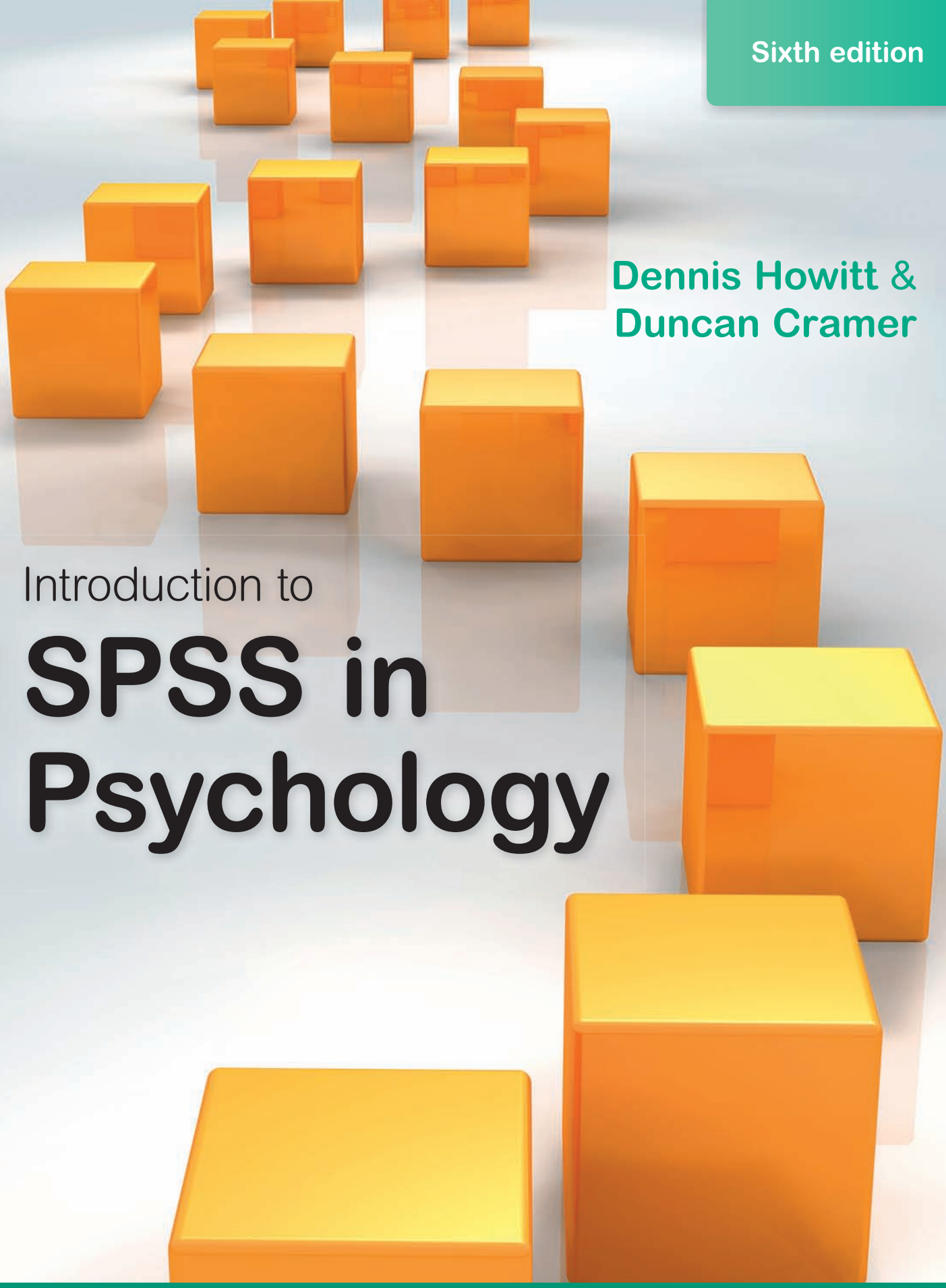
Sixth edition

Dennis Howitt &
Duncan Cramer



Introduction to

SPSS in Psychology

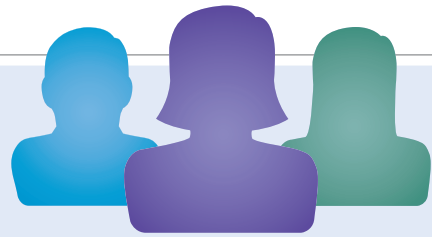
ALWAYS LEARNING

PEARSON

Introduction to SPSS in Psychology

Visit the *Introduction to SPSS in Psychology*, *sixth edition* Companion Website at www.pearsoned.co.uk/howitt to find valuable student learning material, including:

- Learning objectives for each chapter
- Datasets and syntax
- Annotated links to relevant sites on the web



Introduction to SPSS in Psychology

For Version 22 and earlier

Sixth edition

Dennis Howitt, Loughborough University
Duncan Cramer, Loughborough University

PEARSON

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Auckland • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Pearson Education Limited
Edinburgh Gate
Harlow CM20 2JE
United Kingdom
Tel: +44 (0)1279 623623
Web: www.pearson.com/uk

First published 1997 (print)
Second edition published 2000 (print)
Second revised edition 2002 (print)
Third edition 2005 (print)
Fourth edition 2008 (print)
Fifth edition 2011 (print)
This edition published 2014 (print and electronic)

ISBN: 978-1-292-00069-5 (print)
978-1-292-00070-1 (CSeTxt)
978-1-292-00073-2 (PDF)

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for the print edition is available from the Library of Congress

10 9 8 7 6 5 4 3 2 1
18 17 16 15 14

Print edition typeset in 9.5/12pt Sabon LT Std by 35
Printed by DZS Grafic, Slovenia

NOTE THAT ANY PAGE CROSS REFERENCES REFER TO THE PRINT EDITION

Summary of contents

Contents

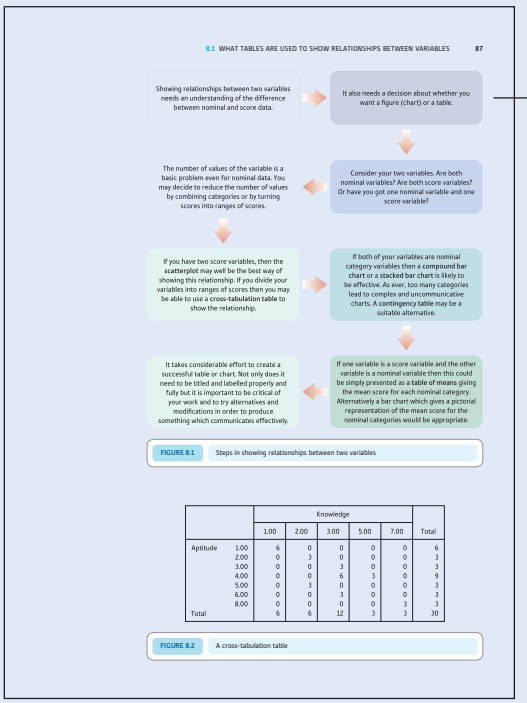
Companion Website

For open-access **student resources** specifically written to complement this textbook and support your learning, please visit **www.pearsoned.co.uk/howitt**

The website contains:

- Datasets relating to exercises from the book
- Additional datasets with questions
- A set of research scenarios with multiple choice questions
- Flashcards

Guided tour



8.1 WHAT TABLES ARE USED TO SHOW RELATIONSHIPS BETWEEN VARIABLES 87

Showing relationships between two variables needs an understanding of the difference between nominal and score data.

It also needs a decision about whether you want a figure (chart) or a table.

Consider your two variables. Are both nominal variables? Are both score variables? Or have you got one nominal variable and one score variable?

The number of values of the variable is a basic problem even for nominal data. You may decide to reduce the number of values by combining categories or by turning scores into ranges of scores.

If you have two score variables, then the scatterplot may well be the best way of showing this relationship. If you divide your variables into ranges of scores then you may be able to use a cross-tabulation table to show the relationship.

If both of your variables are nominal category variables then a compound bar chart or a stacked bar chart is likely to be effective. As ever, too many categories lead to complex and uncommunicative charts. A contingency table may be a suitable alternative.

It takes considerable effort to create a successful table or chart. Not only does it need to be titled and labeled properly and fully but it is important to be critical of your work and to try alternatives and modifications in order to produce something which communicates effectively.

If one variable is a score variable and the other variable is a nominal variable then this could be simply presented as a table of means giving the mean score for each nominal category. Alternatively a bar chart which gives a pictorial representation of the mean score for the nominal categories would be appropriate.

FIGURE 8.1 Steps in showing relationships between two variables

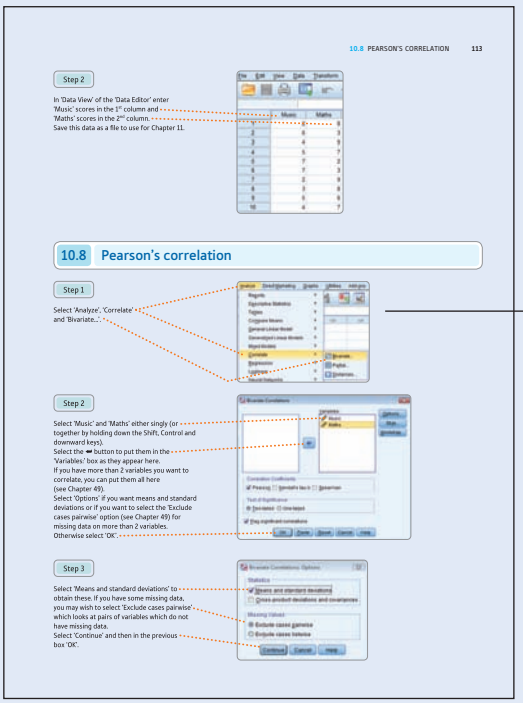
		Knowledge					
		1.00	2.00	3.00	5.00	7.00	Total
Aptitude	1.00	6	0	0	0	0	6
	2.00	0	3	0	0	0	3
	3.00	0	0	3	0	0	3
	4.00	0	0	6	3	0	9
	5.00	0	3	0	0	0	3
	6.00	0	0	3	0	0	3
	8.00	0	0	0	0	3	3
Total		6	6	12	3	3	30

FIGURE 8.2 A cross-tabulation table

Background details

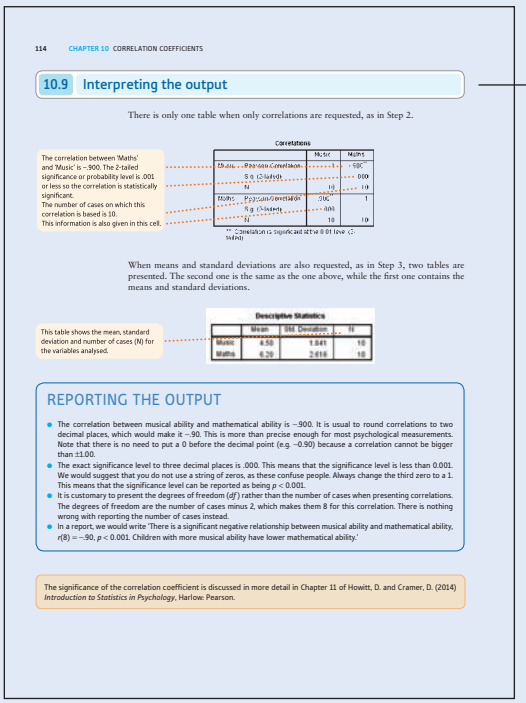
Outlines the background to the techniques discussed in each chapter to encourage a deeper understanding. This includes details of:

- What the technique is
- When you should use it
- When you should not use it
- The data required for the analysis
- Typical problems to be aware of



Step by step illustrations and screenshots of SPSS

This presents the stages of data entry and data analysis visually to help you gain confidence in the processes and procedures of SPSS.



114 CHAPTER 10 CORRELATION COEFFICIENTS

10.9 Interpreting the output

There is only one table when only correlations are requested, as in Step 2.

The correlation between 'Maths' and 'Music' is −.900. The 2-tailed significance or probability level is .001 or less so the correlation is statistically significant.
The number of cases on which this correlation is based is 10.
This information is also given in this cell.

When means and standard deviations are also requested, as in Step 3, two tables are presented. The second one is the same as the one above, while the first one contains the means and standard deviations.

This table shows the mean, standard deviation and number of cases (N) for the variables analysed.

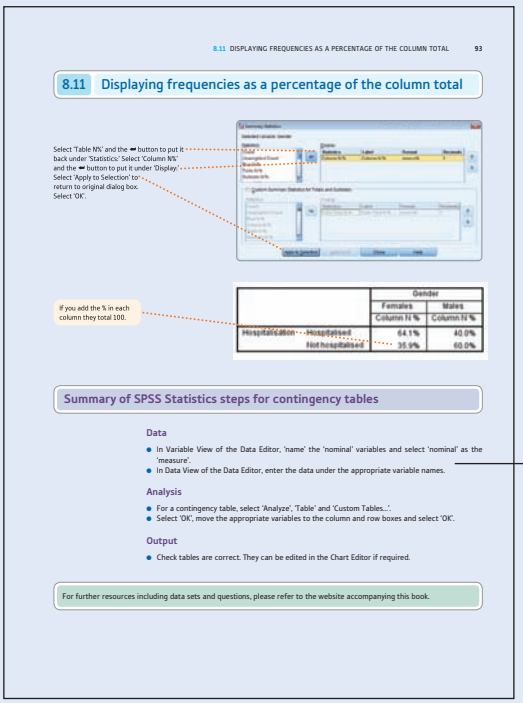
REPORTING THE OUTPUT

- The correlation between musical ability and mathematical ability is −.900. It is usual to round correlations to two decimal places, which would make it −.90. This is more than precise enough for most psychological measurements. Note that there is no need to put a 0 before the decimal point (e.g. −0.90) because a correlation cannot be bigger than ±1.00.
- The exact significance level to three decimal places is .000. This means that the significance level is less than 0.001. We would suggest that you do not use a string of zeros, as these confuse people. Always change the third zero to a 1. This means that the significance level can be reported as being $p < 0.001$.
- It is customary to present the degrees of freedom (df) rather than the number of cases when presenting correlations. The degrees of freedom are the number of cases minus 2, which makes them 8 for this correlation. There is nothing wrong with reporting the number of cases instead.
- In a report, we would write 'There is a significant negative relationship between musical ability and mathematical ability, $r(8) = -.90$, $p < 0.001$. Children with more musical ability have lower mathematical ability.'

The significance of the correlation coefficient is discussed in more detail in Chapter 11 of Howitt, D. and Cramer, D. (2014) *Introduction to Statistics in Psychology*, Harlow: Pearson.

Interpreting the output

Offers a simple explanation of what the important parts of the output mean. SPSS statistical output is presented exactly as it appears on screen to help you become familiar with it.



Summary of SPSS Statistics steps

Summarises the important steps so that you can easily see how to input, analyse and present data.

Introduction

Statistical Package for the Social Sciences (**SPSS**) was initially developed in 1965 at Stanford University in California. Since then it has become the leading data analysis package in the field and available all over the world in universities and elsewhere. Modern computing developments have allowed it to be used on home computers. Because of its popularity and universality, using SPSS is one of the most readily transferable of all research skills. Once learnt, SPSS can be used virtually anywhere in the world. SPSS is constantly being updated, both in terms of the range of statistical techniques covered and the detail of the output.

This book is a stand-alone, step-by-step approach to statistical analysis using SPSS for Windows and is applicable to Releases 10 to 21. It is suitable for students and researchers wishing to analyse psychological, sociological, criminological, health and similar data. SPSS does change with time, but this book will also be helpful to those using earlier releases of SPSS as the changes which affect this book are generally hard to notice. Although the last six releases of SPSS have Statistics in the title and releases 17 and 18 were called **PASW Statistics**, we shall generally refer to all versions as SPSS unless we are speaking about particular versions – in which case we will give their release numbers. This is what is generally done by most users. The official name of the latest release at the time of publication is IBM SPSS Statistics Version 22 (Release 22.0.0).

This book updates the fifth edition of *Introduction to SPSS Statistics in Psychology* to cover recent changes in SPSS. The structure provides the fastest possible access to computerised data analysis with SPSS, even when using the most advanced techniques. Each statistical technique is carefully described, step-by-step, using screenshots of SPSS data analysis and output. The user's attention is focused directly on the screenshots, what each of them signifies, and why they are important. In other words, it is as close as is possible in a textbook to face-to-face individual instruction. Users with little or no previous computer skills will be able to quickly analyse quite complex data and appreciate what the output means.

The chapters have a common pattern. The computer steps (which keys to press) are given in exact sequence. However, this is not the end of any data analysis, and so there are also explanations of how to interpret and report the SPSS output. For this new edition, we have added one new SPSS chapter on two-way repeated measures. SPSS is excellent but it does not include certain procedures which a good researcher may need. So we have included instructions in how to carry out a few additional things. The section on power analysis and meta-analysis which uses software freely available on the internet is now included in the book as the eighth section.

The basic structure of the major chapters is:

- A brief bulleted overview of the use of the procedure or statistical technique. This will often be sufficient to get a clear idea of where and when to use the techniques.
- An account of what the technique is for and what needs to be known in preparation for doing the SPSS analysis. It also gives information about when the technique is

used, when it should not be used, the data required for the analysis, and typical problems that we know from long experience cause users difficulties.

- An illustrative example is given of the appropriate sorts of data for each statistical technique. These examples allow the user to work through our computations, and to gain confidence before moving on to their own data.
- Data entry for a particular statistical analysis is presented visually and explained in adjacent text.
- This is followed by a step-by-step, screenshot-by-screenshot, description of how a particular statistical analysis is done using SPSS for Windows.
- The SPSS statistical output is included exactly as it appears on the monitor screen and in printouts of the analysis. This is crucial – SPSS output can be confusing and unclear at first.
- The key features of the statistical output are highlighted on the output itself, together with simple explanations of what the important parts of the output mean – SPSS output is infamous for its over-inclusiveness.
- Suggestions are made on reporting the statistical findings in reports, theses and publications. These include samples of how to describe research findings and present tables clearly. The form followed is that recommended by the American Psychological Association (APA), which is also widely used by other publishers.

This book is based on the latest version of *SPSS Statistics for Windows* (that is, Release 22); but remains suitable for Releases 10 to 21 because of their similarity. Notes after this Introduction describe the main differences between these releases. Although SPSS is updated every year or so, usually there should be little difficulty in adapting knowledge gained on the older versions to the new version.

Introduction to SPSS in Psychology is an excellent single source for data analysis. It is complete in itself and contains many features not available elsewhere. Unlike other SPSS books, it meets the needs of students and researchers at all levels. However, it is also part of a package of methodology books by the same authors designed to be comprehensive, authoritative and exhaustive. The three volumes in this series are closely tied to each other. The other two are:

- *Introduction to Statistics in Psychology* (2014) (6th edition) (Pearson Education: Harlow): This is a thorough introduction to statistics for all students. It consists of a basic introduction to key psychological statistics and also covers many intermediate and advanced techniques in detail, while maintaining its accessibility to students. It contains chapters on topics, such as meta-analysis, which are seldom covered in other statistics texts. Importantly, the structure of the statistics textbook is closely linked to this book. Thus, anyone following a chapter in the statistics book will, where appropriate, find an equivalent chapter in this book with details of how to do the analysis using SPSS. Similarly, anyone using this book will be able to find a detailed account of the technique in the statistics textbook.
- *Introduction to Research Methods in Psychology* (2014) (4th edition) (Pearson Education: Harlow): This is a major textbook on research methods in psychology. It covers both quantitative and qualitative methods. There are major chapters on report writing, ethics in psychology and searching the literature. All aspects of experimental, field study, survey and questionnaire construction are covered, and guidance is given on qualitative data collection and analysis. There are numerous cross-references to this book and *Introduction to Statistics in Psychology*.

In other words, the three books offer a comprehensive introduction to conducting research in psychology. They may be used independently or in any combination.

Introduction to SPSS in Psychology can be used alongside virtually any statistics textbook to support a wide variety of statistics and practical courses. The range of statistical techniques covered is large and includes the simplest as well as the most important advanced statistical techniques. The variety of techniques described and the relative ease of using SPSS Statistics for Windows ensure that this guide can be used at introductory, intermediate and advanced levels of statistics teaching. The structure of the book is such that statistical procedures are described more or less in order of conceptual difficulty. Generally speaking, computing with SPSS is as easy for advanced statistical techniques as it is for simple ones.

Chapter 2 is essential reading, as it explains data entry and basic computer operating. However, the subsequent chapters can be used on a stand-alone basis if desired. Users with insufficient time to work through the guide chapter by chapter should find enough detail in the relevant chapters to complete an SPSS analysis successfully. Table 1.1, at the end of Chapter 1, states which chapter is most appropriate for which purpose, thereby enabling the reader to move directly to that part of the book.

Those who work steadily through the book will profit by doing so. They will have a much better overview of SPSS computing procedures. For most readers, this is possible in a matter of hours, especially if they have prior knowledge of statistics.

SPSS has an extensive catalogue of statistical procedures – far more than could be included. We have selected those suitable for most purposes when the range of possibilities is likely to confuse the average reader. The quickness and ease of SPSS mean that more advanced users can explore the alternatives by using the menus and dialog boxes. Most users will find our coverage more than sufficient.

The data and statistical analyses carried out in this book correspond almost always to those in the authors' accompanying statistics text, *Introduction to Statistics in Psychology* (2014) (6th edition) (Pearson Education: Harlow). This book is referred to as *ISP*, followed by the corresponding chapter or table number.

Dennis Howitt
Duncan Cramer

Acknowledgements

Our names may appear on the cover of this book but it is the talent and hard work of other people that makes it what it is. The team that Pearson provides to do all of the clever, difficult stuff makes all of the difference between a dull manuscript full of errors and this book which, we believe, is a credit to us but undeservedly so. These back-room stars deserve most of the praise and this is our opportunity to thank them. The team for this edition is international and much of the scarily efficient coordination of everything was in the hands of Priyadharshini Dhanagopal. She was the Managing Project Editor working in Chennai. Regrettably, the Internet ensured that we had no place to run and no place to hide. Without her doing all manner of things in the background which we cannot hope to understand, our lives would have been more difficult and more disorganized than ever. We can only wonder what her job specification is and how one person can fulfil it. But she can and we are more than grateful for that. Angela Hawksbee was the Production Controller. She liaised with type-setters, printers and the like and made sure that everything happened when it should happen. Fortunately, this did not involve us directly as she is formidable in her work judging by the smooth way with which everything progressed.

Janey Webb was the Acquisitions Editor though now she has the title of Publisher. We are not sure what this is but roughly translated it seems to mean 'tower of strength'. We have been working with Janey for quite a few editions now and it never fails to amaze us how she always has the energy and determination to get the best new edition out of us that we are capable of. This is all done in ways far too subtle for two psychologists to understand. Somehow, we found ourselves wanting to do what we did not want to do and feeling happy about it. How she gets inside of the topic in question even for arcane things like SPSS and statistics will never cease to amaze us.

If you like the way that this book looks (and we think it looks great) there were two main people to thank. The Cover Designer was Nicola Woowat. How she comes up with fresh ideas which make SPSS look so exciting we do not know but we are glad that she can. Kevin Ancient designed the text. His work is so important to a book like this one. It gives a coherent structure to each chapter, which the manuscript alone cannot. It is far easier to read the book than the manuscript because of this. Kevin's master plan was executed by the Copy-editor Anne Henwood. She has the manuscript in one hand, the design in another hand, and brings them both together in another hand. This, of course, makes her super human too. She is enormously patient and polite even when we are probably being very annoying. The trouble is that she is always right. While doing all of this, she still managed to spot all sorts of errors. When the copy has been turned into the page proofs then the Proof Reader goes to work getting rid of all of those awkward glitches, typos, and mistakes which can make a textbook hard to read and very irritating. Brian Burge is our new best friend for doing this so well and thoroughly. He could spot things that we could not even see even after he told us exactly where they were! Annette Musker was the Indexer. She makes it so much easier to find one's way

around the book. It is a job that anyone can make a mess of but needs a special talent to do well. Annette gets the difficult balance needed for an index exactly right. Of course, there are lots of other people that we ought to thank but can only write how much their contributions are appreciated by us.

Thanks

Dennis Howitt and Duncan Cramer

Publisher acknowledgements

We are grateful to the following for permission to reproduce copyright material:

The screenshots in this book are reprinted by permission of IBM SPSS Inc. Screenshots in this book are copyright IBM SPSS Inc. IBM SPSS is a registered trademark and the other product names are registered trademarks of IBM SPSS Inc.

Key differences between IBM SPSS Statistics 22 and earlier versions

SPSS Statistics 21

There seem to be very few differences between 22 and 21 and between 21 and 20 for the procedures described in this book.

SPSS Statistics 20

There seem to be very few differences between 20 and 19 for the procedures described in this book.

SPSS Statistics 19

The name of the software became SPSS Statistics. There seem to be very few differences between 19 and 18 for the procedures described in this book.

PASW Statistics 18

There seem to be very few differences between 18 and 17 for the procedures described in this book.

PASW Statistics 17

This version of SPSS was called PASW, which stands for Predictive Analytic Software. Otherwise, there appear to be very few differences between 17 and 16 for the procedures described in this book. Data Reduction which includes Factor Analysis is now called Dimension Reduction. Right clicking on the keys in the 'Select If: If' box no longer gives a description of what the keys do.

SPSS 16

There seem to be very few differences between 16 and 15 for the procedures described in this book. Basic tables are no longer available in the Tables procedure, so the tables in Chapter 8 are produced with Custom Tables. In the dialog boxes the OK, Paste,

Reset, Cancel and Help options are on the bottom of the box rather than on the right-hand side while the analyses options are on the right-hand side rather than at the bottom of the box.

SPSS 15

There also appear to be very few differences between 15 and 14. The Graph menu in 14 displays all available options. The procedure for Chart Builder is described in Chapter 4 of this book as it was not available in releases before 14. A slightly alternative procedure is shown in Chapters 6, 9 and 11.

SPSS 14

Similarly, there seem to be very few differences between 14 and 13. Release 13 does not have Chart Builder, which was introduced in 14. In 14 the Properties dialog box of the Chart Editor has separate boxes for Text Style and Text Layout.

SPSS 13

The major differences between 13 and 12 are to 'Compute Variable . . .', 'Scatter/Dot . . .' and the 'Chart Editor'. Also the plots in the output of 13 are shaded.

In 12 the 'Compute Variable' dialog box has a single 'Functions' menu from which options can be chosen. 'Scatter/Dot . . .' is called 'Scatter . . .', the 'Scatter/Dot' dialog box is called 'Scatterplot' and there is no 'Dot' option.

In 12, to label the slices of a pie diagram and add the percentages of cases in each, double click anywhere in the 'Chart Editor', double click on the pie diagram (to open the 'Properties' dialog box), select 'Data Value Labels' (in the 'Properties' dialog box), select 'Count' in the 'Contents' box, select the red 'X' (to put 'Count' in the 'Available' box), select the variable name (e.g. 'Occupation'), select the curved upward arrow (to put 'Occupation' in the 'Contents' box), select 'Percent' and the curved upward arrow (to put 'Percent' in the 'Contents' box), select 'Apply' and then 'Close'.

To fit a regression line to a scatterplot, click on a dot in the chart of the 'Chart Editor' so that the circles in the plot become highlighted, select 'Chart', select 'Add Chart Element', select 'Fit Line at Total' (which opens the 'Properties' dialog box). Assuming that the 'Fit Line' tab is active, select 'Linear' (this is usually the default) and then 'Close'.

SPSS 12

The major differences between 12 and 11 also apply to 10. They are relatively few. In 11 and 10 variable names cannot begin with a capital letter and are restricted to eight characters. The 'Data' and 'Transform' options are not available in the 'Viewer' or 'Output' window. Some output, such as partial correlation and reliability, is not organised into tables. The 'Chart Editor' works differently. To fit a regression line to a scatterplot, double click anywhere in the scatterplot to open the 'Chart Editor', select 'Chart', select 'Options . . .' (which opens the 'Scatterplot Options' dialog box), select 'Total' under 'Fit Line' and then 'OK'.

PART 1

Introduction to SPSS Statistics

CHAPTER 1

A brief introduction to statistics

Overview

- Unfortunately, there is no gain without some effort in statistics. There are a small number of statistical concepts which need to be understood in order to speed up learning how to use statistics in research. Each of these is discussed and explained in this chapter.
- Key ideas covered in this chapter include score variables versus nominal (category) variables; unrelated versus **related designs**; descriptive versus inferential statistics; and significance testing. With a knowledge of each of these it is possible to quickly develop a working knowledge of statistical analysis using SPSS Statistics.
- The appropriate statistical analysis for data depends very much on the particular type of research design employed. You need a basic understanding of what statistics are appropriate for your research design.
- The chapter provides detailed advice on how to select a statistical technique for the analysis of psychological data.

1.1 Basic statistical concepts essential in SPSS Statistics analyses

The basics of statistics are quite simple. The problem is in putting these elements together. Nobody can become expert in statistical analysis overnight but a very small amount of knowledge can lead to quite sophisticated analyses being carried out by even the most inexperienced of researchers. Mathematical ability has very little role to play in data analysis. Much more important is that the researcher understands some basic principles of research design. There are close links between different research designs and what the appropriate statistical analysis methods for one's data are. At the most basic level, there are two broad classes of research design – the comparative and the correlational designs. Of course, each of these has any number of variants. The type of research design involved in the study lays down broadly the sort of statistical tests, etc. which are needed for the analysis of the data from that study. Sometimes the personal preferences of the researcher play a part since, quite often, there are several ways of achieving much the same ends.

But before we can discuss research designs, there are two basic concepts we need to understand as they are part of the jargon of statistics and SPSS Statistics:

Variable: A variable is any concept that can be measured and which varies. Variables are largely inventions of the researcher and they can be very different from study to study. There are a few fairly standard variables, such as age and gender that are very commonly measured. Typically, however, the variables used tend to be specific to particular topics of study. Variables are the means by which psychologists attempt to measure the concepts that they use in their research – a variable, generally, cannot perfectly measure a concept and so is an approximation to the concept. For this reason, it is important to understand that data which involves variables and theory which involves concepts do not always map perfectly with each other.

Cases: A case is simply a member of the sample. In psychology a case is usually a person (i.e. an individual participant in the research). Cases are very much SPSS Statistics jargon. It is a wider and more embracing term than the participants which psychologists talk about.

Variables normally appear in SPSS analyses as the columns of the data spreadsheet. Cases (normally) appear in SPSS analyses as the rows of the data spreadsheet though it is possible to reverse these. In other words, variables and cases can be set out in the form of a **matrix** or a two-dimensional array. The size of the matrix depends on the number of variables and cases involved.

1.2 Basic research designs: comparative versus correlational designs

Comparative designs

The basic comparative design compares the typical or average score of a group of participants with that of another group. This might involve comparing a group of men with a group of women or comparing an experimental group with a control group in an experimental study. This design is illustrated in Table 1.1. Usually, in this sort of design, the comparison is between the average score for one group and the average score in the other group. Usually, what most people refer to as the *average* is called by

Table 1.1 Simple comparative design such as an experiment

Participant (case)	GROUP A (e.g. experimental group)	Participant (case)	GROUP B (e.g. control group)
1	13	11	5
2	12	12	8
3	10	13	6
4	7	14	9
5	5	15	3
6	9	16	6
7	5	17	5
8	14	18	4
9	12		
10	16		
Mean =	10.30	Mean =	5.75

statisticians the *mean*. So the design can be used to assess whether, say, the average time taken by males getting ready for a first date is different from the average time taken by females.

This is the basic version of a whole range of statistical procedures which compare the average scores in different groups in more complex research designs. The **analysis of variance** (ANOVA) involves a whole family of different research designs based on this basic principle. Look at Figure 1.1 for more information.

The correlational design

The basic correlational design is one in which the researcher measures several different things at the same time, using a single group of participants. These things which are measured might be gender, age, IQ, extraversion and dogmatism. This basic correlational design is illustrated in Table 1.2.

The statistical analysis of this sort of design is usually based on the **correlation coefficient** or some other closely related statistical procedure based on the correlation coefficient. A correlation coefficient is a numerical index of the size of the relationship between two measures. The data from a correlational design may be analysed using a variety of statistics, as can be seen in Figure 1.1.

Correlational designs are sometimes called cross-sectional studies. They can be more complex, for example, when the researcher adds a time (temporal) dimension to the research design. There are special statistics to deal with these more complex designs (e.g. causal modelling such as **LISREL**) but these are essentially correlational in nature.

It would be misleading to pretend that the above covers every available statistical technique but a surprising range of statistics can be better understood if the underlying research design is clear to the researcher. Also remember that statistics is a mature discipline in its own right so it is unrealistic to assume that there are instant shortcuts to mastery of statistics in psychology. Getting basic concepts clear goes a long way towards this mastery, as does some experience.

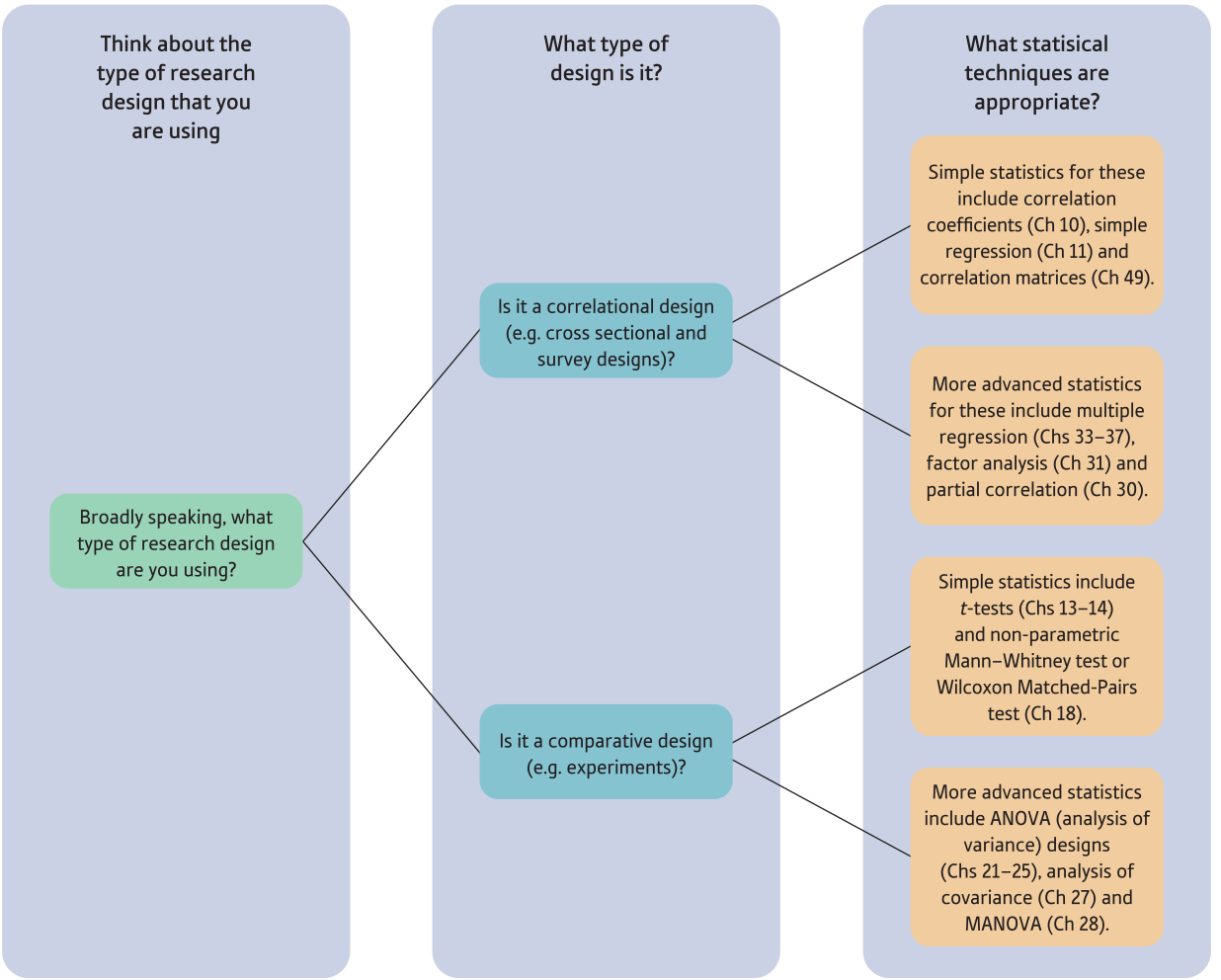


FIGURE 1.1 Fundamentals of design and statistical analysis

Table 1.2 The basic correlational design

PARTICIPANT	GENDER	AGE	IQ	EXTRAVERSION	DOGMATISM
1	Female	26	110	15	9
2	Male	31	130	19	6
3	Female	25	160	22	4
4	Female	22	110	34	8
5	Male	33	170	12	3
6	Female	28	140	17	7
7	Male	29	90	16	6
8	Male	34	130	22	5
9	Female	23	80	26	4
10	Male	27	70	11	2

1.3 The different types of variables in statistics

One's ability to use statistics in a practical context will be made much easier if some basic stuff is learnt about the fundamental different types of variables in statistics. Different types of variables require different kinds of statistical techniques for their analysis. So there are two basic questions that need to be asked:

- What types of variables do I have?
- What statistical tests analyse the data from these variables in the way that I want?

Fortunately, there are just two main types of data, so this is relatively straightforward. On the other hand, there are many different statistical tests and techniques. Of course, the way to learn about each of these is to gain some experience trying each of them out by working through the chapters which follow in this book. Most of the chapters in this book cover just one statistical technique or test in each chapter. The important thing is that each chapter tells you exactly what sorts of data (variables) are appropriate for that test or technique – and then how to do the analysis using a computer.

Types of variable

For all practical purposes, variables can be classified as being of *two* types (see Figure 1.2):

- *Score variables* Some variables are scores. A score is when a numerical value is given to a variable for each case in the sample. This numerical value indicates the quantity or amount of the characteristic (variable) in question. So age is a score variable since the numerical value indicates an increasing amount of the variable age. One could also describe this variable as quantitative.
- *Nominal or **category or categorical variables*** Some variables are measured by classifying cases into one of several named categories. These are also known as nominal, categorical or category variables. A better name for them might be qualitative variables because they measure the qualities of things rather than their quantities. For example, gender has two named categories – male and female. Nationality is another example: English, Welsh, Irish and Scottish are the nationalities of people of Britain. They have *no* numerical implications as such. To say that a person is Scottish is simply to put them into that named category. There is one risk of confusion – categories such as gender are usually entered into SPSS Statistics using different numbers to represent the different categories. For example, the variable gender has two categories – males could be represented by the number 1 and females by the number 2 (or vice versa). The numbers used are arbitrary – it could be 1002 and 2005 if the researcher desired. It is vital not to confuse these numbers, which merely represent different coding categories or qualities, with scores, which indicate the quantity of a variable which characterizes each participant in the research. For this reason, it is important to label the different values of nominal variables in full in the SPSS data spreadsheet since the number codes, in themselves, mean nothing. This is easily done, as is shown on pages 25–26.

The alternative traditional classification system

Sometimes variables are classified as nominal, ordinal, interval and ratio. Most textbooks go into arcane explanations about the difference between ordinal, interval and **ratio data**. This is mainly of conceptual interest and of little practical significance in

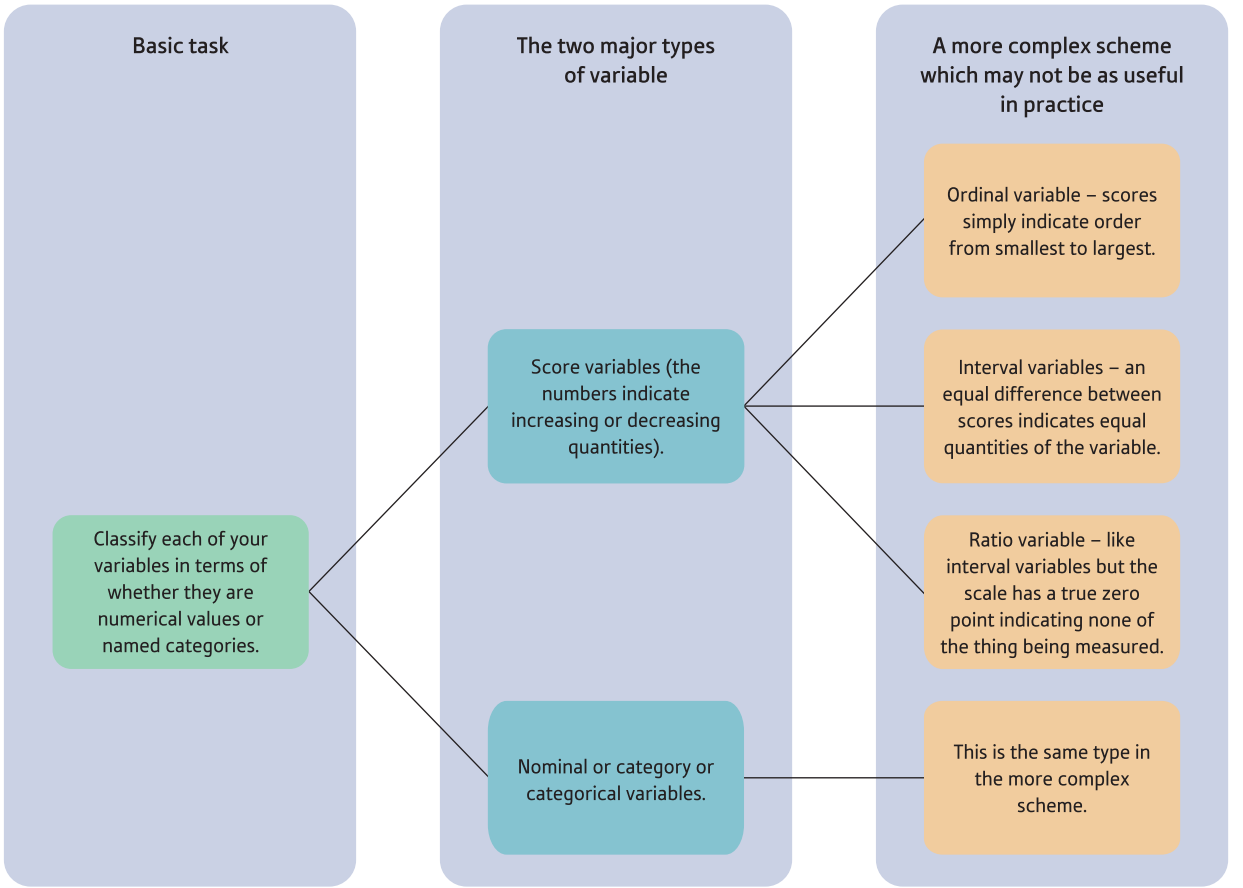


FIGURE 1.2 The two schemes for different types of variables

selecting appropriate statistics. Generally speaking, we would advise that this system should be ignored because it does not correspond with modern practice and it causes great confusion. *Nominal* is exactly the same as our classification of nominal (category) data and is important since a particular set of statistical techniques is called for to analyse category data. Of the other three, *interval* measurement is the most important. Interval measurement is where the steps on the scale of measurement are equal (just as centimetre steps on a rule are equal). Some psychologists are inclined to the view that this scale of measurement should reflect the underlying psychological variable being measured. Unfortunately, it is very difficult (i.e. impossible) to identify whether a psychological measure has equal intervals but, nevertheless, it is a sort of holy grail to them. Others, ourselves included, take the view that, so long as the numerical scale on which variables are measured has equal intervals (which is always the case except for nominal or category data, of course, from this perspective), then there is no problem as it is these numbers on which the statistical calculation is based and not some mystical underlying measurement scale. However, as a concession, we have mentioned equality of intervals as being desirable from time to time in the text. *Ratio* measures have equal intervals and a zero point, which means one can calculate ratios and make statements such as: 'One score is twice as big as another score.' Unfortunately, yet again, it is impossible to identify

any psychological variables which definitely are measured on a ratio measurement scale. Finally, **ordinal data** is data which does not have equal intervals, so scores only give the rank order of scores. Since the size of the intervals do not matter for ordinal data, then it is assumed that any psychological score data corresponds to the ordinal measurement scale at a minimum. For this reason, some psychologists have advocated the use of non-parametric (distribution-free) statistics for the analysis of much psychological data. The problem is that these techniques are not so powerful or flexible as most statistics employed in psychological research. You will find an extended discussion of ordinal, interval and ratio scales of measurement in Howitt, D. and Cramer, D. (2014) *Introduction to Research Methods in Psychology*, Harlow: Pearson.

Importance of deciding the types of variable involved

It is essential to decide for each of your variables whether it is a nominal (category) variable or a score variable. Write a list of your variables and classify each of them if you are a beginner. Eventually, you will do it somewhat automatically and usually without much thought. The statistical techniques which are appropriate for score variables are generally inappropriate for nominal or category variables because they measure qualities. So, for example, it is appropriate to calculate the mean (numerical average) of any variable which is a score (e.g. average age). On the other hand, it is totally inappropriate to calculate the mean (average) for variables which consist of categories. It would be nonsense to say that the average nationality is 1.7 since nationality is not a score. The problem is that SPSS works with the numbers in the data spreadsheet and does not know whether they are scores or numerical codes for different categories. (Though SPSS does allow you to classify your variables as ordinal or nominal.)

Do not be surprised to find that all or nearly all of your variables are scores. Psychologists have a predilection for measuring their variables as scores and it is likely that this rubs off on psychology students. What this means is that statistics for nominal or category data are often not needed, thus simplifying the task of data analysis.

1.4 Descriptive and inferential statistics compared

The difference between descriptive and inferential statistics

There are two main types of statistical techniques – *descriptive* and *inferential* statistics:

- **Descriptive statistics** chiefly describe the main features of individual variables. So calculating the average age of a sample of people is an example of descriptive statistics. Counting the number of English people would be another example of descriptive statistics. If one variable is considered at a time this is known as **univariate** statistics. **Bivariate** statistics are used when the relationship between two (or more) variables is being described.
- Inferential statistics is a totally distinct aspect of statistics. It only addresses the question of whether one can rely on the findings based on a *sample* of cases rather than *all* cases. The use of samples is characteristic of nearly all modern research. The problem with samples is that some of them are not similar to the populations