.00 175.256 13.24 17 76 59 16.00 .375 .169	.361 .709
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	36.47 33.14 39.80 36.58 35.50 102.634 10.13 18 53 35 16.25 124 906	.383 .750
ANXIETY2	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	41.70 38.13 45.27 41.29 41.00 134.549 11.60 20 73 53 19.00 .466 015	.361 .709
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	33.53 29.59 37.46 32.88 30.50 143.499 11.98 17 64 47 18.75 .860 164	1.943 .383 .750

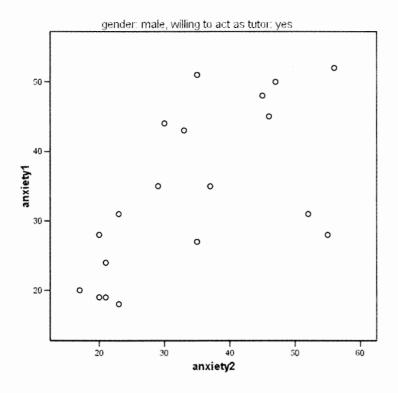
#### **Descriptives**

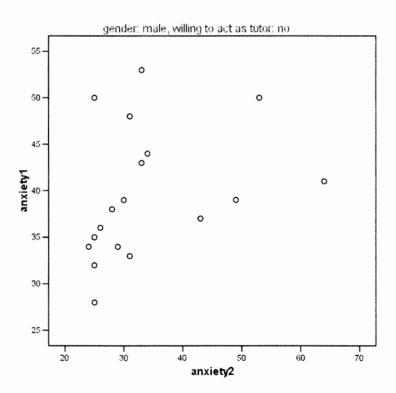
	willing to act as tutor			Statistic	Std. Error
ANXIETY1	yes	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	39.33 34.76 43.89 38.72 38.50 204.020 14.28 17 74 57 20.75 .438 168	.374 .733
	no	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	45.24 41.74 48.75 44.39 44 123.489 11.11 28 76 48 13.50 1.094 1.264	.369 .724
ANXIETY2	yes	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	35.90 31.83 39.97 35.28 34.50 161.682 12.72 17 73 56 21.00 .659 .241	.374 .733
	no	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	39.78 36.01 43.55 39.26 39.00 142.426 11.93 24 65 41 20.50 .469 902	.369 .724

Skewness and kurtosis statistics indicate that the distributions of *anxiety1* and *anxiety2* are relatively normal across both gender and tutor groups.









The scatterplots indicate a reasonable level of linearity between the dependent variable (anxiety2) and the covariate (anxiety1) for both pairs of groups. In relation to the homogeneity of regression slopes assumption, it appears that the slopes of each plot are also similar across groups.

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: anxiety2

F	df1	df2	Sig.	
1.412	1	79	.238	

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+anxiety1+gender

Levene's test shows that the assumption of homogeneity of variance has not been violated.

#### Tests of Between-Subjects Effects

Dependent Variable: anxiety2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	5250.664 <sup>b</sup>	2	2625.332	29.018	.000	.427	58.036	1.000
Intercept	958.990	1	958.990	10.600	.002	.120	10.600	.895
anxiety1	3903.702	1	3903.702	43.148	.000	.356	43.148	1.000
gender	45.390	1	45.390	.502	.481	.006	.502	.108
Error	7056.842	78	90.472					
Total	128437.000	81						
Corrected Total	12307.506	80						

- a. Computed using alpha = .05
- b. R squared = .427 (adjusted R squared = .412)

The main effect for the ANCOVA is not significant and therefore, we can conclude that when pretest anxiety is statistically controlled, no differences exist between males and females. That is, the reduction in computer anxiety is the same for men and women — F(1,78) = .502, p > .05.

Once again in relation to the next research question, the assumptions have not been violated.

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: anxiety2

F	df1	df2	Sig.
.788	1	79	.377

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+anxiety1+tutors

#### Tests of Between-Subjects Effects

Dependent Variable: anxiety2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	5206.303 <sup>b</sup>	2	2603.151	28.593	.000	.423	57.186	1.000
Intercept	917.842	1	917.842	10.082	.002	.114	10.082	.880
anxiety1	4901.421	1	4901.421	53.837	.000	.408	53.837	1.000
tutors	1.029	1	1.029	.011	.916	.000	.011	.051
Error	7101.203	78	91.041					
Total	128437.000	81						
Corrected Total	12307.506	80						
Corrected Total	12307.506	80						

- a. Computed using alpha = .05
- b. R squared = .423 (adjusted R squared = .408)

Again, this ANCOVA reveals no significant main effect. That is, the reduction in computer anxiety is the same for those who acted as tutors and those who did not.

## Reliability analysis

There are a number of different reliability coefficients. One of the most commonly used is Cronbach's alpha, which is based on the average correlation of items within a test if the items are standardised. If the items are not standardised, it is based on the average covariance among the items. Because Cronbach's alpha can be interpreted as a correlation coefficient, it ranges in value from 0 to 1. SPSS output also provides a standardised item alpha that is the value that would be obtained if all the items were standardised. Items usually possess comparable variances so there is little difference between these two alphas.

In addition to Cronbach's alpha, SPSS allows us to compute several other reliability models:

- · Split-half reliability
- Guttman
- Parallel
- Strictly parallel.

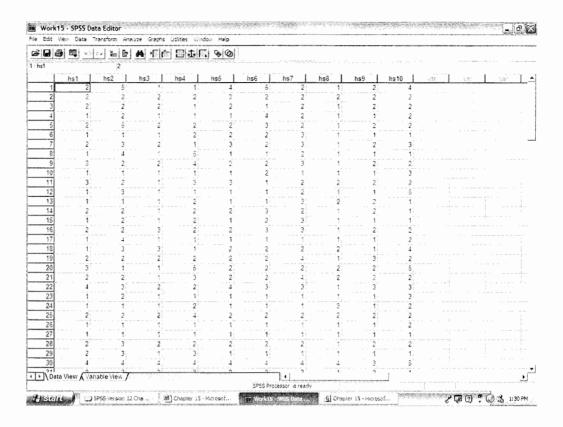
However, discussion in this chapter is limited to Cronbach's alpha.

## Working example



One-hundred-and-five members of the community completed a ten-item attitudes-to-help-seeking instrument. You wish to determine the internal consistency of this scale using Cronbach's alpha. Item 5 was negatively worded and was recoded before analysis.

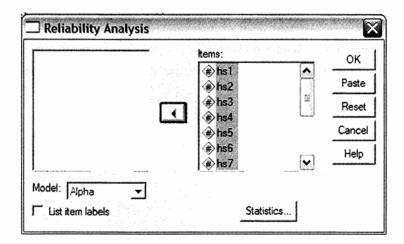
The data file can be found in Work15.sav on the website that accompanies this title and is shown in the following figure.





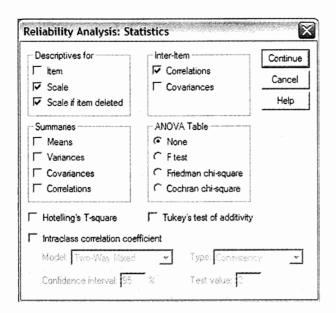
#### To conduct a reliability analysis

- 1 Select the **Analyze** menu.
- 2 Click on Scale and then Reliability Analysis... to open the Reliability Analysis dialogue box.
- 3 Select the variables you require (i.e. *hs1* to *hs10*) and click on the ▶ button to move the variables into the **Items:** box.
- 4 Ensure that *Alpha* is displayed in the **Model:** dropdown list.



5 Click on the **Statistics...** command pushbutton to open the **Reliability Analysis: Statistics** sub-dialogue box.

- 6 In the **Descriptives for** box, select the **Scale** and **Scale if item deleted** check boxes.
- 7 In the Inter-Item box, select the Correlations check box.



#### 8 Click on Continue and OK.

RELIABILITY
/VARIABLES=hs1 hs2 hs3 hs4 hs5 hs6 hs7 hs8 hs9 hs10
/FORMAT=NOLABELS
/SCALE(ALPHA)=ALL/MODEL=ALPHA
/STATISTICS=SCALE CORR
/SUMMARY=TOTAL.

#### **Reliability Statistics**

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N of items
.768	.792	10

#### Inter-Item Correlation matrix

	hs1	hs2	hs3	hs4	hs5	hs6	hs7	hs8	hs9	hs10
hs1	1.000	.297	.552	.454	.542	.355	.179	.390	.538	.148
hs2	.297	1.000	.449	.087	.376	.225	.101	.013	.107	.199
hs3	.552	.449	1.000	.297	.541	.377	.287	.418	.445	.238
hs4	.454	.087	.297	1.000	.274	.242	.197	.323	.271	.027
hs5	.542	.376	.541	.274	1.000	.314	.288	.266	.442	.313
hs6	.355	.225	.377	.242	.314	1.000	.170	.157	.331	.143
hs7	.179	.101	.287	.197	.288	.170	1.000	.063	.200	057
hs8	.390	.013	.418	.323	.266	.157	.063	1.000	.385	.278
hs9	.538	.107	.445	.271	.442	.331	.200	.385	1.000	.167
hs10	.148	.199	.238	.027	.313	.143	057	.278	.167	1.000

The covariance matrix is calculated and used in the analysis.

#### **Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Correct Item- Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
hs1	17.94	21.862	.648	.528	.720
hs2	17.39	23.048	.345	.301	.764
hs3	18.19	22.348	.693	.533	.719
hs4	17.72	22.836	.379	.259	.758
hs5	17.75	21.977	.646	.464	.721
hs6	17.86	23.220	.425	.209	.749
hs7	17.39	25.029	.246	.158	.772
hs8	18.45	24.692	.415	.339	.752
hs9	18.15	24.265	.526	.389	.743
hs10	17.44	23.710	.253	.193	.781

When you examine the characteristics of a scale you want to look at the individual items, the overall scale, and the relationship between the individual items and the overall scale.

You could have obtained means and standard deviations for each item by selecting the **Item** check box in the **Descriptives for** box. However, to minimise the output, this option was not chosen.

In relation to the Item-total Statistics you have a number of important columns:

- Scale Mean if Item Deleted this column tells us the average score for the scale if the item were excluded from the scale.
- Scale Variance if Item Deleted this column tells us the scale variance if the item were eliminated.
- Corrected Item-Total Correlation this column gives the Pearson correlation
  coefficient between the score on the individual item and the sum of the scores on the
  remaining items.
- **Squared Multiple Correlation** this column gives the result of a multiple regression equation with the item of interest as the dependent variable and all of the other items as independent variables.
- Cronbach's Alpha if Item Deleted this column gives the alpha coefficient that would result if the item were removed from the scale.

An examination of the items comprising the attitudes to help-seeking scale indicates that items 7 and 10 have the lowest corrected item-total correlations. The Cronbach's alpha for the overall scale is equal to .768. If these two items were removed from the scale, the **Alpha if Item Deleted** column shows that overall reliability would increase slightly. When you remove these items and recalculate the reliability coefficient, Cronbach's alpha is raised to .7845. Therefore, deletion of these items may be considered appropriate.

## Practice example



A researcher wished to determine the reliability of four independent personality scales — namely, hope, optimism, locus of control and self-esteem. The hope and optimism scales consisted of eight items each, the locus of control scale consisted of nine items and the self-esteem scale consisted of ten items. A five-point Likert scale response format was used for each of these scales. Items 1, 3, 4 and 7 of the optimism scale and items 3, 5, 8, 9 and 10 of the self-esteem scale were negatively worded. Three hundred and sixty-three people responded to each of the four scales, but not all scales were fully completed. Given the data in Prac15.sav, your tasks are to:

- 1 Determine the Cronbach's alpha coefficient for each scale.
- 2 Make recommendations about the inclusion or exclusion of items.

## **Solutions**



#### **Syntax**

```
RELIABILITY
 /VARIABLES=hope1 hope2 hope3 hope4 hope5 hope6 hope7 hope8
 /FORMAT=NOLABELS
 /SCALE(ALPHA)=ALL/MODEL=ALPHA
 /STATISTICS=SCALE
 /SUMMARY=TOTAL .
RELIABILITY
 /VARIABLES=loc1 loc2 loc3 loc4 loc5 loc6 loc7 loc8 loc9
 /FORMAT=NOLABELS
 /SCALE(ALPHA)=ALL/MODEL=ALPHA
 /STATISTICS=SCALE
 /SUMMARY=TOTAL .
RECODE
 opt1 opt3 opt4 opt7 se3 se5 se8 se9 se10 (1=5) (2=4) (4=2) (5=1) .
EXECUTE .
RELIABILITY
 /VARIABLES=se1 se2 se3 se4 se5 se6 se7 se8 se9 se10
 /FORMAT=NOLABELS
 /SCALE(ALPHA)=ALL/MODEL=ALPHA
 /STATISTICS=SCALE
 /SUMMARY=TOTAL .
RELIABILITY
 /VARIABLES=opt1 opt2 opt3 opt4 opt5 opt6 opt7 opt8
 /FORMAT=NOLABELS
 /SCALE(ALPHA)=ALL/MODEL=ALPHA
 /STATISTICS=SCALE
 /SUMMARY=TOTAL .
```

#### Output

#### Hope

#### **Reliability Statistics**

Cronbach's	Cronbach's Alpha Based on Standardised	
Alpha	Items	N of items
.749	.751	8

The Cronbach's alpha coefficient for the hope scale is .75.

#### **Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Correct Item- Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
hope1	25.51	13.315	.499	.333	.712
hope2	25.25	13.549	.429	.274	.726
hope3	25.30	13.474	.493	.264	.714
hope4	25.56	13.929	.445	.225	.723
hope5	25.55	13.168	.527	.342	.707
hope6	25.36	14.258	.303	.142	.751
hope7	25.45	13.997	.397	.215	.732
hope8	25.61	13.627	.475	.311	.717

Inspection of the **Alpha if Item Deleted** column reveals that deletion of *item 6* would increase reliability marginally.

#### **Locus of Control**

#### **Reliability Statistics**

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N of items
718	.716	9
./ 10	., 10	

The reliability coefficient for the locus of control scale is .718.

**Item-Total Statistics** 

	Scale Mean if	Scale Variance if Item Deleted	Correct Item- Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
loc1	21.07	21.967	.267	.165	.714
loc2	21.11	20.668	.301	.171	.713
loc3	21.49	19.013	.543	.339	.664
loc4	21.32	19.513	.512	.397	.671
loc5	21.12	19.563	.514	.320	.671
loc6	20.35	19.836	.489	.321	.676
loc7	21.97	21.993	.217	.097	.725
loc8	20.83	19.718	.488	.374	.676
loc9	21.84	22.526	.213	.077	.722

If *item* 7 were deleted, and the reliability analysis re-run, then the alpha coefficient would increase fractionally to .725.

Having removed this item you would then reassess the new output to determine whether any further items, such as item 9, should also be removed.

#### Self-Esteem

**Reliability Statistics** 

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N of items
.872	.881	10

The Cronbach's alpha coefficient for the self-esteem scale is .872.

**Item-Total Statistics** 

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Correct Item- Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
se1	32.57	37.380	.687	.590	.854
se2	32.58	39.362	.555	.460	.864
se3	32.39	36.036	.717	.540	.851
se4	32.76	39.461	.475	.298	.868
se5	32.63	36.892	.574	.363	.861
se6	32.96	35.858	.762	.622	.848
se7	33.11	36.006	.676	.556	.853
se8	33.58	36.464	.515	.365	.867
se9	33.44	35.224	.562	.381	.864
se10	33.30	35.481	.535	.387	.867

An examination of the item-total statistics indicates that no removal of items would enhance this reliability measure. This is to be expected for such a well-established scale.

#### **Optimism**

#### Reliability Statistics

ndardised	
	onbach's a Based on

The Cronbach's alpha coefficient for the optimism scale is .777.

**Item-Total Statistics** 

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Correct Item- Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
opt1	25.09	19.434	.350	.140	.776
opt2	24.31	18.382	.513	.345	.747
opt3	24.67	18.395	.572	.386	.738
opt4	24.55	18.124	.518	.319	.746
opt5	24.35	18.266	.581	.399	.736
opt6	24.31	18.942	.506	.315	.749
opt7	24.53	20.937	.263	.137	.785
opt8	24.47	18.199	.542	.342	.742

An examination of the item-total statistics indicates that the removal of *item 7* would increase the alpha coefficient only marginally to .785.



## Factor analysis

Factor analysis is a data reduction technique used to reduce a large number of variables to a smaller set of underlying factors that summarise the essential information contained in the variables. More frequently, factor analysis is used as an exploratory technique when the researcher wishes to summarise the structure of a set of variables. However, for testing a theory about the structure of a particular domain, confirmatory factor analysis is appropriate.

When the researcher's goal is to construct a reliable test, factor analysis is an additional means of determining whether items are tapping into the same construct.

The Factor menu in SPSS allows seven methods of factor extraction:

- Principal components (PC)
- · Unweighted least squares
- · Generalised least squares
- Maximum likelihood
- Principal axis factoring (PAF)
- · Alpha factoring
- Image factoring.

The most frequently used of these methods are PC and PAF; and there is much debate in the literature over which method is the most appropriate. This chapter demonstrates only PAF.

A number of steps comprise a factor analytic procedure:

- 1 Computation of the correlation matrix to determine the appropriateness of the factor analytic model.
- **Factor extraction** to determine the number of factors necessary to represent the data.
- **Rotation** to make the factor structure more interpretable. Rotation may be orthogonal (factors are uncorrelated with one another) or oblique (factors are correlated). The choice of rotation is both empirically and theoretically driven. The criteria for making this selection can be found in any good multivariate statistics text.

## Assumption testing



There are a number of assumptions and practical considerations underlying the application of PAF and PC.

**Sample size** — a minimum of five subjects per variable is required for factor analysis. A sample of 100 subjects is acceptable, but sample sizes of 200+ are preferable.

- **Normality** factor analysis is robust to assumptions of normality. However, if variables are normally distributed, then the solution is enhanced.
- 3 Linearity because factor analysis is based on correlation, linearity is important. If linearity is not present, the solution may be degraded.
- 4 Outliers among cases factor analysis is sensitive to outlying cases. These cases need to be identified and either removed from the data set or brought into the distribution by transformation or recode options.
- Multicollinearity and singularity this assumption is not relevant for PC. However, in the case of PAF, singularity and multicollinearity can be identified if any of the squared multiple correlations are near or equal to 1. If this is the case, the inclusion of the offending variables needs to be reconsidered.
- Factorability of the correlation matrix a correlation matrix that is appropriate for factor analysis will have several sizeable correlations. Inspect the matrix for correlations in excess of .3 and, if none are found, reconsider the use of PAF. The anti-image correlation matrix is used to assess the sampling adequacy of each variable. The measures of sampling adequacy are displayed on the diagonal of the anti-image correlation matrix. Variables with a measure of sampling accuracy that falls below the acceptable level of .5 should be excluded from the analysis. Bartlett's test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy are both tests that can be used to determine the factorability of the matrix as a whole. If Bartlett's test of sphericity is large and significant, and if the Kaiser-Meyer-Olkin measure is greater than .6, then factorability is assumed.
- 7 Outliers among variables a variable with a low squared multiple correlation with all other variables, and low correlations with all important factors, is an outlier among the variables. These outliers may need to be deleted from the analysis.

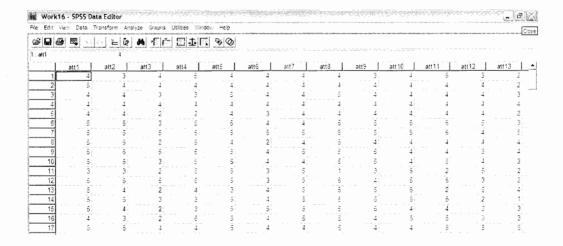
## Working example



Three hundred and eighty-eight participants responded to a survey regarding organ donation. They completed a 16-item scale that was designed to measure attitudes toward organ donation. The researcher wished to determine the underlying factor structure of this scale and decided to conduct a PAF analysis.

A sample size of 388 for a 16-item scale more than meets the desired cases-tovariables ratio for PAF. Furthermore, you are assuming that assumptions of normality and linearity have been satisfied before analysis. No outlying cases were detected. The remainder of the assumptions will be examined as they occur in the output.

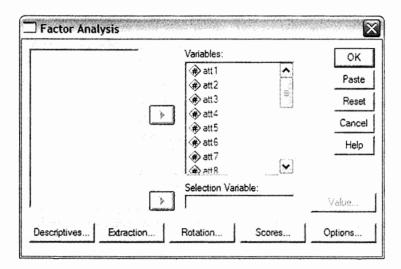
The data file can be found in Work16.sav on your data disk and is shown in the following figure.



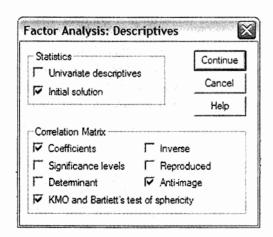


#### To conduct a principal axis factor analysis

- 1 Select the **Analyze** menu.
- 2 Click on **Data Reduction** and then on **Factor...** to open the **Factor Analysis** dialogue box.
- 3 Select the variables you require (i.e. *att1* to *att16*) and click on the ▶ button to move the variables into the **Variables:** box.

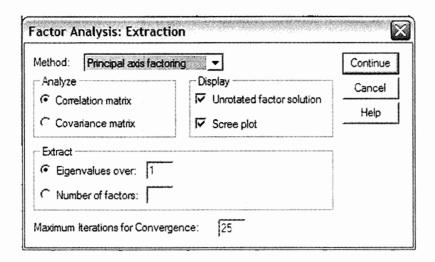


- 4 Click on the **Descriptives...** command pushbutton to open the **Factor Analysis: Descriptives** sub-dialogue box.
- 5 In the **Statistics** box, ensure that the **Initial solution** check box has been selected.
- In the Correlation Matrix box, select the Coefficients, KMO and Bartlett's test of sphericity and Anti-image check boxes.

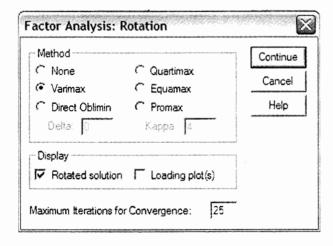


- 7 Click on Continue.
- 8 Click on the Extraction... command pushbutton to open the Factor Analysis: Extraction sub-dialogue box.
- 9 From the **Method:** drop-down menu, select *Principal axis factoring*.
- 10 Ensure the Correlation matrix radio button is selected in the Analyze box.
- In the **Extract** box, ensure that the **Eigenvalues over:** radio button has been selected and that *I* is displayed in the box.

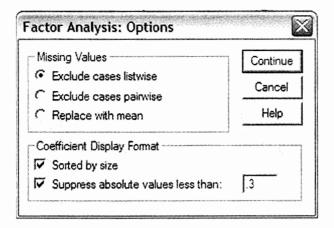
12 In the **Display** box, ensure that the **Unrotated factor solution** and **Scree plot** check boxes are selected.



- 13 Click on Continue.
- Click on the **Rotation...** command pushbutton to open the **Factor Analysis: Rotation** sub-dialogue box.
- 15 In the **Method** box, select the **Varimax** radio button.
- In the **Display** box, ensure that the **Rotated solution** check box has been selected.



- 17 Click on Continue.
- 18 Click on the **Options...** command pushbutton to open the **Factor Analysis: Options** sub-dialogue box.
- In the Coefficient Display Format box, select the Sorted by size and Suppress absolute values less than: check boxes. With the latter check box you are then prompted to enter the appropriate value you require, which is usually .3. You will also notice that in the previous sub-dialogue box you have the option of replacing missing values with the mean of each of the items. This procedure is useful when you have an incomplete data set. By default, missing cases are deleted listwise, meaning that the whole case is excluded from the analysis.



#### 20 Click on Continue and then OK.

You will be aware that there are many more options available within factor analysis. However, only the basic commands required will be addressed in this chapter. For example, a number of other rotation options were given in the **Rotation** sub-dialogue box. These included Equamax, Quartimax, Direct Oblimin and Promax. Similarly, the option to save factor loadings as variables is available in the **Scores** sub-dialogue box.

FACTOR
/VARIABLES att1 att2 att3 att4 att5 att6 att7 att8 att9 att10 att11 att12
att13 att14 att15 att16/MISSING LISTWISE /ANALYSIS att1 att2 att3 att4
att5 att6 att7 att8 att9 att10 att11 att12 att13 att14 att15 att16
/PRINT INITIAL CORRELATION KMO AIC EXTRACTION ROTATION
/FORMAT SORT BLANK(.3)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PAF
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/METHOD=CORRELATION .

An examination of the correlation matrix (see page 159) indicates that a considerable number of correlations exceed .3 and so the matrix is suitable for factoring.

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of	Sampling Adequacy.	.914
Bartlett's Test of Sphericity	Approx. Chi-Square	2491.010
	df	120
	Sig.	.000

You can see that the Bartlett test of sphericity is significant and that the Kaiser-Meyer-Olkin measure of sampling adequacy is far greater than .6.

Measures of sampling adequacy are printed on the diagonal. Inspection of the antiimage correlation matrix reveals that all our measures of sampling adequacy are well above the acceptable level of .5.

×
Ξ
₹
Σ
_
ō
₹
ä
5
'n
ŏ
_

							Correlation Matrix	JU Matrix								
	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12	att13	att14	att15	att16
orrelation att	_	.664	.250	.435	490	.315	.378	.328	.574	.336	575	338	176	.436	.379	.560
att		1.000	.383	909.	444	.456	.345	.260	.525	.316	.468	414	.320	.533	.480	.674
att3	3 .250	.383	1.000	.457	.210	.321	.216	.054	.217	.206	.231	.225	429	.425	314	.296
att		.506	.457	1.000	.351	.352	.336	.240	.415	.352	405	.416	.331	.558	.439	.529
att		.444	.210	.351	1.000	210	.318	194	.303	.216	.603	.330	.188	.296	.238	.352
att		.456	.321	.352	210	1.000	.358	128	379	.475	.329	.290	.276	.421	.311	.486
att		.345	.216	.336	.318	.358	1.000	.256	.373	.344	.332	.320	175	.333	.265	397
att		.260	.054	.240	194	.128	.256	1.000	.348	.209	.215	.128	.128	.200	.231	.265
att		.525	.217	.415	:303	.379	.373	.348	1.000	.437	368	.383	.203	.492	398	609
att1		.316	.206	.352	.216	.475	.344	.209	.437	1.000	366	.296	181	.325	.289	.419
att1		.468	.231	.405	.603	.329	.332	.215	368	366	1.000	.338	176	.382	.333	.445
att1		.414	.225	.416	330	.290	.320	.128	.383	.296	338	1.000	.186	377	.266	.386
att1.		.320	.429	.331	.188	.276	.175	.128	.203	.181	176	.186	1.000	.391	.233	.318
att1		.533	.425	.558	.296	.421	.333	.200	.492	.325	.382	.377	.391	1.000	.428	.579
att1:		.480	314	.439	.238	.311	.265	.231	398	588	.333	.266	.233	.428	1.000	.559
att1		.674	.296	.529	.352	.486	.397	.265	609	.419	.445	.386	.318	629	.559	1.000

	l
ces	
Matri	١
паде м	l
===	I
Ĭ	I
	I

							HIG-II	And-mage Manices	saci								
		att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12	att13	att14	att15	att16
Anti-image Covariance	att1	339	141	001	012	051	.047	042	058	112	005	120	.029	.048	002	.014	016
	att2	-141	366	057	013	054	080	030	010	004	.051	.016	090'-	030	024	045	102
	att3	001	057	.647	135	004	062	023	.075	.021	001	900.	910	189	074	990'-	.063
	att4	012	013	135	.519	027	.021	017	052	.01	046	024	760	019	107	057	047
	att5	051	054	004	027	.570	.040	068	-010	.010	.034	224	063	038	.021	.028	600
	att6	.047	080	062	.021	.040	.605	060:-	.041	010	183	039	004	033	046	.017	060
	att7	042	030	023	017	068	060	.719	680	024	061	002	075	800	015	.00	035
	att8	058	010	920.	052	010	.041	680	.817	102	032	.002	.053	055	.016	054	.017
	att9	112	004	.021	.011	.010	010	024	102	.482	660	.039	071	.022	072	013	095
	att10	005	.051	-00	046	.034	183	061	032	660	.647	082	040	900:-	.023	014	029
	att11	120	.016	900	024	224	039	002	.002	.039	082	.491	025	.020	029	034	015
	att12	.029	090'-	.016	097	063	004	075	.053	071	040	025	.711	900	038	.01	.004
	att13	.048	030	189	019	038	033	800	055	.022	900	.020	900	.737	094	.016	040
	att14	002	024	074	107	.021	046	015	.016	072	.023	029	038	094	.501	026	064
	att15	.014	045	990:-	057	.028	.017	.001	054	013	014	034	.011	.016	026	.632	125
	att16	016	-102	.063	047	600.	060	035	.017	095	029	015	.004	040	064	125	.362
Anti-image Correlation	att1	.897 <sup>a</sup>	-369	002	026	107	960.	6.00-	101	256	600:-	271	.055	680.	005	.028	042
	att2	369	.911ª	118	030	118	169	.059	019	008	.105	.037	-118	058	055	095	280
	att3	002	118	.865 <sup>a</sup>	234	900'-	660	034	104	760.	001	.011	.024	274	130	-104	.131
	att4	026	030	234	.939 <sup>a</sup>	050	.037	027	620	.023	620	047	- 159	031	209	-100	107
	att5	107	118	900:-	050	.875ª	690	106	015	.018	950.	423	860	058	660.	.047	.021
	att6	960.	169	660	.037	690	.907 <sup>a</sup>	136	.058	019	293	072	- 007	049	083	.027	127
	att7	079	.059	034	027	106	136	$.950^{a}$	117	041	680	003	-106	.01	025	.002	070
	att8	-101	019	104	620	015	.058	117	.894ª	162	044	.003	690.	071	.024	075	.032
	att9	256	-008	.037	.023	.018	019	041	162	.921	177	080	121	980.	147	023	229
	att10	600	.105	001	079	.056	293	089	044	177	.901 <sup>a</sup>	145	059	008	.041	022	061
	att11	271	.037	.011	047	423	072	003	.003	080	145	.883 <sup>a</sup>	042	.034	057	090'-	036
	att12	.055	118	.024	159	860	007	106	690.	121	059	042	.944ª	600	063	.016	800.
	att13	680.	058	274	031	058	049	.011	071	.036	008	.034	600	.887 <sup>a</sup>	154	.023	078
	att14	005	055	130	209	.039	083	025	.024	147	.041	057	063	154	.946ª	047	150
	att15	.028	095	- 104	-100	.047	.027	.002	075	023	022	060	.016	.023	047	.943ª	262
	att16	042	280	.131	107	.021	127	070	.032	229	061	036	800	078	150	262	.922ª

a. Measures of Sampling Adequacy (MSA)

The table below displays the communality of the items and you will notice that ATT8 has the lowest communality.

Communalities

	Initial	Extraction
att1	.601	.617
att2	.634	.606
att3	.353	.526
att4	.481	.514
att5	.430	.645
att6	.395	.360
att7	.281	.278
att8	.183	.164
att9	.518	.598
att10	.353	.308
att11	.509	.576
att12	.289	.274
att13	.263	.320
att14	.499	.550
att15	.368	.356
att16	.638	.682

Extraction Method: Principal Axis Factoring

**Total Variance Explained** 

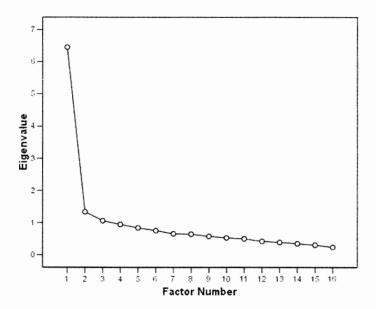
		Initial Eigenv	alues/	Extract	ion Sums of Sq	uared Loadings	Rotation	on Sums of Squ	ared Loadings
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.452	40.324	40.324	5.95	37.243	37.243	3.346	20.915	20.915
2	1.340	8.373	48.697	9	5.206	42.449	2.150	13.438	34.353
3	1.062	6.639	55.336	.833	3.637	46.086	1.877	11.733	46.086
4	.951	5.942	61.278	.582					
5	.841	5.253	66.531						
6	.756	4.727	71.257						
7	.656	4.101	75.359						
8	.643	4.017	79.376						
9	.577	3.608	82.985						
10	.528	3.298	86.283						
11	.499	3.118	89.401						
12	.421	2.633	92.033						
13	.389	2.431	94.464						
14	.348	2.176	96.640						
15	.302	1.889	98.529						
16	.235	1.471	100.000						

Extraction Method: Principal Axis Factoring.

The table above displays the total variance explained at three stages. At the initial stage, it shows the factors and their associated eigenvalues, the percentage of variance explained and the cumulative percentages. In reference to the eigenvalues, you would expect three factors to be extracted because they have eigenvalues greater than 1. If three factors were extracted, then 46 per cent of the variance would be explained.

The following scree plot graphically displays the eigenvalues for each factor and suggests that there is one predominant factor.

#### Scree Plot



The factor matrix is a matrix of loadings or correlations between the variables and factors. Pure variables have loadings of .3 or greater on only one factor. Complex variables may have high loadings on more than one factor, and they make interpretation of the output difficult. Rotation may therefore be necessary.

Factor Matrix<sup>a</sup>

		Factor	
	1	2	3
att16	.797		
att2	.778		
att1	.725	301	
att14	.702		
att9	.696		324
att4	.688		
att11	.643	333	
att15	.581		
att6	.569		,
att5	.562	401	.410
att10	.526		
att12	.522		
att7	.519		
att3	.487	.441	.308
att13	.412	.345	
att8	.352		

Extraction Method: Principal Axis Factoring.

a. Three factors extracted. 18 iterations required.

#### Rotated Factor Matrix<sup>a</sup>

		Factor	
	1	2	3
att9	.732		
att16	.710	.359	
att1	.567		.525
att2	.555	.391	.382
att10	.498		
att6	.464	.366	
att15	.462	.344	
att7	.424		
att8	.370		
att12	.361		
att3		.709	
att13		.545	
att14	.470	.544	
att4	.403	.527	
att5			.770
att11	.341		.655

Extraction Method: Principal Axis Factoring.

Rotation Method: Varimax with Kaiser Normalisation.

a. Rotation converged in six iterations.

Varimax rotation, where the factor axes are kept at right angles to each other, is most frequently chosen. Ordinarily, rotation reduces the number of complex variables and improves interpretation. However, in the example, the rotated solution still includes several complex variables. Factor 1 comprises ten items with factor loadings ranging from .36 to .73. Factor 2 comprises four items with factor loadings ranging from .53 to .71. Factor 3 comprises two items with loadings of .65 and .77. Some items have dual or triple loadings greater than .3 on more than one factor. These items must be

interpreted with caution because simple structure is not apparent.

**Factor Transformation Matrix** 

Factor	1	2	3
1	.717	.515	.470
2	113	.751	650
3	688	.412	.597

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalisation.

It is not surprising that this ambiguous structure has been obtained from varimax/ orthogonal rotation. Items 1–16 were designed to measure a single construct and therefore it is expected that the factors extracted would be highly correlated. In this instance an oblique rotation (direct oblimin) would have been a more appropriate choice.

#### **FACTOR**

/VARIABLES att1 att2 att3 att4 att5 att6 att7 att8 att9 att10 att11 att12 att13 att14 att15 att16/MISSING LISTWISE /ANALYSIS att1 att2 att3 att4 att5 att6 att7 att8 att9 att10 att11 att12 att13 att14 att15 att16 /PRINT INITIAL CORRELATION KMO AIC EXTRACTION ROTATION /FORMAT SORT BLANK(.3) /PLOT EIGEN /CRITERIA MINEIGEN(1) ITERATE(25) /EXTRACTION PAF /CRITERIA ITERATE(25) DELTA(0) /ROTATION OBLIMIN /METHOD=CORRELATION.

Pattern Matrix<sup>a</sup>

		Factor	
	1	2	3
att9	.837		
att16	.767		
att10	.552		
att1	.511		.410
att2	.506		
att6	.494		
att15	.477		
att14	.445	.409	
att7	.419		
att8	.409		
att12	.318		,
att3		.724	
att13		.535	
att4	.329	.408	
att5			.836
att11			.640

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalisation. a. Rotation converged in nine iterations.

Structure Matrix

	_	Factor	
	1	2	3
att16	.818	.483	.468
att9	.767		.418
att2	.734	.506	.581
att1	.707		.681
att14	.648	.624	.392
att4	.607	.606	.455
att15	.571	.424	.333
att6	.564	.441	
att10	.554		.307
att7	.515		.388
att12	.487	.339	.403
att8	.388		
att3	.336	.722	
att13	.313	.563	
att5	.420		.800
att11	.551		.745

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalisation.

**Factor Correlation Matrix** 

Factor	1	2	3
1	1.000	.473	.579
2	.473	1.000	.289
3	.579	.289	1.000

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalisation.

The oblique rotation provides a far more interpretable solution than that of the varimax rotation. Two matrices are produced: a pattern matrix and a structure matrix. The difference between high and low loadings is more apparent in the pattern matrix, so this matrix is interpreted. The loadings in the pattern matrix represent the unique relationship between the factor and the variable. As illustrated in the output, the pattern matrix has fewer complex variables and simpler structure. The factor correlation matrix indicates the relationship between factors. All factors appear moderately related.

The final step in factor analysis involves determining how many factors to interpret and then assigning a label to these factors. The number of factors to be interpreted largely depends on the underlying purpose of the analysis. In the present study, the purpose was to confirm the factor structure of a scale. In other instances, factor analysis may be used to summarise the data set.

From the output you can see that a single factor solution may be more appropriate in this analysis. You will remember from examining the Total Variances explained table (earlier in the output) that factor 1 had an eigenvalue of 5.96, whereas factors 2 and 3 had eigenvalues below 1.000 (.833 and .582 respectively).

The scree plot also confirmed the dominance of a single factor represented by 11 items.

Examination of the items indicates that these items represent a conceptually distinct aspect of attitudes to organ donation that can be labelled 'altruistic issues'. A final step would be to determine Cronbach's alpha coefficient of internal consistency to ensure that the items comprising factor 1 produce a reliable scale. Reliability and factor analysis are complimentary procedures in scale construction and definition.

## Practice example



You will be familiar with this data set because it was used in Chapter 15 to work through Reliability Analysis.

A researcher wished to determine the factor structure of four independent personality scales — namely, hope, optimism, locus of control and self-esteem. The hope and optimism scales consisted of eight items each, the locus of control scale consisted of nine items and the self-esteem scale consisted of ten items. A five-point Likert scale response format was used for each of these scales. Items 1, 3, 4 and 7 of the optimism scale and items 3, 5, 8, 9 and 10 of the self-esteem scale were negatively worded. Three hundred and sixty-three people responded to each of the four scales, but not all scales were completed fully. Given the data in Prac16.sav, your tasks are to:

- 1 Determine the factor structure of each of the four scales independently.
- 2 Make recommendations regarding the inclusion and exclusion of items from each scale.

## **Solutions**



Given the size of the output, please note that the syntax and output below has been displayed for the hope scale only.

### **Syntax**

DESCRIPTIVES
VARIABLES=hope1 hope2 hope3 hope4 hope5 hope6 hope7 hope8
/STATISTICS=MEAN STDDEV KURTOSIS SKEWNESS.

FACTOR
/VARIABLES hope1 hope2 hope3 hope4 hope5 hope6 hope7 hope8 /MISSING LISTWISE
ANALYSIS hope1 hope2 hope3 hope4 hope5 hope6
hope7 hope8
/PRINT INITIAL CORRELATION KMO AIC EXTRACTION ROTATION
/FORMAT SORT BLANK(.3)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PAF
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/METHOD=CORRELATION.

#### **Output**

#### **Assumption testing**

The 363 subjects are well within the required sample size for each scale.

**Descriptive Statistics** 

	N	Mean	Std.	Skewness		Kurt	tosis
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
hope1	363	3.56	.869	454	.128	.049	.255
hope2	363	3.83	.907	766	.128	.322	.255
hope3	362	3.78	.845	621	.128	.484	.256
hope4	362	3.51	.806	428	.128	281	.256
hope5	362	3.53	.875	472	.128	.183	.256
hope6	362	3.72	.928	694	.128	.188	.256
hope7	360	3.63	.847	741	.129	.736	.256
hope8	362	3.47	.839	696	.128	.278	.256
Valid N (listwise)	360						

An examination of the skewness and kurtosis statistics indicates that all items on the hope scale are reasonably normally distributed. You will note that several items have missing cases. Due to the number of boxplots required to identify outliers among cases, this output has been excluded. However, the testing of this assumption should not ordinarily be omitted.

The output for the hope scale is given below. In the **Factor Analysis: Options** dialogue box, we selected the **Replace with Mean** radio button because we noted earlier that there were missing cases. This function replaces all missing data with the mean for that variable and thus increases the sample size for analysis.

Correlation Matrix

		hope1	hope2	hope3	hope4	hope5	hope6	hope7	hope8
Correlation	hope1	1.000	.474	.381	.393	.271	.162	.194	.251
	hope2	.474	1.000	.361	.295	.226	.134	.184	.182
	hope3	.381	.361	1.000	.307	.276	.175	.304	.298
	hope4	.393	.295	.307	1.000	.306	.178	.185	.252
	hope5	.271	.226	.276	.306	1.000	.363	.318	.467
İ	hope6	.162	.134	.175	.178	.363	1.000	.146	.184
	hope7	.194	.184	.304	.185	.318	.146	1.000	.404
	hope8	.251	.182	.298	.252	.467	.184	.404	1.000

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure	.798	
Bartlett's Test of Sphericity	Approx. Chi-Square	542.236
	df	28
	.000	

The Bartlett test of sphericity is significant and the Kaiser-Meyer-Olkin measure of sampling adequacy is greater than .6. Therefore, it is appropriate to proceed with factor analysis. Measures of sampling adequacy are printed on the diagonal in the correlation matrix overleaf. Because all these values are above the acceptable level of .5, there is further evidence of the factorability of the items.

#### **Anti-image Matrices**

***		hope1	hope2	hope3	hope4	hope5	hope6	hope7	hope8
Anti-Image Covariance	hope1	.667	238	118	164	033	020	001	043
	hope2	238	.726	133	063	033	009	028	.015
	hope3	118	133	.736	084	026	040	121	071
	hope4	164	063	084	.775	093	036	008	043
	hope5	033	033	026	093	.658	213	080	220
	hope6	020	009	040	036	213	.858	011	.007
	hope7	001	028	121	008	080	011	.785	197
	hope8	043	.015	071	043	220	.007	197	.689
Anti-Image Correlation	hope1	.780ª	342	169	228	049	026	002	064
	hope2	342	.782ª	182	084	047	012	038	.021
	hope3	169	182	.853ª	111	037	051	160	100
3	hope4	228	084	111	.856ª	130	044	010	060
	hope5	049	047	037	130	.770 <sup>a</sup>	283	112	327
	hope6	026	012	051	044	283	.774 <sup>a</sup>	013	.009
	hope7	002	038	160	010	112	013	.813 <sup>a</sup>	267
	hope8	064	021	100	060	327	009	267	.772 <sup>a</sup>

a. Measures of Sampling Adequacy(MSA)

#### Communalities

	Initial	Extraction
hope1	.333	.542
hope2	.274	.412
hope3	.264	.334
hope4	.225	.282
hope5	.342	.502
hope6	.142	.143
hope7	.215	.264
hope8	.311	.451

Extraction Method: Principal Axis Factoring.

#### **Total Variance Explained**

		Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
1	2.949	36.867	36.867	2.345	29.314	29.314	1.472	18.399	18.399	
2	1.160	14.501	51.369	.585	7.313	36.627	1.458	18.229	36.627	
3	.929	11.608	62.977							
4	.735	9.183	72.159							
5	.647	8.082	80.241							
6	.597	7.462	87.704							
7	.513	6.412	94.115							
8	.471	5.885	100.000							

Extraction Method: Principal Axis Factoring.

Examination of the initial statistics reveals that two factors will be extracted. The final statistics outline that these two factors account for 37 per cent of the variance. It does appear that factor 1 is predominant with an eigenvalue of 2.34. The scree plot (not displayed) also suggests that a one-factor solution is most appropriate.

Factor 1 comprises items 5, 8, 7 and 6, while factor 2 is made up of items 1, 2, 3 and 4. Item 3 is a complex variable with dual loadings on both factors. The inclusion of this item is at the discretion of the analyst.

Factor Matrix<sup>a</sup>

	Factor				
	1	2			
hope1	.630	381			
hope5	.626	.332			
hope8	.582	.336			
hope3	.568				
hope2	.536	352			
hope4	.517				
hope7	.468				
hope6	.349				

Extraction Method: Principal Axis Factoring. a. Two factors extracted. 11 iterations required.

Rotated Factor Matrix<sup>a</sup>

	Factor				
	1	2			
hope5	.678				
hope8	.650				
hope7	.482				
hope6	.349				
hope1		.714			
hope2		.628			
hope3	.326	.478			
hope4		.449			

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalisation.

a. Rotation converged in six iterations.

The factor transformation matrix indicates a relatively high correlation between factors 1 and 2, and again suggests that an oblique rotation may have been more appropriate.

**Factor Transformation Matrix** 

Factor	1	2
1	.710	.704
2	.704	710

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalisation.

Given that you did so well on this practice example, now repeat the procedure for each of the other scales — optimism, locus of control and self-esteem. Before you do, remember that items 1, 3, 4 and 7 of the optimism scale and items 3, 5, 8, 9 and 10 of the self-esteem scale were negatively worded and will require recoding.



# Multiple regression

Multiple regression is an extension of bivariate correlation. The result of regression is an equation that represents the best prediction of a dependent variable from several independent variables. Regression analysis is used when independent variables are correlated with one another and with the dependent variable. Independent variables can be either continuous or categorical. However, in the latter case these variables must be coded as dummy variables. In contrast, the dependent variable must be measured on a continuous scale. If the dependent variable is not continuous, then discriminant function analysis is appropriate.

There are three major regression models — namely, standard or simultaneous regression, hierarchical regression and stepwise regression. These models differ in two ways: first, in the treatment of overlapping variability due to correlation of the independent variables, and second, in terms of the order of entry of the independent variables into the equation.

In the standard or simultaneous model, all independent variables enter the regression equation at once because you want to examine the relationship between the whole set of predictors and the dependent variable. In hierarchical multiple regression, you determine the order of entry of the independent variables based on theoretical knowledge.

In stepwise regression, the number of independent variables entered and the order of entry are determined by statistical criteria generated by the stepwise procedure. Method of entry can be forward, backward or a combination of both. Forward selection involves the entry of predictors one at a time. The order of entry and whether the predictor is eventually accepted are decided on the basis of whether the F-test exceeds a certain critical value (FIN) and whether a critical alpha level (PIN) is met. Backward selection starts with all the variables in the equation and gradually deletes poor performers on the basis of whether the partial F-value is less than a critical value (FOUT). The default criterion (POUT) must also be met. Stepwise selection is a combination of the forward and backward procedures. It allows for the later removal of variables that were previously entered.

The choice of technique largely depends on the researcher's goals.

## Assumption testing



A number of assumptions underpin the use of regression:

- 1 Ratio of cases to independent variables the number of cases needed depends on the type of regression model to be used. For standard or hierarchical regression you should ideally have twenty times more cases than predictors, whereas even more cases are required for stepwise regression.
  - The minimum requirement is to have at least five times more cases than independent variables.
- Outliers extreme cases have considerable impact on the regression solution and should be deleted or modified to reduce their influence. Univariate outliers can be detected during data screening as described in chapter 3. Multivariate outliers can be detected using statistical methods such as Mahalanobis distance (see chapter 18) and graphical methods such as residual scatterplots.
  - The decision to remove outliers from the data set must be made with care because their deletion often results in the generation of further outlying cases.
- Multicollinearity and singularity multicollinearity refers to high correlations among the independent variables, whereas singularity occurs when perfect correlations among independent variables exist. These problems affect how you interpret any relationships between the predictors (IVs) and the dependent variable, and they can be detected by examining the correlation matrix, squared multiple correlations and tolerances. Most computer programs have default values for multicollinearity and will not admit variables that are a problem.
- Normality, linearity, homoscedasticity and independence of residuals an examination of residual scatterplots allows us to test the above assumptions. It is assumed that the differences between the obtained and predicted dependent variable scores are normally distributed. Furthermore, it is assumed that the residuals have a linear relationship with the predicted dependent variable scores, and that the variance of the residuals is the same for all predicted scores. Mild deviations from linearity are not serious. Moderate to extreme deviations may lead to a serious underestimation of a relationship.

Assumption 1 relates to research design. Assumptions 2, 3 and 4 are assessed through regression analysis.

## Working example

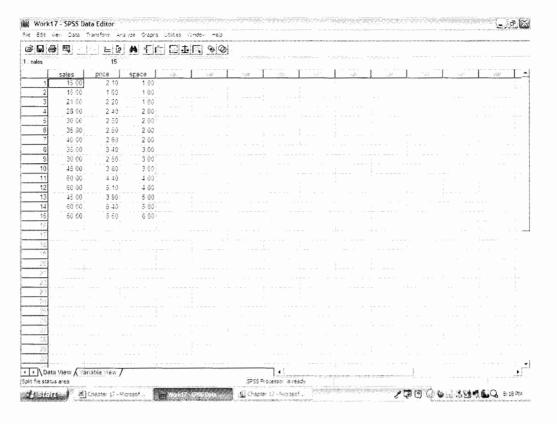


A marketing manager of a large supermarket chain wanted to determine the effect of shelf space and price on the sales of pet food. A random sample of 15 equal-sized shops was selected, and the sales, shelf space in square metres and price per kilogram were recorded.

To demonstrate the application of the three regression models you will use the same data set to ask three questions:

- What contribution do both shelf space and price make to the prediction of sales of pet food?
- Which is the best predictor of sales of pet food?
- 3 Previous research has suggested that shelf space is the salient predictor of sales of pet food. Is this hypothesis correct?

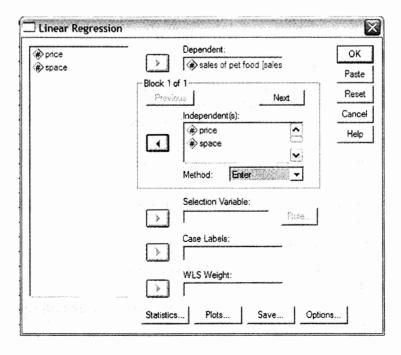
The data file can be found in Work17.sav on the website that accompanies this title and is shown in the following figure.



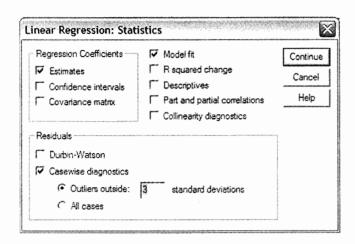


## To conduct a standard (simultaneous) regression analysis

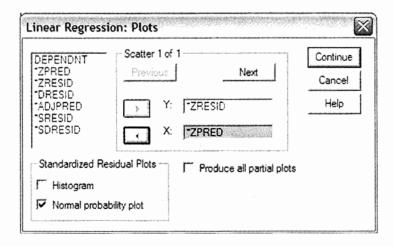
- 1 Select the Analyze menu.
- 2 Click on Regression and then on Linear... to open the Linear Regression dialogue box.
- 3 Select the dependent variable (i.e. *sales of pet food*) and click on the ▶ button to move the variable into the **Dependent:** box.
- 4 Select the independent variables (i.e. *space* and *price*) and click on the ▶ button to move the variables into the **Independent(s):** box.
- 5 In the **Method:** drop-down list, ensure that **Enter** is selected.



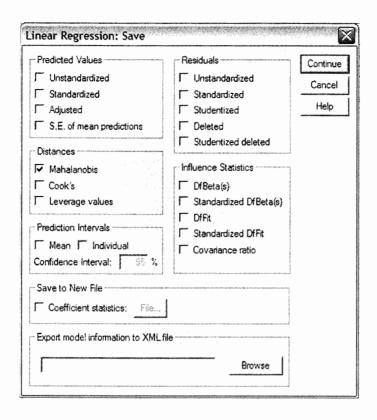
Click on the Statistics... command pushbutton to open the Linear Regression: Statistics sub-dialogue box and ensure the Estimates and Model fit check boxes are selected. In the Residuals box, select the Casewise diagnostics check box. Ensure the Outliers outside radio button has been selected. The default value of three standard deviations is most frequently used.



- 7 Click on Continue.
- 8 Click on the **Plots...** command pushbutton to open the **Linear Regression: Plots** sub-dialogue box.
- 9 Select \*ZRESID and click on the ▶ button to move the item into the Y: box.
- 10 Select \*ZPRED and click on the ▶ button to move the item into the X: box.
- In the Standardized Residual Plots box, select the Normal probability plot check box.



- 12 Click on Continue.
- 13 Click on the **Save...** command pushbutton to open the **Linear Regression: Save** sub-dialogue box.
- 14 In the Distances box, select the Mahalanobis check box.



## 15 Click on Continue and then OK.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT sales
/METHOD=ENTER space price
/SCATTERPLOT=(\*ZRESID ,\*ZPRED )
/RESIDUALS NORM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE MAHAL .

#### Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	space, price <sup>a</sup>		Enter

- a. All requested variables entered.
- b. Dependent Variable: sales of pet food

## Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.922a	.850	.825	6.05904

- a. Predictors: (Constant), space, price
- b. Dependent Variable: sales of pet food

Both independent variables together explain 85 per cent of the variance (R Square) in sales of pet food, which is highly significant as indicated by the F-value of 34.081 below.

**ANOVA<sup>b</sup>** 

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2502.390 440.543	2 12	1251.195 36.712	34.081	.000 <sup>a</sup>
	Residual Total	2942.933	14	30.712		

a. Predictors: (Constant), space, priceb. Dependent Variable: sales of pet food

An examination of the T-values indicates that price contributes to the prediction of sales.

Coefficients<sup>a</sup>

		Unstand Coeffi	dardised cients	Standardised Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1	(Constant)	2.029	5.126		.396	.699
	price	10.500	3.262	.916	3.219	.007
	space	.057	2.613	.006	.022	.983

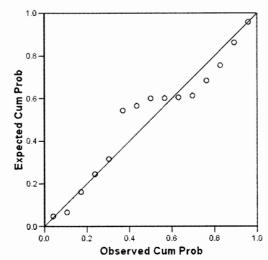
a. Dependent Variable: sales of pet food

Because no univariate outliers were found, casewise plots were not necessary. If they had been produced, then these plots would have identified outlying cases with standard deviations greater than three.

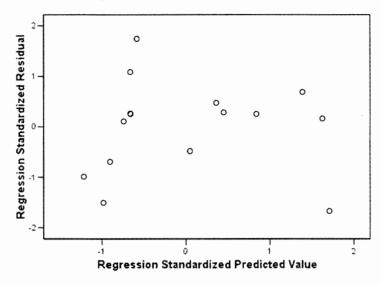
From the scatterplot of residuals against predicted values, you can see that there is no clear relationship between the residuals and the predicted values, consistent with the assumption of linearity.

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: sales of pet food

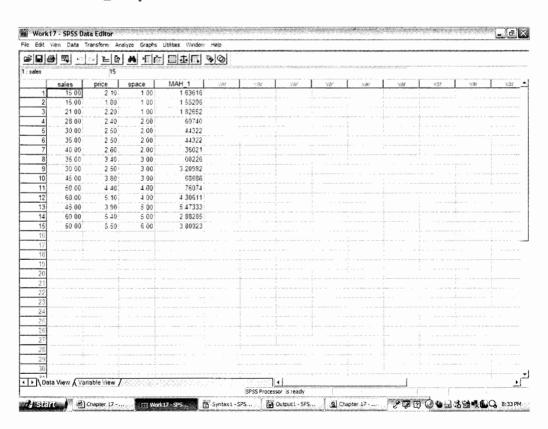






The normal plot of regression standardised residuals for the dependent variable also indicates a relatively normal distribution.

If you go back to your data file, you will also notice that SPSS has added another new variable MAH\_1 to your data file.



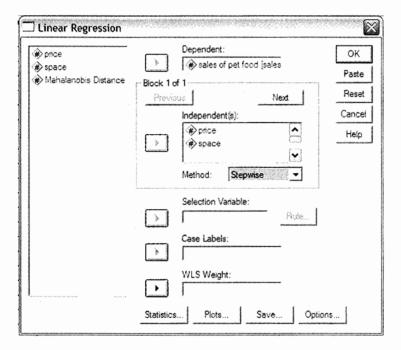
An examination of the Mahalanobis distance values indicates that there are no multivariate outliers among the independent variables; that is, no values are greater than or equal to the critical chi-square value of 13.8 at an alpha level of .001.

Therefore, in answer to research question 1, you can say that price significantly predicts sales of pet food — F(2,12) = 34.081, p < .05. However, the shelf space allocated is not a significant predictor.



To conduct a stepwise regression analysis

- 1 Select the **Analyze** menu.
- 2 Click on Regression and then on Linear... to open the Linear Regression dialogue box.
- 3 Select the dependent variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Dependent:** box.
- 4 Select the independent variables (i.e. *space* and *price*) and click on the ▶ button to move the variables into the **Independent(s)**: box.
- 5 In the **Method:** drop-down list, ensure **Stepwise** is selected.



- 6 Click on the Statistics... command pushbutton to open the Linear Regression: Statistics sub-dialogue box and ensure the Estimates and Model fit check boxes are selected.
- 7 Click on Continue and then OK.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT sales
/METHOD=STEPWISE price space
/SCATTERPLOT=(\*ZRESID ,\*ZPRED )
/RESIDUALS NORM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE MAHAL .

## Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	price		Stepwise (Criteria: Probability-of- F-to-enter <=.050, Probability-of- F-to-remove >=.100).

a. Dependent Variable: sales of pet food

## Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.922ª	.850	.839	5.82145

a. Predictors: (Constant), price

b. Dependent Variable: sales of pet food

## ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
[	Regression	2502.373	1	2502.373	73.840	.000ª
ı	Residual	440.561	13	33.889		
l	Total	2942.933	14			

a. Predictors: (Constant), price

b. Dependent Variable: sales of pet food

## Coefficientsa

			dardised cients	Standardised Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1 (Constant) price		1.977 10.566	4.373 1.230	.922	.452 8.593	.659 .000

a. Dependent Variable: sales of pet food

## Excluded Variables<sup>b</sup>

					Partial	Collinearity Statistics
Mode	l	Beta In	t	Sig.	Correlation	Tolerance
1	space	.006 <sup>a</sup>	.022	.983	.006	.154

a. Predictors in the Model: (Constant), price

b. Dependent Variable: sales of pet food

You will notice from the output that only the variable price has been entered into the regression equation, and this variable explains 85 per cent of the variability in sales of pet food, F(1,13) = 73.84, p < .05. The second independent variable, shelf space, failed to meet the selection criteria, as indicated by the nonsignificant t-value (p > .05).



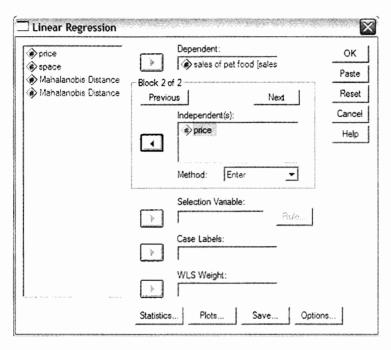
## To conduct a hierarchical regression analysis

Select the Analyze menu.

2 Click on Regression and then on Linear... to open the Linear Regression dialogue box.

Select the dependent variable (i.e. sales) and click on the  $\triangleright$  button to move the variable into the **Dependent:** box.

- 4 Select the independent variable you have chosen to enter first (i.e. space) and click on the ▶ button to move the variable into the Independent(s): box.
- 5 Click on Next.
- Select the next independent variable you have chosen to enter (i.e. *price*) and click on the button to move the variable into the **Independent(s)**: box. You will notice that the box above in your output reads **Block 2 of 2**.



- 7 Click on the **Statistics** command pushbutton to open the **Linear Regression:**Statistics sub-dialogue box and ensure the **Estimates**, **Model fit** and **R**squared change check boxes are selected.
- 8 Click on Continue and then OK.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT sales
/METHOD=STEPWISE space/METHOD=ENTER price
/SCATTERPLOT=(\*ZRESID ,\*ZPRED )
/RESIDUALS NORM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE MAHAL .

### Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1			Stepwise
			(Criteria:
			Probability-of-
			F-to-enter
	space		<=.050,
			Probability-of-
1			F-to-remove
			>=.100).
2	price <sup>a</sup>		Enter

- a. All requested variables entered.
- b. Dependent Variable: sales of pet food

#### Model Summary<sup>c</sup>

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1 2	.849 <sup>a</sup> .922 <sup>b</sup>	.721 .850	.700 .825	7.94653 6.05904	.721 .129	33.604 10.361	1	13 12	.000 .007

a. Predictors: (Constant), space

b. Predictors: (Constant), space, price

c. Dependent Variable: sales of pet food

You will notice that shelf space on its own contributes 72 per cent of the variance in sales of pet food and is a significant predictor. At the second step you will notice from the R Square Change statistic and the Sig. F Change value that price makes a significant unique contribution of 13 per cent to the variance of sales of pet food after shelf space.

#### **ANOVA<sup>c</sup>**

Mode	el	Sum of Squares	df	Mean Square	F	Sig.
1	Regression Residual Total	2122.017 820.916 2942.933	1 13 14	2122.017 63.147	33.604	.000ª
2	Regression Residual Total	2502.390 440.543 2942.933	2 12 14	1251.195 36.712	34.081	.000 <sup>b</sup>

a. Predictors: (Constant), space

b. Predictors: (Constant), space, price

c. Dependent Variable: Sales of pet food

## Coefficients<sup>a</sup>

			dardised cients	Standardised Coefficients		
Mode	ı	В	Std. Error	Beta	t	Sig.
1	(Constant) space	14.405 7.794	4.446 1.344	.849	3.240 5.797	.006 .000
2	(Constant) space price	2.029 .057 10.500	5.126 2.613 3.262	.006 .916	.396 .022 3.219	.699 .983 .007

a. Dependent Variable: sales of pet food

However, examination of the above table indicates that shelf space is no longer a significant predictor when both independent variables are entered into the regression equation. In other words, these variables must be significantly correlated such that price subsumes shelf space.

Excluded Variables<sup>b</sup>

					Partial	Collinearity Statistics
Mod	lel	Beta In	t	Sig.	Correlation	Tolerance
1	price	.916ª	3.219	.007	.681	.154

a. Predictors in the Model: (Constant), space

b. Dependent Variable: sales of pet food

Therefore, in answer to the third research question, you can say that shelf space on its own is a salient predictor of sales of pet food, F(1,13) = 33.604, p < .05. However, in combination with price, its effect is insignificant.

# Practice example



In a study concerning the relationship between the volume of wood provided by a forest area and various characteristics of the area, an experimental forest containing mixed soft-wood trees was divided into plots, of which 25 were selected at random. Measurements were made on each of the plots at the beginning of the study: initial wood volume, number of trees, average age of trees and average volume of trees. Five years later at the end of the study, the final wood volume was measured. Given the data in Prac17.sav, your tasks are to:

- 1 Assess the data for violation of assumptions.
- Test the hypothesis indicated by previous forestry research that initial wood volume is the best predictor of final wood volume, followed by the number of trees in the plot.

## **Solutions**



## **Syntax**

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT finalvol
/METHOD=ENTER initvol /METHOD=ENTER notrees /METHOD=ENTER avage /METHOD=ENTER volume
/SCATTERPLOT=(\*ZRESID ,\*ZPRED )
/RESIDUALS NORM(ZRESID)

/CASEWISE PLOT(ZRESID) OUTLIERS(3)

## Output

## **Assumption testing**

/SAVE MAHAL.

Having only 25 cases, there is barely the minimum requirement (which is at least five times more cases than independent variables). However, the analysis can proceed despite this limitation. The remainder of the assumptions can be tested through the regression analysis that follows.

#### Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	initial wood volumea		Enter
2	number of trees <sup>a</sup>		Enter
3	average age of trees <sup>a</sup>		Enter
4	average volume of trees <sup>a</sup>		Enter

a. All requested variables entered.

b. Dependent Variable: final wood volume

## Model Summary<sup>e</sup>

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.984 <sup>a</sup>	.968	.966	20.702	.968	692.839	1	23	
2	.984 <sup>b</sup>	.969	.966	20.864	.001	.645	1	22	
3	.986°	.973	.969	19.817	.004	3.385	1	21	
4	.987 <sup>d</sup>	.974	.969	19.989	.001	.640	1	20	

- a. Predictors: (Constant), initial wood volume
- b. Predictors: (Constant), initial wood volume, number of trees
- c. Predictors: (Constant), initial wood volume, number of trees, average age of trees
- d. Predictors: (Constant), initial wood volume, number of trees, average age of trees, average volume of trees
- e. Dependent Variable: final wood volume

The independent variable, initial volume, explains 96.8 per cent of the variance in final wood volume and is highly significant, as indicated by the significant F-value. An examination of the Model Summary table indicates that no other variable adds to the predictive power of the equation.

**ANOVA<sup>e</sup>** 

Mode	el	Sum of Squares	df	Mean Square	F	Sig.
1	Regression Residual Total	296935.0 9857.270 306792.2	1 23 24	296934.970 428.577	692.839	.000ª
2	Regression Residual Total	297215.9 9576.314 306792.2	2 22 24	148607.963 435.287	341.402	.000 <sup>b</sup>
3	Regression Residual Total	298545.3 8246.974 306792.2	3 21 24	99515.089 392.713	253.404	.000°
4	Regression Residual Total	298800.9 7991.365 306792.2	4 20 24	74700.219 399.568	186.952	.000 <sup>d</sup>

- a. Predictors: (Constant), initial wood volume
- b. Predictors: (Constant), initial wood volume, number of trees
- c. Predictors: (Constant), initial wood volume, number of trees, average age of trees
- d. Predictors: (Constant), initial wood volume, number of trees, average age of trees, average volume of trees
- e. Dependent Variable: final wood volume

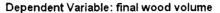
## Coefficients<sup>a</sup>

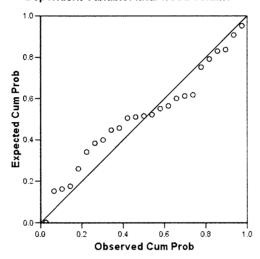
		Unstandardised Coefficients		Standardised Coefficients		
Mode	el	В	Std. Error	Beta	t	Sig.
1	(Constant) initial wood volume	29.441 .986	7.610 .037	.984	3.868 26.322	.001 .000
2	(Constant) initial wood volume number of trees	23.994 .951 .357	10.236 .057 .445	.949 .046	2.344 16.580 .803	.029 .000 .430
3	(Constant) initial wood volume number of trees average age of trees	44.817 1.048 .160 520	14.921 .076 .436 .283	1.046 .021 103	3.004 13.844 .367 -1.840	.007 .000 .717 .080
4	(Constant) initial wood volume number of trees average age of trees average volume of trees	27.535 .946 .725 504 3.000	26.332 .149 .832 .286 3.751	.944 .093 100	1.046 6.365 .872 -1.763 .800	.308 .000 .394 .093 .433

a. Dependent Variable: final wood volume

The beta values in the Coefficients table at step 4 confirm that initial wood volume is the best predictor of final wood volume. Having obtained this solution, it is necessary to check that assumptions have not been violated that could modify the interpretation of the output. No univariate outliers were found, so casewise plots were not necessary.

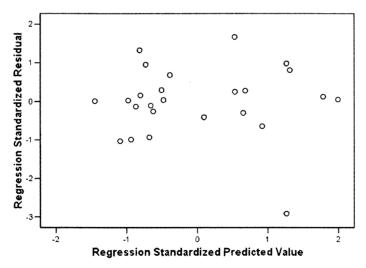
Normal P-P Plot of Regression Standardized Residual





Scatterplot

Dependent Variable: final wood volume



The normal plot of regression standardised residuals for the dependent variable, final wood volume, indicates a relatively normal distribution. From the scatterplot of residuals against predicted values it can be seen that there is no clear relationship consistent with the assumption of linearity.

The Mahalanobis distance values are displayed in the data file as a new variable. No distance is greater than the critical value of chi-square at an alpha level of .001 (that is, 18.47), so there are no multivariate outliers. The critical value for chi-square is obtained from a statistics text.

Having confirmed that no assumptions have been violated, it can be concluded that the above interpretation is sound and that initial volume is the best predictor of final wood volume, F(1,23) = 692.839, p < .05, as indicated by previous research.

# Multiple analysis of variance (MANOVA)

The extension of univariate analysis of variance to the involvement of multiple dependent variables is termed multivariate analysis of variance (MANOVA). The hypotheses tested with MANOVA are similar to those tested with ANOVA except that sets of means replace the individual means specified in ANOVA. In a one-way MANOVA design, for example, the hypothesis tested is that the populations from which the samples are selected have the same means for all dependent variables.

The MANOVA command also allows you to perform a stepdown analysis if you have theoretical reasons for ordering your dependent variables. The ordering of these dependent variables can be tested using prior dependent variables as covariates. The choice between univariate and stepdown F may be difficult. When there is little correlation among the dependent variables, univariate F is acceptable. However, when dependent variables are highly correlated, stepdown F is preferable.

Bartlett's test of sphericity may help you decide whether stepdown analysis is appropriate. If Bartlett's test is significant at an alpha level of .05, this indicates that the dependent variables are related and thus MANOVA with stepdown analysis should be conducted. If Bartlett's test is not significant, this indicates that the dependent variables are unrelated and that univariate ANOVAs, with adjustment for familywise error, would be more appropriate.

# Assumption testing



A number of assumptions underpin the use of MANOVA:

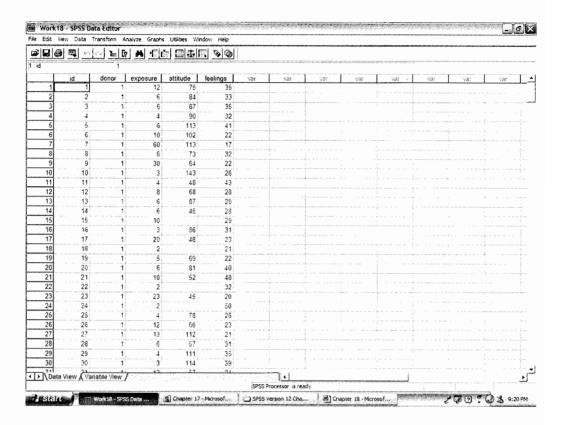
- Cell sizes it is necessary to have more subjects in each cell than the number of dependent variables. When cell size is greater than 30, assumptions of normality and equal variances are of little concern. If cell sizes are small and unequal, then assumption testing becomes more critical. Although equal cell size is ideal, it is not essential. However, ratios of smallest to largest size greater than 1:1.5 may cause problems.
- Univariate and multivariate normality MANOVA is sensitive to violations of univariate and multivariate normality. To test for univariate normality, the techniques described in chapter 3 can be used for each group or level of the independent variable using the Split File option. Multivariate outliers which influence normality can be identified using Mahalanobis distance in the Regression sub-menu. Mahalanobis distance is evaluated as chi-square with degrees of freedom equal to the number of dependent variables. The critical chi-square levels can be found in any critical values of chi-square table. An alpha level of .001 is recommended.
- 3 Linearity linear relationships among all pairs of dependent variables must be assumed. Within-cell scatterplots must be conducted to test this assumption.
- 4 Homogeneity of regression this assumption is related to stepdown analysis and must be tested if stepdown analysis is required. It is assumed that the relationship between covariates and dependent variables in one group is the same as the relationship in other groups.
- Homogeneity of variance—covariance matrices this assumption is similar to the assumption of homogeneity of variance for individual dependent variables. In multivariate designs, this assumption is more complex. At the univariate level, Cochrans C and Bartlett-Box F tests can be used. If these tests are not significant (p > .05), then homogeneity of variance is assumed. Multivariate homogeneity of the variance—covariance matrices is tested using Box's M test, which must also be nonsignificant (p > .001). This test is very sensitive so the alpha level of .001 is recommended.
- **Multicollinearity and singularity** when correlations among dependent variables are high, problems of multicollinearity and singularity exist. When the determinant of the within-cell correlation matrix is near zero (< .0001) or when the log (determinant) is less than –9.21034, singularity or multicollinearity may be present.

# Working example



A social scientist wished to compare those respondents who had lodged an organ donor card with those who had not. Three-hundred-and-eighty-eight new drivers completed a questionnaire that measured their attitudes towards organ donation, their feelings about organ donation and their previous exposure to the issue. It was hypothesised that individuals who agreed to be donors would have more positive attitudes towards organ donation, more positive feelings towards organ donation and greater previous exposure to the issue. Therefore, the independent variable was whether a donor card had been signed, and the dependent variables were attitudes towards organ donation, feelings towards organ donation and previous exposure to organ donation. Attitudes and feelings were measured on traditional scales with a Likert scale response format. Exposure was measured in terms of media exposure and personal experience. Conceptually and theoretically these dependent variables were believed to be related and thus MANOVA was the analysis of choice.

The data file can be found in Work18.sav on the website that accompanies this title and is shown in the following figure.



# Data screening

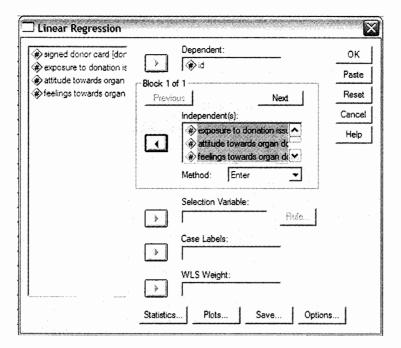


- Cell sizes equality of cell size is often regulated by the type of design chosen. When using experimental designs, for example, the researcher is able to assign a set number of subjects to each condition. However, in quasiexperimental designs, equal n across cells is outside of the researcher's control. In this research example, cell sizes were approximately equal.
- 2 (a) *Univariate normality* the data for each dependent measure by group were screened using the **Explore** option described in chapter 3. Assumptions of univariate normality were not violated.
  - (b) *Multivariate outliers* these were assessed using Mahalanobis distance in the **Regression** sub-menu. This is achieved by using the identification number in the data file as the dependent variable, in a simultaneous regression model, with all other variables as independent variables.

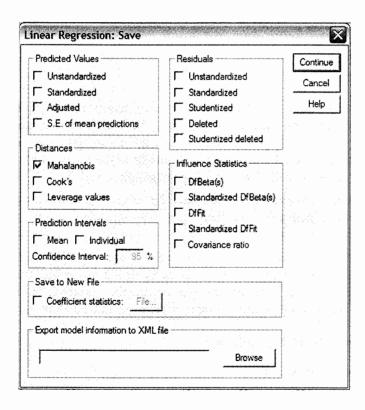


## To detect multivariate outliers

- 1 Select the **Analyze** menu.
- 2 Click on **Regression** and then **Linear...** to open the **Linear Regression** dialogue box.
- Select the dependent variable (i.e. id) and click on the  $\triangleright$  button to move the variable into the **Dependent:** box.
- 4 Select the independent variables (i.e. *attitude*, *exposure*, *feelings*) and click on the **▶** button to move the variables into the **Independent(s)**: box.



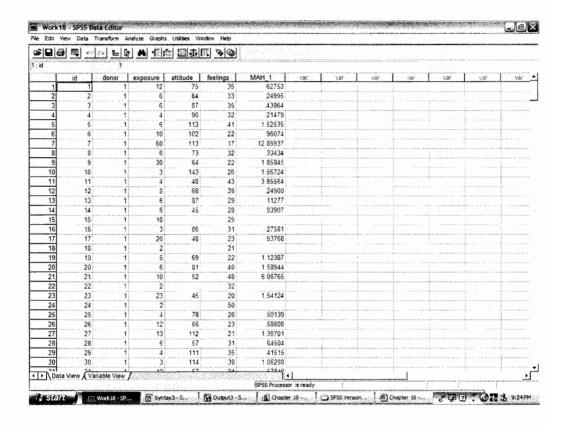
- 5 Click on the **Save...** command pushbutton to open the **Linear Regression: Save** sub-dialogue box.
- 6 In the **Distances** box, select the **Mahalanobis** check box.



7 Click on Continue and then OK.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT id
/METHOD=ENTER exposure attitude feelings
/SAVE MAHAL.

You will notice that these distances have been added to your data file.



The critical value of chi-square, for three dependent variables, at an alpha level of .001 is 16.2. Using this value you have seven outlying cases, which is not unexpected in a sample size of 388. Consequently, these outliers were retained in the data set. If a large number of outlying cases were identified, then their inclusion would need to be considered carefully.

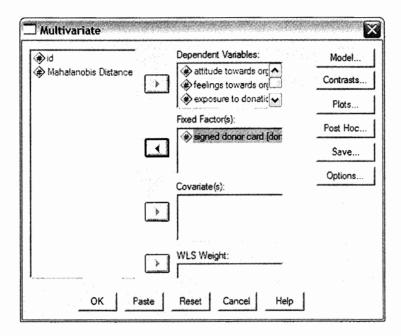
- 3 Linearity linearity among dependent measures was confirmed using scatterplots among pairs of dependent variables across groups.
- 4 Homogeneity of regression this assumption is tested using the general factorial ANOVA model sub-menu, and it involves a fairly arduous process of modifying syntax files. Because you have no theoretical grounds for ordering the dependent variables, stepdown analysis will not be carried out in this example.

Assumptions 5 and 6 will be tested in the MANOVA analysis.

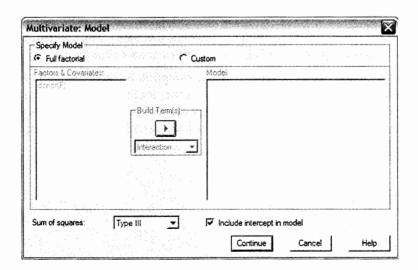


## To conduct a MANOVA

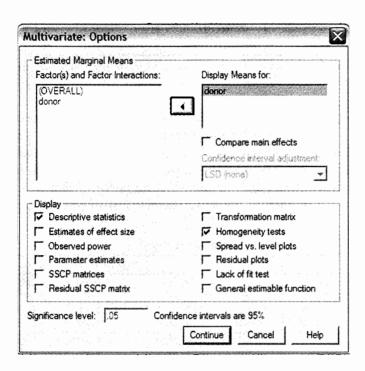
- 1 Select the Analyze menu.
- 2 Click on General Linear Model and then Multivariate... to open the Multivariate dialogue box.
- 3 Select the dependent variables (i.e. *attitudes*, *feelings*, *exposure*) and click on the button to move the variables into the **Dependent Variables:** box.
- 4 Select the independent variable(s) (i.e. *donor*) and click on the ▶ button to move the variable into the **Fixed Factor(s)**: box.



- 5 Click on the **Model...** command pushbutton to open the **Multivariate: Model** sub-dialogue box.
- In the **Specify Model** box, ensure the **Full factorial** radio button is selected and **Type III** is selected from the **Sum of squares:** drop-down list. The full factorial model contains all main effects, covariate main effects and all factor-by-factor interactions, but does not contain covariate interactions. Type III is the default and the most commonly used method of calculating sum of squares for balanced/ unbalanced models with no missing cells.



- 7 Click on Continue.
- 8 Click on the **Options...** command pushbutton to open the **Multivariate Options** sub-dialogue box.
- 9 In the Estimated Marginal Means box, under the heading Factor(s) and Factor Interactions:, click on the independent variable (i.e. donor) and click on the ▶ button to move the variable into the Display Means for: box.
- 10 In the **Display** box, select the **Descriptive statistics** and the **Homogeneity tests** check boxes.



## 11 Click on Continue and then OK.

GLM
attitude feelings exposure BY donor
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/EMMEANS = TABLES(donor)
/PRINT = DESCRIPTIVE HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = donor.

## Box's Test of Equality of Covariance Matrices<sup>a</sup>

Box's M	19.260
F	3.182
df1	6
df2	1018790
Sig.	.004

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+donor

Box's M tests the homogeneity of the variance—covariance matrices. You have homogeneity of variance because this test is not significant at an alpha level of .001.

The univariate tests for homogeneity of variance for each of the dependent measures indicate that homogeneity of variance has not been violated for exposure to organ donation issues and feelings towards organ donation. However, for attitude towards organ donation, the Levene's test of equality of error variances is significant. If the univariate F-test for this variable is also significant, then you must interpret this finding at a more conservative alpha level.

Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
attitude towards organ donation	15.346	1	375	.000
feelings towards organ donation	1.284	1	375	.258
exposure to donation issues	2.936	1	375	.087

Tests the null hypothesis that the error variance of the dependent variable is equal across groups. a. design: Intercept+donor

The multivariate tests of significance test whether there are significant group differences on a linear combination of the dependent variables. You will notice there are a number of multivariate statistics available. Pillai's Trace criterion is considered to have acceptable power and to be the most robust statistic against violations of assumptions. Having obtained a significant multivariate effect for donor — that is, a significance of F < .05 — you can now interpret the univariate/between-subject effects given below.

Multivariate Testsb

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace Wilks' Lambda Hotelling's Trace	.935 .065 14.402	1790.688 <sup>a</sup> 1790.688 <sup>a</sup> 1790.688 <sup>a</sup>	3.000 3.000 3.000	373.000 373.000 373.000	.000 .000 .000
	Roy's Largest Root	14.402	1790.688 <sup>a</sup>	3.000	373.000	.000
donor	Pillai's Trace Wilks' Lambda Hotelling's Trace Roy's Largest Root	.033 .967 .034 .034	4.255 <sup>a</sup> 4.255 <sup>a</sup> 4.255 <sup>a</sup> 4.255 <sup>a</sup>	3.000 3.000 3.000 3.000	373.000 373.000 373.000 373.000	.006 .006 .006

a. Exact statistic

An examination of the univariate F-tests for each dependent variable indicates which individual dependent variables contribute to the significant multivariate effect. It is advisable to evaluate these effects using a Bonferroni-type adjustment. Adjusting for familywise or experimentwise error decreases the chance of type I error. The simple formula that may be applied is: alpha / number of tests. In this example the adjusted alpha is equal to .017 (0.05/3). Using this alpha level, you have a significant univariate main effect for feelings towards organ donation (p < .017).

b. Design: Intercept+donor

Tests of Between-Subject Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	attitude towards organ donation feelings towards organ donation exposure to donation issues	12372.705 <sup>a</sup> 922.187 <sup>b</sup> 903.925 <sup>c</sup>	1 1 1	12372.705 922.187 903.925	3.960 11.100 3.830	.047 .001 .051
Intercept	attitude towards organ donation feelings towards organ donation exposure to donation issues	3105144.376 331042.346 39489.506	1 1 1	3105144.376 331042.346 39489.506	993.910 3984.772 167.331	.000 .000 .000
donor	attitude towards organ donation feelings towards organ donation exposure to donation issues	12372.705 922.187 903.925	1 1 1	12372.705 922.187 903.925	3.960 11.100 3.830	.047 .001 .051
Error	attitude towards organ donation feelings towards organ donation exposure to donation issues	1171563.820 31153.824 88498.590	375 375 375	3124.170 83.077 235.996		
Total	attitude towards organ donation feelings towards organ donation exposure to donation issues	4288063.000 363028.000 128924.000	377 377 377			
Corrected Total	attitude towards organ donation feelings towards organ donation exposure to donation issues	1183936.525 32076.011 89402.515	376 376 376			

a. R Squared = .010 (Adjusted R Squared = .008)

Therefore, you can conclude that a person's decision to act as a donor is significantly influenced by their feelings towards organ donation, F(1,375) = 11.100, p < .017. No significant main effects were found for the other dependent measures (exposure to or attitude towards organ donation).

An examination of the estimated marginal means for feelings towards organ donation and the variable donor indicates that those individuals who had signed a donor card had more positive feelings towards organ donation than those individuals who had not signed a donor card.

signed donor card

				95% Confidence Interval	
Dependent Variable	signed donor card	Mean	Std. Error	Lower Bound	Upper Bound
attitude towards organ donation	yes	85.026	4.066	77.032	93.021
	no	96.484	4.077	88.468	104.500
feelings towards organ donation	yes	28.069	.663	26.765	29.372
	no	31.197	.665	29.890	32.504
exposure to donation issues	yes	11.783	1.117	9.586	13.980
	no	8.686	1.120	6.483	10.889

When independent variables involve more than two levels, further analytic comparisons are needed to find out which levels are significantly different on the individual dependent variable. These comparisons can be planned (when you have prior reason for anticipating specific or planned differences) or post hoc (when you have no prior reason for anticipating differences but wish to explore the data further).

More than one independent variable and the presence of an interaction effect complicate the process of making comparisons and are beyond the scope of this chapter. However, the new general linear model procedure does allow you to perform contrasts using the **Contrasts...** and **Post Hoc...** command pushbuttons in the **General Factorial** dialogue box.

b. R Squared = .029 (Adjusted R Squared = .026)

c. R Squared = .010 (Adjusted R Squared = .007)

It is important to remember that MANOVA is an intricate analysis and is more straightforward when there is only one independent variable and only a few dependent variables. As the number of independent and dependent variables increase, the analysis becomes more complex. When the dependent variables are measured on the same scale, a special form of MANOVA (referred to as profile analysis) is more appropriate.

# Practice example



Three hundred and sixty-four male and female subjects completed a personality inventory comprising a self-esteem, optimism and hope scale. The researcher is aware that these sub-scales are highly correlated. Given the data in Prac18.sav, your tasks are to:

- 1 Check for violations of assumptions.
- 2 Determine whether gender differences exist across the combination of scales.

## **Solutions**



## Syntax

```
EXAMINE
 VARIABLES=hope esteem optimism BY gender
 /PLOT BOXPLOT STEMLEAF
 /COMPARE GROUP
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
COMPUTE id = $casenum.
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT id
 /METHOD=ENTER esteem hope optimism
 /SAVE MAHAL.
GRAPH
 /SCATTERPLOT(BIVAR)=hope WITH esteem
 /MISSING=LISTWISE.
GRAPH
 /SCATTERPLOT(BIVAR)=hope WITH optimism
 /MISSING=LISTWISE.
 /SCATTERPLOT(BIVAR)=esteem WITH optimism
 /MISSING=LISTWISE.
CORRELATIONS
 /VARIABLES=esteem hope optimism
 /PRINT=TWOTAIL NOSIG
 /MISSING=PAIRWISE.
 esteem hope optimism BY gender
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /EMMEANS = TABLES(gender)
 /PRINT = DESCRIPTIVE HOMOGENEITY
 /CRITERIA = ALPHA(.05)
 /DESIGN = gender.
```

## Output

The data for each dependent measure by group (gender) were screened using the **Explore** option and by selecting the **Outliers** check box.

An examination of the skewness and kurtosis statistics indicated that the distributions across groups were relatively normal. You will notice that the **Outliers** option provides a list of cases that may be extreme in the upper and lower boundaries of the distribution. Due to the size of the current sample a few outlying cases would be expected, and because the skewness and kurtosis statistics are normal, there is no need to bring outliers into the distribution.

## **Descriptives**

	gender			Statistic	Std. Error
hope	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	29.5598 28.9868 30.1327 29.5966 30.0000 15.516 3.93898 17.00 39.00 22.00 5.00 213 161	.29039 .179 .356
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	28.6444 28.0117 29.2772 28.7593 29.0000 18.510 4.30230 14.00 40.00 26.00 6.00 450 .397	.32067 .181 .360
esteem	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	31.6467 31.2210 32.0725 31.6316 32.0000 8.569 2.92720 22.00 39.00 17.00 3.00 028	.21580 .179 .356

(continued)

## Descriptives (continued)

	gender			Statistic	Std. Error
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	31.0500 30.5786 31.5214 30.9753 31.0000 10.271 3.20488 24.00 41.00 17.00 4.00 .332 060	.23888 .181 .360
optimism	female	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	25.4728 25.0834 25.8623 25.4698 25.0000 7.169 2.67743 18.00 33.00 15.00 3.00 .004	.19738 .179 .356
	male	Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median Variance Std. Deviation Minimum Maximum Range Interquartile Range Skewness Kurtosis	Lower Bound Upper Bound	25.3722 24.8968 25.8476 25.3704 25.0000 10.447 3.23222 17.00 35.00 18.00 4.75 028 033	.24092 .181 .360

## Extreme Values

	gender			Case Number	Value
hope	female	Highest	1	108	39.00
		_	2	35	38.00
			3	58	38.00
			4	10	37.00
			5	50	37.00 <sup>a</sup>
		Lowest	1	144	17.00
			2	34	20.00
			3	328	21.00
			4	120	21.00
			5	362	22.00 <sup>b</sup>
	male	Highest	1	175	40.00
l .		_	2	13	37.00
			3	17	37.00
			4	11	36.00
			5	149	36.00 <sup>c</sup>

(continued)

## Extreme Values (continued)

	gender			Case Number	Value
		Lowest	1 2 3 4 5	143 227 295 107 251	14.00 17.00 18.00 18.00 20.00 <sup>d</sup>
esteem	female	Highest	1 2 3 4 5	1 74 86 116 9	39.00 39.00 39.00 38.00 37.00 <sup>a</sup>
		Lowest	1 2 3 4 5	280 328 316 2 320	22.00 23.00 26.00 26.00 27.00 <sup>e</sup>
	male	Highest	1 2 3 4 5	93 141 233 355 249	41.00 39.00 39.00 39.00 38.00
		Lowest	1 2 3 4 5	341 99 289 285 179	24.00 24.00 25.00 25.00 25.00
optimism	female	Highest	1 2 3 4 5	280 222 226 14 35	33.00 32.00 32.00 31.00 31.00
		Lowest	1 2 3 4 5	360 362 344 30 342	18.00 19.00 19.00 19.00 20.00 <sup>d</sup>
	male	Highest	1 2 3 4 5	29 51 69 71 119	35.00 32.00 32.00 32.00 32.00 <sup>g</sup>
		Lowest	1 2 3 4 5	325 145 67 336 191	17.00 17.00 18.00 19.00 19.00 <sup>h</sup>

a. Only a partial list of cases with the value 37.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 22.00 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 36.00 are shown in the table of upper extremes.

d. Only a partial list of cases with the value 20.00 are shown in the table of lower extremes.

e. Only a partial list of cases with the value 27.00 are shown in the table of lower extremes.

f. Only a partial list of cases with the value 31.00 are shown in the table of upper extremes.

g. Only a partial list of cases with the value 32.00 are shown in the table of upper extremes. h. Only a partial list of cases with the value 19.00 are shown in the table of lower extremes.

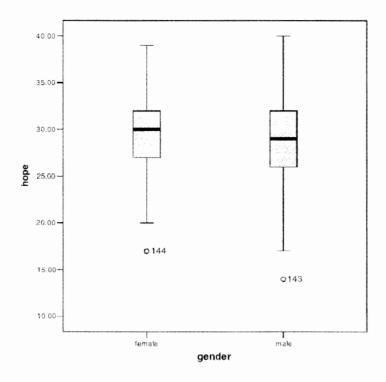
# Hope

# hope Stem-and-Leaf Plot for gender = female

Frequency	Ster	m & Leaf
1.00	Extremes	(=<17.0)
1.00	20 .	0
2.00	21 .	00
2.00	22 .	00
6.00	23 .	000000
5.00	24 .	00000
14.00	25 .	000000000000
14.00	26 .	000000000000
14.00	27 .	000000000000
11.00	28 .	0000000000
15.00	29 .	0000000000000
18.00	30 .	0000000000000000
13.00	31 .	00000000000
29.00	32 .	000000000000000000000000000000000000000
12.00	33 .	0000000000
10.00	34 .	000000000
6.00	35 .	000000
5.00	36 .	00000
3.00	37 .	000
2.00	38 .	00
1.00	39 .	0
Stem width:		1.00
Each leaf:		one case

# hope Stem-and-Leaf Plot for gender = male

Frequency	Ster	m & Leaf
2.00	Extremes	(=<17.0)
.00	1.	
2.00	1.	88
8.00	2.	00011111
12.00	2.	222233333333
13.00	2 .	444555555555
29.00	2.	6666666666667777777777777
34.00	2.	888888888899999999999999999999999
34.00	3 .	00000000000000001111111111111111
25.00	3 .	2222222222233333333333
14.00	3.	4444444555555
6.00	3.	666677
.00	3.	
1.00	4 .	0
Stem width:		10.00
Each leaf:		one case



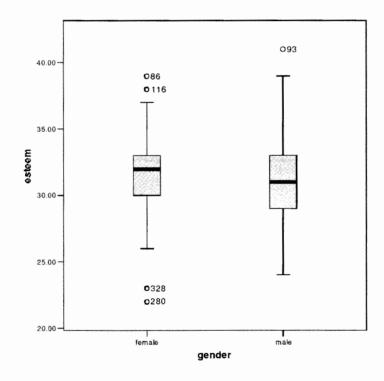
## **Esteem**

# esteem Stem-and-Leaf Plot for gender = female

Frequency	Ster	m & Leaf
2.00	Extremes	(=<23.0)
2.00	26 .	00
11.00	27 .	0000000000
9.00	28 .	00000000
18.00	29 .	00000000000000000
19.00	30 .	00000000000000000
29.00	31 .	000000000000000000000000000000000000000
25.00	32 .	00000000000000000000000
26.00	33 .	000000000000000000000000
12.00	34 .	0000000000
13.00	35 .	00000000000
8.00	36 .	0000000
6.00	37 .	000000
4.00	Extremes	(>=38.0)
Stem width:		1.00
Each leaf:		one case

# esteem Stem-and-Leaf Plot for gender = male

Frequency	Ster	m & Leaf
2.00	24 .	00
3.00	25 .	000
4.00	26 .	0000
16.00	27 .	00000000000000
16.00	28 .	00000000000000
24.00	29 .	000000000000000000000000000000000000000
16.00	30 .	00000000000000
22.00	31 .	000000000000000000000000000000000000000
14.00	32 .	000000000000
23.00	33 .	000000000000000000000000000000000000000
19.00	34 .	000000000000000000
7.00	35 .	000000
3.00	36 .	000
6.00	37 .	000000
1.00	38 .	0
4.00	Extremes	(>=39.0)
Stem width:		1.00
Each leaf:		one case



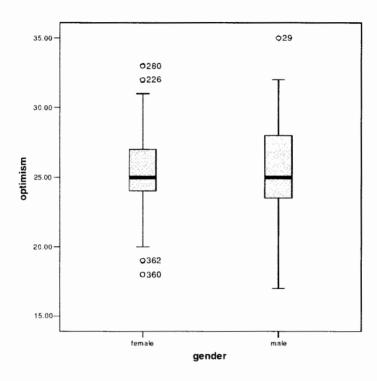
## **Optimism**

# optimism Stem-and-Leaf Plot for gender = female

Frequency	Ster	m & Leaf
4.00	Extremes	(=<19.0)
2.00	20 .	00
8.00	21 .	0000000
7.00	22 .	000000
17.00	23 .	000000000000000
26.00	24 .	0000000000000000000000000000000
30.00	25 .	000000000000000000000000000000000000000
28.00	26 .	000000000000000000000000000000000000000
25.00	27 .	00000000000000000000000000000
14.00	28 .	000000000000
11.00	29 .	0000000000
5.00	30 .	00000
4.00	31 .	0000
3.00	Extremes	(>=32.0)
Stem width:		1.00
Each leaf:		one case

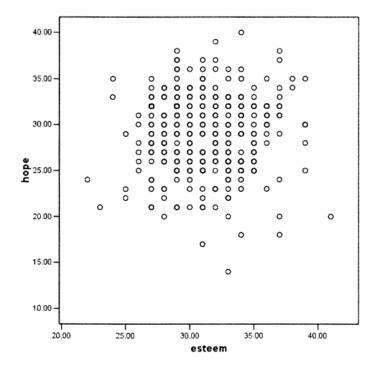
# esteem Stem-and-Leaf Plot for gender = male

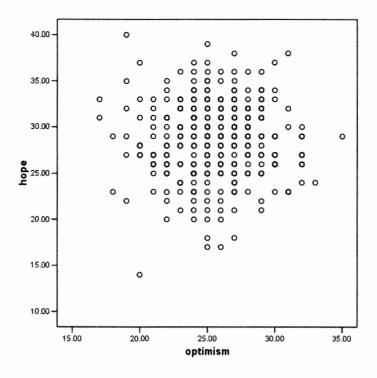
Frequency	Ster	m & Leaf
2.00	17.	00
1.00	18 .	0
3.00	19 .	000
7.00	20 .	000000
9.00	21 .	00000000
14.00	22 .	000000000000
9.00	23 .	00000000
21.00	24 .	00000000000000000000
31.00	25 .	000000000000000000000000000000000000000
18.00	26 .	00000000000000000
19.00	27 .	000000000000000000
15.00	28 .	0000000000000
14.00	29 .	000000000000
9.00	30 .	00000000
1.00	31 .	0
6.00	32 .	000000
1.00	Extremes	(>=35.0)
Stem width:		1.00
Each leaf:		one case

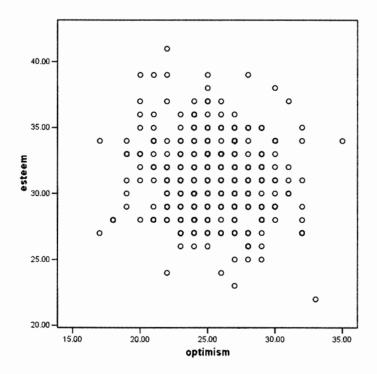


Multivariate outliers were tested for using the regression analysis option, requesting Mahalanobis distance. To create a suitable dependent variable for this analysis, a variable called *id* was created using the **Compute** command (see syntax file).

Given that there are three independent variables in this analysis, the critical chi-square value at p < .001 was 16.2 — the same as in the working example. An examination of the newly created variable in the data file reveals that there are no cases that exceed this critical value and thus all cases will be retained for further analysis. Although the sample is outlier free, remember that, as noted earlier in the chapter, it is not unusual to have some outlying cases, particularly with large sample sizes.







In testing the assumption of linearity it is obvious from the scatterplots and the correlation matrix that our dependent variables are not significantly related. The only correlation that is significant is that between esteem and optimism (p < .05). However, the MANOVA analysis can proceed because there is a theoretical relationship between these variables.

#### Correlations

		hope	esteem	optimism
hope	Pearson Correlation Sig. (2-tailed) N	1 364	.013 .808 364	017 .754 364
esteem	Pearson Correlation Sig. (2-tailed) N	.013 .808 364	1 364	116* .026 364
optimism	Pearson Correlation Sig. (2-tailed) N	017 .754 364	116* .026 364	364

<sup>\*.</sup> Correlation is significant at the 0.05 level (2-tailed).

The Box's M test for homogeneity of variance—covariance matrices indicates that this assumption has not been violated at an alpha level of .001.

#### Box's Test of Equality of Covariance Matrices<sup>a</sup>

	T
Box's M	12.859
F	2.124
df1	6
df2	948209.9
Sig.	.047

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

The Levene test of equality of error variances, for each of the dependent measures, indicates that hope, esteem and optimism have not violated this assumption at an alpha level of .01. However, at the .05 level, optimism does violate this assumption.

Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
hope	.319	1	362	.572
esteem	2.330	1	362	.128
optimism	5.605	1	362	.018

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

## Multivariate Tests<sup>b</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.996	30610.736a	3.000	360.000	.000
l	Wilks' Lambda	.004	30610.736 <sup>a</sup>	3.000	360.000	.000
	Hotelling's Trace	255.089	30610.736a	3.000	360.000	.000
	Roy's Largest Root	255.089	30610.736 <sup>a</sup>	3.000	360.000	.000
donor	Pillai's Trace	.022	2.737 <sup>a</sup>	3.000	360.000	.043
	Wilks' Lambda	.978	2.737 <sup>a</sup>	3.000	360.000	.043
	Hotelling's Trace	.023	2.737 <sup>a</sup>	3.000	360.000	.043
	Roy's Largest Root	.023	2.737 <sup>a</sup>	3.000	360.000	.043

a. Exact statistic

The multivariate tests of significance reveal that there are significant differences in the dependent variables across gender.

Examination of our univariate/between-subjects tests for each dependent variable indicates that hope is the only variable that differs significantly across gender when significance is measured at an alpha level of .05, F(1,362) = 4.485, p < .05.

a. Design: Intercept+gender

a. Design: Intercept+gender

b. Design: Intercept+gender

## Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	hope	.76.235ª	1	76.235	4.485	.035
	esteem	32.401 <sup>b</sup>	1	32.401	3.443	.064
	optimism	.921°	1	.921	.105	.746
Intercept	hope	308246.388	1	308246.388	18136.305	.000
	esteem	357666.983	1	357666.983	38007.369	.000
	optimism	235226.514	1	235226.514	26761.156	.000
donor	hope	76.235	1	76.235	4.485	.035
	esteem	32.401	1	32.401	3.443	.064
	optimism	.921	1	.921	.105	.746
Error	hope	6152.587	362	16.996		
	esteem	3406.588	362	9.410		
	optimism	3181.925	362	8.790		
Total	hope	314619.000	364			
	esteem	361224.000	364			
	optimism	238448.000	364			
Corrected Total	hope	6228.821	363			
	esteem	3438.989	363			
	optimism	3182.846	363			

a. R Squared = .012 (Adjusted R Squared = .010)

An examination of the means for the variable hope reveals that females have significantly higher hope scores (mean of 29.56) than those of males (28.64).

## gender

				95% Confidence Interval	
Dependent Variable	gender	Mean	Std. Error	Lower Bound	Upper Bound
hope	female	29.560	.304	28.962	30.157
	male	28.644	.307	28.040	29.249
esteem	female	31.647	.226	31.202	32.091
	male	30.151	.229	30.600	31.500
optimism	female	25.473	.219	25.043	25.903
	male	25.372	.221	24.938	25.807

b. R Squared = .009 (Adjusted R Squared = .007)

c. R Squared = .000 (Adjusted R Squared = .002)



# Nonparametric techniques

When you have serious violations of the distribution assumptions of parametric tests, then there are alternative nonparametric techniques that can be used. These tests tend to be less powerful than their parametric counterparts; however, some nonparametric tests are appropriate for data measured on scales that are not interval or ratio.

SPSS has a wide selection of nonparametric techniques available, of which the following will be discussed:

- · Chi-square test for goodness of fit
- · Chi-square test for independence or relatedness
- Mann-Whitney test (Wilcoxon rank sum)
- Wilcoxon signed-rank test
- · Kruskal-Wallis test
- · Friedman test
- · Spearman's rank order correlation.

Assumption testing for nonparametric techniques is not as critical as for parametric methods. However, a number of generic assumptions apply:

- 1 Random sampling
- 2 Similar shape and variability across distributions
- 3 Independence for between-subjects designs, independence (that is, subjects appear in only one group and the groups are not related in any way) must be ensured.

## Chi-square tests



There are two main types of chi-square test. The chi-square test for goodness of fit applies to the analysis of a single categorical variable, and the chi-square test for independence or relatedness applies to the analysis of the relationship between two categorical variables.

## **Assumption testing**

There are three assumptions you need to address before conducting chi-square tests.

- **Random sampling** observations should be randomly sampled from the population of all possible observations.
- 2 Independence of observations each observation should be generated by a different subject and no subject is counted twice.

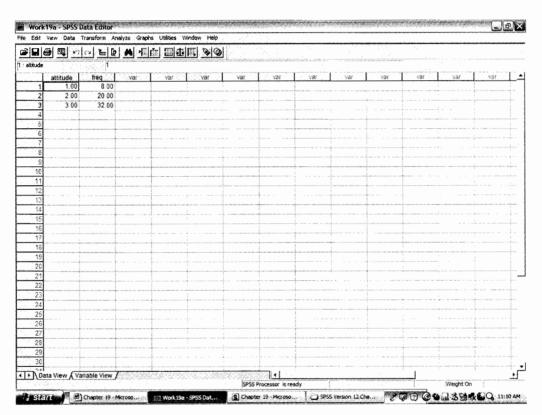
3 Size of expected frequencies — when the number of cells is less than ten and particularly when the total sample size is small, the lowest expected frequency required for a chi-square test is five. However, the observed frequencies can be any value, including zero.

## Working example — chi-square test for goodness of fit

The following table outlines the attitudes of sixty people towards US military bases in Australia. A chi-square test for goodness of fit will allow us to determine if differences in frequency exist across response categories.

Attitude towards US military bases in Australia	Frequency of response
In favour	8
Against	20
Undecided	32

The data file can be found in Work19a.sav on the website that accompanies this title.

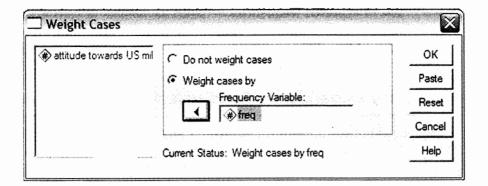


Initially, you have to tell the program that the data for frequency are in the form of frequency counts not scores. Using the **Weight Cases** option, you can do this quickly and easily and then you can perform the chi-square test.

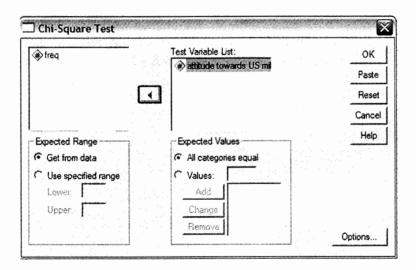


## To conduct a chi-square test for goodness of fit

- 1 Select the **Data** menu.
- 2 Click on Weight Cases... to open the Weight Cases dialogue box.
- 3 Click on the Weight cases by radio button.
- 4 Select the variable you require (i.e. *freq*) and click on the ▶ button to move the variable into the **Frequency Variable:** box.



- 5 Click on **OK**. The message **Weight On** should appear on the status bar at the bottom right of the application window.
- 6 Select the Analyze menu.
- 7 Click on Nonparametric Tests and then Chi-Square... to open the Chi-Square Test dialogue box.
- 8 Select the variable you require (i.e. *attitude*) and click on the ▶ button to move the variable into the **Test Variable List:** box.



9 Click on OK.

WEIGHT
BY freq.
NPAR TEST
/CHISQUARE=attitude
/EXPECTED=EQUAL
/MISSING ANALYSIS.

#### attitude towards US military bases

	Observed N	Expected N	Residual
in favour	8	20.0	-12.0
against	20	20.0	.0
undecided	32	20.0	12.0
Total	60		

#### **Test Statistics**

	attitude towards US military bases
Chi-Square <sup>a</sup>	14.400
df	2
Asymp. Sig.	.001

a. 0 cells (.0%) have expected frequencies less than five. The minimum expected cell frequency is 20.0.

You can see from the output that the chi-square value is significant (p < .05).

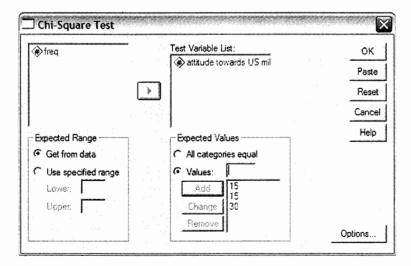
Thus it can be concluded that there are significant differences in the frequency of attitudes towards military bases in Australia, and the results show that people are largely undecided on this issue,  $X^2(2, \underline{N}=60) = 14.4$ , p < .05.

It should be noted that the expected frequencies in the above example represent a 1/3:1/3:1/3 split. Sometimes the expected frequencies are not evenly balanced across categories. Say the expected frequency for each category was 15, 15 and 30, for example. SPSS allows you to specify expected values that may not be equivalent. This technique will be demonstrated using already weighted cases.



To conduct a chi-square test for goodness of fit with unequal expected frequencies

- 1 Select the **Analyze** menu.
- 2 Click on Nonparametric Tests and then Chi-Square... to open the Chi-Square Test dialogue box.
- 3 Select the variable you require (i.e. *attitude*) and click on the ▶ button to move the variable into the **Test Variable List:** box.
- 4 In the Expected Values box, click the Values: radio button.
- 5 Type 15 in the box and click on Add.
- **6** Type another 15 in the box and click on **Add**.
- 7 Type 30 in the box and click on Add.



8 Click on **OK**.

NPAR TEST
/CHISQUARE=attitude
/EXPECTED=15 15 30
/MISSING ANALYSIS.

#### attitude towards US military bases

	Observed N	Expected N	Residual
in favour	8	15.0	-7.0
against	20	15.0	5.0
undecided	32	30.0	2.0
Total	60		

#### **Test Statistics**

	attitude towards US military bases
Chi-Square <sup>a</sup>	5.067
df	2
Asymp. Sig.	.079

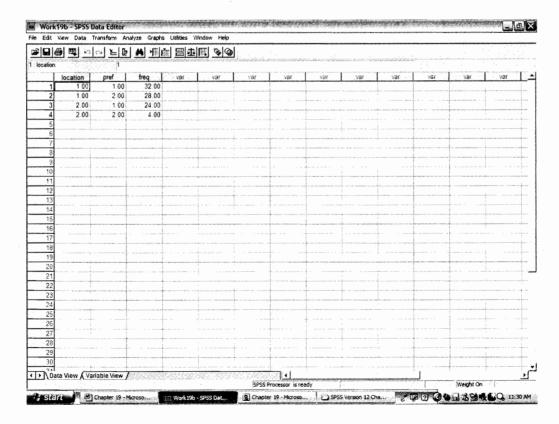
 a. 0 cells (.0%) have expected frequencies less than five. The minimum expected cell frequency is 15.0.

You will notice that with the different cell frequencies, the chi-square statistic is no longer significant (p > .05).

# Working example — chi-square test for relatedness or independence

A magazine publisher wished to determine whether preference for certain publications depended on geographic location of the reader. Of those preferring the *Financial Review*, 32 were urban readers and 24 were from the country. Of those who chose *Newsweek*, 28 were city readers while four were rural readers.

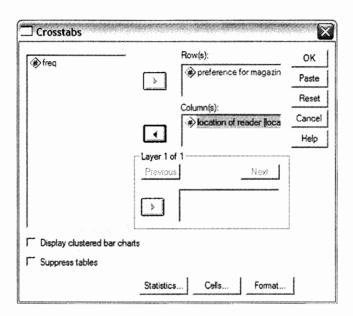
The data file can be found in Work19b.sav.



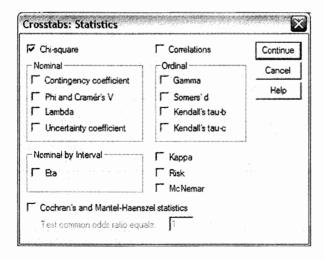


#### To conduct chi-square test for relatedness or independence

- 1 Select the **Analyze** menu.
- 2 Click on Descriptive Statistics and then on Crosstabs... to open the Crosstabs dialogue box.
- 3 Select a row variable (i.e. *pref*) and click on the ▶ button to move the variable into the **Row**(s): box.
- 4 Select a column variable (i.e. *location*) and click on the ▶ button to move the variable into the **Column(s)**: box.



- 5 Click on the **Statistics...** command pushbutton to open the **Crosstabs: Statistics** sub-dialogue box.
- 6 Click on the Chi-square check box.



- 7 Click on Continue.
- 8 Click on the Cells... command pushbutton to open the Crosstabs: Cell Display sub-dialogue box.
- 9 In the Counts box, click on the Observed and Expected check boxes.
- 10 In the Percentages box, click on the Row, Column and Total check boxes.

Counts	Continue
✓ Observed	Cancel
Expected     ■     Expected     Expect	Help
Percentages	Residuals
<b>▼</b> Row	☐ Unstandardized
√ Column	☐ Standardized
▼ Total	☐ Adjusted standardized
Noninteger Weig	nts
<ul> <li>Round cell of</li> </ul>	ounts Round case weights
C Truncate ce	counts C Truncate case weights
○ No adjustme	nts

#### 11 Click on Continue and then OK.

CROSSTABS
/TABLES=pref BY location
/FORMAT= AVALUE TABLES
/STATISTIC=CHISQ
/CELLS= COUNT EXPECTED ROW COLUMN TOTAL
/COUNT ROUND CELL .

#### preference for magazine \* location of reader Crosstabulation

			location of reader		
			urban	rural	Total
preference for magazine	Financial Review	Count	32	24	56
		Expected Count	38.2	17.8	56.0
		% within preference for magazine	57.1%	42.9%	100.0%
		% within location of reader	53.3%	85.7%	63.6%
		% of Total	36.4%	27.3%	63.6%
	Newsweek	Count	28	4	32
		Expected Count	21.8	10.2	32.0
		% within preference for magazine	87.5%	12.5%	100.0%
		% within location of reader	46.7%	14.3%	36.4%
		% of Total	31.8%	4.5%	36.4%
Total	11.	Count	60	28	88
		Expected Count	60.0	28.0	88.0
		% within preference for magazine	68.2%	31.8%	100.0%
		% within location of reader	100.0%	100.0%	100.0%
		% of Total	68.2%	31.8%	100.0%

#### **Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square Continuity Correction <sup>a</sup> Likelihood Ratio Fisher's Exact Test Linear-by-Linear Association N of Valid Cases	8.650 <sup>b</sup> 7.308 9.487 8.552 88	1 1 1	.003 .007 .002	.004	.003

a. Computed only for a 2×2 table

b. 9 cells (.0%) have expected count less than five. The minimum expected count is 10.18.

To interpret the chi-square printout you need to look at the Pearson statistic, which is short for Pearson's chi-square. In this example, Pearson has a value of 8.650 with a significance of .003. This significance value is well below the alpha level of .05 and is thus significant.

You can also see in the notes below the Chi-square Tests table that the minimum expected cell frequency is 10, which is > 5, and therefore you can be confident that you have not violated one of the main assumptions of chi-square.

In examining the observed cell frequencies, it can be concluded that the rural readers prefer to read the *Financial Review* while the urban readers do not show a marked preference for either publication,  $X^2(1, N=88) = 8.65$ , p < .05.

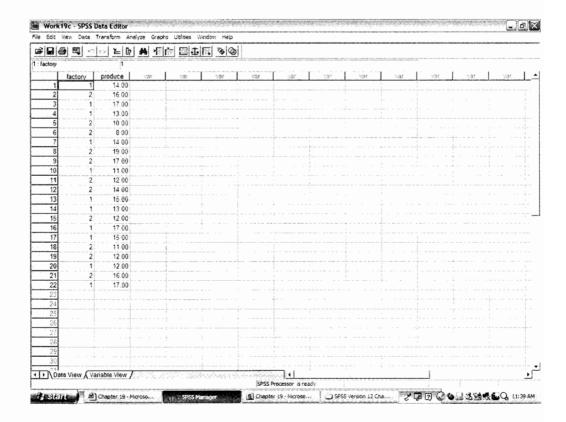
# Mann-Whitney U test (Wilcoxon rank sum W test)



The Mann-Whitney U test tests the hypothesis that two independent samples come from populations having the same distribution. This test is equivalent to the independent groups t-test.

## Working example

The productivity levels of factories A and B are to be compared. The monthly output, in tonnes of produce, was recorded for 22 consecutive months. The data violate the stringent assumptions of an independent groups t-test, so you have decided to perform a Mann-Whitney test. The data file can be found in Work19c.sav.

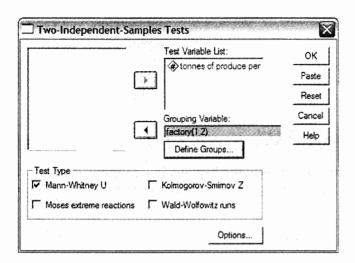




#### To conduct a Mann-Whitney U test

- 1 Select the **Analyze** menu.
- 2 Click on **Nonparametric Tests** and then on **2 Independent Samples...** to open the **Two-Independent-Samples Tests** dialogue box.

- 3 Select the dependent variable (i.e. *produce*) and click on the ▶ button to move the variable into the **Test Variable List:** box.
- 4 Select the independent variable (i.e. *factory*) and click on the ▶ button to move the variable into the **Grouping Variable:** box.



- 5 Click on the **Define Groups...** command pushbutton to open the **Two Independent Samples: Define Groups...** sub-dialogue box.
- In the **Group 1:** box, enter the first value for the independent variable (*sex*) (i.e. 1), then tab. Enter the second value for the independent variable (i.e. 2) in the **Group 2:** box.
- 7 Click on Continue.
- 8 Ensure the Mann-Whitney U check box has been selected.
- 9 Click on OK.

NPAR TESTS
/M-W= produce BY factory(1 2)
/MISSING ANALYSIS.

#### Ranks

	factory	Ν	Mean Rank	Sum of Ranks
tonnes of produce per month	Α	11	12.77	140.50
	В	11	10.23	112.50
	Total	22		

#### Test Statistics<sup>b</sup>

	tonnes of produce per month
Mann-Whitney U	46.500
Wilcoxon W	112.500
Z	927
Asymp. Sig. (2-tailed)	.354
Exact Sig. [2*(1-tailed Sig.)]	.365ª

a. Not corrected for ties.

b. Grouping Variable: factory

To interpret the output from the Mann-Whitney U test you need to consider the Z-score and two-tailed P-value, which have been corrected for ties.

The output indicates that the result, with correction for ties and Z-score conversion, was not significant,  $\underline{z} = -.927$ , p > .05, and therefore no significant differences in productivity exist between factories A and B.

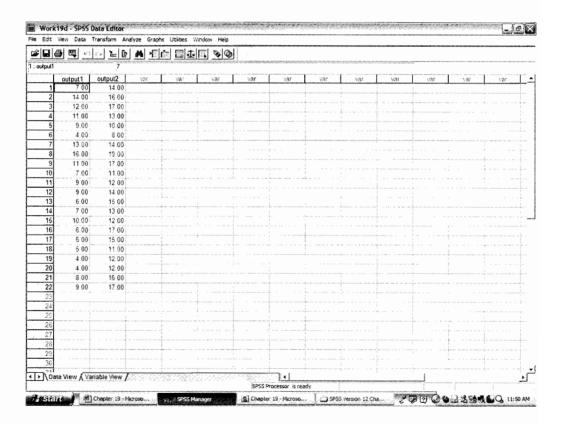
# Wilcoxon signed-rank test



The Wilcoxon signed-rank test, also referred to as the Wilcoxon T-test, is used when you would use a repeated measures or paired t-test — that is, when the same participants perform under each level of the independent variable.

## Working example

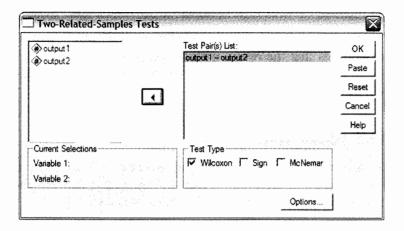
A factory manager wished to compare productivity in his factory for the first and second halves of the year. He observed the productivity from 22 work stations and recorded their output in tonnes. The data violate the stringent assumptions of a paired t-test, so he decided to perform a Wilcoxon signed-rank test. The data file can be found in Work19d.sav.



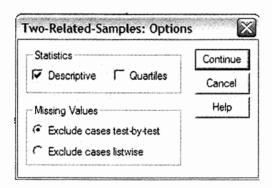


#### To conduct a Wilcoxon signed-rank test

- 1 Select the **Analyze** menu.
- 2 Click on Nonparametric Tests and then on 2 Related Samples... to open the Two-Related-Samples Tests dialogue box.
- 3 Select the variables you require (i.e. *output1* and *output2*) and click on the ▶ button to move the variables into the **Test Pair(s) List:** box.
- 4 Ensure that the **Wilcoxon** check box has been selected.



- 5 Click on the **Options...** command pushbutton to open the **Two-Related-Samples: Options** sub-dialogue box.
- 6 Ensure that the **Descriptives** check box has been selected.



7 Click on Continue and then OK.

NPAR TEST
/WILCOXON=output1 WITH output2 (PAIRED)
/STATISTICS DESCRIPTIVES
/MISSING ANALYSIS.

#### Ranks

		N	Mean Rank	Sum of Ranks
output2 - output1	Negative Ranks	0 <sup>a</sup>	.00	.00
	Positive Ranks	22 <sup>b</sup>	11.50	253.00
	Ties	0°		
	Total	22		

- a. output2 < output1
- b. output2 > output1
- c. output2 = output1

#### Test Statistics<sup>b</sup>

	output2 - output1
Z	-4.113 <sup>a</sup>
Asymp. Sig. (2-tailed)	.000

- a. Based on negative ranks.
- b. Wilcoxon Signed-Rank Test

To interpret the Wilcoxon signed-rank test you need to examine the Z-score value and the two-tailed P-value.

**Descriptive Statistics** 

	Ν	Mean	Std. Deviation	Minimum	Maximum
output1	22	8.5000	3.33452	4.00	16.00
output2	22	13.8636	2.74808	8.00	19.00

The output indicates that there is a significant difference in factory output for the first and second halves of the year,  $\underline{z} = -4.113$ , p < .05, with higher productivity in the second part of the year.

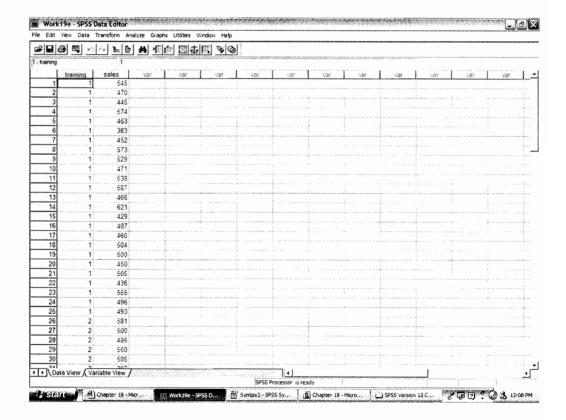
## Kruskal-Wallis test



The Kruskal-Wallis test is equivalent to the one-way between-groups ANOVA and therefore allows possible differences between two or more groups to be examined.

## Working example

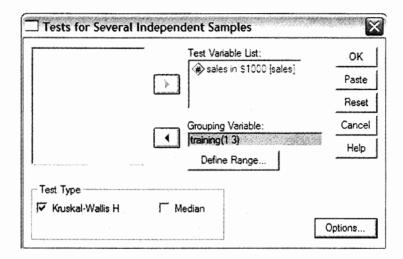
A personnel manager of a large insurance company wished to evaluate the effectiveness of three different sales training programs that had been designed for new employees. Seventy-five new graduates were randomly assigned to one of the programs, and then their annual sales figures (in \$1000s) were compared 12 months later. The data violate the stringent assumptions of a one-way ANOVA, so the manager decided to perform a Kruskal-Wallis test. The data file can be found in Work19e.sav.





#### To conduct a Kruskal-Wallis test

- 1 Select the **Analyze** menu.
- 2 Click on Nonparametric Tests and then on K Independent Samples... to open the Tests for Several Independent Samples box.
- 3 Select the dependent variable (i.e. *sales*) and click on the ▶ button to move the variable into the **Test Variable List:** box.
- 4 Select the independent variable (i.e. *training*) and click on the ▶ button to move the variable into the **Grouping Variable:** box.



- 5 Click on the Define Range... command pushbutton to open the Several Independent Samples: Define Range sub-dialogue box.
- 6 Enter the first value for the independent variable (*training*) (i.e. 1) in the **Minimum:** box, then tab. Enter the greatest value for the independent variable (i.e. 3) in the **Maximum:** box.
- 7 Click on **Continue**.
- 8 Ensure the **Kruskal-Wallis** H check box has been selected.
- 9 Click on OK.

Ranks

	training method	N	Mean Rank
sales in \$1000	Α	25	31.38
	В	25	38.98
	С	25	43.64
	Total	25	

NPAR TESTS
/K-W=sales BY training(1 3)
/MISSING ANALYSIS.

Test Statisticsa,b

	sales in \$1000
Chi-Square df	4.032
Asymp. Sig.	.133

- a. Kruskal-Wallis Test
- b. Grouping Variable: training method

To interpret the output from the Kruskal-Wallis test you need to look at the chi-square value, degree of freedom (df) and significance, which has been corrected for ties.

These values indicate that sales do not significantly differ across the three training programs,  $X^2(2, N=75) = 4.032$ , p > .05.

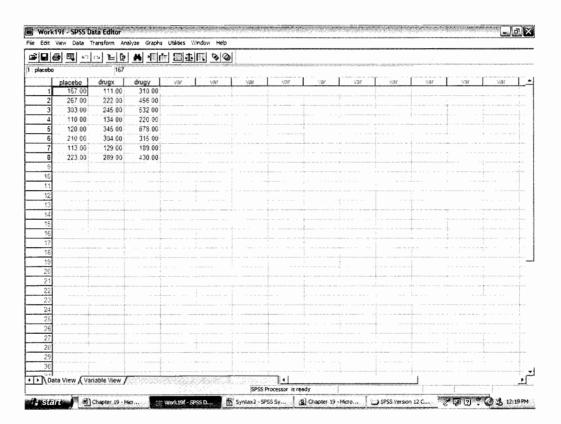
## Friedman test



The Friedman test is used to compare two or more related samples, and is equivalent to the repeated measures or within-subjects ANOVA.

## Working example

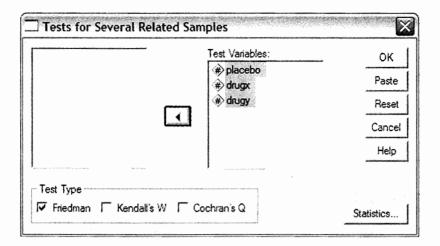
Reaction times for eight subjects were measured under a placebo condition, a drug X condition and a drug Y condition. It was hypothesised that reaction times would differ significantly across drug conditions. The data violate the stringent assumptions of a repeated measures ANOVA, so it was decided that a Friedman test should be performed. The data file can be found in Work19f.say.





#### To conduct a Friedman test

- 1 Select the **Analyze** menu.
- 2 Click on Nonparametric Tests and then on K Related Samples... to open the Tests for Several Related Samples box.
- Select the variables you require (i.e. *drugx*, *drugy* and *placebo*) and click on the button to move the variables into the **Test Variable List:** box.
- 4 Ensure the **Friedman** check box has been selected.



#### 5 Click on OK.

NPAR TESTS
/FRIEDMAN = placebo drugx drugy
/MISSING LISTWISE.

Ranks

	Mean Rank		
placebo	1.38		
drugx	1.63		
drugy	3.00		

Test Statistics<sup>a</sup>

N	8
Chi-Square	12.250
df	3
Asymp. Sig.	.002

a. Friedman Test

The results of the Friedman test indicate that significant differences do exist in reaction time across drug conditions,  $X^2(2, \underline{N}=8) = 12.25$ , p < .05, and that drug Y appears to slow reaction time considerably.

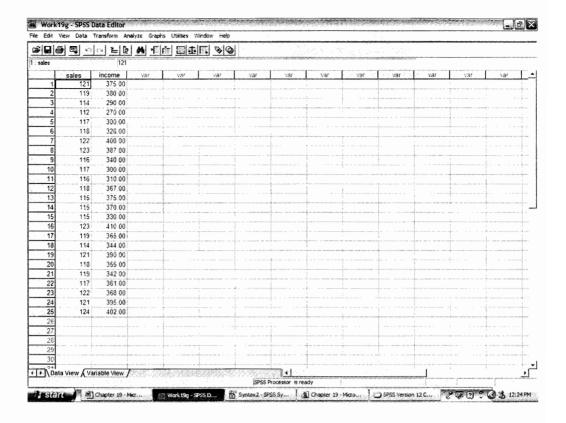
## Spearman's rank order correlation



A nonparametric alternative to the parametric bivariate correlation (Pearson's r) is Spearman's rho.

## Working example

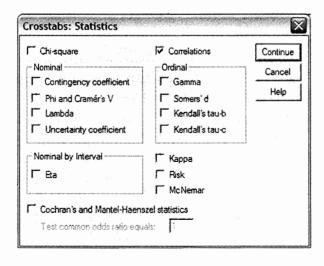
To examine the relationship between sales performance and employee income, data were collected from 20 sales representatives. The data violate the stringent assumptions of a Pearson's r, so it was decided a Spearman's rho should be performed. The data file can be found in Work19g.sav.





#### To conduct a Spearman's rank order correlation

- 1 Select the **Analyze** menu.
- 2 Click on **Descriptive Statistics** and then on **Crosstabs...** to open the **Crosstabs** dialogue box.
- Select the first variable (i.e. sales) and click on the  $\triangleright$  button to move the variable into the **Row**(s): box.
- 4 Select the second variable (i.e. *income*) and click on the ▶ button to move the variable into the Column(s): box.
- 5 Click on the **Statistics...** command pushbutton to open the **Crosstabs: Statistics** sub-dialogue box.
- 6 Select the **Correlations** check box.



7 Click on Continue and then OK.

CROSSTABS
/TABLES=sales BY income
/FORMAT= AVALUE TABLES
/STATISTIC=CORR
/CELLS= COUNT
/COUNT ROUND CELL .

#### **Symmetric Measures**

	Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Interval by Interval Pearson's R Ordinal by Ordinal Spearman Correlation N of Valid Cases	.767 .757 25	.083 .110	5.728 5.554	.000°

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.
- c. Based on normal approximation.

You can see from the output that Spearman's rank order correlation is significant,  $\underline{r}$  (25) = .757, p < .05, and so you can conclude that higher sales are associated with higher incomes.

Spearman's rho can also be obtained through the **Correlate-Bivariate** option in the **Analyze** menu.

# Practice examples



You will remember that nonparametric tests, or distribution-free tests, do not rely on parameter estimation or assumptions about parameters or the shape of distributions. Assuming that these assumptions cannot be met, perform the appropriate nonparametric test for each of the practice examples below.

## Practice example 1

A coffee company is about to launch a range of specialty coffee bags and would like to know if some varieties are likely to sell better than others. Four of their new varieties are placed on a supermarket shelf for one week. The number of boxes sold for each variety are: 20 Mocha Kenya; 40 Colombian; 30 Lebanese; and ten Swiss. Given the data in Prac19a.sav, your task is to:

1 Determine whether the distribution of sales suggests that some varieties of coffee bags are more popular than others.

## Practice example 2

A researcher was interested in determining whether drinking preference was gender related. Given the data in Prac19b.sav, your task is to:

1 Determine whether drinking preference is gender related — that is, whether most men prefer to drink beer rather than wine.

## Practice example 3

A psychologist is interested in the relationship between personality types and religious affiliation. Fifteen Baptist and 15 Anglican churchgoers were randomly selected from their respective communities and asked to complete a personality questionnaire measuring introversion–extroversion on a scale from 0 to 100 — the higher the score on the scale, the more extroverted the individual. Given the data in Prac19c.sav, your task is to:

1 Determine whether personality types differ significantly across the two religions.

## Practice example 4

A human factors specialist working for an airline company wishes to examine the effect of temperature on performance. The performances of a random sample of 15 pilots were recorded following participation in an air simulation task, in cockpits of differing temperatures (20 degrees Celsius and 35 degrees Celsius). Given the data in Prac19d.sav, your task is to:

1 Determine whether temperature influences pilot performance.

## Practice example 5

A physical education teacher observes that a large number of sporting injuries result from ball sports such as football, basketball and rugby. Over the period of one month, he records the number of injuries sustained by 45 boys actively involved in one of the three sports. Given the data in Prac19e.sav, your task is to:

1 Determine whether the number of injuries sustained differs according to the type of sport played.

## Practice example 6

A veterinarian wishing to start a new dog-breeding business wants to determine which of four breeds of small dog is more companionable: the silky terrier, the Jack Russell, the Bichon Frise or the pug. He randomly samples 20 people who regularly attend his clinic and asks them to complete a perceived companionability scale — the higher the score, the more companionable the breed of dog. Given the data in Prac19f.sav, your task is to:

Determine whether certain breeds of small dog are more companionable than others.

## Practice example 7

A truck manufacturer notices that a truck's fuel consumption depends on the load it is carrying. He keeps a log of diesel consumption and load for 20 trucks over a six-month period. Given the data in Prac19g.sav, your tasks are to:

- 1 Plot the relationship between the two variables.
- 2 Determine whether there is a relationship between load and diesel consumption.

## **Solutions**



## Practice example 1 — chi-square test for goodness of fit

#### **Syntax**

WEIGHT
BY freq :
NPAR TEST
/CHISQUARE=coffee
/EXPECTED=EQUAL
/MISSING ANALYSIS.

#### Output

#### type of coffee bag

	Observed N	Expected N	Residual
Mocha Kenyan	20	25.0	-5.0
Columbian	40	25.0	15.0
Lebanese	30	25.0	5.0
Swiss	10	25.0	-15.0
Total	100		

#### **Test Statistics**

	type of coffee bag
Chi-Square <sup>a</sup>	20.000
df	3
Asymp. Sig.	.000

 a. 0 cells (.0%) have expected frequencies less than five. The minimum expected cell frequency is 25.0.

The chi-square test for goodness of fit is significant,  $X^2(3, \underline{N}=100)=20.0$ , p<.05, and suggests that the brands are not equally popular. Inspection of the observed cases reveals that the Colombian coffee bags are the most popular and the Swiss coffee bags least popular.

# Practice example 2 — chi-square test for relatedness or independence

#### **Syntax**

WEIGHT
BY freq .
CROSSTABS
/TABLES=drink BY gender
/FORMAT= AVALUE TABLES
/STATISTIC=CHISQ
/CELLS= COUNT EXPECTED
/COUNT ROUND CELL .

#### Output

#### drinking preference \* gender Crosstabulation

			gender		
			male	female	Total
drinking preference	wine	Count Expected Count	16 20.0	21 17.0	37 37.0
	beer	Count Expected Count	31 27.0	19 23.0	50 50.0
Total		Count Expected Count	47 47.0	40 40.0	87 87.0

#### **Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.012 <sup>b</sup>	1	.083		
Continuity Correction <sup>a</sup>	2.304	1	.129		
Likelihood Ratio	3.022	1	.082		
Fisher's Exact Test				.127	.064
Linear-by-Linear Association	2.977	1	.084		
N of Valid Cases	87				

a. Computed only for a 2×2 table

A review of the output indicates that the Pearson's statistic is not significant (p > .05). The results indicate that drinking preference is not gender related,  $X^{2}(1, N=87) =$ 3.012, p > .05, and men do not prefer to drink beer rather than wine.

## Practice example 3 - Mann-Whitney U test (Wilcoxon rank sum W test)

#### Syntax

NPAR TESTS /M-W= iescore BY religion(1 2) /MISSING ANALYSIS.

#### Output

#### Ranks

	religious affiliation	Ν	Mean Rank	Sum of Ranks
introversion-extraversion score	Baptist	15	10.00	150.00
İ	Anglican	15	21.00	315.00
	Total	30		

#### Test Statistics<sup>b</sup>

	introversion- extraversion score
Mann-Whitney U	30.000
Wilcoxon W	150.000
Z	-3.425
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.000 <sup>a</sup>

a. Not corrected for ties.

b. 0 cells (.0%) have expected count less than five. The minimum expected count is 17.01.

b. Grouping Variable: religious affiliation

In interpreting the Z-score and two-tailed P-value, which has been corrected for ties, it is found that there is a significant result.

Therefore, it can be concluded that there is a relationship between religious affiliation and personality type,  $\underline{z} = -3.425$ , p < .05, with Anglicans considered to be more extroverted than Baptists.

## Practice example 4 - Wilcoxon signed-rank test

#### **Syntax**

NPAR TEST

/WILCOXON=cockpit1 WITH cockpit2 (PAIRED) /MISSING ANALYSIS.

#### Output

#### Ranks

		N	Mean Rank	Sum of Ranks
9	Negative Ranks	O <sup>a</sup>	.00	.00
performance at 20 degrees C F	Positive Ranks	10 <sup>b</sup>	5.50	55.00
Т	Ties	0°		
Ţ	Total .	10		

- a. performance at 35 degrees C < performance at 20 degrees C
- b. performance at 35 degrees C > performance at 20 degrees C
- c. performance at 35 degrees C = performance at 20 degrees C

#### Test Statistics<sup>b</sup>

	performance at 35 degrees C - performance at 20 degrees C
Z	-2.836 <sup>a</sup>
Asymp. Sig. (2-tailed)	.005

- a. Based on negative ranks.
- b. Wilcoxon Signed-Rank Test

#### **Descriptive Statistics**

	N	Mean	Std. Deviation	Minimum	Maximum
performance at 20 degrees C	10	9.9000	1.91195	7.00	13.00
performance at 35 degrees C	10	11.9000	1.96921	9.00	15.00

An examination of the Z-score value and two-tailed P-value indicates that the test is significant.

Therefore it can be concluded that temperature does influence pilot performance,  $\underline{z} = -2.836$ , p < .05, with better performance exhibited by pilots in the warmer cockpit.

## Practice example 5 - Kruskal-Wallis test

#### **Syntax**

NPAR TESTS
/K-W=injuries BY sport(1 3)
/MISSING ANALYSIS.

#### Output

Ranks

	type of sport	N	Mean Rank
number of injuries	football	15	13.23
	rugby	15	34.23
	basketball	15	21.53
	Total	45	

#### Test Statisticsa,b

	number of injuries			
Chi-Square	19.593			
df	2			
Asymp. Sig.	.000			

- a. Kruskal-Wallis Test
- b. Grouping Variable: type of sport

Examination of the chi-square value, which has been corrected for ties, indicates that the test is significant.

Therefore it can be concluded that the number of injuries sustained varies across the nature of the sport,  $X^2(2, \underline{N} = 45) = 19.593$ , p < .05, with rugby players' suffering more injuries (Mean = 34.23) than are received by either basketball (Mean = 21.53) or football players (Mean = 13.23).

## Practice example 6 — Friedman test

#### **Syntax**

NPAR TESTS

/FRIEDMAN = breed1 breed2 breed3 breed4 /MISSING LISTWISE.

#### Output

#### Ranks

	Mean Rank
silky terrier	3.28
Jack Russell	1.43
Bichon Frise	3.48
Pug	1.83

#### Test Statistics<sup>a</sup>

N	20
Chi-Square	38.528
df	3
Asymp. Sig.	.000

a. Friedman Test

The results of the Friedman test indicate that there are significant differences in companionability across different breeds of dog,  $X^2(3, \underline{N}=20)=38.528$ , p<.05. Bichon Frise were seen to be the most companionable breed, followed by silky terriers and pugs. Jack Russells were perceived to be the least companionable.

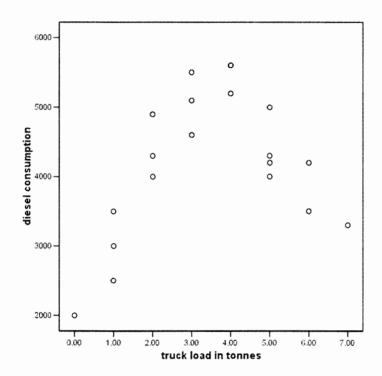
## Practice example 7 - Spearman's rank order correlation

#### **Syntax**

```
GRAPH
/SCATTERPLOT(BIVAR)= load WITH diesel
/MISSING=LISTWISE .
CROSSTABS
/TABLES=load BY diesel
/FORMAT= AVALUE TABLES
/STATISTIC=CORR
/CELLS= COUNT EXPECTED
/COUNT ROUND CELL .
```

#### Output

The following scatterplot indicates a curvilinear relationship between the amount of diesel consumption and truck load.



An examination of Spearman's correlation coefficient indicates that this relationship is not significant.

**Symmetric Measures** 

	Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Interval by Interval Pearson's R Ordinal by Ordinal Spearman Correlation N of Valid Cases	.331 .232 20	.222 .277	1.488 1.011	.154° .326°

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.
- c. Based on normal approximation.

Therefore, our analysis suggests that there is no relationship between truck load and diesel consumption,  $\underline{r}$  (20) = .232, p > .05.

# Multiple response and multiple dichotomy analysis

Multiple response and multiple dichotomy analysis are commonly used in the analysis of questionnaire or survey data.

# Multiple response analysis



Multiple response analysis is commonly used for the analysis of open-ended questions. For example, a questionnaire item may be:

Q: 'What do you think are the important things that need to be considered when developing higher density housing in your neighbourhood?'

You must do two things on receiving all your questionnaires back from the field, before analysing the information. These are:

- Look at the responses to the open-ended question and determine the maximum number of responses to the question. In this example, are people giving one, two, three or more responses to the question?
- Make sure you have some understanding of the range of possible responses. Will the number of different responses be less than 100, in which case you can code all responses from 01 to 99? If there appears to be more than 100 different responses, then you may have to use a three-digit code and code from 001 to 999.

Let's assume that in the example above you have decided that there is a maximum of four responses given to the question and that there are fewer than 100 different responses. Therefore, it is possible to use a two-digit code for these data. The next step is to code all your questionnaires. This coding system is defined using the **Value Labels** option. However, it is more efficient to use the **Template** option, given that there is likely to be multiple variables. Possible codes, for the above example, could include:

- 01 Privacy of residents
- 02 Sewerage
- 03 Noise from traffic
- 04 Access to public transport.

In acknowledging a maximum of four possible responses to this open-ended question, that four variables will be coded is accepted. Within SPSS these variables could be defined as *open1*, *open2*, *open3* and *open4*.

# Working example



Three-hundred-and-sixty-four farmers were asked to identify the selection criteria they used when choosing rams for their breeding program. The maximum number of responses obtained from an individual was four, and 14 possible criteria were identified. If an individual identified only two criteria, then a code of 88 would be used for variables 3 and 4. The data from the first four participants may look as follows:

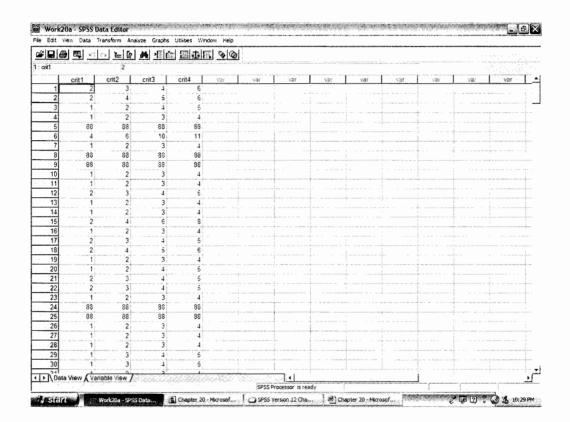
Participant 1 01040288 (This participant has given three responses.)

Participant 2 05888888 (This participant has given one response.)

Participant 3 06011107 (This participant has given four responses.)

Participant 4 88888888 (This participant did not respond to the question.).

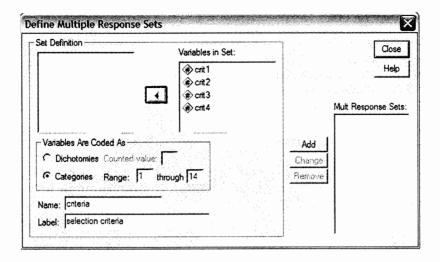
You wish to obtain a frequency analysis of the multiple responses of your entire sample for the question outlined on page 227. This data file can be found in Work20a.sav on the website that accompanies this title and is shown in the following figure.



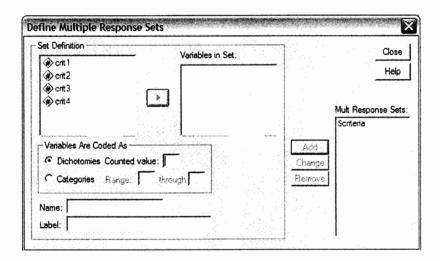


#### To conduct a multiple response analysis

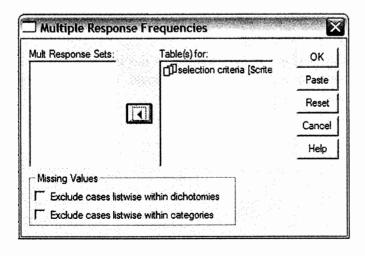
- 1 Select the **Analyze** menu.
- 2 Click on Multiple Response and then on Define Sets... to open the Define Multiple Response Sets dialogue box.
- 3 Select the four variables you require (i.e. *crit1* to *crit4*) and click on the button to move the variables into the **Variables in Set:** box.
- 4 In the Variables Are Coded As box, click on the Categories radio button.
- In the **Range:** box, type the lowest code (i.e. 1), press tab and then type the highest code (i.e. 14).
- 6 In the Name: box, type a suitable variable name (i.e. *criteria*) and in the Label: box type a description of this variable (i.e. *Selection Criteria*).



7 Click on the **Add** command pushbutton; you will notice that a new variable has been created (i.e. \$criteria) in the **Multiple Response Sets:** box.



- 8 Click on Close.
- 9 Select the **Analyze** menu.
- 10 Click on Multiple Response and Frequencies... to open the Multiple Response Frequencies dialogue box.
- Select the variable (i.e. *\$criteria*) and click on the ▶ button to move the variable into the **Table(s) for:** box.



#### 12 Click on OK.

MULT RESPONSE GROUPS=\$criteria 'selection criteria' (crit1 crit2 crit3 crit4 (1,14)) /FREQUENCIES=\$criteria .

Group Scriteri selection criteria

			Pct of	Pct of
Category label	Code	Count	Responses	Cases
Fleece style	1	33	4.3	12.0
Fibre diameter	2	85	11.0	30.8
Greasy fleece weight	3	85	11.0	30.8
Clean fleece weight	4	180	23.3	65.2
Live weight	5	48	6.2	27.4
Reproduction records	6	76	9.8	27.5
Disease resistance	7	15	1.9	5.4
Frame size	8	47	6.1	17.0
Constitution	9	42	5.4	15.2
Skin charactersitics	10	79	10.2	28.6
Staple strength	22	15	1.9	5.4
Feet	12	37	4.8	13.4
Hocks, jaws	13	26	3.4	9.4
Wool handle	14	4	.5	1.4
	Total responses	772	100.0	279.7

88 missing cases; 276 valid cases

Abbreviated Extended Name Name

Scriteri Scriteria

The frequency table of the multiple response set indicates that clean fleece weight is the most frequently used selection criterion.

Percentage of responses refers to the proportion of a given response in relation to the count; for example, 33 people cited fleece style as a selection criterion, and this is 33/772 per cent of the total responses (4.3%).

Percentage of cases refers to the proportion of a given response in relation to the number of valid cases; for example, 79 people cited skin characteristics as a selection criterion and this is 79/276 per cent of the valid cases (28.6%).

# Multiple dichotomy analysis



Multiple dichotomy analysis is very similar to the multiple response analysis. The following questionnaire item, for example, could be analysed using a multiple dichotomy analysis.

'Would you object to any of the following developments being built next to your property?'

- Single house
- Group housing duplex
- Group housing villa units (four units or fewer)
- Group housing villa units (five or more units)
- Group housing townhouses (four townhouses or fewer)
- Group housing townhouses (five or more)
- Flats or home units (three storeys or fewer)
- Flats or home units (four storeys or more)

In this example each item would be given a variable label and a code. The codes normally used in these cases are 1 if the item is ticked and 0 if it is not ticked. If the person failed to place a tick alongside at least one item, then the missing value code of 9 might be assigned to all items. The data from the first four participants may look as follows:

Participant 1 00001111 (This participant did not tick the first four items.)

Participant 2 10000000 (This participant ticked only the first item.)

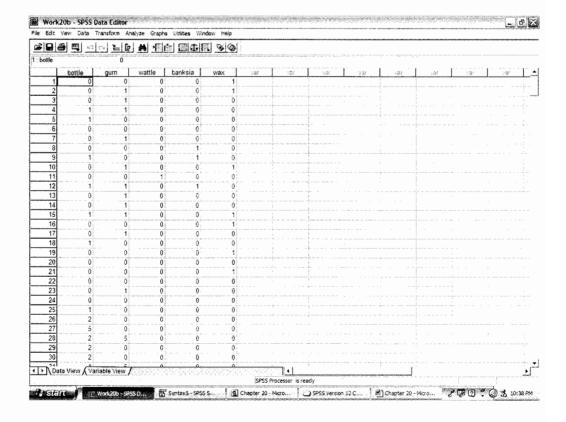
Participant 3 01111011 (This participant ticked variably across items.)

Participant 4 11111111 (This participant ticked all eight items.).

# Working example



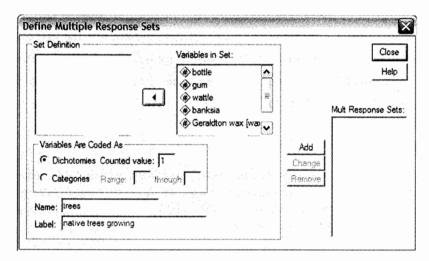
A sample of 388 home owners were asked to indicate which native trees were growing in their gardens. They were given five species to select from. Determine the species most frequently chosen by home owners. The data file can be found in Work20b.sav and is shown in the following figure.



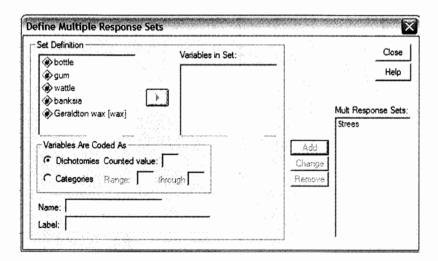


## To conduct a multiple dichotomy analysis

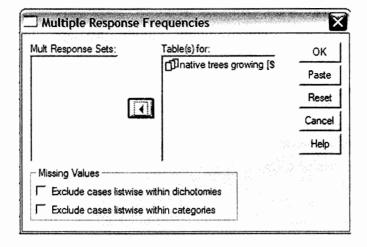
- 1 Select the **Analyze** menu.
- 2 Click on Multiple Response and Define Sets... to open the Define Multiple Response Sets dialogue box.
- 3 Select the five variables (i.e. banksia, bottle, gum, wattle, wax) and click on the 
  ▶ button to move the variables into the Variables in Set: box.
- 4 In the Variables Are Coded As box, ensure the Dichotomies radio button is selected.
- 5 In the Counted value: box, type the value that you have assigned to those items that were ticked by respondents (i.e. 1).
- In the Name: box, type a suitable variable name (i.e. *trees*) and in the Label: box, type a description of this variable (i.e. *native trees growing*).



7 Click on the **Add** command pushbutton. You will notice that a new variable has been created (i.e. *\$trees*) in the **Mult Response Sets:** box.



- 8 Click on Close.
- 9 Select the Analyze menu.
- 10 Click on Multiple Response and Frequencies... to open the Multiple Response Frequencies dialogue box.
- Select the variable (i.e. *\$trees*) and click on the **▶** button to move the variable into the **Table(s) for:** box.



#### 12 Click on OK.

MULT RESPONSE
GROUPS=\$trees 'native trees growing' (bottle gum wattle banksia wax (1))
/FREQUENCIES=\$trees .

Group Strees native trees growing (Value tabulated = 1)

Dichotomy label	Name	Count	Pct of Responses	
	bottle	281	92.7	95.3
	gum	11	3.6	3.7
	wattle	<u>1</u>	.3	.3
	banksia	3	1.0	1.0
Geraldton wax	wax	7	2.3	2.4
Tot	al responses	303	100.0	102.7

93 missing cases; 295 valid cases

Inspection of the frequency table for the multiple dichotomous set indicates that bottle brushes are the most frequently chosen native species of tree in the garden.

Having defined multiple response or multiple dichotomous sets, it is possible to also perform crosstabs analysis, by means of which response sets or categorical variables can be presented in tabular form.

# Practice examples



## Practice example 1

A counselling psychologist was interested in determining why people marry. The maximum number of reasons given by any one person was three. These reasons were coded as follows:

01 Love05 Security02 Company06 To have children03 Money07 Familiarity04 Sex

Given the data in Prac20a.sav, your tasks are to:

- 1 Determine the most frequently quoted reason for marriage.
- 2 Determine the most frequently quoted reason for marriage given by men.

## Practice example 2

A random sample of drivers was asked to indicate factors that influenced their decision when purchasing a used car. The following criteria were specified:

- 01 Year of manufacture
- 02 Mileage
- 03 State of the interior
- 04 State of the chassis
- 05 State of the duco
- 06 Other

Given the data in Prac20b.sav, your task is to:

1 Determine the most influential criterion on vehicle purchase.

## **Solutions**



## Practice example 1

#### **Syntax**

```
MULT RESPONSE
GROUPS=$reasons 'Reasons for marriage' (reason1 reason2 reason3 (1,7))
/FREQUENCIES=$reasons .

MULT RESPONSE
GROUPS=$reasons 'Reasons for marriage' (reason1 reason2 reason3 (1,7))
/VARIABLES=gender(1 2)
/TABLES=$reasons BY gender
/CELLS=ROW COLUMN TOTAL
/BASE=CASES .
```

#### Output

Group Sreasons Reasons for marriage

Category label	Code	Count	Pot of Responses	<del>-</del>
love	2	9	20.9	45.0
company	2	8	18.6	40.0
money	3	9	20.9	45.0
зех	4	3	7.9	15.0
security	5	8	18.6	40.0
have children	6	2	4.7	10.0
familiarity	7	4	9.3	20.0
Total	responses	43	100.0	215.0

0 missing cases: 20 valid cases

An examination of the frequency table for the multiple dichotomous set indicates that love and money are the most frequently quoted reasons for marriage. When an analysis by gender is undertaken, men most frequently quote money as the main reason for marriage.

```
Page 1 of 2
                gender
          Count
                ⇔female male
          Row pct ⇔
                                Row
          Col pct
                    1 ⇔ 2
          Tab pct
                \Leftrightarrow
          $reasons
               1 ⇔
                    6 ⇔ 3 ⇔ 9
love
                 ⇔ 66.7 ⇔ 33.3 ⇔ 45.0
                 ⇔ 60.0 ⇔ 30.0

⇔ 30.0 ⇔ 15.0

                2 \Leftrightarrow 4 \Leftrightarrow 4 \Leftrightarrow 8

⇔ 50.0 ⇔ 50.0 ⇔ 40.0

company

⇔ 40.0 ⇔ 40.0

⇔ 20.0 ⇔ 20.0 ⇔

                3 ⇔
                    4 ⇔
                          5 ⇔

⇔ 44.4 ⇔ 55.6 ⇔ 45.0

monev

⇔ 40.0 ⇔ 50.0 ⇔

⇔ 20.0 ⇔ 25.0 ⇔

                 4 ⇔
                          1 ⇔ 3
                   2 ⇔

⇔ 66.7 ⇔ 33.3 ⇔ 15.0

sex

⇔ 20.0 ⇔ 10.0 ⇔

⇔ 10.0 ⇔ 5.0 ⇔

                5 ⇔
                   4 ⇔ 4 ⇔ 8

⇔ 50.0 ⇔ 50.0 ⇔ 40.0

security

⇔ 40.0 ⇔ 40.0

                 ⇔ 20.0 ⇔ 20.0
                6 ⇔ 1 ⇔ 1 ⇔ 2

⇔ 50.0 ⇔ 50.0 ⇔ 10.0

have children

⇔ 10.0 ⇔ 10.0 ⇔

                 ⇔ 5.0 ⇔ 5.0 ⇔
                Column
                   10
                          10 20
            Total 50.0
                        50.0
                             100.0
Percentages and totals based on respondents
       * * * CROSSTABULATION * * *
 $reasons (group) Reasons for marriage by gender
Page 2 of 2
                gender
          Count
                ⇔female male
          Row pct ⇔
          Col pct
                \Leftrightarrow
                    1 ⇔ 2
          Tab pct ⇔
          $reasons
                    1 ⇔
familiarity

⇔ 25.0 ⇔ 75.0 ⇔ 20.0

⇔ 10.0 ⇔ 30.0

⇔ 5.0 ⇔ 15.0 ⇔

                Column 10
                         10
                                 20
            Total 50.0
                        50.0 100.0
Percentages and totals based on respondents
```

\* \* \* CROSSTABULATION \* \* \*
\$reasons (group) Reasons for marriage by gender

20 valid cases; 0 missing cases

## Practice example 2

#### **Syntax**

MULT RESPONSE

GROUPS=\$carpick 'criteria for purchasing a car' (year mileage interior chassis duco other (1)) /FREQUENCIES=\$carpick .

#### Output

Group Scarpick criteria for purchasing a car (Value tabulated = 1)

Dichotomy label		Name	Count	Pot of Responses	
year of manufacture		year	12	25.0	60.0
		mileage	10	20.8	50.0
state of the interior		interior	7	14.6	35.0
state of chassis		chassis	7	14.6	35.0
state of duco		duco	8	16.7	40.0
		other	4	8.3	20.0
	Total :	responses	48	100.0	240.0

0 missing cases; 20 valid cases

Inspection of the frequency table for the multiple dichotomous set indicates that the year of manufacture is the most important consideration when buying a car, with 60% of the sample selecting this criterion.

# **Multidimensional** scaling

Multidimensional Scaling (MDS) starts with a data set of 'proximities', which indicate the degree of 'similarity' or 'dissimilarity' among elements in a defined set. The proximity measures may be either 'direct' or 'derived' measures of proximity. If the measures are direct, then respondents have usually provided a direct estimate of the degree of (dis)similarity between any pair of elements, either through a rating or classification task. Derived proximities, on the other hand, usually consist of some measure of association (e.g. correlation) among elements within a set.

The objective of MDS is to take the proximities data and represent the elements in multidimensional space, such that the distances among the elements in the space accurately represent the original proximity measures. This chapter provides a basis overview of the concept of MDS. SPSS also provides several alternative data analysis options that can be further developed and explored after a basic knowledge of MDS has been developed.

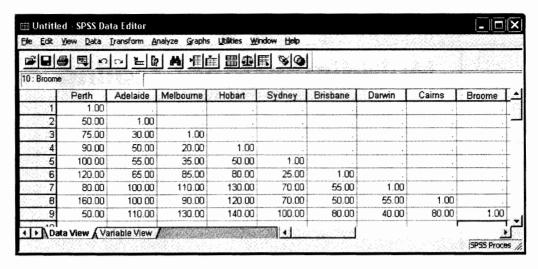
# Working example



As part of a study into tourist perceptions of Australia, you are interested in understanding how tourists from other countries comprehend travel distances among some of the major cities in Australia.

You provide a group of three visiting tourists with the names of eight Australian cities they are likely to visit. You indicate to the group that the distance between Sydney and Perth is 100 units and that they are to make judgements about what they think are the distances between all the remaining pairs of towns relative to the distance between Sydney and Perth.

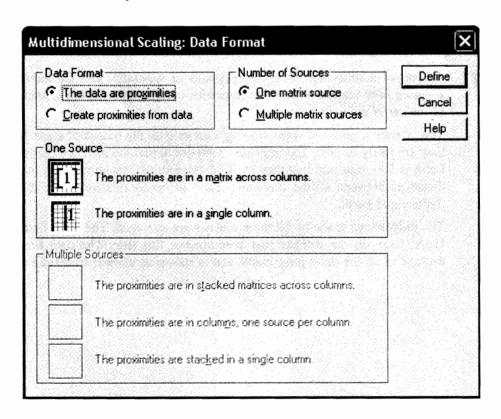
The judgements made by all three tourists are averaged. The data file can be found in Work21.sav on the website that accompanies this title. The data file represents the average of all the three judgements and is shown as a lower triangular matrix.



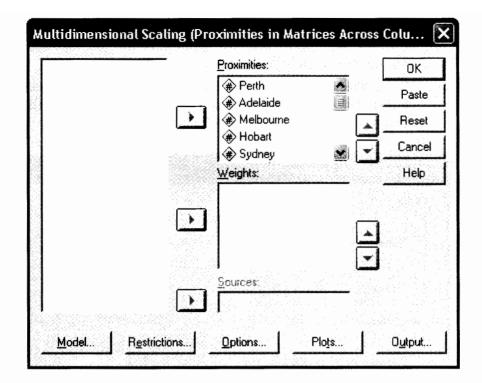


#### To conduct a MDS analysis

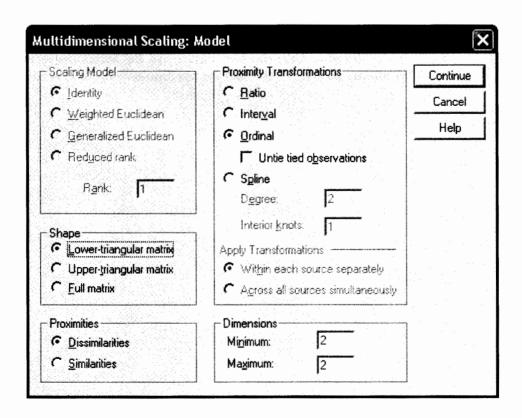
- 1 Select the Analyze menu.
- 2 Click on Scale and Multidimensional Scaling... (PROXSCAL) to open the Multidimensional Scaling: Data Format dialogue box.
- 3 In the Data Format box, select The Data are Proximities radio button.
- 4 In the **Number of Sources** box, make sure that the **One matrix source** radio button is also selected.
- 5 In the One Source box, select The proximities are in a matrix across columns option.



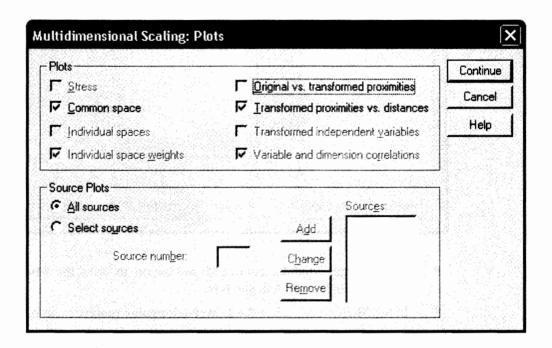
- 6 Click on the **Define** command pushbutton to open the **Multidimensional** Scaling dialogue box.
- 7 Select the nine variables representing each of the towns and click on the button to move the variables into the **Proximities:** box.



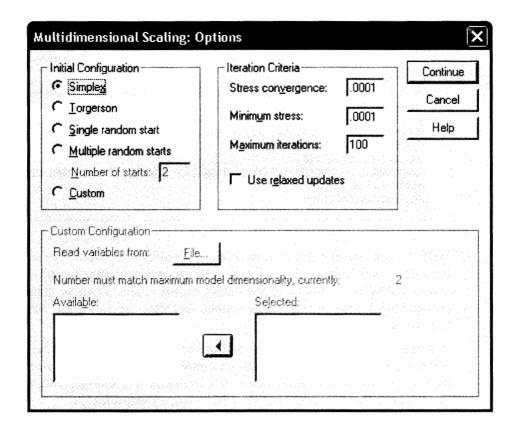
- 8 Click on the **Model...** command pushbutton to open the **Multidimensional** Scaling: Model sub-dialogue box.
- 9 In the Shape: box, select the Lower-triangular matrix radio button.
- 10 In the **Proximities:** box, select the **Dissimilarities** radio button.
- In the **Proximity Transformations:** box, make sure that the **Ordinal** radio button is selected.
- In the **Dimensions:** box ensure that you have entered 2 for the **Minimum Dimensions** and 2 for the **Maximum Dimensions**.



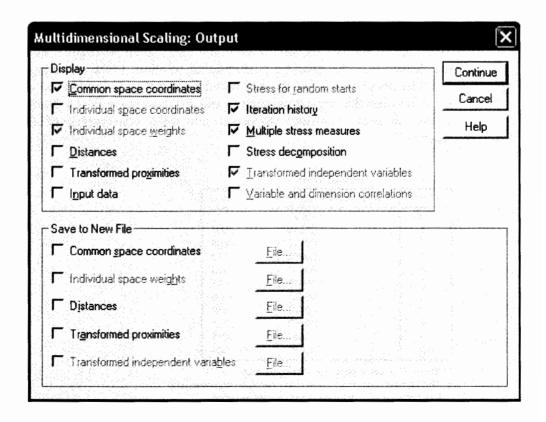
- 13 Click on Continue to return to the Multidimensional Scaling dialogue box.
- 14 Click on the **Plots** command pushbutton to open the **Multidimensional** Scaling: **Plots** sub-dialogue box.
- In the Plots: box, ensure that the Common Space and Transformed proximities vs. distances check boxes are both selected. The All sources radio button should also be selected in the Source Plots: box.



- 16 Click on Continue once again to return to the Multidimensional Scaling dialogue box.
- 17 Click on the **Options** command pushbutton to open the **Multidimensional Scaling: Options:** sub-dialogue box.
- 18 In the **Initial Configuration:** box, ensure that the **Simplex** radio button is selected.
- In the **Iteration Criteria:** box, ensure that the **Stress convergence:** is .0001, **Minimum stress:** is .0001 and the **Maximum iterations:** is 100.



- 20 Again, click on Continue to return to the Multidimensional Scaling dialogue box.
- Click on the **Output** command pushbutton to open the **Multidimensional Scaling: Output:** sub-dialogue box.
- In the **Display:** box, ensure that you select the **Common space coordinates**, **Iteration history** and **Multiple stress measures** check boxes.



VARIABLES=Perth Adelaide Melbourne Hobart Sydney Brisbane Darwin Cairns Broome /SHAPE=LOWER

/INITIAL=SIMPLEX

/TRANSFORMATION=ORDINAL (KEEPTIES)

/PROXIMITIES=DISSIMILARITIES

/ACCELERATION=NONE

/CRITERIA=DIMENSIONS(2,2) MAXITER(100) DIFFSTRESS(.0001) MINSTRESS(.0001)

/PRINT=COMMON HISTORY STRESS

/PLOT=COMMON RESIDUALS (ALL) .

MDS analysis has the potential to produce considerable output. The analysis that has been described is for a simple MDS solution in a two-dimensional space. Once an understanding is developed of the output of this simple solution, commands may be changed to produce more extended output, including increasing the dimensional space to three or more dimensions.

It is important before any inspection of the analyses to confirm that the data have been entered and analysed correctly. The case-processing summary shown below indicates there were nine cases (towns), one source or matrix was analysed and, when excluding the diagonal, this matrix consisted of 36 proximity measures.

Case-Processing Summary

Cases		9
Sources		1
Objects		9
Proximities	Total Proximities	<sup>b</sup> 36
	Missing Proximities	0
	Active Proximities <sup>a</sup>	36

- a. Active proximities include all nonmissing proximities.
- b. Sum of all strictly lower-triangular proximities.

The iteration history with the Normalised Raw Stress values is shown below. The Raw Stress values are an indication of how well the original proximities are represented by the distances in the final multidimensional solution. From a starting configuration, Proxscal iteratively moves the data points so that there is continuous improvement in the fit between the original proximities and distances in the final solution. After 10 iterations the iteration history stopped as the stress convergence of .0001 had been reached.

Iteration History

Iteration	Normalised Raw Stress	Improvement
0	<sup>a</sup> .18077	
1 ,	.01011	.17067
2	.00716	.00295
3	.00606	.00110
4	.00551	.00054
5	.00516	.00035
6	.00491	.00025
7	.00472	.00019
8	.00458	.00014
9	.00447	.00011
10	.00438	<sup>b</sup> .00009

- a. Stress of initial configuration: simplex start.
- The iteration process has stopped because improvement has become less than the convergence criterion.

There are several stress measures and measures of fit that indicate how well the original proximities are represented as distances in a multidimensional space. The most commonly used measure of fit is Stress-I or 'Kruskal's Stress'. It is this value that needs to be minimised, as a lower stress value indicates a better fit between the original proximities and the derived distances. As a rule of thumb, when Kruskal's Stress is less than 0.10 the fit is considered excellent and anything over 0.10 is becoming less acceptable. While an attempt is made to minimise stress values, measures of the 'Dispersion Accounted For' and Tucker's Coefficient of Congruence when high (greater than 0.90), indicate a reasonably good fit of the original proximities to the derived distances. In the working example, Kruskal's Stress I and Tucker's Coefficient of Congruence indicate there is a good fit between the original proximity measures and derived distances.

#### Stress and Fit Measures

Normalised Raw Stress	.00438
Stress-I	a.06620
Stress-II	<sup>a</sup> .18386
S-Stress	<sup>b</sup> .01127
Dispersion Accounted For	.99562
Tucker's Coefficient of Congruence	.99781

PROXSCAL minimises Normalised Raw Stress.

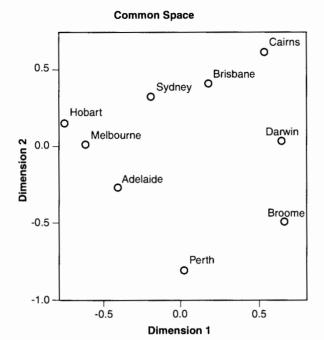
- a. Optimal scaling factor = 1.004.
- b. Optimal scaling factor = .997.

The coordinates of each of the nine data points (towns) for a two-dimensional solution are illustrated in the Final Coordinates output below. These are the coordinates that SPSS uses to plot the common space.

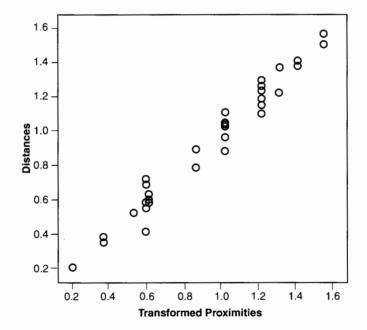
**Final Coordinates** 

	Dime	nsion
	1	2
Perth	.019	803
Adelaide	416	269
Melbourne	624	.011
Hobart	766	.153
Sydney	206	.325
Brisbane	.163	.413
Darwin	.640	.040
Cairns	.527	.616
Broome	.666	485

Although stress and other goodness of fit measures may be used to provide an indication of how well the final configuration fits the original proximities data, one of the most important criteria is the interpretability of the final solution. The common space, illustrated in the chart, is certainly interpretable. If the space is rotated anticlockwise 90 degrees and then flipped, the nine towns approximate a map of Australia. In other words the MDS analysis has taken the original proximity data and represented these proximities as distances, in such a way that the towns now approximate the locations that would be found on a map of Australia.



Further evidence of the reasonably high level of fit between the original proximities and derived distances is also shown in the Shepard Diagram below, which plots the original, but transformed proximities, against the derived distances. As the fit improves, this relationship should increasingly approximate a straight line relationship with minimal dispersion.



### Practice example



In the previous working example three tourists judged the distances among nine towns in Australia. Their judgements were averaged and their average scores were used to develop a single matrix of proximities representing their judgements of the distances among towns.

In the practice example the three tourists are asked to use a five-point scale to make a judgement of the similarity among the nine towns. A score of 1 would indicate the pair of towns were very similar and a score of 5 would indicate they were very dissimilar. They are instructed they can base their judgement on any criteria they like.

In this example solutions will be derived in not only two dimensions but two-to-five dimensions. The data for the practice example are shown in Prac21.sav.

In undertaking the analysis, in the **Multidimensional Scaling: Plots:** sub-dialogue box, make sure that you have entered 2 for the **Minimum dimensions** and 2 for the **Maximum dimensions**. Run the analysis using the same configuration as you did before and record the value of Stress I.

Repeat the process for three through to five dimensions and record the Stress I values for each dimensional solution.

#### **Syntax**

PROXSCAL
VARIABLES=Perth Adelaide Melbourne Hobart Sydney Brisbane Darwin Cairns Broome
/SHAPE=LOWER
/INITIAL=SIMPLEX
TRANSFORMATION=ORDINAL (KEEPTIES)
PROXIMITIES=DISSIMILARITIES
/ACCELERATION=NONE
/CRITERIA=DIMENSIONS(3,3) MAXITER(100).

#### Output

Two dimensions

#### Stress and Fit Measures

Normalised Raw Stress	.02404
Stress-I	<sup>a</sup> .15504
Stress-II	a.44666
S-Stress	b.07219
Dispersion Accounted For	.97596
Tucker's Coefficient of Congruence	.98791

PROXSCAL minimises Normalised Raw Stress.

- a. Optimal scaling factor = 1.025.
- b. Optimal scaling factor = .977.

#### Three Dimensions

#### Stress and Fit Measures

Normalised Raw Stress	.00460
Stress-I	<sup>a</sup> .06782
Stress-II	a.26209
S-Stress	b.01602
Dispersion Accounted For	.99540
Tucker's Coefficient of Congruence	.99770

PROXSCAL minimises Normalised Raw Stress.

- a. Optimal scaling factor = 1.005.
- b. Optimal scaling factor = .995.

Stress and Fit Measures

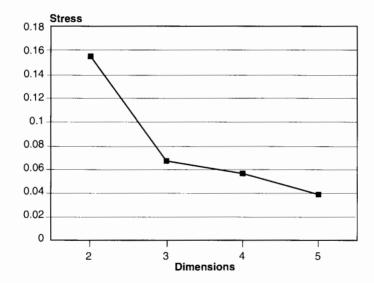
.00332
<sup>a</sup> .05760
<sup>a</sup> .26067
<sup>b</sup> .01101
.99668
.99834

PROXSCAL minimises Normalised Raw Stress.

- a. Optimal scaling factor = 1.003.
- b. Optimal scaling factor = .996.

When you have run the analyses, plot the values as a 'scree plot' yourself with each dimension on the X axis and the Stress I value on the Y axis.

You should have a plot similar to the following, which has an 'elbow' at the three-dimensional solution. This indicates that there is significant improvement in the fit between the two- and three-dimensional solution, but much less improvement between a three- and four-dimensional solution and between a four- and five-dimensional solution.



Start by looking at the plot for the three-dimensional solution. The final coordinates of each of the data points (towns) on the three dimensions are shown in the Final Coordinates output. It is a good idea to draw a cube and use the coordinates to plot the location of each of the data points (towns) on the three axes of the cube.

**Final Coordinates** 

		Dimension	
	1	2	3
Perth	.060	.526	.361
Adelaide	.333	.476	086
Melbourne	.531	437	013
Hobart	.470	286	542
Sydney	.409	.007	.328
Brisbane	279	434	.461
Darwin	642	016	345
Cairns	571	361	.036
Broome	312	.526	200

Although not as visually accurate as plotting the data points in a cube, you can look at the coordinates of each of the towns on each of the three dimensions. For example on dimension one Darwin, Cairns, Broome and Brisbane are at one end of the dimension, while Melbourne, Hobart, Sydney and Adelaide are at the other.

Now the question can be asked what is it that distinguishes the towns on this dimension. Is climate a useful label for dimension one, with one pole of this dimension being tropical/subtropical and the other pole being towns with temperate climates?

What labels would you apply to dimensions two and three? Is there a dimension that distinguishes towns on the basis of air quality, isolation, rainfall or the composition of the population?

This is where art meets science in the interpretation of MDS solutions!

## Working with output

This chapter provides an overview of how to work with output within the SPSS program.

## Editing output in the SPSS Viewer



The SPSS Viewer or Output Navigator Window, as it has been referred to in previous versions, provides a table of contents of all your output in a session. You can use this outline to navigate through your output and control the output display. Most actions undertaken in this pane have a corresponding effect on the contents pane. The outline pane is useful if you want to browse your results, to show or hide particular tables and charts, to change the display or ordering of results by moving particular items, or to move between the outline pane and other applications.

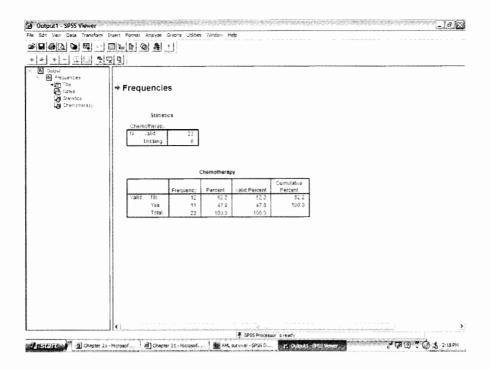
To start working with some output we first need to access some data. When the SPSS program is installed it comes with particular data sets. For the purpose of this chapter we will use one of the data sets available within the program.



#### To access the data and obtain some output

- 1 Double-click on the SPSS program icon to begin your SPSS session.
- The program will open to the SPSS for Windows dialogue box which will give you a number of choices, such as Run the tutorial, Type in data, etc. Ensure the Open an existing data source radio button is selected.
- 3 You will notice a number of different files are listed. Click on the filename *AML Survival*, then click on **OK**.
- 4 Click on the Analyze menu.
- 5 Click on Descriptive Statistics and then on Frequencies to open the Frequencies: dialogue box.
- 6 Select the variable you require (i.e. *Chemotherapy*) and click on the ▶ button to move the variable into the **Variable(s)**: box.
- 7 Click on OK.

You should now be in the SPSS Viewer and the output of the procedure just run should be displayed.

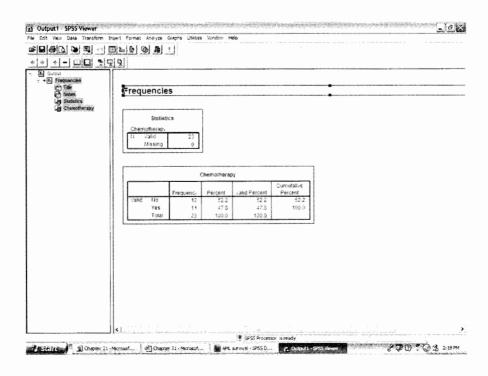


The output results can now be browsed by expanding, collapsing, hiding and showing particular aspects of the output.



#### To collapse and expand outline items

- Click on the small yellow box to the left of the outline item you want to collapse (i.e. the small yellow box next to the **Frequencies** item). You will notice that a small red arrow has appeared to the left of **Frequencies**, and that all sub-items beneath this output have been highlighted.
- 2 Select the **View** menu and click on **Collapse**. You will notice that the output in the contents pane has been collapsed and is no longer displayed.
- 3 To expand the item, repeat the above procedure, but this time click on **Expand** not **Collapse**. The output is again displayed.

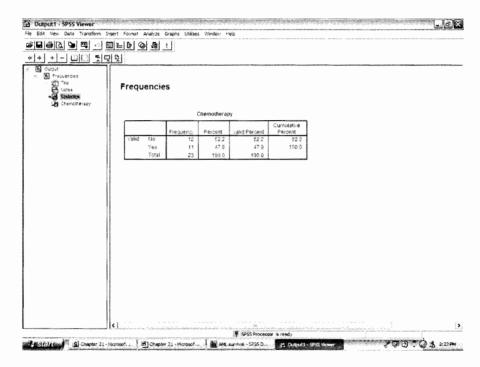


It is also possible to expand and collapse outline items without using the menu system. This is as simple as clicking the small box to the left of the outline item you wish to expand or collapse. Alternatively, you can use the **Collapse** button (-) and **Expand** button (+) on the toolbar.



#### To hide and show outline items

- Click on the outline sub-item titled *Statistics* within the **Frequencies** item. Again, you will notice that a small red arrow has appeared next to the *Statistics* sub-item and that this arrow is also highlighting the corresponding statistics table in the contents pane. You will also notice that the symbol (a book) next to the *Statistics* sub-item is open, indicating that the corresponding table is present in the contents pane. However, you will see that the book for the *Notes* sub-item is closed.
- 2 Select the **View** menu and click on **Hide**. You will notice that the corresponding output in the contents pane has disappeared and is no longer displayed.



3 To show or display the output again, repeat the above procedure but click on **Show** not Hide. The output is once again displayed.

Now it's your turn: show or display the output corresponding to *Notes* in the contents pane using the above procedure.

As was the case with collapsing and expanding output, it is possible to hide and show output without using the menu system. This is done by double-clicking your mouse on the book icon that corresponds to each outline item. Remember, if the book icon is open, then the output is displayed; if the book is closed, then the output has been temporarily hidden from view. Alternatively, show and hide buttons are available on the toolbar.

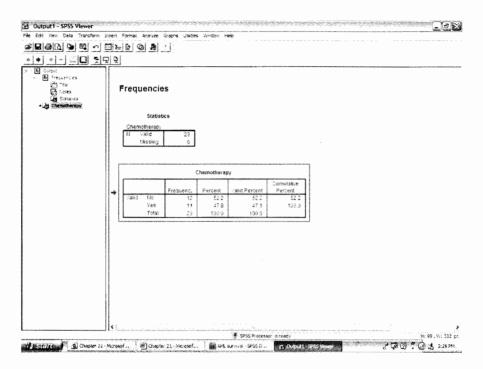
It is also possible to change the level of an outline item in the outline pane — that is, to promote or demote an item.



#### To change the level of an outline item

1 Click on the outline sub-item titled *Chemotherapy* within the **Frequencies** item. Again, you will notice that a small red arrow has appeared next to this sub-item and that this arrow is also highlighting the corresponding table in the contents pane. You will also notice that the book icon is open.

2 Select the **Edit** menu and click on **Outline** and then **Promote**. You will notice that the **Frequencies** table for *Chemotherapy* has now been promoted to an item of its own.

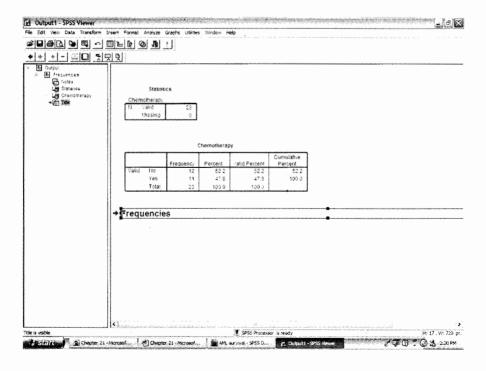


3 To demote the item, repeat the above procedure but click on **Demote** not **Promote**. Items can also be promoted and demoted by using the left arrow  $\rightleftharpoons$  and right arrow  $\rightleftharpoons$  keys on the SPSS Viewer toolbar.



#### To move outline items in the SPSS Viewer

- 1 Click an item in the outline or contents pane to select it (i.e. *title*).
- 2 Use the mouse to click and drag the item to the bottom of the sub-item listing.
- Release the mouse button just above where you want to place the moved item. You will notice that the title has also moved in the contents pane.

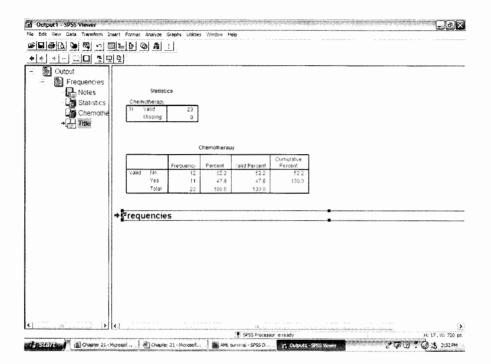


Items can also be cut, pasted or deleted in this manner using the **Edit** menu. To delete the *Notes* sub-item, select it and then press the delete button.



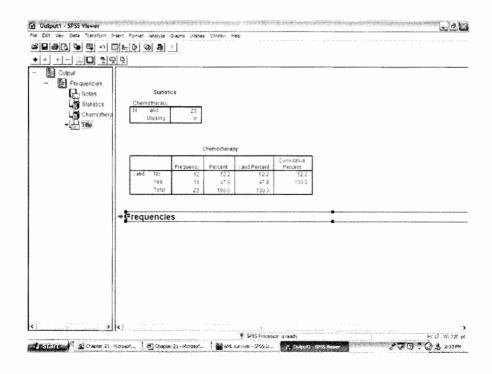
#### To change the size and font of items in the outline

- 1 Select the **View** menu and select **Outline Size**.
- Click on Medium. You will notice that the font size in the outline pane has increased in size.



- 3 Select the View menu again and select Outline Font.
- 4 Select **Arial Narrow** from the list of fonts and click on **OK**.

You will notice that the font style has now changed.



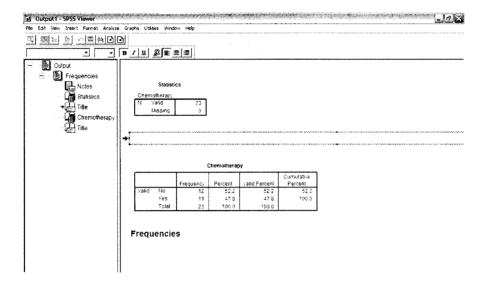
Because the font style and size has been changed it is now necessary to increase the size of the outline pane to be able to fully read the text in the outline pane. Using your mouse, position the cursor on the vertical bar separating the outline pane from the contents pane. Click on this bar and, holding the left mouse button down, move the bar to the right until the text in the outline pane is fully displayed, then release.



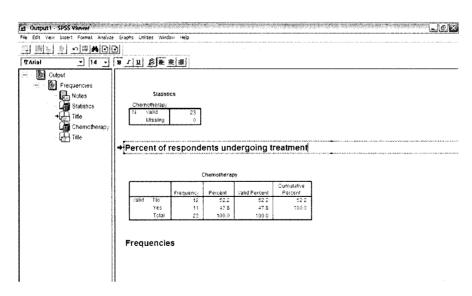
#### To add a new title or text item to the outline contents pane

You may wish to add new titles or text items to your output that are not connected to particular tables or charts.

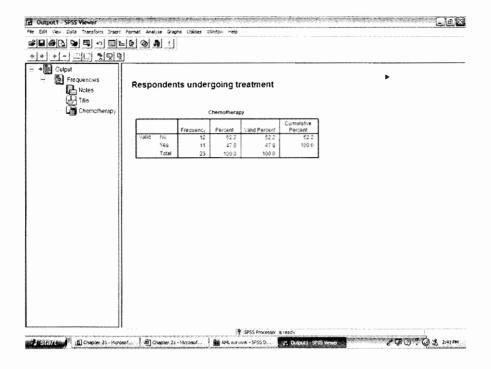
- 1 Click on the table, chart or object in the contents pane that will precede the title or text (e.g. *Statistics*). This means that the new title will be inserted beneath the *Statistics* table in the contents pane.
- 2 Select the **Insert** menu and click on **New Title**.



- 3 Double-click on the new object/box that has been created in the contents pane.
- 4 Enter the text you want at this location (i.e. *Respondents undergoing treatment*). You will notice that a new *Title* sub-item has been listed in the outline pane.



Now with two titles, use the delete key to delete the old title *Frequencies* and the *Statistics* table.



New text and new headings can also be added in this manner using the Insert menu.

## Modifying and improving charts for presentation



Now let us consider some chart outputs. After creating a chart, you may wish to modify it, either to provide more information about the data or to enhance the chart for presentation. This section will largely deal with enhancing charts for presentation.

Although the charts generated by SPSS contain the requested information in a logical format, you may find that some changes are required for presentational purposes to enhance the charts you have created. Such changes may include inserting titles and subtitles; adding a 3-D effect to a bar chart; changing the intercluster spacing between bars on a bar chart; removing axis titles; altering the orientation of labels; enlarging scale axes; and adding annotations such as average scores and other descriptive statistics.

As outlined earlier in the Preface, SPSS version 12.0 has a number of improved charting features including:

- · Better default chart appearance
- Support for long text strings and automatic text wrapping
- Control of default scale ranges using chart templates
- 3-D effects for pie and bar charts
- Improved choice of colour patterns.

Many of these functions can be achieved using the **Chart Editor** menu bar, which replaces the main menu bar when a Chart window is active. Remember you can tell when a window is active by looking for the red exclamation mark! at the bottom of the window in which you are operating. The **Chart Editor** menu bar contains ten menu items, including **File**, **Edit**, **View**, **Gallery**, **Chart**, **Series**, **Format**, **Statistics**, **Graph** and **Help**.

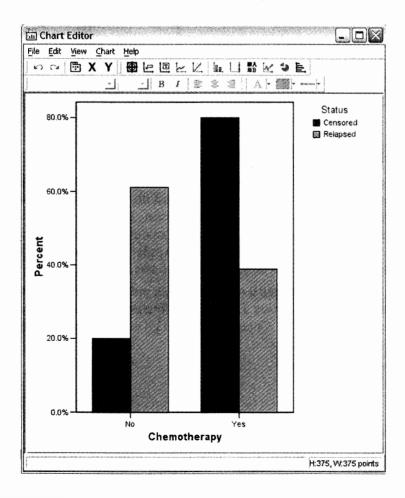
Let's first generate some output with which to work. You will remember that when working through the section of this chapter on 'Editing output in the SPSS Viewer', you accessed a data file already available within the program. If you haven't still got this file open, access it again.



#### To access the data and obtain a chart output

- 1 If you are not currently in the SPSS program, double-click on the SPSS program icon to begin your SPSS session.
- 2 Ensure the **Open an existing data source** radio button is highlighted.
- 3 You will notice a number of different files are listed. Click on **More files** and on **OK**. This will bring up more files in the program folder. Click on the filename *AML Survival* and then click on **OK**.
- 4 Select the **Graphs** menu and click on **Bar**.
- 5 Click on Clustered. Also ensure the Summaries for groups of cases radio button has been checked in the Data in chart are box.
- 6 Click on Define.
- 7 Select the variable which you require (i.e. *Chemotherapy*) and click on the button to move the variable into the **Category Axis:** box.
- 8 Select the variable by which you wish to define the above variable (i.e. *status*) and click on the button to move the variable into the **Define clusters by:** box.
- 9 In the Bars Represent box, ensure that the % of cases radio button has been checked.
- 10 Click on the Options pushbutton and uncheck the Display groups defined by missing values check box by clicking on this check box with your mouse.
- 11 Click on Continue and then OK.

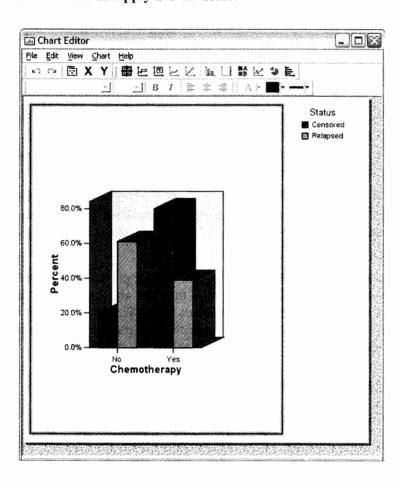
You should now be in the SPSS Viewer, where a bar graph of your results is displayed. Let's now open the Chart Editor to modify the chart and improve its presentation. This is achieved by double-clicking on the chart in the Viewer window to open the Chart Editor window. Also highlight the bars in the bar graph by clicking on the bar with your left mouse button. Now you're ready to modify the bar chart.





#### To create a 3-D effect

- 1 Click on the **Show Properties Window** tool located on the toolbar to open the **Properties** dialogue box.
- 2 Click on **Depth and Angle** tab and then click on the **3-D** radio button in the **Effect** box. You will notice that the bars in the chart have now changed.
- 3 Click on Apply and on Close.

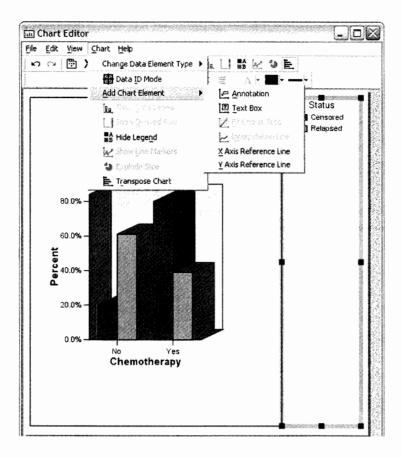




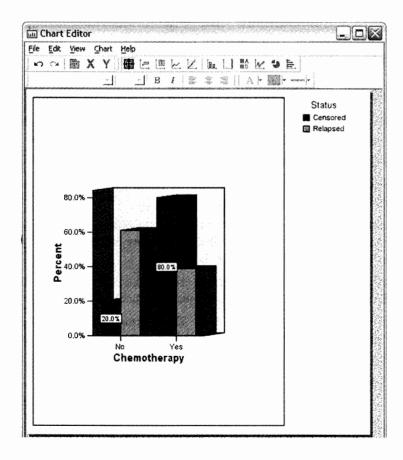
#### To modify and add chart attributes

Modifying and adding chart attributes is easy in the Chart Editor window. By double-clicking on any aspect of the chart, you can open the relevant **Properties** dialogue box to alter chart size, text, axis labels, number format, scale, fill, borders and so on. Take a few minutes to experiment with the chart currently in the editor. Always remember when modifying chart attributes to click on **Apply** to change the particular attribute and then to click on **Close** to close the dialogue box.

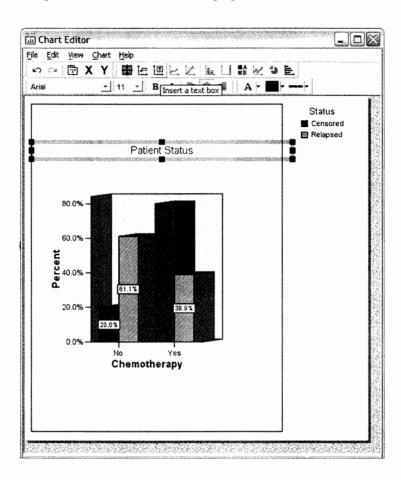
Select the **Chart** menu to open a range of further options such as **Add Chart Element** (annotations, text boxes), **Hide Legend** and **Transpose Chart**. These functions are also available on the toolbar. Once again have a go at using these functions yourself on the chart in the Chart Editor.



Another very useful tool within the chart editor is the Data ID Mode. If you click on this tool on the toolbar, and then click on a bar on the chart, the tool will attach the data value for that particular bar. Do the same for the other bars of the chart. To turn this function off, click on the Data ID Mode tool again.



Now create a title for the chart by inserting a text box above the chart. Just click on the Insert a text box tool and a box will appear in the chart window; type in a chart title and place the title above the bar graph.



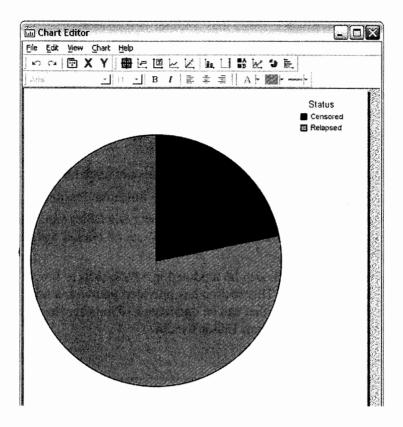
It really is very simple to make your charts look fabulous.

Now create a pie chart to modify.

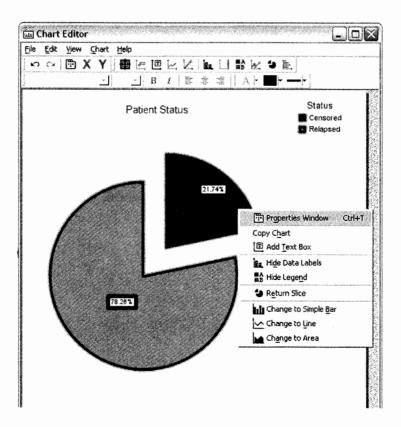


#### To create a pie chart

- 1 Select the **Graphs** menu. You will notice this can also be accessed from within the SPSS Viewer window as well as in the Data Editor window.
- 2 Click on **Pie...** to open the **Pie Charts** dialogue box.
- 3 Ensure the **Summaries for groups of cases** radio button is checked and then click on the **Define** pushbutton.
- 4 Select the variable you require from the variable list (i.e. *status*) and click on the 
  ▶ button to move the variable into the **Define Slices by:** box. Also make sure that the **% of cases** radio button is selected.
- 5 Click on the **Options** pushbutton and make sure that the **Display groups defined by missing values** check box is unchecked.
- 6 Click on **Continue** and then **OK**. This output has now been added to your Output Navigator/Viewer window, and your pie chart is displayed.



Try modifying the chart using the tools in the chart editor. Remember to double-click anywhere on the pie chart to activate the Chart Editor window – you will notice a few extra tools available to you, such as the Explode slice tool. To make changes, just double-click on the aspect of the chart you are trying to modify or click your right mouse button to open the menu of tools available to you, as illustrated in the figure below.



Now copy the chart and paste it into another application.



#### To copy and paste a chart into another application

- 1 Ensure the Chart Editor window is active.
- Select the **Edit** menu and click on **Copy Chart**. The chart has now been saved to the clipboard even though you cannot see it. You can view the chart in the clipboard viewer available in Windows if you wish.
- 3 Open a blank or existing file in a new application (for example, Microsoft Word).
- 4 Select the **Edit** menu in the new application and click on **Paste**.
- 5 Save your Word file and return to SPSS to finish the session.

SPSS version 12.0 has the capability to embed both pivot tables and charts interactively into other applications; that is, tables and charts can be copied and pasted into other applications.

Obviously, many other charts can be modified in SPSS, such as boxplots, scatterplots, histograms and line graphs. This section has provided you with a brief introduction to some of the basic procedures that can be undertaken to improve the presentation of bar graphs and pie charts in the Chart Editor window.

#### **Practice example**

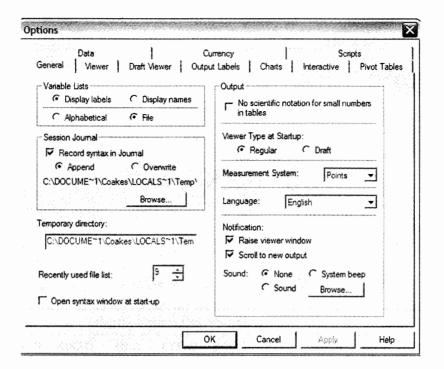
Using the data in the 'carpet' file, which is a sample data file available in SPSS version 12.0, use a bar chart and a pie chart to appropriately represent two of the variables in this data file. Experiment with modifying these charts for presentation now that you are more familiar with some of the functions. Then practise copying both charts into a Microsoft Word document.

## Viewer and Draft Viewer output



SPSS version 12.0 also provides you with the option of working in an SPSS Viewer or Draft Viewer format. This chapter has provided an overview of the SPSS Viewer function of working with pivot tables and charts, but the Draft Viewer function allows you to edit your output directly. Many regular users of SPSS who produce lots of data output prefer the ease of editing output using the Draft Viewer; however, if you are wanting to produce high-quality graphs and charts for reports, then the viewer function is probably preferable.

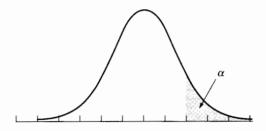
To change output type, select the **Edit** menu and click on **Options**. Ensure the **General** tab is selected and in the **Viewer Type at Start-Up** box select the **Draft** radio button, not the **Regular** radio button, which is the default option, as shown opposite. Click on **Apply** and then **OK**. For the Draft Viewer to be enabled, you will need to restart your SPSS program.





## **Appendix**

TABLE 1 + CRITICAL VALUES



				Upper	Tail Proi	oability (	α)		
Degrees of Freedom	0.15	0.10	0.05	0.025	0.015	0.01	0.005	0.001	0.0005
1	1.963	3.078	6.314	12.706	21.205	31.821	63.657	318.309	1273.155
2	1.386	1.886	2.920	4.303	5.643	6.965	9.925	22.327	44.703
3	1.250	1.638	2.353	3.182	3.896	4.541	5.841	10.215	16.326
4	1.190	1.533	2.132	2.776	3.298	3.747	4.604	7.173	10.305
5	1.156	1.476	2.015	2.571	3.003	3.365	4.032	5.893	7.976
6	1.134	1.440	1.943	2.447	2.829	3.143	3.707	5.208	6.788
7	1.119	1.415	1.895	2.365	2.715	2.998	3.499	4.785	6.082
8	1.108	1.397	1.860	2.306	2.634	2.896	3.355	4.501	5.617
9	1.100	1.383	1.833	2.262	2.574	2.821	3.250	4.297	5.291
10	1.093	1.372	1.812	2.228	2.527	2.764	3.169	4.144	5.049
11	1.088	1.363	1.796	2.201	2.491	2.718	3.106	4.025	4.863
12	1.083	1.356	1.782	2.179	2.461	2.681	3.055	3.930	4.717
13	1.079	1.350	1. <i>77</i> 1	2.160	2.436	2.650	3.012	3.852	4.597
14	1.076	1.345	1.761	2.145	2.415	2.625	2.977	3.787	4.499
15	1.074	1.341	1 <i>.75</i> 3	2.131	2.397	2.602	2.947	3.733	4.417
16	1.071	1.337	1.746	2.120	2.382	2.583	2.921	3.686	4.346
1 <i>7</i>	1.069	1.333	1.740	2.110	2.368	2.567	2.898	3.646	4.286
18	1.067	1.330	1.734	2.101	2.356	2.552	2.878	3.611	4.233
19	1.066	1.328	1.729	2.093	2.346	2.539	2.861	3.579	4.187
20	1.064	1.325	1.725	2.086	2.336	2.528	2.845	3.552	4.146
21	1.063	1.323	1.721	2.080	2.328	2.518	2.831	3.527	4.109
22	1.061	1.321	1.717	2.074	2.320	2.508	2.819	3.505	4.077
23	1.060	1.319	1.714	2.069	2.313	2.500	2.807	3.485	4.047
24	1.059	1.318	1.711	2.064	2.307	2.492	2.797	3.467	4.021
25	1.058	1.316	1.708	2.060	2.301	2.485	2.787	3.450	3.997
26	1.058	1.315	1.706	2.056	2.296	2.479	2.779	3.435	3.974
27	1.057	1.314	1.703	2.052	2.291	2.473	2.771	3.421	3.954
28	1.056	1.313	1. <i>7</i> 01	2.048	2.286	2.467	2.763	3.408	3.935
29	1.055	1.311	1.699	2.045	2.282	2.462	2.756	3.396	3.918
30	1.055	1.310	1.697	2.042	2.278	2.457	2.750	3.385	3.902
40	1.050	1.303	1.684	2.021	2.250	2.423	2.704	3.307	3.788
50	1.047	1.299	1.676	2.009	2.234	2.403	2.678	3.261	3.723
60	1.045	1.296	1.671	2.000	2.223	2.390	2.660	3.232	3.681
120	1.041	1.289	1.658	1.980	2.196	2.358	2.617	3.160	3.578
Z critical value	1.036	1.282	1.645	1.960	2.170	2.326	2.576	3.090	3.290
Level of Significance for a one-tailed test	0.15	0.10	0.05	0.025	0.015	0.01	0.005	0.001	0.0005
Level of Significance for a two-tailed test	0.30	0.20	0.10	0.05	0.03	0.02	0.01	0.002	0.001

TABLE 2 CHI-SQUARE DISTRIBUTION TABLE

The entries in this table give the critical value of  $X^2$  for the specified number of degrees of freedom and areas in the right tail.



1, 1000         0.099         0.999         <	Degrees						Upper Tail	il Areas					
00000         00000         00001         00000 <th< th=""><th>Freedom</th><th>0.995</th><th>0.99</th><th>0.975</th><th>0.95</th><th>6.0</th><th>0.75</th><th>0.25</th><th>0.1</th><th>0.05</th><th>0.025</th><th>0.01</th><th>0.005</th></th<>	Freedom	0.995	0.99	0.975	0.95	6.0	0.75	0.25	0.1	0.05	0.025	0.01	0.005
0.077         0.115         0.216         0.526         0.524         1.217         4.08         4.525         7.615         7.616         1.245           0.077         0.017         0.026         0.02	- 6	0.000	0.000	0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879
0.207         0.227         0.44         0.711         1 0.64         1 92.3         5.885         7.779         9 488         11.143         1 1.02           0.676         1.687         1.688         1.687         1.687         1.688         1.687         1.688         1.687         1.688	<b>4</b> 69	0.072	0.020	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838
0.676         0.854         0.831         1.145         1.610         2.675         6.225         1.100         1.2 832         1.5 00           0.676         0.876         0.875         1.237         1.639         1.640         1.610         1.2 832         1.600           0.876         0.876         0.876         1.287         1.640         1.610         1.610         1.610         1.610         1.610         1.610         1.600         1.	4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860
0.6%         0.87         1.237         1.635         2.204         3.455         7.841         1.0645         16.37	10	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.070	12.832	15.086	16.750
1,345   1,239   1,680   2,733   3,490   5,071   10,219   13,362   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,071   15,072   15,007	9	9.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548
1,334   1,647   2,180   2,733   3,480   5,677   10,199   11,332   1,648   1,535   20,080   1,344   1,647   2,180   2,733   4,168   5,897   11,339   14,684   16,919   19,032   2,0060   2,633   3,624   4,644   6,725   6,304   6,326   1,344   1,647   1,248   1,24	_	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278
1,735         2,088         2,700         3,325         4,168         5,899         11,389         14,644         16,919         19,033         21,666           2,603         3,053         3,816         4,575         5,578         7,84         13,701         17,275         19,675         20,433         23,272         24,725         26,687         27,84         19,812         21,920         24,725         26,887         27,84         19,812         21,920         24,725         26,887         21,898         19,812         21,920         24,725         26,887         21,898         19,812         21,920         24,725         26,887         21,898         19,812         21,926         24,725         21,898         21,924         21,	∞ •	1.344	1.647	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955
2.156         2.558         3.240         4.865         5.64         1.549         18.54         1.549         18.54         1.548         2.420           2.156         3.558         3.141         4.575         5.578         5.74         1.544         18.54         11.02         2.192         2.220           3.054         4.055         4.00         5.226         6.504         8.484         18.845         18.549         2.1026         2.337         2.220           4.055         4.60         5.229         6.577         7.790         10.165         17.117         2.1062         2.347         2.248         2.247         2.248         3.248 <th>٥ ;</th> <th>1.735</th> <th>2.088</th> <th>2.700</th> <th>3.325</th> <th>4.168</th> <th>5.899</th> <th>11.389</th> <th>14.684</th> <th>16.919</th> <th>19.023</th> <th>21.666</th> <th>23.589</th>	٥ ;	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589
3.57         3.57         4.04         5.22         6.30         6.48         1.84         1.84         1.87         21.02         23.37         26.72           3.57         4.60         5.22         6.20         6.57         7.04         9.299         15.84         1.81         21.02         23.37         26.11           4.60         5.22         6.26         7.24         6.22         7.24         8.47         1.037         18.24         22.307         24.96         23.47         26.11           4.60         5.22         6.22         7.24         8.54         1.1037         18.24         22.307         24.96         27.88         30.191         32.00           5.67         6.40         7.92         1.1037         18.24         22.307         24.96         27.88         30.191         33.00           6.44         7.63         8.64         7.01         8.24         9.30         10.10         11.452         22.307         24.96         28.88         31.52         26.11         33.409           6.45         7.01         8.231         10.81         12.443         12.46         22.30         24.96         28.88         31.52         26.11         33.40	2:	2.156	2.558	3.24/	3.940	4.865	6./3/	12.549	15.98/	18.307	20.483	23.209	25.188
35.65         4.107         5.009         5.892         7.041         9.299         15.984         19.812         22.362         24.756         27.688           4.075         5.629         6.571         7.790         10.165         17.17         21.064         23.985         24.176         27.688           5.142         5.229         6.269         6.571         7.790         10.165         17.17         21.064         24.996         27.488         30.500           5.142         5.229         6.265         7.640         8.742         10.085         22.307         24.996         27.888         30.191         30.000           6.265         7.015         8.221         10.085         12.645         22.308         22.388         23.685         31.800         30.191         31.000           6.265         7.015         8.231         10.085         11.541         11.451         31.040         32.88         30.191         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800         31.800	12	3.074	3.571	4.404	5.226	6.304	8.438	14.845	18.549	21.026	23.337	26.217	28.300
4075         4660         5.629         6.571         7790         10 165         17117         21 064         23 685         26 119         29 141           4,01         51229         6.522         7.261         8.547         11 037         18 245         22 307         24 996         22 488         30 578           5,47         5847         6.568         7.564         8.672         10 085         12 792         20 489         24 789         27 88         30 191           6,587         7.633         8.931         9.390         10 865         12 472         20 489         24 789         27 88         30 191           6,844         7.633         8.907         10.117         11 651         14 562         22 718         22 204         26 48         30 581           6,844         7.635         8.034         8.043         9.30         10.117         11 651         14 562         22 718         22 204         30 141         31 409           8 034         8.034         8.034         9.30         10.117         11 641         14 562         22 718         22 704         31 170         31 190           8 034         8.035         10.082         11 60.83         14 40	13	3.565	4.107	5.009	5.892	7.041	9.299	15.984	19.812	22.362	24.736	27.688	29.819
4 601         5 229         6 262         7 261         8 547         11 037         18 245         22 307         24 996         27 488         30 578           5.642         5.684         7.564         8.672         10.085         12.792         20.489         22.307         24.996         27 488         30.191           5.687         5.687         5.688         7.564         8.672         10.085         12.792         20.489         22.396         28.845         30.191         33.000           6.286         7.015         8.231         8.997         10.117         11.651         14.562         22.188         27.847         32.849         33.898         30.191         33.000         34.805         34.805         30.191         33.190         32.877         33.409         33.191         34.170         34.170         37.547         35.479         38.932         36.191         37.72         38.057         35.479         38.932         36.191         37.72         38.057         35.479         38.932         36.781         38.932         36.781         38.932         36.781         38.932         36.781         38.932         36.781         38.932         36.781         38.932         36.781         36.781	14	4.075	4.660	5.629	6.571	7.790	10.165	17.117	21.064	23.685	26.119	29.141	31.319
5.142         5.812         6.908         7.962         9.312         11.912         10.948         23.542         26.266         28.845         32.000           5.647         6.408         7.564         8.008         7.962         9.312         11.912         10.948         21.542         26.26         28.845         30.191         33.000           6.265         7.015         8.231         10.375         21.605         25.989         28.845         30.191         34.805           6.265         7.015         8.231         10.342         21.605         25.989         28.869         31.524         34.805           8.034         8.234         8.260         9.591         10.817         11.243         12.443         24.935         29.615         31.610         34.805         34.806         34.806         34.806         3	15	4.601	5.229	6.262	7.261	8.547	11.037	18.245	22.307	24.996	27.488	30.578	32.801
5.697         6.408         7.564         8.672         10.085         12.792         20.489         24.769         27.587         30.191         33.409           6.265         7.015         8.677         10.61         11.651         11.651         12.792         20.489         27.287         30.191         33.409           6.264         7.633         8.907         10.117         11.651         12.443         15.422         22.714         30.144         32.885         30.140         34.170         34.885           8.034         8.624         10.882         11.591         11.240         16.344         22.981         22.041         32.817         34.79         38.895         34.79         38.784         40.289           8.634         9.540         10.882         12.401         16.441         17.240         26.033         33.924         36.781         40.289           9.260         10.186         11.689         13.091         14.488         18.137         22.141         32.007         35.723         36.781         40.289           9.860         10.186         13.448         18.37         22.441         32.007         35.782         40.644         44.394           10.520	91	5.142	5.812	806.9	7.962	9.312	11.912	19.369	23.542	26.296	28.845	32.000	34.267
6.265         7.015         8.231         9.390         10.865         13.675         21.605         25.989         28.869         31.526         34.805           6.265         7.015         8.231         9.390         10.117         11.651         14.652         22.718         27.204         30.144         32.852         34.919           7.434         8.266         9.591         10.881         12.443         16.452         23.828         28.415         32.877         35.172         37.170         37.879         37.889         37.814         37.8	17	5.697	6.408	7.564	8.672	10.085	12.792	20.489	24.769	27.587	30.191	33.409	35.718
6.844         7.633         8.907         10.117         11.651         14.552         22.718         27.204         30.144         32.832         36.191           6.844         7.633         8.907         10.117         11.651         12.403         13.240         16.344         24.935         29.615         30.141         32.832         36.191           8.034         8.897         10.283         11.591         13.240         16.344         24.935         29.615         30.671         34.70         35.647         34.70         36.749         36.791         36.791         36.791         36.792         36.791         36.791         36.791         36.792         36.791         36.791         36.792         36.791         36.793         36.791         36.791         36.791         36.791         36.791         42.993         36.731         36.731         42.993         36.731         36.731         42.993         36.741         40.113         42.994         42.994           11.160         11.524         13.624         19.037         28.241         36.741         42.995         36.741         42.996         46.942         42.995         46.942         42.996         46.942         42.996         42.996         42.996 </th <th><b>8</b> 9</th> <th>6.265</th> <th>7.015</th> <th>8.231</th> <th>9.390</th> <th>10.865</th> <th>13.675</th> <th>21.605</th> <th>25.989</th> <th>28.869</th> <th>31.526</th> <th>34.805</th> <th>37.156</th>	<b>8</b> 9	6.265	7.015	8.231	9.390	10.865	13.675	21.605	25.989	28.869	31.526	34.805	37.156
7.434         8 260         9.591         10.851         12.443         15.422         23.828         28.412         31.410         34.70         37.506           8.034         8 826         10.283         11.591         13.240         15.344         26.039         30.813         33.924         35.772         38.932           8.643         9.542         10.085         12.338         14.041         17.240         26.039         30.813         33.924         36.781         40.280           9.260         10.196         11.689         13.091         14.848         18.137         27.141         32.007         35.172         38.932         40.286         41.638           9.260         10.196         11.689         13.091         14.848         18.137         27.141         32.007         35.172         38.932         41.638         40.286         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         41.638         42.640         42.860         42.640         42.860         42.642         40.646         42.860         42.862         40.646         42.860         42.862	6	6.844	7.633	8.907	10.117	11.651	14.562	22.718	27.204	30.144	32.852	36.191	38.582
8.034         9.889/1 (1.284)         10.283         11.591         17.240         16.344         24.735         29.615         32.671         35.479         38.932           9.260         10.1085         12.338         14.041         17.240         26.039         30.813         33.924         36.781         40.289           9.260         10.1085         12.401         13.848         15.659         19.037         28.241         33.196         36.415         39.364         42.980           10.520         11.524         13.120         14.611         16.473         19.939         29.339         34.382         37.652         40.289         42.980           10.520         11.160         12.198         14.611         16.473         19.839         29.339         34.481         42.980         42.980           11.100         12.198         14.573         16.151         18.114         21.244         31.623         36.415         39.344         42.980           11.100         12.198         14.621         16.5379         17.292         20.843         30.435         36.415         39.344         42.980           11.100         12.198         16.928         18.739         22.657         37.652 <th>200</th> <th>7.434</th> <th>8.260</th> <th>9.591</th> <th>10.851</th> <th>12.443</th> <th>15.452</th> <th>23.828</th> <th>28.412</th> <th>31.410</th> <th>34.170</th> <th>37.566</th> <th>39.997</th>	200	7.434	8.260	9.591	10.851	12.443	15.452	23.828	28.412	31.410	34.170	37.566	39.997
8.643         7.542         10.982         12.338         14.041         17.240         26.039         30.813         33.924         36.781         40.289           9.266         10.196         11.689         13.091         14.848         18.137         22.141         33.007         33.172         38.076         41.638           9.886         10.196         11.689         12.093         29.339         34.382         37.652         40.646         44.314           10.520         11.524         13.120         14.611         16.473         19.939         29.339         34.382         37.652         40.646         44.314           11.160         12.198         13.844         15.379         17.292         20.843         30.435         33.645         40.646         44.314           11.160         12.198         13.844         15.379         17.292         20.843         30.435         33.641         40.143         42.980           11.160         12.198         12.844         15.379         17.729         20.843         30.435         33.482         41.923         41.964         42.940           11.160         12.198         13.8493         22.657         32.657         33.711         <	22	8.034	8.897	10.283	11.591	13.240	16.344	24.935	29.615	32.671	35.479	38.932	41.401
7.260         10.876         11.089         15.071         14.848         18.13         27.141         35.172         38.070         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         41.036         42.980         42.981         42.980         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.981         42.282         42.982         42.981 <th>22</th> <th>8.643</th> <th>9.542</th> <th>10.982</th> <th>12.338</th> <th>14.04</th> <th>17.240</th> <th>26.039</th> <th>30.813</th> <th>33.924</th> <th>36.781</th> <th>40.289</th> <th>42.796</th>	22	8.643	9.542	10.982	12.338	14.04	17.240	26.039	30.813	33.924	36.781	40.289	42.796
7.880         10.550         12.401         13.848         15.037         17.037         25.176         35.176         35.413         37.554         42.700           10.50         11.524         11.524         15.379         17.292         20.843         30.435         35.652         40.646         44.314           11.60         12.18         13.844         15.379         17.292         20.843         30.435         35.633         38.885         41.923         45.042           11.808         12.878         14.573         16.151         18.114         21.749         31.528         36.741         40.113         43.195         46.963           11.808         12.878         16.572         20.879         22.657         32.620         37.916         41.923         49.548           12.461         18.793         22.657         23.560         24.478         36.877         44.461         48.278           13.121         14.256         16.591         18.793         22.657         32.300         35.887         41.422         44.985         45.772         49.588           13.12         16.551         18.291         20.072         22.271         26.390         36.973         42.585 <td< th=""><th>2 2</th><th>7.260</th><th>10.196</th><th>10,401</th><th>13.091</th><th>14.848</th><th>18.13/</th><th>27.141</th><th>32.00/</th><th>30.172</th><th>38.070</th><th>41.038</th><th>44.181</th></td<>	2 2	7.260	10.196	10,401	13.091	14.848	18.13/	27.141	32.00/	30.172	38.070	41.038	44.181
10.520	4 6	988.6	10.856	12.401	13.848	15.659	19.037	28.241	33.196	36.415	39.364	42.980	42.038
11.808         12.878         15.844         15.577         17.272         20.843         30.453         30.803         41.723         43.195         45.043           11.808         12.878         14.573         16.151         18.114         21.749         31.528         36.741         40.113         43.195         46.963           11.808         12.875         16.278         16.791         18.939         22.657         32.620         37.916         40.137         44.461         48.278         48.278         48.278         48.278         48.278         48.278         48.278         48.278         48.278         48.278         48.278         46.979         50.892         52.191         49.800         40.256         43.773         46.979         50.892         52.191         49.800         40.256         43.773         46.979         50.892         52.191         40.805         40.805         46.097         46.097         50.892         52.191         40.258         44.780         46.979         50.892         52.191         47.400         50.725         52.191         47.400         50.725         52.191         47.400         50.725         52.191         47.400         50.725         52.475         52.445         52.445	7 6	0.520	1.324	13.120	14.01	17.202	7.434	27.337	34.362	200.70	40.040	44.3.4	40.720
12.461         13.565         15.30         16.728         18.939         22.657         32.620         37.916         41.337         44.725         48.278           13.121         14.256         16.047         17.708         19.768         23.567         33.62         37.916         41.337         44.925         49.588           13.121         14.256         16.791         18.493         20.599         24.478         34.800         40.256         43.773         46.979         50.892           13.131         14.458         15.655         17.539         19.281         21.434         25.390         35.887         41.422         44.985         46.979         50.892           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.134         16.501         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.644         24.903         46.059         49.802         53.203         57.342           20.707         22.164	2,0	11.808	12.878	13.644	15.379	18 11 4	20.043	31.528	34.741	30.003	41.723	45.042	40.270
13.121         14.256         16.047         17.708         19.768         23.567         33.711         39.087         42.557         45.722         49.588           13.121         14.256         16.047         17.708         19.768         23.567         33.711         39.087         42.557         45.722         49.588           13.787         14.953         16.739         19.281         21.434         25.390         35.887         41.422         44.985         46.979         50.892           14.458         15.655         17.539         19.281         21.434         25.390         35.887         41.422         44.985         46.979         50.892           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.815         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.644         24.903         48.602         51.96         55.746         55.758         59.342         63.691           17.192         18.509	28	12.461	13.565	15.308	16.928	18.939	22.657	32.620	37.916	41.337	44.46]	48.278	50.994
13.787         14.953         16.791         18.493         20.599         24.478         34.800         40.256         43.773         46.979         50.892           14.458         15.655         17.539         19.281         21.434         25.390         35.887         41.422         44.985         48.232         52.191           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.815         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.644         23.952         28.136         40.223         46.059         49.802         51.966         56.061           17.192         18.509         20.569         22.054         40.223         46.059         49.802         53.203         57.342           20.707         22.164         24.459         52.294	5	13.121	14.256	16.047	17.708	19.768	23.567	33.711	39.087	42.557	45.722	49.588	52.335
14.458         15.655         17.539         19.281         21.434         25.390         35.887         41.422         44.985         48.232         52.191           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486         53.486           15.815         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.644         23.952         28.136         39.141         44.903         48.602         51.966         56.061           17.192         18.509         20.569         22.054         40.223         46.059         49.802         53.203         57.342           20.707         22.164         24.433         26.509         29.051         33.60         45.616         51.805         59.342         63.691           35.534         37.485         40.482         46.459	30	13.787	14.953	16.791	18.493	20.599	24.478	34.800	40.256	43.773	46.979	50.892	53.672
15.134         16.362         18.291         20.072         22.271         26.304         36.973         42.585         46.194         49.480         53.486           15.815         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.664         23.952         28.136         39.141         44.903         48.602         51.966         56.061           17.192         18.509         20.569         22.465         24.797         29.054         46.059         49.802         53.203         57.342           20.707         22.164         24.433         26.509         29.051         33.60         45.616         51.805         59.342         63.691           35.534         37.485         40.482         44.459         52.294         66.981         74.397         79.082         83.298         88.379           83.852         86.923         91.573         95.705         100.624         109.220         130.055         140.233         146.567         152.211         158.950         1	31	14.458	15.655	17.539	19.281	21.434	25.390	35.887	41.422	44.985	48.232	52.191	55.002
15.815         17.073         19.047         20.867         23.110         27.219         38.058         43.745         47.400         50.725         54.775           16.501         17.789         19.806         21.664         23.952         28.136         39.141         44.903         48.602         51.966         56.061           17.192         18.509         20.569         22.465         24.797         29.054         40.223         46.059         49.802         53.203         57.342           20.707         22.164         24.433         26.509         29.051         33.660         45.616         51.805         55.758         59.342         63.691           35.534         37.485         40.482         43.188         46.459         52.294         66.981         74.397         79.082         83.298         88.379           83.852         86.923         91.573         95.705         100.624         109.220         130.055         140.233         146.567         152.211         158.950         1	32	15.134	16.362	18.291	20.072	22.271	26.304	36.973	42.585	46.194	49.480	53.486	56.328
16.501         17.789         19.806         21.664         23.952         28.136         39.141         44.903         48.602         51.966         56.061           17.192         18.509         20.569         22.465         24.797         29.054         40.223         46.059         49.802         53.203         57.342           20.707         22.164         24.433         26.509         29.051         33.660         45.616         51.805         55.758         59.342         63.691           35.534         37.485         40.482         44.459         52.294         66.981         74.397         79.082         83.298         88.379           83.852         86.923         91.573         95.705         100.624         109.220         130.055         140.233         146.567         152.211         158.950         1	33	15.815	17.073	19.047	20.867	23.110	27.219	38.058	43.745	47.400	50.725	54.775	57.648
17.192         18.509         20.569         22.465         24.797         29.054         40.223         46.059         49.802         53.203         57.342           20.707         22.164         24.433         26.509         29.051         33.660         45.616         51.805         55.758         59.342         63.691           35.534         37.485         40.482         43.188         46.459         52.294         66.981         74.397         79.082         83.298         88.379           83.852         86.923         91.573         95.705         100.624         109.220         130.055         140.233         146.567         152.211         158.950         1	34	16.501	17.789	19.806	21.664	23.952	28.136	39.141	44.903	48.602	51.966	56.061	58.964
20,707 22.164 24,433 26.509 29.051 33.660 45.616 51.805 55,758 59.342 63.691 35.534 37.485 40,482 43.188 46.459 52.294 66.981 74.397 79.082 83.298 88.379 83.852 86.923 91.573 95.705 100.624 109.220 130.055 140.233 146.567 152.211 158.950 1	35	17.192	18.509	20.569	22.465	24.797	29.054	40.223	46.059	49.802	53.203	57.342	60.275
35,534 37,485 40,482 43,188 46,459 52.294 66,981 74,397 79,082 83,278 88,379 83,852 86,923 91,573 95,705 100,624 109,220 130,055 140,233 146,567 152,211 158,950 1	9	20.707	22.164	24.433	26.509	29.051	33.660	45.616	51.805	55.758	59.342	63.691	66.766
83.832 86.923 71.373 73.703 100.624 107.220 130.033 140.233 146.367 132.211 138.730	98	35.534	37.485	40.482	43.188	46.459	52.294	130.051	74.397	79.082	83.298	88.3/9	91.952
	2	83.832	80.723		607.6%	100.024	109.220	130.055	140.233	140.30/	137.761	138.730	103.048

TABLE 3 THE F DISTRIBUTION TABLE

a. Area in the Right Tail under the F Distribution Curve = 0.01



Degrees of							Deg	Degrees of Freedom – Numerator	– mopaa	Numerat	or						
Penominator Denominator	-	2	3	4	2	9	7	80	6	10	=	12	13	14	15	16	17
-	4052.185	4999.340	5403.534	5624.257	5763.955	5858.950	5928.334	5980.954	6022.397	6055.925	6083.399	6106.682	6125.774	6143.004	6156.974	6170.012	6181.188
7	98.502	99.000	99.164	99.251	99.302		99.357	99.375	99.390	99.397	99.408	99.419	99.422	99.426	99.433	99.437	99.441
က	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228	27.132	27.052	26.983	26.924	26.872	26.826	26.786
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452	14.374 -	14.306	14.249	14.198	14.154	14.114
2	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.963	9.888	9.825	9.770	9.722	9.680	9.643
9	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790	7.718	7.657	7.605	7.559	7.519	7.483
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538	6.469	6.410	6.359	6.314	6.275	6.240
œ	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734	5.667	5.609	5.559	5.515	5.477	5.442
6	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178	5.111	5.055	5.005	4.962	4.924	4.890
0	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772	4.706	4.650	4.601	4.558	4.520	4.487
=	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462	4.397	4.342	4.293	4.251	4.213	4.180
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220	4.155	4.100	4.052	4.010	3.972	3.939
13	9.074	6.701	5.739	5.205	4.862	4.620	4.44]	4.302	4.191	4.100	4.025	3.960	3.905	3.857	3.815	3.778	3.745
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864	3.800	3.745	3.698	3.656	3.619	3.586
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730	3.666	3.612	3.564	3.522	3.485	3.452
91	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.616	3.553	3.498	3.451	3.409	3.372	3.339
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593	3.518	3.455	3.401	3.353	3.312	3.275	3.242
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.434	3.371	3.316	3.269	3.227	3.190	3.158
16	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.360	3.297	3.242	3.195	3.153	3.116	3.084
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294	3.231	3.177	3.130	3.088	3.051	3.018
12	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.236	3.173	3.119	3.072	3.030	2.993	2.960
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.184	3.121	3.067	3.019	2.978	2.941	2.908
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.137	3.074	3.020	2.973	2.931	2.894	2.861
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.094	3.032	2.977	2.930	2.889	2.852	2.819
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056	2.993	2.939	3.892	2.850	2.813	2.780
76	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	3.021	2.958	2.904	2.857	2.815	2.778	2.745
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.988	2.926	2.872	2.824	2.783	2.746	2.713
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.959	2.896	2.842	2.795	2.753	2.716	2.683
56	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.931	2.868	2.814	2.767	2.726	2.689	2.656
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.906	2.843	2.789	2.742	2.700	2.663	2.630
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727	2.665	2.611	2.563	2.522	2.484	2.451
9	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.559	2.496	2.442	2.394	2.352	2.315	2.281
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.399	2.336	2.282	2.234	2.191	2.154	2.119

or         18         19         20         21         22           6191.432         6200.746         6208.662         6216.113         6223.097           99.444         99.448         99.451         99.455           26.751         26.79         26.604         26.63           26.751         26.79         26.604         26.63           26.751         26.79         26.644         26.539           26.79         26.80         26.644         26.63           26.79         26.80         26.644         26.63           26.79         26.80         26.64         26.63           2609         5.80         9.553         9.528         9.506           27.451         7.422         7.396         7.372         7.311           4.860         4.833         4.808         4.786         4.765           4.860         4.123         4.099         4.057         4.057           4.150         4.123         4.099         4.057         4.057           4.457         4.430         4.405         4.383         4.363           3.716         3.689         3.836         3.816           3.716         3.689 <th>Degrees of</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>Degrees</th> <th>Degrees of Freedom</th> <th>m - Numerator</th> <th>rerator</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	Degrees of							Degrees	Degrees of Freedom	m - Numerator	rerator						
6191.432         6200.746         6208.662         6216.113         6223.097           99.444         99.448         99.445         99.445         99.445           26.751         26.890         26.639         26.639         13.944         13.970           9.609         9.580         9.553         9.528         7.351         6.639           7.451         7.422         7.396         7.372         7.351         6.209         6.181         6.132         6.111         5.56         7.351         6.209         6.181         6.158         4.786         4.765         4.765         4.786         4.765         4.363         3.816         3.326         3.330         3.310         3.265         3.237         3.216         3.321         3.326	Freedom Denominator	81	91	20	21	22	23	24	25	26	27	28	29	30	40	09	120
99,444 99,448 99,448 99,451 26,751 26,719 26,690 26,644 14,079 14,048 14,019 13,994 9,609 9,580 9,553 9,528 7,451 7,422 7,396 7,372 6,209 6,181 6,155 6,132 5,412 5,384 5,339 4,077 3,910 3,883 3,858 3,836 3,716 3,689 3,665 3,643 3,716 3,689 3,665 3,483 3,716 3,689 3,665 3,483 3,423 3,396 3,355 3,310 3,283 3,356 3,310 3,283 3,356 3,310 3,283 3,357 3,310 3,283 2,916 2,931 2,904 2,880 2,916 2,931 2,904 2,880 2,916 2,931 2,904 2,880 2,576 2,751 2,762 2,738 2,716 2,751 2,762 2,738 2,716 2,751 2,762 2,738 2,716 2,751 2,762 2,738 2,716 2,751 2,762 2,738 2,716 2,751 2,762 2,738 2,716 2,751 2,764 2,699 2,677 2,751 2,688 2,664 2,677 2,751 2,688 2,664 2,552 2,600 2,573 2,549 2,574 2,251 2,231 2,394 2,366 2,251 2,223 2,175	_	l	l		6216.113	_	6228.685	6234.273	6239.861	6244.518	6249.174	6252.900	6257.091	6260.350	6286.427	6312.970	6339.513
26.751     26.751     26.751     26.640     26.664       14.079     14.048     14.019     13.994       9.609     9.580     9.553     9.528       7.451     7.422     7.396     7.372       6.209     6.181     6.155     6.132       6.209     6.181     6.155     6.132       6.209     6.181     6.155     6.132       4.860     4.833     4.808     4.786       4.457     4.430     4.405     4.383       4.150     4.123     4.099     4.077       3.910     3.883     3.858     3.836       3.716     3.689     3.665     3.643       3.423     3.396     3.556     3.483       3.129     3.162     3.139       3.12     3.186     3.162     3.139       3.12     3.186     3.162     3.139       3.12     3.186     3.162     3.130       2.989     2.962     2.938     2.916       2.989     2.962     2.938     2.916       2.931     2.904     2.880     2.576       2.751     2.724     2.699     2.677       2.751     2.751     2.626     2.632     2.609       2.	7	99.444	99.448	99.448	99.451	99.455	99.455	99.455	99.459	99.462	99.462	99.462	99.462	99.466	99.477	99.484	99.491
14,079       14,048       14,019       13,994         9,609       9,580       9,553       9,528         7,451       7,422       7,396       7,372         6,209       6,181       6,155       6,132         6,209       6,181       6,155       6,132         6,209       6,181       6,155       6,132         4,860       4,833       4,808       4,786         4,457       4,430       4,405       4,383         4,150       4,123       4,099       4,077         3,910       3,883       3,858       3,836         3,716       3,689       3,665       3,643         3,716       3,689       3,665       3,643         3,716       3,883       3,858       3,836         3,716       3,689       3,665       3,643         3,716       3,899       3,665       3,643         3,123       3,186       3,162       3,137         3,128       3,101       3,077       3,055         3,128       3,101       3,077       3,055         3,054       2,962       2,938       2,916         2,897       2,897       2,805	ဗ	26.751	26.719	26.690	26.664	26.639	26.617	26.597	26.579	26.562	26.546	26.531	26.517	26.504	26.411	26.316	26.221
9.609       9.580       9.523       9.528         7.451       7.422       7.396       7.372         6.209       6.181       6.155       6.132         6.209       6.181       6.155       6.132         4.860       4.833       4.808       4.786         4.457       4.430       4.405       4.383         4.150       4.123       4.099       4.077         3.910       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.716       3.689       3.665       3.483         3.716       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.716       3.689       3.655       3.483         3.716       3.883       3.858       3.836         3.718       3.101       3.077       3.055         3.128       3.101       3.077       3.055         3.054       3.027       3.03       2.916         2.989       2.962       2.938       2.916         2.789       2.720       2.738       2.716         2.775       2.724       2.697	4	14.079	14.048	14.019	13.994	13.970	13.949	13.929	13.911	13.894	13.878	13.864	13.850	13.838	13.745	13.652	13.558
7.451 7.422 7.396 7.372 6.209 6.181 6.155 6.132 5.412 5.384 5.359 5.336 4.860 4.833 4.808 4.786 4.457 4.430 4.405 4.383 4.150 4.123 4.099 4.077 3.910 3.883 3.858 3.836 3.716 3.689 3.665 3.643 3.556 3.529 3.505 3.483 3.423 3.396 3.372 3.350 3.310 3.283 3.259 3.237 3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 2.038 2.916 2.931 2.904 2.880 2.857 2.879 2.962 2.938 2.716 2.751 2.704 2.880 2.778 2.751 2.704 2.890 2.752 2.607 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.574 2.600 2.573 2.526 2.201 2.394 2.369 2.346 2.251 2.223 2.198 2.175	ĸ	609.6	9.580	9.553	9.528	9.506	9.485	9.466	9.449	9.433	9.418	9.404	9.391	9.379	9.291	9.202	9.112
6.209 6.181 6.155 6.132 5.412 5.384 5.359 5.336 4.860 4.833 4.808 4.786 4.457 4.430 4.405 4.383 4.150 4.123 4.099 4.077 3.910 3.883 3.858 3.836 3.716 3.689 3.665 3.643 3.556 3.529 3.505 3.483 3.423 3.396 3.372 3.350 3.310 3.283 3.259 3.237 3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.738 2.716 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.677 2.751 2.724 2.699 2.579 2.653 2.656 2.632 2.609 2.653 2.656 2.579 2.600 2.573 2.549 2.552 2.201 2.334 2.346 2.251 2.223 2.175	9	7.451	7.422	7.396	7.372	7.351	7,331	7.313	7.296	7.281	7.266	7.253	7.240	7.229	7.143	7.057	6.969
5.412       5.384       5.359       5.336         4.860       4.833       4.808       4.786         4.457       4.430       4.405       4.383         4.150       4.123       4.099       4.077         3.910       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.729       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.350         3.128       3.101       3.077       3.055         3.054       3.027       3.03       2.981         2.989       2.962       2.938       2.916         2.931       2.904       2.880       2.857         2.879       2.762       2.738       2.716         2.789       2.762       2.738       2.716         2.751       2.724       2.699       2.677         2.751       2.724       2.699       2.607         2.653       2.656       2.602       2.579         2.600       2.574       2.552         2.600       2.751       2.251         2.251 <t< th=""><th>^</th><th>6.209</th><th>6.181</th><th>6.155</th><th>6.132</th><th>6.111</th><th>6.092</th><th>6.074</th><th>6.058</th><th>6.043</th><th>6.029</th><th>910.9</th><th>6.003</th><th>5.992</th><th>5.908</th><th>5.824</th><th>5.737</th></t<>	^	6.209	6.181	6.155	6.132	6.111	6.092	6.074	6.058	6.043	6.029	910.9	6.003	5.992	5.908	5.824	5.737
4.860       4.833       4.808       4.786         4.457       4.430       4.405       4.383         4.150       4.123       4.099       4.077         3.910       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.729       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.237         3.128       3.101       3.077       3.055         3.054       3.027       3.003       2.981         2.989       2.962       2.938       2.916         2.931       2.904       2.880       2.857         2.879       2.762       2.738       2.716         2.789       2.762       2.738       2.716         2.771       2.688       2.644       2.647         2.632       2.652       2.602       2.579         2.633       2.656       2.574       2.552         2.600       2.574       2.552         2.600       2.573       2.346         2.251       2.223       2.175 <tb< th=""><th><b>∞</b></th><th>5.412</th><th>5.384</th><th>5.359</th><th>5.336</th><th>5.316</th><th>5.297</th><th>5.279</th><th>5.263</th><th>5.248</th><th>5.234</th><th>5.221</th><th>5.209</th><th>5.198</th><th>5.116</th><th>5.032</th><th>4.946</th></tb<>	<b>∞</b>	5.412	5.384	5.359	5.336	5.316	5.297	5.279	5.263	5.248	5.234	5.221	5.209	5.198	5.116	5.032	4.946
4.457       4.430       4.405       4.383         4.150       4.123       4.099       4.077         3.910       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.423       3.396       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.237         3.128       3.101       3.077       3.055         3.054       3.027       3.003       2.981         2.989       2.962       2.938       2.916         2.931       2.904       2.880       2.857         2.879       2.762       2.738       2.716         2.789       2.762       2.738       2.716         2.751       2.724       2.699       2.677         2.751       2.626       2.632       2.609         2.632       2.644       2.642       2.579         2.633       2.656       2.632       2.609         2.640       2.574       2.552         2.600       2.574       2.552         2.600       2.734       2.346         2.251       <	٥	4.860	4.833	4.808	4.786	4.765	4.746	4.729	4.713	4.698	4.684	4.672	4.660	4.649	4.567	4.483	4.398
4.150       4.123       4.099       4.077         3.910       3.883       3.858       3.836         3.716       3.689       3.665       3.643         3.423       3.529       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.237         3.128       3.101       3.077       3.055         3.054       3.027       3.003       2.981         2.989       2.962       2.938       2.916         2.931       2.904       2.880       2.857         2.879       2.827       2.805         2.879       2.762       2.738       2.716         2.751       2.724       2.699       2.677         2.751       2.688       2.644       2.642         2.653       2.656       2.632       2.609         2.653       2.656       2.677         2.600       2.579       2.574       2.552         2.600       2.574       2.526         2.601       2.346       2.346         2.251       2.253       2.175         2.251       2.253       2.175 <th>01</th> <th>4.457</th> <th>4.430</th> <th>4.405</th> <th>4.383</th> <th>4.363</th> <th>4.344</th> <th>4.327</th> <th>4.311</th> <th>4.296</th> <th>4.283</th> <th>4.270</th> <th>4.258</th> <th>4.247</th> <th>4.165</th> <th>4.082</th> <th>3.996</th>	01	4.457	4.430	4.405	4.383	4.363	4.344	4.327	4.311	4.296	4.283	4.270	4.258	4.247	4.165	4.082	3.996
3.910 3.883 3.858 3.836 3.716 3.689 3.665 3.643 3.556 3.529 3.505 3.483 3.423 3.396 3.372 3.350 3.310 3.283 3.259 3.237 3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.738 2.716 2.751 2.764 2.699 2.677 2.751 2.764 2.699 2.677 2.751 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.656 2.632 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.346 2.251 2.223 2.198 2.175	=	4.150	4.123	4.099	4.077	4.057	4.038	4.021	4.005	3.990	3.977	3.964	3.952	3.941	3.860	3.776	3.690
3.716       3.689       3.665       3.643         3.556       3.529       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.35         3.212       3.186       3.162       3.13         3.128       3.101       3.077       3.05         3.054       3.027       3.003       2.981         2.989       2.962       2.938       2.916         2.931       2.904       2.880       2.857         2.879       2.852       2.827       2.805         2.832       2.805       2.738       2.716         2.751       2.724       2.699       2.677         2.751       2.688       2.644       2.642         2.632       2.632       2.609         2.633       2.656       2.632       2.609         2.653       2.656       2.672       2.579         2.600       2.574       2.552         2.600       2.574       2.526         2.601       2.369       2.346         2.251       2.253       2.175         2.251       2.253       2.175	12	3.910	3.883	3.858	3.836	3.816	3.798	3.780	3.765	3.750	3.736	3.724	3.712	3.701	3.619	3.535	3.449
3.556       3.529       3.505       3.483         3.423       3.396       3.372       3.350         3.310       3.283       3.259       3.372         3.212       3.186       3.162       3.139         3.128       3.101       3.077       3.055         3.054       3.027       3.003       2.981         2.931       2.904       2.880       2.857         2.879       2.852       2.827       2.805         2.832       2.805       2.738       2.716         2.751       2.724       2.699       2.677         2.751       2.688       2.644       2.642         2.653       2.656       2.632       2.609         2.653       2.656       2.632       2.609         2.653       2.656       2.677       2.579         2.600       2.574       2.552         2.600       2.573       2.549       2.576         2.600       2.573       2.549       2.526         2.600       2.574       2.552         2.600       2.573       2.346         2.251       2.253       2.175           2.251       <	13	3.716	3.689	3.665	3.643	3.622	3.604	3.587	3.571	3.556	3.543	3.530	3.518	3.507	3.425	3.341	3.255
3.423 3.396 3.372 3.350 3.310 3.283 3.259 3.237 3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.526 2.251 2.233 2.198 2.175	14	3.556	3.529	3.505	3.483	3.463	3.444	3.427	3.412	3.397	3.383	3.371	3.359	3.348	3.266	3.181	3.094
3.310 3.283 3.259 3.237 3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.526 2.251 2.223 2.198 2.175	15	3.423	3.396	3.372	3.350	3.330	3.311	3.294	3.278	3.264	3.250	3.237	3.225	3.214	3.132	3.047	2.959
3.212 3.186 3.162 3.139 3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.632 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.526 2.251 2.223 2.198 2.175	91	3.310	3.283	3.259	3.237	3.216	3.198	3.181	3.165	3.150	3.137	3.124	3.112	3.101	3.018	2.933	2.845
3.128 3.101 3.077 3.055 3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.656 2.632 2.609 2.653 2.656 2.632 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.526 2.251 2.223 2.198 2.175	17	3.212	3.186	3.162	3.139	3.119	3.101	3.083	3.068	3.053	3.039	3.026	3.014	3.003	2.920	2.835	2.746
3.054 3.027 3.003 2.981 2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.632 2.609 2.653 2.626 2.632 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.251 2.223 2.198 2.175	81	3.128	3.101	3.077	3.055	3.035	3.016	2.999	2.983	2.968	2.955	2.942	2.930	2.919	2.835	2.749	2.660
2.989 2.962 2.938 2.916 2.931 2.904 2.880 2.857 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.656 2.632 2.609 2.653 2.626 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.526 2.251 2.253 2.198 2.175	16	3.054	3.027	3.003	2.981	2.961	2.942	2.925	2.909	2.894	2.880	2.868	2.855	2.844	2.761	2.674	2.584
2.931 2.904 2.880 2.857 2.805 2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.751 2.751 2.754 2.699 2.677 2.715 2.688 2.664 2.642 2.653 2.653 2.656 2.653 2.650 2.650 2.579 2.600 2.573 2.549 2.552 2.600 2.573 2.549 2.552 2.251 2.251 2.253 2.198 2.175	20	2.989	2.962	2.938	2.916	2.895	2.877	2.859	2.843	2.829	2.815	2.802	2.790	2.778	2.695	2.608	2.517
2.879 2.852 2.827 2.805 2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.632 2.609 2.656 2.579 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	12	2.931	2.904	2.880	2.857	2.837	2.818	2.801	2.785	2.770	2.756	2.743	2.731	2.720	2.636	2.548	2.45
2.832 2.805 2.780 2.758 2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	22	2.879	2.852	2.827	2.805	2.785	2.766	2.749	2.733	2.718	2.704	2.691	2.679	2.667	2.583	2.495	2.40
2.789 2.762 2.738 2.716 2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	23	2.832	2.805	2.780	2.758	2.738	2.719	2.702	2.686	2.671	2.657	2.644	2.632	2.620	2.536	2.447	2.35
2.751 2.724 2.699 2.677 2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	24	2.789	2.762	2.738	2.716	2.695	2.676	2.659	2.643	2.628	2.614	2.601	2.589	2.577	2.492	2.403	2.310
2.715 2.688 2.664 2.642 2.683 2.656 2.632 2.609 2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	25	2.751	2.724	2.699	2.677	2.657	2.638	2.620	2.604	2.589	2.575	2.562	2.550	2.538	2.453	2.364	2.27(
2.683 2.656 2.632 2.609 2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	76	2.715	2.688	2.664	2.642	2.621	2.602	2.585	2.569	2.554	2.540	2.526	2.514	2.503	2.417	2.327	2.233
2.653 2.626 2.602 2.579 2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	27	2.683	2.656	2.632	2.609	2.589	2.570	2.552	2.536	2.521	2.507	2.494	2.481	2.470	2.384	2.294	2.1%
2.626 2.599 2.574 2.552 2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	28	2.653	2.626	2.602	2.579	2.559	2.540	2.522	2.506	2.491	2.477	2.464	2.451	2.440	2.354	2.263	2.167
2.600 2.573 2.549 2.526 2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	29	2.626	2.599	2.574	2.552	2.531	2.512	2.495	2.478	2.463	2.449	2.436	2.423	2.412	2.325	2.234	2.138
2.421 2.394 2.369 2.346 2.251 2.223 2.198 2.175	30	2.600	2.573	2.549	2.526	2.506	2.487	2.469	2.453	2.437	2.423	2.410	2.398	2.386	2.299	2.208	2.111
2.251 2.223 2.198 2.175	40	2.421	2.394	2.369	2.346	2.325	2.306	2.288	2.271	2.256	2.241	2.228	2.215	2.203	2.114	2.019	1.917
	9	2.251	2.223	2.198	2.175	2.153	2.134	2.115	2.098	2.083	2.068	2.054	2.041	2.028	1.936	1.836	1.726
2.089 2.060 2.035 2.011	120	2.089	2.060	2.035	2.011	1.989	1.969	1.950	1.932	1.916	1.901	1.886	1.873	1.860	1.763	1.656	1.533

TABLE 3 (continued)

b. Area in the Right Tail under the F Distribution Curve = 0.05



	Degrees of							Dec	Degrees of Freedom	reedom –	Numerator	ōr						
18.51   19.00   19.4   19.2	Preedom Denominator	-	7	e	4	2	9	^	æ	6	0	=	12	13	14	15	91	17
10128   9520   9144   91247   91328   91329   91321   91321   91328   91345   91428   91421   91412	-	161.446	199.499	215.707	224.583	230.160	233.988	236.767	238.884	240.543	241.882	242.981	243.905	244.690	245.363	245.949	246.466	246.917
10.10.8         9.552         9.277         9.11/7 </th <th>7</th> <th>18.513</th> <th>19.000</th> <th>19.164</th> <th>19.247</th> <th>19.296</th> <th>19.329</th> <th>19.353</th> <th>19.371</th> <th>19.385</th> <th>19.396</th> <th>19.405</th> <th>19.412</th> <th>19.419</th> <th>19.424</th> <th>19.429</th> <th>19.433</th> <th>19.437</th>	7	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.433	19.437
7709         6.944         6.599         6.126         6.126         6.041         5.994         5.954         5.997         5.894         6.996         6.944         6.990         6.936         6.041         5.994         5.954         6.599         5.143         4.299         4.950         4.818         4.284         4.007         4.004         4.007         4.069         4.007         4.007         4.079         4.007         4.007         4.079         4.007	က	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.692	8.683
6.608         5.786         5.786         5.409         5.192         5.090         4.876         4.818         4.772         4.774         4.078         4.669         4.669         5.786         4.819         4.080         4.080         4.669         5.786         4.819         4.080         4.887         4.819         4.080         4.887         4.819         4.080         4.887         4.887         3.680         3.781         3.782         3.680         3.781         3.783         3.681         3.781         3.782         3.682         3.881         3.888         3.881         3.888         3.881         3.882         3.281         3.282 <th< th=""><th>4</th><th>7.709</th><th>6.944</th><th>6.591</th><th>6.388</th><th>6.256</th><th>6.163</th><th>6.094</th><th>6.041</th><th>5.999</th><th>5.964</th><th>5.936</th><th>5.912</th><th>5.891</th><th>5.873</th><th>5.858</th><th>5.844</th><th>5.832</th></th<>	4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.844	5.832
5987         5143         4757         4284         4207         4147         4099         4064         4007         4007         3976         3978         3938         3938         3938         3938         3938         3938         3949         4737         4170         4180         4064         4070         3976         3978         3078 <th< th=""><th>2</th><th>809.9</th><th>5.786</th><th>5.409</th><th>5.192</th><th>5.050</th><th>4.950</th><th>4.876</th><th>4.818</th><th>4.772</th><th>4.735</th><th>4.704</th><th>4.678</th><th>4.655</th><th>4.636</th><th>4.619</th><th>4.604</th><th>4.590</th></th<>	2	809.9	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.604	4.590
5591         4773         4347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41347         41348         31383         3388         3189         3189         3189         3189         3189         3189         3189         3189         3189         3189         3188         3188         3188         3188         3188         3188         3188         3188         3188         3189         3189         3189         3189         3	9	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.922	3.908
5318         4459         4066         3888         3.581         3.347         3.317         3.317         3.329         3.229         3.229         3.347         3.317         3.317         3.318         3.347         3.217         3.289         3.289         3.289         3.347         3.318         3.348         3.148         3.248         3.059         3.078         2.089         2.848         2.8	7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.494	3.480
5117         4.256         3.863         3.643         3.442         3.374         3.293         3.179         3.107         3.102         3.004         3.008         2.989           4.846         4.103         3.086         3.487         3.374         3.374         3.123         3.107         3.102         3.003         3.008         2.898           4.846         4.103         3.087         3.326         3.075         3.026         2.891         2.896         2.843         2.796         2.873         2.717         2.887         2.845         2.847         2.738         2.793         2.845         2.845         2.846         2.847         2.763         2.647         2.788         2.747         2.887         2.846         2.847         2.767         2.641         2.853         2.647         2.847         2.848         2.844 <th><b>∞</b></th> <th>5.318</th> <th>4.459</th> <th>4.066</th> <th>3.838</th> <th>3.688</th> <th>3.581</th> <th>3.500</th> <th>3.438</th> <th>3.388</th> <th>3.347</th> <th>3.313</th> <th>3.284</th> <th>3.259</th> <th>3.237</th> <th>3.218</th> <th>3.202</th> <th>3.187</th>	<b>∞</b>	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.202	3.187
4,865         4 103         3.708         3.478         3.127         3.135         3.072         3.072         2.943         2.944         2.943         2.943         2.944         2.943         2.944         2.943         2.944         2.944         2.944         2.944         2.944         2.944         2.944         2.944         2.944         2.944         2.944         2.945         2.944         2.944         2.945         2.944         2.944         2.945 <th< th=""><th>6</th><th>5.117</th><th>4.256</th><th>3.863</th><th>3.633</th><th>3.482</th><th>3.374</th><th>3.293</th><th>3.230</th><th>3.179</th><th>3.137</th><th>3.102</th><th>3.073</th><th>3.048</th><th>3.025</th><th>3.006</th><th>2.989</th><th>2.974</th></th<>	6	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.989	2.974
4844         3982         3.387         3.374         3.095         3.01         2.948         2.884         2.884         2.884         2.884         2.894         2.884         2.894         2.894         2.894         2.894         2.894         2.894         2.894         2.994         2.494         2.994         2.997         2.494         2.	9	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.828	2.812
4.747         3885         3.490         3.259         3.106         2.996         2.913         2.849         2.775         2.717         2.687         2.667         2.687         2.687         2.617         2.589         2.849         2.796         2.776         2.647         2.671         2.637         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.648         2.648         2.648         2.648         2.648         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.647         2.648         2.644         2.647         2.648         2.644         2.647         2.648         2.644         2.647         2.647         2.648         2.647         2.648         2.447         2.447         2.447         2.348         2.444         2.648         2.447         2.448         2.349         2.348         2.448         2.447         2.448         2.348         2.449         2.648	=	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.701	2.685
4667         3896         3411         3179         3025         2915         2842         2767         2714         2631         2635         2640         2554         2534         2534         2534         2534         2533         2513         2513         2513         2513         2513         2513         2513         2513         2514         2445         4460         3.882	13	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.599	2.583
4600         3739         3344         3112         2988         2848         2764         2649         2640         2650         2554         2507         2448         2443         3444         3112         2988         2344         2670         2641         2588         2644         2650         2445         2450         2445         2450         2445         2450         2446         2450         2446         2670         2446         2680         2641         2588         2449         2450         2445         2450         2445         2450         2446         2450         2446         2460         2446         2641         2589         2441         2567         2441         2567         2441         2567         2441         2567         2441         2567         2441         2577         2442         2460         2442         2447         2446         2440         2446         2440         2446         2440         2446         2440         2446         2440         2440         2440         2440         2440         2440         2440         2440         2440         2440         2440         2440         2446         2446         2445         2440         2440 <th< th=""><th>13</th><th>4.667</th><th>3.806</th><th>3.411</th><th>3.179</th><th>3.025</th><th>2.915</th><th>2.832</th><th>2.767</th><th>2.714</th><th>2.671</th><th>2.635</th><th>2.604</th><th>2.577</th><th>2.554</th><th>2.533</th><th>2.515</th><th>2.499</th></th<>	13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.515	2.499
4.543         3.682         3.287         3.056         2.901         2.707         2.641         2.588         2.544         2.507         2.475         2.549         2.549         2.549         2.547         2.449         2.647         2.474         2.687         2.494         2.456         2.475         2.449         2.445         2.475         2.445         2.475         2.473         2.373         2.373         2.332         2.333         2.348         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.445         2.347         2.342         2.347         2.349         2.440         2.456         2.440         2.450         2.442         2.440         2.440         2.445         2.442         2.442         2.440 <th< th=""><th>14</th><th>4.600</th><th>3.739</th><th>3.344</th><th>3.112</th><th>2.958</th><th>2.848</th><th>2.764</th><th>2.699</th><th>2.646</th><th>2.602</th><th>2.565</th><th>2.534</th><th>2.507</th><th>2.484</th><th>2.463</th><th>2.445</th><th>2.428</th></th<>	14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.445	2.428
4444         3 634         3 239         3 007         2 882         2 741         2 58         2 494         2 456         2 435         2 337         2 337         2 332         2	15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.385	2.368
4451         3552         3197         2.965         2.810         2.699         2.614         2.449         2.445         2.341         2.332         2.329         2.308         2.289           4.414         3.555         3.160         2.978         2.747         2.510         2.445         2.412         2.344         2.342         2.314         2.290         2.266         2.277         2.510         2.445         2.346         2.340         2.308         2.280         2.278         2.310         2.280         2.280         2.280         2.278         2.310         2.280         2.290         2.280         2.347         2.348         2.308         2.280         2.293         2.286         2.290         2.284         2.477         2.487         2.348         2.308         2.280         2.290         2.278         2.348         2.308         2.290         2.278         2.347         2.348         2.309         2.278         2.291         2.789         2.440         2.375         2.326         2.278         2.292         2.278         2.178         2.178         2.179         2.178         2.179         2.178         2.179         2.178         2.179         2.179         2.179         2.179         2.179<	91	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.333	2.317
4.414         3.555         3.160         2.928         2.773         2.510         2.456         2.412         2.374         2.342         2.314         2.290         2.269         2.250           4.381         3.522         3.127         2.895         2.740         2.477         2.423         2.378         2.340         2.308         2.280         2.256         2.234         2.15           4.381         3.522         3.127         2.895         2.741         2.497         2.493         2.340         2.308         2.260         2.256         2.234         2.276         2.250         2.250         2.256         2.234         2.276         2.262         2.279         2.276         2.275         2.283         2.250         2.275         2.279         2.276         2.275         2.283         2.260         2.725         2.279         2.131 <th>17</th> <th>4.451</th> <th>3.592</th> <th>3.197</th> <th>2.965</th> <th>2.810</th> <th>2.699</th> <th>2.614</th> <th>2.548</th> <th>2.494</th> <th>2.450</th> <th>2.413</th> <th>2.381</th> <th>2.353</th> <th>2.329</th> <th>2.308</th> <th>2.289</th> <th>2.272</th>	17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.289	2.272
4.381         3.522         3.127         2.895         2.740         2.628         2.544         2.477         2.423         2.378         2.340         2.308         2.286         2.213         2.156         2.134 <th< th=""><th>8</th><th>4.414</th><th>3.555</th><th>3.160</th><th>2.928</th><th>2.773</th><th>2.661</th><th>2.577</th><th>2.510</th><th>2.456</th><th>2.412</th><th>2.374</th><th>2.342</th><th>2.314</th><th>2.290</th><th>2.269</th><th>2.250</th><th>2.233</th></th<>	8	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.250	2.233
4.351         3.493         3.098         2.866         2.711         2.599         2.514         2.447         2.393         2.348         2.210         2.225         2.225         2.293         2.173         2.184           4.325         3.467         3.072         2.840         2.685         2.573         2.488         2.420         2.366         2.321         2.283         2.250         2.197         2.197         2.156         2.136         2.136         2.137         2.136         2.137         2.289         2.226         2.198         2.173         2.136         2.148         2.137         2.286         2.236         2.236         2.236         2.148         2.146         2.337         2.286         2.236         2.148         2.144         2.337         2.2	19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.215	2.198
4.325         3.467         3.072         2.840         2.685         2.573         2.488         2.420         2.366         2.321         2.283         2.220         2.197         2.156         2.156           4.301         3.443         3.049         2.817         2.641         2.549         2.442         2.375         2.297         2.259         2.256         2.198         2.173         2.151         2.131           4.301         3.443         3.049         2.817         2.640         2.528         2.442         2.375         2.297         2.259         2.226         2.198         2.173         2.131         2.131         2.131         2.131         2.131         2.131         2.132         2.137         2.131         2.197         2.132         2.130         2.132         2.136         2.136         2.130         2.136         2.130         2.136         2.130 </th <th>20</th> <th>4.351</th> <th>3.493</th> <th>3.098</th> <th>2.866</th> <th>2.711</th> <th>2.599</th> <th>2.514</th> <th>2.447</th> <th>2.393</th> <th>2.348</th> <th>2.310</th> <th>2.278</th> <th>2.250</th> <th>2.225</th> <th>2.203</th> <th>2.184</th> <th>2.167</th>	20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.184	2.167
4.301         3.443         3.049         2.817         2.641         2.342         2.342         2.259         2.256         2.198         2.173         2.151         2.131           4.279         3.422         3.049         2.817         2.640         2.528         2.442         2.375         2.320         2.259         2.226         2.198         2.175         2.150         2.150         2.109           4.279         3.422         3.028         2.796         2.640         2.528         2.442         2.375         2.320         2.275         2.236         2.195         2.195         2.195         2.195         2.195         2.196         2.098         2.099 </th <th>12</th> <th>4.325</th> <th>3.467</th> <th>3.072</th> <th>2.840</th> <th>2.685</th> <th>2.573</th> <th>2.488</th> <th>2.420</th> <th>2.366</th> <th>2.321</th> <th>2.283</th> <th>2.250</th> <th>2.222</th> <th>2.197</th> <th>2.176</th> <th>2.156</th> <th>2.139</th>	12	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.156	2.139
4.279         3.422         3.028         2.796         2.640         2.528         2.442         2.375         2.320         2.275         2.236         2.175         2.150         2.178         2.109           4.260         3.403         3.009         2.776         2.621         2.508         2.423         2.355         2.300         2.255         2.216         2.183         2.155         2.130         2.108         2.088           4.240         3.385         2.991         2.759         2.603         2.490         2.405         2.326         2.286         2.183         2.165         2.183         2.195         2.191         2.094         2.089         2.069         2.052         2.479         2.286         2.220         2.181         2.148         2.111         2.089         2.062         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.052         2.050         2.050         2.052         2.050         2.052         2.030         2.052         2.030         2.052         2.030         2.052         2.050         2.052 </th <th>22</th> <th>4.301</th> <th>3.443</th> <th>3.049</th> <th>2.817</th> <th>2.661</th> <th>2.549</th> <th>2.464</th> <th>2.397</th> <th>2.342</th> <th>2.297</th> <th>2.259</th> <th>2.226</th> <th>2.198</th> <th>2.173</th> <th>2.151</th> <th>2.131</th> <th>2.114</th>	22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.131	2.114
4.260         3.403         3.009         2.776         2.621         2.508         2.423         2.355         2.20         2.183         2.155         2.130         2.108         2.088           4.242         3.385         2.991         2.759         2.603         2.490         2.405         2.337         2.286         2.198         2.165         2.136         2.111         2.089         2.069           4.225         3.386         2.975         2.743         2.587         2.474         2.388         2.321         2.265         2.181         2.148         2.119         2.094         2.072         2.052           4.225         3.369         2.975         2.743         2.475         2.474         2.388         2.321         2.265         2.181         2.118         2.193         2.119         2.193         2.193         2.037         2.035         2.036         2.270         2.181         2.18         2.084         2.056         2.036         2.036         2.036         2.036         2.036         2.036         2.036         2.036         2.036         2.036         2.037         2.037         2.011         2.087         2.040         2.166         2.18         2.044         2.044	23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.109	2.091
4.242         3.385         2.991         2.759         2.603         2.490         2.045         2.282         2.286         2.198         2.165         2.136         2.111         2.089         2.069           4.225         3.369         2.975         2.743         2.587         2.474         2.388         2.321         2.265         2.20         2.181         2.148         2.119         2.094         2.072         2.052           4.210         3.354         2.960         2.778         2.475         2.459         2.373         2.305         2.206         2.181         2.192         2.103         2.094         2.072         2.052           4.196         3.340         2.947         2.714         2.558         2.445         2.359         2.291         2.236         2.166         2.151         2.118         2.087         2.043         2.037         2.031           4.183         3.328         2.934         2.701         2.545         2.432         2.278         2.223         2.177         2.138         2.104         2.075         2.050         2.050         2.050         2.050         2.050         2.050         2.050         2.050         2.050         2.050         2.050	24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.216	2.183	2.155	2.130	2.108	2.088	2.070
4.225         3.369         2.975         2.743         2.587         2.474         2.388         2.321         2.265         2.20         2.181         2.148         2.119         2.094         2.072         2.052           4.210         3.354         2.960         2.728         2.572         2.459         2.373         2.305         2.250         2.164         2.132         2.103         2.078         2.056         2.036           4.196         3.340         2.947         2.714         2.558         2.445         2.359         2.291         2.236         2.190         2.151         2.118         2.089         2.064         2.041         2.036           4.183         3.328         2.947         2.714         2.545         2.432         2.278         2.223         2.177         2.138         2.104         2.075         2.050         2.077         2.007           4.171         3.316         2.922         2.690         2.534         2.421         2.366         2.212         2.105         2.052         2.050         2.057         2.015         1.995           4.085         3.232         2.839         2.606         2.449         2.334         2.147         2.040         <	25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.136	2.111	2.089	2.069	2.051
4.210         3.354         2.960         2.728         2.572         2.459         2.373         2.355         2.250         2.250         2.166         2.132         2.103         2.078         2.056         2.036           4.196         3.340         2.947         2.714         2.558         2.445         2.359         2.291         2.236         2.190         2.151         2.118         2.089         2.064         2.041         2.021           4.183         3.328         2.934         2.701         2.545         2.432         2.346         2.278         2.223         2.177         2.138         2.104         2.075         2.050         2.027         2.007           4.171         3.316         2.922         2.690         2.534         2.421         2.266         2.211         2.165         2.105         2.053         2.057         2.015         1.995           4.085         3.232         2.839         2.606         2.449         2.334         2.180         2.124         2.040         1.993         1.952         1.917         1.887         1.860         1.836         1.815           4.001         3.150         2.758         2.255         2.368         2.077	<b>3</b> 9	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.181	2.148	2.119	2.094	2.072	2.052	2.034
4.196         3.340         2.947         2.714         2.558         2.445         2.359         2.291         2.236         2.190         2.151         2.118         2.089         2.064         2.041         2.021           4.183         3.328         2.934         2.701         2.545         2.432         2.346         2.273         2.177         2.138         2.104         2.075         2.050         2.027         2.007           4.171         3.316         2.922         2.690         2.534         2.421         2.366         2.211         2.165         2.126         2.092         2.063         2.037         2.015         1.995           4.085         3.232         2.839         2.606         2.449         2.334         2.180         2.124         2.077         2.038         2.003         1.974         1.948         1.924         1.904           4.001         3.150         2.758         2.525         2.368         2.254         2.167         2.040         1.959         1.910         1.869         1.834         1.803         1.775         1.750         1.728	27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.166	2.132	2.103	2.078	2.056	2.036	2.018
4.183       3.328       2.934       2.701       2.545       2.432       2.378       2.223       2.177       2.138       2.104       2.075       2.050       2.007       2.007       2.007       2.007       2.007       2.003       1.974       1.995       1.995         4.001       3.150       2.758       2.525       2.364       2.167       2.040       2.077       2.038       2.003       1.974       1.948       1.904         4.001       3.150       2.758       2.525       2.368       2.254       2.167       2.040       1.993       1.952       1.917       1.887       1.887       1.860       1.815         3.920       3.072       2.680       2.447       2.207       2.016       1.959       1.910       1.869       1.834       1.803       1.775       1.728	28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.151	2.118	2.089	2.064	2.041	2.021	2.003
4.171 3.316 2.922 2.690 2.534 2.421 2.334 2.266 2.211 2.165 2.126 2.092 2.063 2.037 2.015 1.995 1.995 2.085 3.232 2.839 2.606 2.449 2.336 2.249 2.180 2.124 2.077 2.038 2.003 1.974 1.948 1.924 1.904 4.001 3.150 2.758 2.525 2.368 2.254 2.167 2.097 2.040 1.993 1.952 1.917 1.887 1.860 1.836 1.815 1.815 3.920 3.072 2.680 2.447 2.290 2.175 2.087 2.016 1.959 1.910 1.869 1.834 1.803 1.775 1.750 1.728	50	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.138	2.104	2.075	2.050	2.027	2.007	1.989
4.085 3.232 2.839 2.606 2.449 2.336 2.249 2.180 2.124 2.077 2.038 2.003 1.974 1.948 1.924 1.904 1.904 4.001 3.150 2.758 2.525 2.368 2.254 2.167 2.097 2.040 1.993 1.952 1.917 1.887 1.860 1.836 1.815 3.920 3.072 2.680 2.447 2.290 2.175 2.087 2.016 1.959 1.910 1.869 1.834 1.803 1.775 1.750 1.728	30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.063	2.037	2.015	1.995	1.976
4.001 3.150 2.758 2.525 2.368 2.254 2.167 2.097 2.040 1.993 1.952 1.917 1.887 1.860 1.835 1.815 3.920 3.072 2.680 2.447 2.290 2.175 2.087 2.016 1.959 1.910 1.869 1.834 1.803 1.775 1.750 1.728	40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.974	1.948	1.924	1.904	1.885
3.920 3.072 2.680 2.447 2.290 2.175 2.087 2.016 1.959 1.910 1.869 1.834 1.803 1.775 1.750 1.728	9	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.952	1.917	1.887	1.860	1.836	1.815	1.796
	120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.869	1.834	1.803	1.775	1.750	1.728	1.709

Denominator	18	61	20	21	22	23	24	25	26	27	28	29	30	40	09	120
-	247.324	247.688	248.016	248.307	248.579	248.823	249.052	249.260	249.453	249.631	249.798	249.951	250.096	251.144	252.196	253.254
7	19.440	19.443	19.446	19.448	19.450	19.452	19.454	19.456	19.457	19.459	19.460	19.461	19.463	19.471	19.479	19.487
က	8.675	8.667	8.660	8.654	8.648	8.643	8.638	8.634	8.630	8.626	8.623	8.620	8.617	8.594	8.572	8.549
4	5.821	5.811	5.803	5.795	5.787	5.781	5.774	5.769	5.763	5.759	5.754	5.750	5.746	5.717	5.688	5.658
2	4.579	4.568	4.558	4.549	4.541	4.534	4.527	4.521	4.515	4.510	4.505	4.500	4.496	4.464	4.431	4.398
9	3.896	3.884	3.874	3.865	3.856	3.849	3.841	3.835	3.829	3.823	3.818	3.813	3.808	3.774	3.740	3.705
^	3.467	3.455	3.445	3.435	3.426	3.418	3.410	3.404	3.397	3.391	3.386	3.381	3.376	3.340	3.304	3.267
<b>∞</b>	3.173	3.161	3.150	3.140	3.131	3.123	3.115	3.108	3.102	3.095	3.090	3.084	3.079	3.043	3.005	2.967
6	2.960	2.948	2.936	2.926	2.917	2.908	2.900	2.893	2.886	2.880	2.874	2.869	2.864	2.826	2.787	2.748
10	2.798	2.785	2.774	2.764	2.754	2.745	2.737	2.730	2.723	2.716	2.710	2.705	2.700	2.661	2.621	2.580
=	2.671	2.658	2.646	2.636	2.626	2.617	2.609	2.601	2.594	2.588	2.582	2.576	2.570	2.531	2.490	2.448
12	2.568	2.555	2.544	2.533	2.523	2.514	2.505	2.498	2.491	2.484	2.478	2.472	2.466	2.426	2.384	2.341
13	2.484	2.471	2.459	2.448	2.438	2.429	2.420	2.412	2.405	2.398	2.392	2.386	2.380	2.339	2.297	2.252
14	2.413	2.400	2.388	2.377	2.367	2.357	2.349	2.341	2.333	2.326	2.320	2.314	2.308	2.266	2.223	2.178
15	2.353	2.340	2.328	2.316	2.306	2.297	2.288	2.280	2.272	2.265	2.259	2.253	2.247	2.204	2.160	2.114
91	2.302	2.288	2.276	2.264	2.254	2.244	2.235	2.227	2.220	2.212	2.206	2.200	2.194	2.151	2.106	2.059
17	2.257	2.243	2.230	2.219	2.208	2.199	2.190	2.181	2.174	2.167	2.160	2.154	2.148	2.104	2.058	2.011
82	2.217	2.203	2.191	2.179	2.168	2.159	2.150	2.141	2.134	2.126	2.119	2.113	2.107	2.063	2.017	1.968
61	2.182	2.168	2.155	2.144	2.133	2.123	2.114	2.106	2.098	2.090	2.084	2.077	2.071	2.026	1.980	1.930
20	2.151	2.137	2.124	2.112	2.102	2.092	2.082	2.074	2.066	2.059	2.052	2.045	2.039	1.994	1.946	1.896
21	2.123	2.109	2.096	2.084	2.073	2.063	2.054	2.045	2.037	2.030	2.023	2.016	2.010	1.965	1.916	1.866
22	2.098	2.084	2.071	2.059	2.048	2.038	2.028	2.020	2.012	2.004	1.997	1.990	1.984	1.938	1.889	1.838
23	2.075	2.061	2.048	2.036	2.025	2.014	2.005	1.996	1.988	1.981	1.973	1.967	1.961	1.914	1.865	1.813
24	2.054	2.040	2.027	2.015	2.003	1.993	1.984	1.975	1.967	1.959	1.952	1.945	1.939	1.892	1.842	1.790
25	2.035	2.021	2.007	1.995	1.984	1.974	1.964	1.955	1.947	1.939	1.932	1.926	1.919	1.872	1.822	1.768
56	2.018	2.003	1.990	1.978	1.966	1.956	1.946	1.938	1.929	1.921	1.914	1.907	1.901	1.853	1.803	1.749
27	2.002	1.987	1.974	1.961	1.950	1.940	1.930	1.921	1.913	1.905	1.898	1.891	1.884	1.836	1.785	1.731
28	1.987	1.972	1.959	1.946	1.935	1.924	1.915	1.906	1.897	1.889	1.882	1.875	1.869	1.820	1.769	1.714
56	1,973	1.958	1.945	1.932	1.921	1.910	1.901	1.891	1.883	1.875	1.868	1.861	1.854	1.806	1.754	1.698
30	1.960	1.945	1.932	1.919	1.908	1.897	1.887	1.878	1.870	1.862	1.854	1.847	1.841	1.792	1.740	1.683
40	1.868	1.853	1.839	1.826	1.814	1.803	1.793	1.783	1.775	1.766	1.759	1.751	1.744	1.693	1.637	1.577
09	1.778	1.763	1.748	1.735	1.722	1.711	1.700	1.690	1.681	1.672	1.664	1.656	1.649	1.594	1.534	1.467
120	1.690	1.674	1.659	1.645	1.632	1.620	1.608	1.598	1.588	1.579	1.570	1.562	1.554	1.495	1.429	1.352



# SPSS 12.0 for Windows

## Analysis without Anguish

This edition of **SPSS: Analysis without Anguish** is based on SPSS version 12.0 for Windows.

SPSS is a sophisticated piece of software used by social scientists and related professionals for statistical analysis. There are several new features in SPSS version 12.0 that are covered in the new edition of this workbook, including:

- · improved charting capabilities
- · new data management features
- · statistical enhancements.

Always popular in previous editions, *SPSS: Analysis without Anguish Version 12.0 for Windows* continues to address the needs of all users of SPSS software. It serves as a practical introduction to new users and also allows those wishing to undertake more advanced analysis to work their way systematically through each stage.

Extensive use of screen displays and a range of 'step by step' working and practice examples remain as features of the workbook. SPSS data sets for all exercises are also now available at the accompanying website www.johnwiley.com.au/highered/spssV12.

Features new to this edition of the workbook include:

- a reporting box at the end of each relevant chapter to illustrate how results from each of the particular tests should be reported
- a chapter on multidimensional scaling
- an 'at a glance' listing of practice and working examples in the front of the book
- · three common statistical tables in the appendix: f, chi square, and t.

SPSS: Analysis without Anguish Version 12.0 is suitable across a number of disciplines, including business, health, social sciences, environmental science and geography. It can also be value packed with other Wiley textbooks in subjects such as statistics, market research and research methods to save students money.





John Wiley & Sons Australia, Ltd www.johnwiley.com.au

