Ben Baarda
Cor van Dijkum
Martijn de Goede

ENGLISH EDITION

Introduction to Statistics with SPSS

A guide to the processing, analysing and reporting of data

Second fully revised English edition (corresponds to the fifth Dutch edition)

Noordhoff Uitgevers

Introduction to Statistics with SPSS

A guide to the processing, analysing and reporting of data

Ben Baarda

Cor van Dijkum

Martijn de Goede

Second fully revised edition
(corresponds to the fifth Dutch edition)

Noordhoff Uitgevers Groningen/Houten

Cover design: Rocket Industries, Groningen
Cover Image: Getty Images
English translation: Prue Gargano

If you have any comments or queries about this or any other publication, please contact: Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB Groningen, email: info@noordhoff.nl

In regard to some texts and/or illustrations the publisher was not able to trace all possibly entitled copyright holders despite careful efforts. If you are of the opinion that you are the copyright holder of texts and/or illustrations in this book we request you to contact the publisher.

0 / 14

ISBN (ebook) 978-90-01-85744-8
ISBN 978-90-01-83441-8
NUR 916

Preface to the second edition

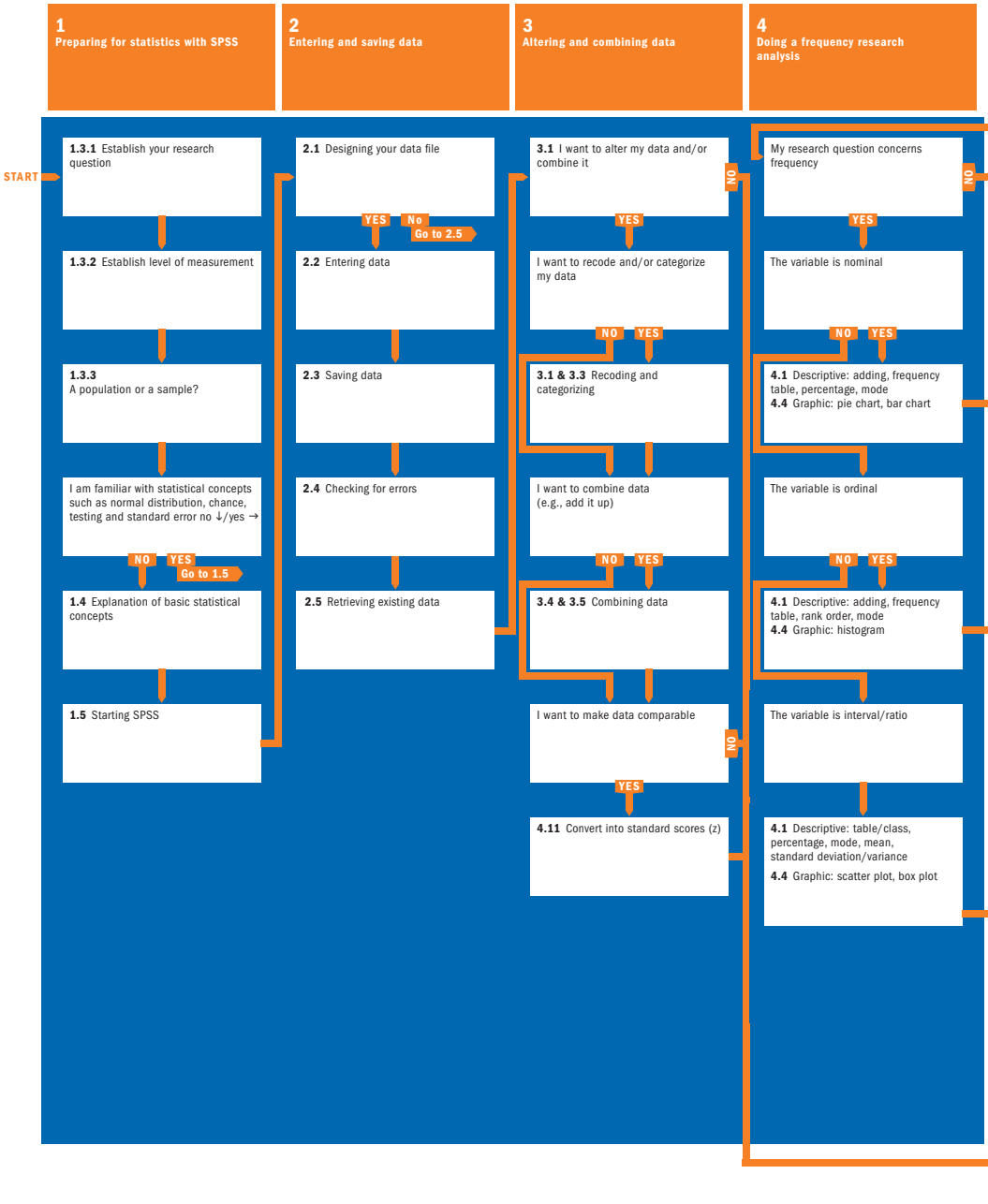
Introduction to Statistics with SPSS contains tips for processing and analysing research data using SPSS. SPSS stands for 'Statistical Products and Service Solutions' and it is among the most used statistical software for entering statistical data and analyzing it. SPSS has become a part of IBM, and as such, its official name is IBM SPSS Statistics. This book will familiarize you with the SPSS package as well as with statistics. We invite you to start working by using a data set on the relationship between money and happiness.
When writing this book we used SPSS version 21.0.
The following revisions have been made to the former edition:

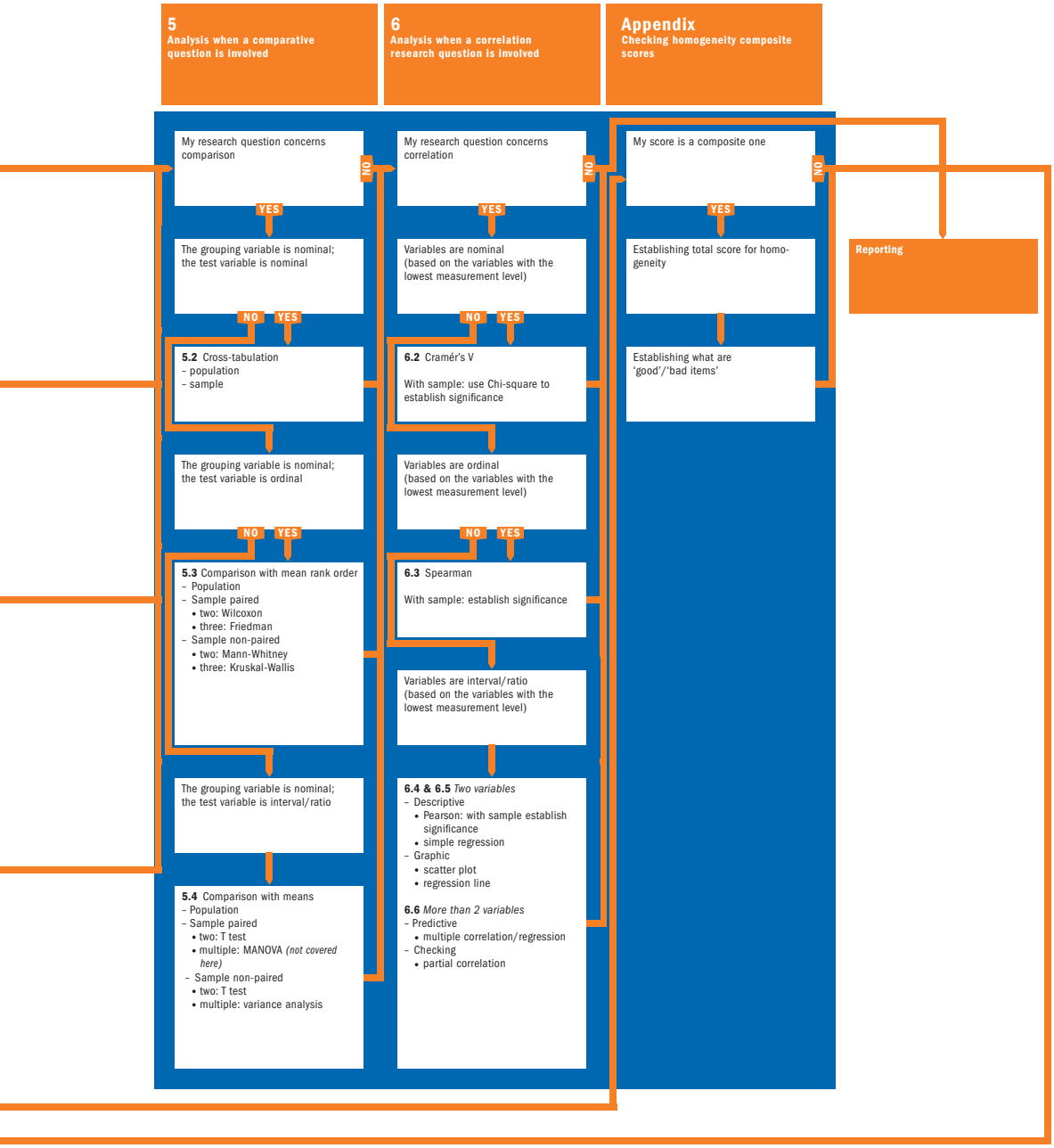
- The screenshots include numbered post-its which indicate the SPSS procedure to follow when you are doing a statistical analysis. The text refers directly to them.
- Appendix II explains how to use a log-linear analysis in situations where there are more than two nominal variables: for example, sex, marital/family status and whether you work part-time or not.
- Appendix III explains how to perform a variance analysis in the event of more than one grouping variable: for example, both sex and marital/family status.

You will make optimal use of *Introduction to Statistics with SPSS* if it is used in conjunction with the books *Basisboek Methoden en Technieken* (in Dutch only) and *Research. This is it!* These two books are practical manuals for designing and conducting research. *Statistiekkwartetspel* (in Dutch only) can be used to compliment the material. It explains the essential statistical concepts and tests your understanding in a playful way. But naturally, *Introduction to Statistics with SPSS* is fully self-explanatory and can be used without reference to the other two books.

Ben Baarda
Cor van Dijkum
Martijn de Goede
Winter 2014

Introduction to Statistics with SPSS





Introduction

Tens of thousands of organisations all over the world use IBM® SPSS® Statistics, in the private and public sector and in academia. The software helps them find answers to key business and research questions and makes it easy for them to access and manage research data, conduct various analyses and share the results with others.

IBM SPSS Statistics consists of a range of solutions that work closely together. These software solutions support the entire analytical process – from research planning through data collection and analysis to reporting on the research results. Advanced analyses and models make it possible to identify patterns emerging from data and predict trends and developments to a high degree of accuracy.

Originally used primarily in social sciences, IBM SPSS Statistics was implemented e.g. to facilitate research in the fields of political science, sociology and psychology. The software package is also used these days in the context of marketing, market research, policy research and fraud detection.

All kinds of companies use IBM SPSS Statistics to support customer-driven business practices. They generate customer profiles and segment target audiences, mapping the current and future behaviour and needs of their customers in order to tailor an effective response. Some companies make use of the software to identify their best customers or most interesting target groups to maximise marketing returns. Organisations also use it to acquire and retain customers.

Media companies, for example, apply IBM SPSS Statistics to explore public opinion about specific TV shows as expressed on social media, using the results to make modifications to boost viewing figures. Sports organisations take advantage of analytics to predict and mitigate the risk of injuries. The police deploy the software to predict and prevent crime, while municipal authorities use it to fight benefit fraud more effectively. Hospitals apply analytics to determine which treatments are most effective.

These are only some of the contexts in which organisations in highly diverse sectors are able to use IBM SPSS Statistics to make better decisions and improve their operating results. The book before you will reveal how it all works, making it possible for you to use IBM SPSS Statistics successfully in your own professional context.

Enjoy!

Laila Fettah

IBM Netherlands

Contents

Using this book effectively

Our intention in *Introduction to Statistics with SPSS* is to familiarize you with the software package as you go. After you have read a general introduction to the basic rules of SPSS you can start entering data into the computer straight away and analysing them. We invite you to refer to a data set on money and happiness. This data set consists of fictitious data collected from 500 women and 500 men via a questionnaire.
We will proceed in the usual research manner:

- You will start by entering the data (Chapter 2)
- You will then learn how you can change and combine data in SPSS (Chapter 3)

After this preparatory work you will learn how to analyze the data. The analytical technique that you will use depends on your specific research question. Your research problem will always consist of at least one specific research question to which you want to find an answer by carrying out research.
We can distinguish the following research questions:

1 Research questions focussing on frequencies (determining how many times something occurs; Chapter 4)
2 Comparative research questions (determining and testing differences between two or more groups in respect of a single construct; Chapter 5)
3 Research questions focussing on relationships (determining and testing the relationship between two variables; Chapter 6)

The Appendix explains a number of techniques that are extensions of the basic statistical techniques, including homogeneity analysis, log-linear analysis and univariate analysis.
Using SPSS responsibly not only requires an understanding of SPSS but also of statistics. In the first chapter we discuss basic concepts such as level of measurement, normal distribution, chance, significance, one and two-tailed hypothesis testing, power and effect size. Chapters 4, 5 and 6 explain that the choice of statistical analysis is not only dependent on the type of research question but also on the level of measurement of the specific variable and whether it concerns a population or sample.

In the Dutch version of SPSS a comma is used as the decimal point. This is contrary to the English and American language use. You can change the decimal point sign in SPSS by clicking in the Variable Type menu om ‘Comma’ or by changing the system language. Because most of the students will use the Dutch version of SPSS we did not change the decimal sign, so the screendumps are comparable to the ones you get with the Dutch version of SPSS.

The learning material is presented in the form of answers to questions posed by a hypothetical user:

1 What statistical analysis technique should I use given my research question?
2 What does this technique entail, what are the assumptions and how do I apply this technique in SPSS?
3 Once the chosen analysis technique has been applied, how do I read the output? What do the results imply for my research question?
4 How do I report on the conclusions in my research report?
5 How do I write up my conclusions?

Each technique is illustrated using an example of SPSS output. We explain how to read it and what it all means. We also show how to report on it. This book is intended to serve an educational purpose and we have paid a lot of attention to the way the material is presented. Each chapter follows a similar pattern:
Prior knowledge: the knowledge which is needed to understand the chapter
Questions: the questions which will be addressed in the chapter
Terms: the main terms discussed in the chapter
Content: the specific content relating to the subject of the chapter
Summary

We used SPSS 21.0 when writing this book. This software package has been installed on most computers at tertiary institutes of professional education and universities. You can also order the software package at a discount at www.surfpot.nl . The data set on which the analyses will be performed can be found at www.introductiontostatisticswithspss.noordhoff.nl

For I don't care too much for money,
Money can't buy me love

— John Lennon & Paul McCartney

Prior knowledge

No prior knowledge is required, though we advise you to consult *Basisboek Methoden en Technieken* (5th revised edition, 2012, in Dutch only) or *Research. This is it!* (2nd edition, 2014) and *Statistiekkwartetspel* (1st edition, 2010, in Dutch only).

1 Preparing for statistics with SPSS 21.0

The following questions will be raised in this chapter:

- What topics will be discussed in this book? (Section 1.1)
- What is the case study 'Money and Happiness' about? (Section 1.2)
- How to deal with the basic matters: (Section 1.3)
  - Does the research focus on frequencies, differences or relationships? (Section 1.3.1)
  - What is the level of measurement? (Section 1.3.1)
  - Does it involve a population or a sample? (Section 1.3.3)
- What do the following statistical terms mean: *normal distribution*, *standard error*, *reliability*, *significance*, *one-tailed and two-tailed hypothesis testing*, *relevance*, *effect size*, *degrees of freedom*? (Section 1.4)
- How does SPSS 21.0 work using Windows? (Section 1.5)

1.1 Introduction

When you are doing research, choosing the right statistical technique for analysing the collected data is one of the key decisions in the long list of decisions you have to make. The ultimate objective is to answer the research question or questions. To demonstrate where data analysis fits in the research cycle as a whole, we will firstly outline the various stages of the research cycle, with each stage of the research cycle shown in the form of a question:

1 What is (are) the research question(s)? What is the research objective?
2 How do I search for information (in sources such as literature reviews)?
3 What type of research should I conduct?
4 How is the research best designed?
5 Should I involve the population in my research or take a sample?
6 What method of data collection should I use?
7 How do I prepare the data for the analysis?
8 How do I analyse my data?
9 How do I report on and evaluate the research?

In this book, stages 7 and 8 and part of 9 will be discussed: preparation, analysis and description of the research data to be analysed using SPSS. The following steps will be discussed by reference to a study into the relationship between money and happiness (see Fig. 1.1):

- How do I prepare the collected data for computer analysis? (Chapter 2)
- How do I alter and adjust data using SPSS? Before beginning the analysis you will have to reverse the values of certain variables first (recode them) or combine the values of the variables to generate a new score (Chapter 3) and then check their reliability (Section 3.4 and Appendix).
- How do you choose the right statistical analysis? To determine this, you will first have to determine what type of research question it is (Section 1.3.1.). Does it involve frequencies (Chapter 4), differences (Chapter 5) or relationships (Chapter 6)? Next, you will have to determine the level of measurement of the data (Section 1.3.2). Finally, you will have to determine whether it involves a population or a sample. On the basis of the list of guidelines in 'How to analyse your data', you should then determine which statistical analysis best suits your research question. You can find these guidelines on the inside of the cover as well as on a loose sheet inside the book itself.
- How do you perform the analysis using SPSS and how do you interpret the results? This will be indicated in relation to each statistical analysis we discuss. We will also discuss how to report the results.

As well, this chapter will discuss a number of important terms, including normal distribution, significance, chance and standard error (Section 1.4). In the last section of this chapter (1.5) we discuss how to start SPSS.

1.2 Money and happiness: introduction to the case study

Before discussing how to use SPSS, we will introduce the case study 'Money and Happiness' (see Fig. 1.1). A representative sample of the Dutch adult population consisting of 500 women and 500 men between

25 and 55 were asked the fictitious research questions shown in Fig. 1.1. The choice of this age group was a conscious one. Many young people are still studying and therefore have no steady income. People aged 55 or older may have retired (or at least partially) from the labour process, which means that their financial situation will have changed. The data relating to this research study can be found on the website www.introductiontostatisticswithspss.noordhoff.nl under 'data1'.

FIGURE 1.1 'Money and happiness' case study

Does money make you happy?

A researcher wants to know whether there is a relationship between money and happiness, and in particular, whether money makes you happy. His central research question was whether there is a positive correlation between personal wealth and happiness. The concept of money was defined broadly by the researcher and extended beyond income and personal wealth to include the financial resources one has access to. Happiness was defined as the extent to which a person is satisfied with the life he or she leads. The researcher operationalized these concepts in the form of a questionnaire. To measure the aforementioned concepts, five items were formulated for both money and happiness.

Money was measured by the items:

1	I own a car	Yes/No
2	I own a house or apartment	Yes/No
3	I own a tablet	Yes/No
4	I receive subsidized health care	Yes/No
5	My rent is subsidized	Yes/No

Happiness was measured by the items:

1 If I had the chance to repeat my life I would have it the same way.

○ Disagree completely ○ Disagree ○ Partly agree/disagree ○ Agree ○ Agree completely

2 Most people have it much better than me.

○ Disagree completely ○ Disagree ○ Partly agree/disagree ○ Agree ○ Agree completely

3 Life is enjoyable.

○ Disagree completely ○ Disagree ○ Partly agree/disagree ○ Agree ○ Agree completely

4 Life is tough.

○ Disagree completely ○ Disagree ○ Partly agree/disagree ○ Agree ○ Agree completely

5 I feel lonely.

○ Disagree completely ○ Disagree ○ Partly agree/disagree ○ Agree ○ Agree completely

Assuming that happiness not only depends on financial means, the researcher asked the respondents questions relating to easily measured factors such as sex, age, level of education and marital/family situation.

Gender ○ Male ○ Female

Age in years ...

Domestic status
- ○ Live alone
- ○ Live with partner
- ○ Live with partner and children

Highest level of education achieved
- ○ Basic level of education (primary school, preparatory technical education, preparatory secondary education)
- ○ Secondary education (general secondary education or secondary vocational education)
- ○ Higher education (university, advanced vocational training, tertiary level professional education)

The researcher subdivided the main research topic into several research questions:

1 How many people own a car, house or tablet?
2 How many people receive subsidized health care?
3 How many people pay subsidized rent?
4 To what extent are people happy with the life they lead?
5 To what extent do people feel lonely?
6 Are there any differences between men and women in regard to car ownership, tablets, subsidized health care and subsidized rent?
7 Is there any difference between the following groups in regard to how satisfied they are with the life they lead?
   - People with and people without a partner
   - People with children and people without children
8 Do men and women differ in the extent to which they feel lonely?
9 Do men and women differ in the extent to which they are satisfied with the life they lead?
10 Is there a relationship between the extent to which people are satisfied and their age?
11 Is there a relationship between money and happiness?

1.3 How to analyse your data: a guide to the process

To determine the correct statistical analysis to apply, you will need to have answers to the following:

1 Are the questions about frequency (how often/to what degree), difference, relationships, or a combination of them all?
2 What is the level of measurement (nominal, ordinal, interval/ratio) of the data you collected?
3 Does it involve a population or a sample?

The list of guidelines in 'How to analyse your data' (found inside the cover and on a loose sheet inside the book itself) has been designed to address these questions. We will deal with them in greater detail in the next section.

1.3.1 What are your specific research questions?

The research questions are the starting point in any research and they will determine which type of statistical analysis is most suitable. A research study should always have at least one research question which needs to be answered.

In general, research questions fall into three categories:

- Research questions focussing on how often and to what extent something happens (*frequency*): for example, 'How happy are people?' or 'What percentage of the population owns a car?'
- *Comparative* research questions: for example, 'Are men happier than women?'
- Research questions focussing on *relationship*s: for example, 'Is there a relationship between money and happiness?'

It will be apparent that questions 1 to 5 of the case study research questions (Section 1.2) can be classified as frequency research questions. A typical research question is how many people own a car. Chapter 4 will give

an example of a frequency analysis. Those research questions 6 to 9 mentioned in Section 1.2 are comparative research questions. How to analyse this type of research question will be discussed in Chapter 5. Research questions 10 and 11 are research questions focussing on relationships. Chapter 6 will describe how to analyse this type of research question.

1.3.2 What is your data measurement level?

Once you have determined the type of research question you are going to study (see first column in the guidelines 'How to analyse your data') you then determine the level of measurement of the variables. In the guidelines, you will find this under frequency, difference and relationships. For each research question, indicate what the level of measurement is for the variables in question.

The level of measurement of the variables in question of research question 9 (difference between men and women in the extent to which they feel happy) differs from the level of measurement of the variables in research question 11 (relationship between money and happiness).
The variable 'sex' has only two values or categories, namely male and female. While there is a difference between them, it is not in the sense of being more or less. A male is different from a female but not more or less. The same holds true for marital/family status: a person could be single or be in a relationship with or without children. This type of response category implies a *nominal level of measurement*. While you can indicate how many men and women have a car, it is not possible to indicate that someone is more male or more female. You are either male or female: you cannot be a degree of male.

Nominal level of measurement

With an *ordinal level of measurement* you can measure differences, though the difference between the categories cannot be expressed numerically. With level of education, for example, there is a difference in level in terms of higher and lower. Higher general secondary education is more advanced than junior secondary vocational education, though it is not possible to express just how much more advanced it is. This also holds true for the medals awarded for a championship. The 100 metre sprinter who won the gold medal will have run faster than the one who won a silver medal. The gold medal does indicate that he has run faster but not how much faster.

Ordinal level of measurement

With *interval/ratio levels of measurement*, this difference between categories in terms of degree can be expressed by a number. A good example is temperature. The difference between 5 and 10 degrees Celsius is equal to the difference between 45 and 50 degrees Celsius. When measured on an interval level of measurement there is no natural *zero point* as is the case with a ratio level of measurement such as weight and length. Zero degrees Celsius is not a natural zero point. The natural zero point of temperature is –273 degrees Celsius, which is equal to zero degrees Kelvin. If you indicate the temperature in Kelvin then it can be considered to be ratio level of measurement as there is a natural zero point.

Interval/ratio levels of measurement

Zero point

This has consequences for the mathematical calculations which are possible. When referring to temperature in Celsius you cannot state that 20 degrees is twice as much as 10 degrees. But if temperature is referred to in Kelvin, you *can* state that 20 degrees is twice as much as 10 degrees, and with a measurement of weight, 20 kilos is twice as much as 10 kilos.

It should be noted that SPSS calls the interval and ratio level of measurement *Scale*. The other two levels of measurement mentioned are *Nominal* and *Ordinal*. Furthermore, we can make a distinction between discrete and continuous variables. *Continuous variables* can be visualised as a line along which the points are continuous: a continuum, therefore. Between two points there will always be an infinite number of other possible values. A person's height, age or IQ are some examples of continuous variables. Variables which only take integers as values are known as *discrete variables*: for example, the number of cars someone owns or the number of children in a family.

Continuous variables

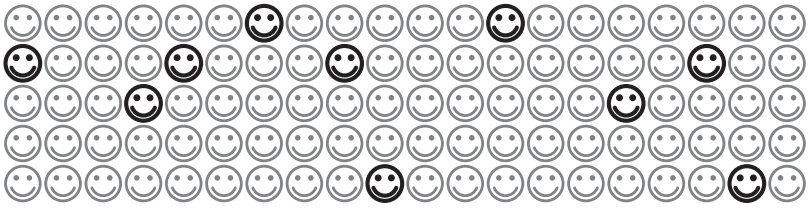
Discrete variables

TABLE 1.1 Overview of levels of measurement, their mathematical possibilities and examples

Level of measurement	Mathematical possibilities	Example
Nominal	Count, percentages (categorization only)	Gender
Ordinal	Count, percentages and ranking (categorization, ranking)	Level of education
Interval	Count, differences expressed numerically, mean, standard deviation (categorization, ranking)	Intelligence
Ratio	Count, differences expressed numerically, mean, standard deviation, calculation of ratios (categorization, ranking)	Age

1.3.3 Does it involve a population or a sample?

FIGURE 1.2 Example of a population and a simple random sample



N=100 population of employees
n=10 random probability sample from the population of employees

Statistics fall into two categories: descriptive statistics and inductive or inferential statistics. *Descriptive statistics* is used when you are studying a population. Having a population implies that every unit of measurement in respect of which statements will be made have been included in the study. For example, if you want to measure work satisfaction, you will interview every employee. In order to reduce the cost you could, however, elect to interview some of the employees only (a sample), drawn randomly from the employees as a whole. When drawing a sample, the objective is to draw conclusions that apply to the population as a whole. In such cases, *inductive/inferential statistics* is your method of choice. The objective is to make claims that apply to the population as a whole on the basis of a single case (a sample).

Descriptive statistics

Inductive/ inferential statistics

Consult a methodology or statistics book for more information. Wikipedia also contains more detailed information about sampling: http://en.wikipedia.org/wiki/Sampling (statistics).

Before you start analysing, ask yourself to which units the claims have to apply. If they have to be applicable to every person or object in your study then a population study is indicated. If the objective is to state claims in regard to persons or objects which are not included in the study but which are represented by the research units that you select, then it will have to be a study based on a sample.

In the next section, we will briefly deal with some key statistical concepts that you will constantly encounter regardless of whether the sample results are based on chance or can be generalized to within a certain margin of error to the population from which the sample was drawn.

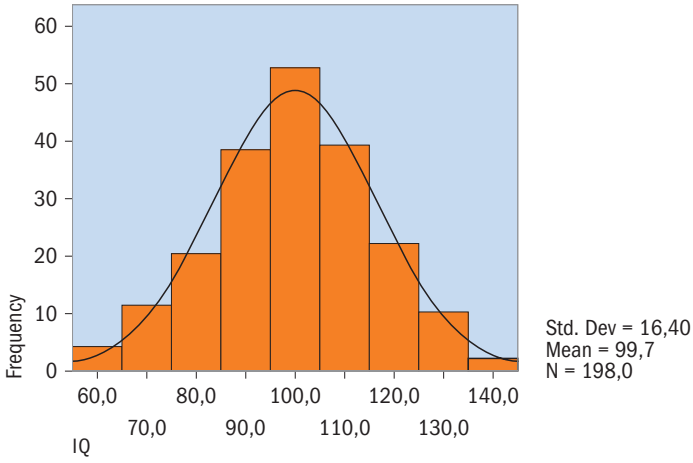
1.4 Some general statistical concepts

As mentioned in the previous sections, there are two types of statistics: descriptive and inductive or inferential. The objective of descriptive statistics is to present your data in an insightful and clear way. If you have determined the work satisfaction of 1000 employees, it does not make any sense to present the data individually. As a rule, the data is combined into a percentage, a mean (Section 4.1) or a histogram (Section 4.1.1). The data is described in a reduced and therefore clear way.

Normal distribution

Data presented in a graphic form is usually in the form of a so-called *normal distribution*.
Figure 1.3 shows a fictitious example of this type.

FIGURE 1.3 IQ levels of 198 employees of the 'Labour' firm



The figure shows the distribution of the scores on an intelligence test which was administered to 198 employees of the 'Labour' firm. This distribution approximates to a normal distribution. For the purpose of comparison, a normal distribution line has been drawn in Figure 1.3. This distribution is named after the mathematician Gauss: the Gauss or Gauss curve. The typical feature of a normal distribution is that it is shaped like a bell curve, with the left and right sides of the curve symmetrical. Using SPSS you can check whether your distribution approximates the normal distribution (Section 4.4).

If the 198 employees of the Labour firm are a simple random sample of the total number of employees (N = 2213) of Labour then inductive statistics needs to be employed. The question is whether the mean IQ of 99.7 is representative of the employee population as a whole. In other words, what are the chances of a mean IQ of 99.7 being obtained if the whole population had been included in the study? It is not likely since the mean in the sample is dependent on a chance composition of the sample. If we were to take a simple random sample time and time again then the mean IQ might well be higher or lower. There will be differences, though these will probably only be slight. You can calculate the standard error in SPSS (Section 4.3).

Standard error

The *standard error* is the extent to which the sample mean is a *reliable* estimator of the population mean. The standard error is determined by the sample size and the homogeneity of the sample. The standard error increases as the sample size becomes smaller and the difference in IQ within the group increases. On the basis of the standard error you can state, for example, that with 95% certainty the population will lie within the intervals of the sample mean minus twice the standard error and the sample mean plus twice the standard error.

Chance is a very important concept in inductive statistics. Even if you are comparing two sample means, it is debatable whether the mean difference also holds true for the population. Suppose the sample taken of the Labour firm consisted of 99 women and 99 men, and the mean IQ of the women is 102.2 and of the men 98.2. Is it correct to conclude that the female employees of the Labour firm are, on average, more intelligent than male employees? Whether the difference is significant can be tested. Chapter 5 will explain how this can be done for differences and Chapter 6 how it can be done in case of relationships.

Significant

When is something *significant*? The rule of thumb is that a difference or relationship in a sample is significant when the probability of finding such a result by coincidence is smaller than 5%, or in case of larger samples (> 1000) is smaller than 1%.

One-tailed or two-tailed testing

The SPSS output will often show whether *one-tailed or two-tailed testing* has been carried out. One-tailed testing is used when you have a hypothesis or expectation. If you have a theory according to which you expect female employees to be more intelligent, *one-tailed testing* can be used. If you do not have a clue as to whether there will be a difference or what direction the difference will take, then *two-tailed testing* should be used. Determining the significance is based on a few features of the sample. These are often the sample size and the homogeneity of the sample. As the sample size increases, the margin of error decreases. As the differences in relation to a

variable within groups (homogeneous groups) decrease, the margin of error in relation to the differences will decrease accordingly.

If a result is significant this does not automatically mean that it is relevant. In the example, the female IQ is higher that the male IQ by 4 points. This difference is indeed significant as the p value is 2%, so smaller than 5%. While the difference in IQ of four points is significant, it is questionable how relevant this is since it tells you very little about the differences between the male and the female employees. Nowadays, research reports often report the *effect size* alongside the significance. The usual statistical measure for an effect size is *Cohen's d*. In this example, Cohen's d is equal to 0.17 which is a negligible effect. Only if d exceeds 0.20 is the effect described as small. Gender does not even explain 1% of the differences in IQ. In Chapter 5 we will discuss how to calculate an effect size.

Effect size

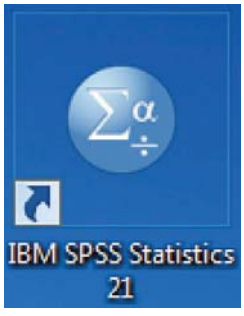
In SPSS output you will often come across the term *degrees of freedom* (df). Degrees of freedom indicate to what extent the scores can vary. If you only know one number out of the two numbers, namely 36, and the mean is 40 then the other number has to be 44. In this case you only have one degree of freedom. If you know one of the numbers then you can know the other. The number of degrees of freedom for many tests (such as a t-test; see Section 5.4.) is equal to the number of elements in the sample minus 1. The number of degrees of freedom for cross-tabulation tables (Section 5.2) is equal to the number of rows minus 1 multiplied by the number of columns minus 1. The number of degrees of freedom for a 2 X 2 cross-tabulation table is 1. If the marginal totals are known and you know the cell frequencies then you can calculate the other frequencies. The number of degrees of freedom is important if you want to estimate the population mean on the basis of a sample. Whether a difference or a relationship in a sample is significant depends on the number of degrees of freedom. With the exception of cross-tabulation tables, the number of degrees of freedom often depends on the sample size.

Degrees of freedom

1.5 SPSS using Windows

As indicated in the introduction, SPSS is a widely used and highly elaborate statistical software package. If you have installed SPSS on your computer properly, the icon as shown in Fig. 1.4 should be shown on your screen.

FIGURE 1.4 The SPSS icon



Open SPSS by double clicking on the icon. If you cannot find the icon, go to 'Start' in the lower left corner and select 'Programs'. Check whether SPSS is in the list of programs. If it is on the list, you can start SPSS by double-clicking on the SPSS icon.

FIGURE 1.5 Opening screen of SPSS

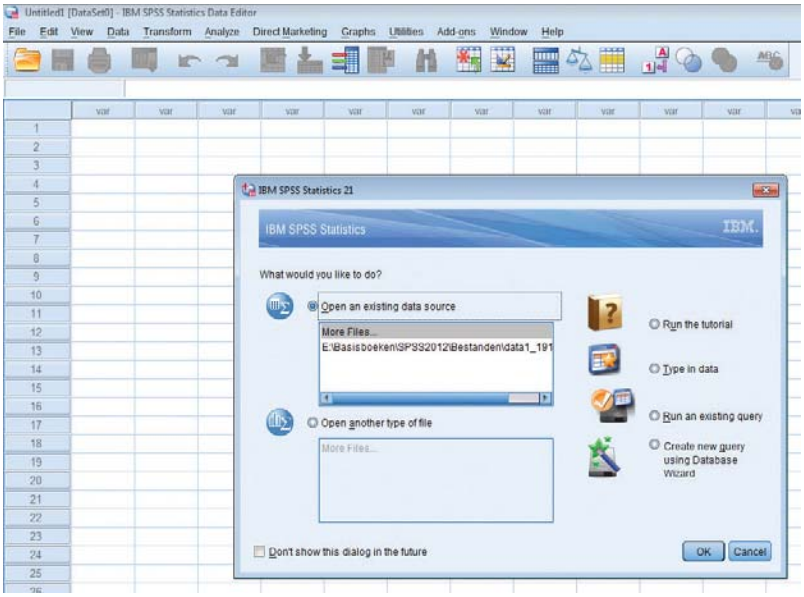
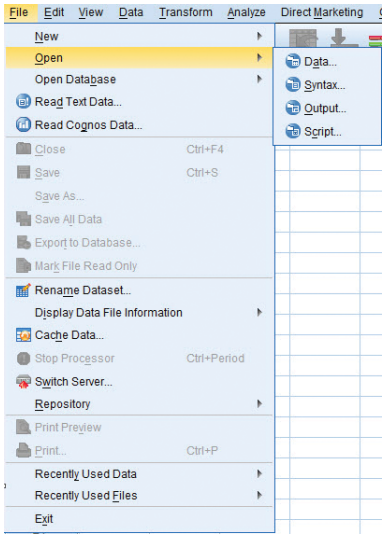


Fig. 1.5 shows the opening screen of SPSS. The opening screen presents several options.

- *Open an existing data source*

If you want to continue using a file you have used previously you will most likely find this under 'Open an existing data source'. Now click on the file in question and then on 'OK'. This box has been ticked as a default setting. If you only click on 'OK' you can indicate where another file you want to open is located, and SPSS will then retrieve it. You can also open files via the more standard approach, by using the file button which you will find in the Data Editor (see Fig. 1.6). You will see that there are several file types to choose from. We will focus on data files. The other types will be discussed later.

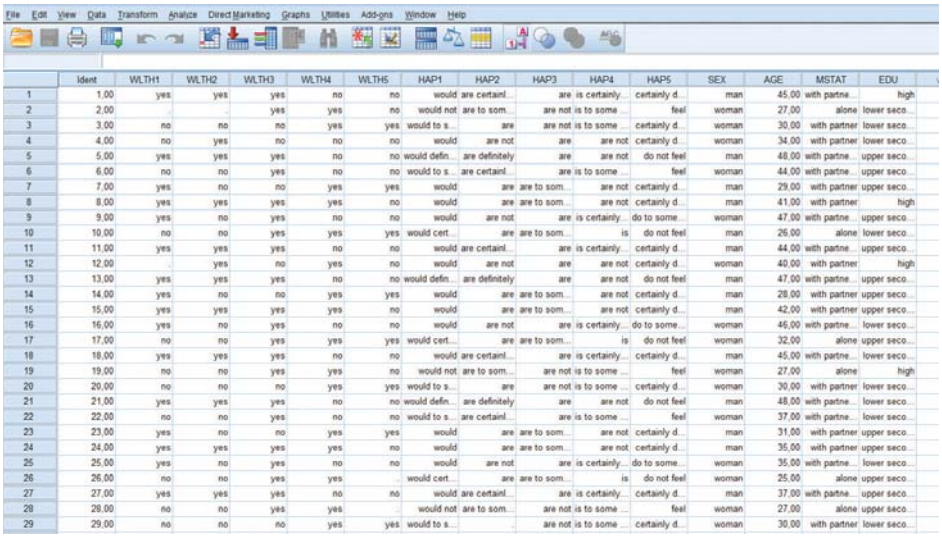
FIGURE 1.6 Opening the file using the file menu



After opening the file, the first thing you will see is the *Data View screen* containing all of the data that you have entered.

Data View screen

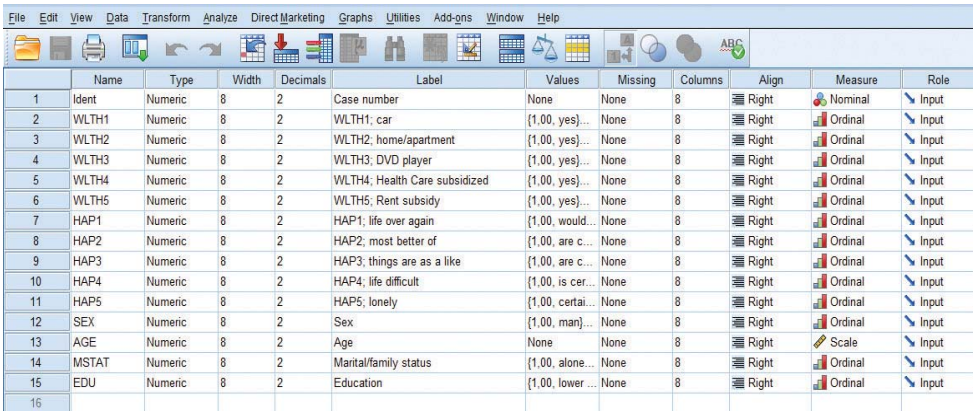
FIGURE 1.7 Data View screen when you open Data1



	Ident	WLTH1	WLTH2	WLTH3	WLTH4	WLTH5	HAP1	HAP2	HAP3	HAP4	HAP5	SEX	AGE	MSTAT	EDU
1	1,00	yes	yes	yes	no	no	would	are certainl...	are	is certainly...	certainly d...	man	45,00	with partne...	high
2	2,00			yes	yes	no	would not	are to som...	are not	is to some ...	feel	woman	27,00	alone	lower seco...
3	3,00	no	no	no	yes	yes	would to s...	are	are not	is to some ...	certainly d...	woman	30,00	with partner	lower seco...
4	4,00	no	yes	no	no	no	would	are not	are	are not	certainly d...	woman	34,00	with partner	lower seco...
5	5,00	yes	yes	yes	no	no	would defin...	are definitely	are	are not	do not feel	man	48,00	with partne...	upper seco...
6	6,00	no	no	yes	no	no	would to s...	are certainl...	are	is to some ...	feel	woman	44,00	with partne...	upper seco...
7	7,00	yes	no	no	yes	yes	would	are	are to som...	are not	certainly d...	man	29,00	with partner	upper seco...
8	8,00	yes	yes	yes	yes	no	would	are	are to som...	are not	certainly d...	man	41,00	with partner	high
9	9,00	yes	no	yes	no	no	would	are not	are	is certainly...	do to some...	woman	47,00	with partne...	upper seco...
10	10,00	no	no	yes	yes	yes	would cert...	are	are to som...	is	do not feel	man	26,00	alone	lower seco...
11	11,00	yes	yes	yes	no	no	would	are certainl...	are	is certainly...	certainly d...	man	44,00	with partne...	upper seco...
12	12,00		yes	no	yes	no	would	are not	are	are not	certainly d...	woman	40,00	with partner	high
13	13,00	yes	yes	yes	no	no	would defin...	are definitely	are	are not	do not feel	man	47,00	with partne...	upper seco...
14	14,00	yes	no	no	yes	yes	would	are	are to som...	are not	certainly d...	man	28,00	with partner	upper seco...
15	15,00	yes	yes	yes	yes	no	would	are	are to som...	are not	certainly d...	man	42,00	with partner	upper seco...
16	16,00	yes	no	yes	no	no	would	are not	are	is certainly...	do to some...	woman	46,00	with partne...	lower seco...
17	17,00	no	no	yes	yes	yes	would cert...	are	are to som...	is	do not feel	woman	32,00	alone	upper seco...
18	18,00	yes	yes	yes	no	no	would	are certainl...	are	is certainly...	certainly d...	man	45,00	with partne...	lower seco...
19	19,00	no	no	yes	yes	no	would not	are to som...	are not	is to some ...	feel	woman	27,00	alone	high
20	20,00	no	no	no	yes	yes	would to s...	are	are not	is to some ...	certainly d...	woman	30,00	with partner	lower seco...
21	21,00	yes	yes	yes	no	no	would defin...	are definitely	are	are not	do not feel	man	48,00	with partne...	lower seco...
22	22,00	no	no	yes	no	no	would to s...	are certainl...	are	is to some ...	feel	woman	37,00	with partne...	upper seco...
23	23,00	yes	no	no	yes	yes	would	are	are to som...	are not	certainly d...	man	31,00	with partner	upper seco...
24	24,00	yes	yes	yes	yes	no	would	are	are to som...	are not	certainly d...	man	35,00	with partner	upper seco...
25	25,00	yes	no	yes	no	no	would	are not	are	is certainly...	do to some...	woman	35,00	with partne...	lower seco...
26	26,00	no	no	yes	yes		would cert...	are	are to som...	is	do not feel	woman	25,00	alone	upper seco...
27	27,00	yes	yes	yes	no	no	would	are certainl...	are	is certainly...	certainly d...	man	37,00	with partne...	upper seco...
28	28,00	no	no	yes	yes		would not	are to som...	are not	is to some ...	feel	woman	27,00	alone	upper seco...
29	29,00	no	no	no	yes	yes	would to s...		are not	is to some ...	certainly d...	woman	30,00	with partner	lower seco...

If you click on 'Variable View' in the lower left corner the variable screen will come up. This screen shows which variables are in the file and what the properties of these variables are. This is shown in Figure 1.8. In Chapter 2 we will look into these properties and how you can change them.
You can now do a number of things. You can process the data. You can perform analyses via 'Analyse' or generate graphs using 'Graphs'. In the next chapters all these possibilities will be explained by way of examples.

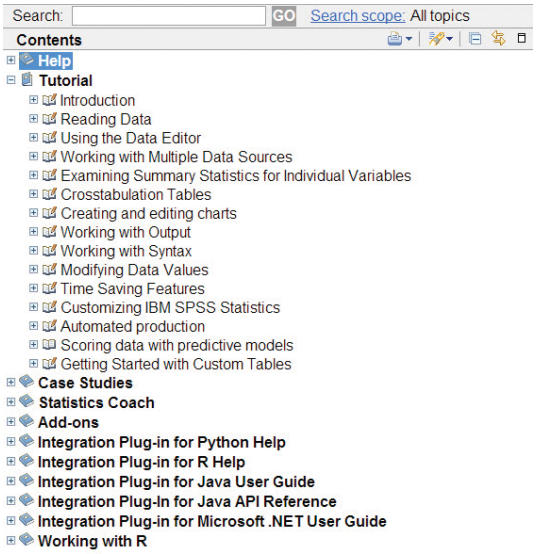
FIGURE 1.8 Variable view when you open Data1



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Ident	Numeric	8	2	Case number	None	None	8	Right	Nominal	Input
2	WLTH1	Numeric	8	2	WLTH1; car	{1,00, yes}...	None	8	Right	Ordinal	Input
3	WLTH2	Numeric	8	2	WLTH2; home/apartment	{1,00, yes}...	None	8	Right	Ordinal	Input
4	WLTH3	Numeric	8	2	WLTH3; DVD player	{1,00, yes}...	None	8	Right	Ordinal	Input
5	WLTH4	Numeric	8	2	WLTH4; Health Care subsidized	{1,00, yes}...	None	8	Right	Ordinal	Input
6	WLTH5	Numeric	8	2	WLTH5; Rent subsidy	{1,00, yes}...	None	8	Right	Ordinal	Input
7	HAP1	Numeric	8	2	HAP1; life over again	{1,00, would...	None	8	Right	Ordinal	Input
8	HAP2	Numeric	8	2	HAP2; most better of	{1,00, are c...	None	8	Right	Ordinal	Input
9	HAP3	Numeric	8	2	HAP3; things are as a like	{1,00, are c...	None	8	Right	Ordinal	Input
10	HAP4	Numeric	8	2	HAP4; life difficult	{1,00, is cer...	None	8	Right	Ordinal	Input
11	HAP5	Numeric	8	2	HAP5; lonely	{1,00, certai...	None	8	Right	Ordinal	Input
12	SEX	Numeric	8	2	Sex	{1,00, man}...	None	8	Right	Ordinal	Input
13	AGE	Numeric	8	2	Age	None	None	8	Right	Scale	Input
14	MSTAT	Numeric	8	2	Marital/family status	{1,00, alone...	None	8	Right	Ordinal	Input
15	EDU	Numeric	8	2	Education	{1,00, lower ...	None	8	Right	Ordinal	Input
16											

- *Run the tutorial* brings up information on topics of all kinds (see Fig. 1.9). If you click on 'All topics' you will get an overview of topics which are available.

FIGURE 1.9 The topics in the Tutorial menu



- *Type in data.* If you have new data you have not entered into SPSS or any other programme such as Excel, this is the best option to choose. SPSS will then open the Data Editor for you (Fig. 1.7). Chapter 2 will explain how to enter your data into the Data Editor.
- *Run an existing query/Generate a new query*. SPSS also allows you to work with data from other statistical software packages such as Excel. In order to do so, you do need to indicate how SPSS should read the data. This is done via a query. You can recognize the query by the extension .spq. If you already have an existing query, click on the third button. If you want to generate a new query, click on the fourth button. SPSS will help you to make a query. It is beyond the scope of this book to go into this in any detail. If you want to know more about how to import Excel data into SPSS, go to the following site: http://www.stattutorials.com/SPSS/TUTORIAL-SPSS-Prepare-Data-Excel.htm

SPSS can be closed by clicking on the white cross in the top right corner. When you do so, SPSS will ask whether to save any changes you have made to the file. If you indicate that the changed file needs to be saved, you will have to give that file a name. If you type in a different name it will be saved alongside the original file. If you do not change the name, the current file will be overwritten. So it is wise to save the file under a new name because if you have made a mistake, you will still be able to access the last version of the file.

Summary

- Section 1.1 has given a rundown of what will be discussed in this book and in the case study 'Money and Happiness' that will be used to illustrate the discussion. (Section 1.2)

- When you are doing research by yourself, ask yourself the following questions before you start:
  - Does the research question focus on frequencies, differences or relationships? (Section 1.3.1)
  - What is the level of measurement of my data: nominal, ordinal, interval or ratio? (Section 1.3.2)
  - Does it involve a population or will a sample be drawn? If it is a sample, inductive or inferential statistics will need to be used. (Section 1.3.3)

- Section 1.4 discussed the following statistical concepts: normal distribution, standard error, reliability, confidence, significance, one-tailed and two-tailed testing, relevance, effect size and degrees of freedom.

- Section 1.5 explained how to start SPSS, its various screens ('Data View' and 'Variable View') as well as where you can obtain help within SPSS (Tutorials).