ENGLISH EDITION

Baarda, De Goede en Van Dijkum

# Introduction to Statistics with SPSS

## A guide to the processing, analysing and reporting of (research) data

Noordhoff Uitgevers

Translation of the second, fully revised edition

## 5
**Analyzing data when frequency research questions are involved**

I have a frequency research question

NO · YES

Variable is nominal

NO · YES

5.1.1 *descriptive:* counting, frequency table, percentage, mode
5.1.2 *graphic:* pie chart, bar chart

Variable is ordinal

NO · YES

5.1.1 *descriptive*: counting, table, percentage, mode, rank order, median
5.1.2 *graphic*: histogram

Variable is interval/ratio

5.1.1 *descriptive*: stem and leaf, frequency table/classes, percentage, mode, median, mean, standard deviation/ variance
5.1.2 *graphic*: histogram and standard deviation

I have sub-groups

NO · YES

**5.2** *descriptive*: using all of the above-mentioned techniques for each sub-group
**5.3** *graphic*: clustered bar chart, box plot

## 6
**Analyzing data when comparative research questions are involved**

I have a comparative research question

NO · YES

Grouping variable is nominal; test variable is nominal

NO · YES

**6.1** Contingency table
– *Population*: **Go to 6.2.1**
– *Sample*: Chi-square (6.1)

Grouping variable is nominal; test variable is ordinal

NO · YES

**6.2** Comparing mean rank orders:
– *Population*: comparing mean rank orders (6.2.1)
– *Dependent sample*:
  - *two*: Wilcoxon (6.2.3)
  - *more*: Friedman (6.2.4)
– *Independent sample*:
  - *two*: Mann-Whitney (6.2.1)
  - *more*: Kruskal Wallis (6.2.2)

Grouping variable is nominal; test variable is interval/ratio

**6.3** Comparing means:
– *Population*: comparing mean rank orders (6.3.1)
– *Dependent sample*:
  - *two*: t-test (6.3.2)
  - *more*: MANOVA (not in this book)
– *Independent sample*:
  - *two*: t-test (6.3.1)
  - *more*: one way ANOVA (6.3.3)

## 7
**Analyzing data when correlation research questions are involved**

I have a correlation research question

NO · YES

Variables are nominal (based on the variables with the lowest level of measurement)

NO · YES

**7.1** Cramér's V

Determining significance when a sample is involved

Variables are ordinal (based on the variables with the lowest level of measurement)

NO · YES

**7.2** Spearman

Determining significance when a sample is involved (7.2)

Variables are interval/ratio (based on the variables with the lowest level of measurement)

*Two variables*
– *discipltive*:
  - Pearson (7.3.2)
    Determining significance when a sample is involved vaststellen (7.3.2)
  - imple regression (7.3.3)
– *graphic* (7.3.1):
  - scattergram
  - regression line

*More then two variables*
– *predictive*:
  - multiple correlation/ regression (7.3.4)
– *controlling for*:
  - partial correlation (7.3.4)

**Making a report**

Introduction to Statistics with SPSS

Dr. D. B. Baarda, Dr. M. P. M. de Goede
en Dr. C. J. van Dijkum

# Introduction to
# **Statistics with SPSS**

*A guide to the processing, analysing and reporting of (research) data*

Second, completely revised edition

Noordhoff Uitgevers bv Groningen/Houten

*Cover design:* Total Identity, Amsterdam
*Text Layout:* Ebel Kuipers, Sappemeer

Any comments and remarks about these or other publications can be sent to: Noordhoff BV, Higher Education Department, Antwoordnummer 13, 9700 VB Groningen, e-mail: info@noordhoff.nl

2 3 4 5 / 13 12 11



ISBN (ebook) 90 018 49887
ISBN 90 207 32978
NUR 916

# Preface

The extent to which a decision is accepted does not merely depend on the position in an organisation that the decision maker occupies, nor on a person's years of experience in the field, the degree of expertise in the area, powers of intuition or the fact that an individual may have a "good nose" for business. Such qualities are no longer sufficient, even if decisions taken on such bases are nevertheless good ones. After all, to be accepted, decisions not only have to be right, they must also be justifiable. Subjectivity is only one of the factors involved.

Sound decisions stem from conclusions based on outcomes derived from research. This involves the collection and analysis of data, as well as the reporting and dissemination of information. The ultimate foundation is provided by the data. Figures and statistics are objective, verifiable and, therefore, convincing.
Such views are currently becoming more widespread. Just open any given business magazine and you will find articles on this subject, all of them concerning the process of taking sound decisions. In this context, frequent use is made of such expressions as "Data Mining", "Business Intelligence" and "Statistical Analysis". What these notions have in common is the fact that they all involve data processing: the search for trends, problems, solutions or new (market) opportunities. Furthermore, they all promote the value of considering both subjective and objective factors whenever taking decisions.

With some pride, we can say that SPSS is by far the most frequently used statistical-analysis program. The *Introduction to Statistics with SPSS* constitutes the logical successor to *Statistiek met SPSS voor Windows*, a Dutch publication that appeared in 1999. As you will discover, SPSS has taken a further step forward in the areas involving ease of use and professionalism with the development of SPSS 12.0. SPSS can now be used to discover relationships and patterns, to detect causes, and to make well-founded decisions.

Good luck!
G.J. Hulzebos RIS (Registered Information Scientist)
*Country Manager Products*
SPSS Benelux B.V.

# Foreword

## What would we like to achieve in this book?

Once you have studied this book, you should:

1 *Have mastered the basic skills enabling you to work with SPSS on your own.*

This book provides practical instructions for processing and analysing research data using SPSS for Windows. SPSS stands for 'Statistical Products and Service Solutions', currently one of the most widely-used software packages for entering and statistically analysing data. The basic SPSS package contains programs for such commonly-used statistical analysis techniques as those involved in calculating a frequency distribution, cross tabulation or correlation coefficient.

In this *Introduction to Statistics with SPSS* for Windows, we hope that you will learn, through experience, to trust the software package. After a general introduction and familiarisation with the basic principles of SPSS, you will immediately be able to enter data into the computer and to analyze it. For this reason, we do not discuss the many extra features of SPSS in this book. The purpose of this book is to give you the basic skills so that you can work independently with SPSS. You will then inevitably learn how you can enhance these skills.

We invite the reader to begin by experimenting with the sample database concerning the relationship between wealth and happiness that we provide. This database contains survey data from 500 men and 500 women. We use it to investigate the usual practices involved in a research project.

- We begin with the entering of data (Chapter 2);
- we then explain how SPSS can be used to modify and to combine the data (Chapter 3);
- and in Chapter 4, we discuss how to establish the homogeneity of composite scores.

After this preparatory work, it is time for analysis. The methods used will depend on several factors including the specific nature of the inquiry's general research question. A well-designed general research question always involves one or more specific issues to which answers are sought by means of research. These specific research questions can further be distinguished into:

- frequency research questions, which ask how often or to what extent something occurs; techniques for describing such variables are investigated in Chapter 5;
- comparative research questions, which establish and test the similarities and differences between two or more groups concerning a given characteristic; they will be discussed in Chapter 6;
- correlation research questions, which establish and test the relationships between two characteristics; these will be discussed in Chapter 7.

2 *Know when to use which uni- or bivariate statistical method.*
Proper use of SPSS not only requires you to have some knowledge of SPSS, but also some expertise in the field of statistics. In the first chapter the basic concepts will be introduced, such as level of measurement, normal distribution, probability, significance, one and two-tailed tests, power and effect size. Chapters 5, 6 and 7 will reveal that the choice of statistical technique is not only dependent on the nature of the research question, but also on the level at which the relevant characteristic is measured and the issue of whether a population or a sample is involved.
Since this book only presents the statistical information that is strictly necessary to analyze data with SPSS using predominantly uni- and bivariate analytical methods, we will only deal with a few multivariate techniques of analysis.

3 *Know the conditions and SPSS procedure involved in a given technique.*
Using both explanation and example, we detail the essential factors in each technique without going too deeply into the underlying mathematics. We additionally indicate when a technique should and should not be used. Finally, there is an explanation of how to perform the technique using SPSS.

4 *Be able to read and to interpret SPSS output, and to issue reports based on it.*
The material will be further presented in terms of the following questions which you, the user, might pose:

- If the chosen method of statistical analysis has been successfully performed using SPSS, how do I read the computer output? What is precisely revealed by the results of the data analysis?
- How should I interpret the output? What does the output mean for my research question?
- How should I present my output in my research report?
- How do I write up my conclusions?

We will provide examples of SPSS output for each presented technique. On the basis of these examples, we will explain how the data should be read and what they mean. To illustrate this latter point, we additionally demonstrate how the output can be reported.

### What are the instructional principles underlying this book?

A lot of consideration has been given to this book's instructional methodology.

Each chapter starts with an introduction that, in a general sense, provides a glimpse of that chapter's contents.

Each chapter ends with a list of key words in the order that they appear in the text. Additionally, the designated SPSS procedures are listed on a section-by-section basis, so that anyone who has an excellent or even reasonable command of SPSS for Windows need not work through the entire chapter again.

The main instructional premise is that you learn by doing. It is also important that you know what you are doing. For this reason, we also briefly explain the essential points of the technique to be used without going into detail. To gain insight into the research material, we advise you first to analyze the data graphically. First construct something like a scatter diagram to see what type of relationship is possibly involved. Depending on the research question, we discuss a few suitable graphic techniques that are available in SPSS.

The learning process is illustrated on the basis of a sample provided for practice entitled "Wealth and Happiness" or "Can money buy happiness?". Nearly all the problems encountered during the processing and analysis of data are explained by means of this practice example.

### How can this book be used?

When you have conducted research and obtained a set of numerical values, you can use this book as a guide to the analysis of the data. It provides both information about statistics and instructions for using SPSS to analyze and interpret the figures. In fact, it is a book on statistics combined with an SPSS user manual. In principle, you do not need a separate statistics textbook and need not have any previous knowledge of the subject.

### What do you need?

In writing this book, we have made use of SPSS for Windows, Version 12.0. This guide can, however, also be used with earlier versions of SPSS for Windows.

After you have worked through this book, you will be capable of quickly entering simple research data in a computer and adequately analysing them. You will likely also be able to perform statistical techniques that are not presented in this book but with which you are already familiar. The menus provide a great deal of information, along with relevant examples.

### Where do I find the sample data files used in this book?

Statistical terms and methods are systematically presented and explained on the basis of concrete examples always related to an explicitly-stated, specific research question.
The sample research questions are developed in a step-by-step manner and illustrated in the boxes intended for that purpose. The data files on which we conduct the analyses can be found on the internet at http://www.basisboekstatistiekmetspss.noordhoff.nl.

Spring 2004

Ben Baarda
Martijn de Goede
Cor van Dijkum

# Required hard- and software

This book explains how research data can be entered, prepared, analyzed and interpreted using SPSS for Windows and how a report can be written about the output.
You must be able to work with Windows (version 98, 2000, ME or XP) and must, of course, have installed SPSS. In this book, we make use of SPSS, Version 12.0.

# Symbols and data files used

*The following symbols are used in this book:*

| | |
|---|---|
| > | Move the mouse pointer to the position indicated on the screen. |
| ▶ | Click the ▶ button in SPSS. |
| ▼ | Single-click the left mouse button. |
| ◎ | Double-click the left mouse button. |

*The following data files are used in this book:*

**data1** raw, unprocessed data, entered according to the scheme of codes presented in Figure 2.1.

**data2** recoded items rwlth1, rwlth2, rwlth3, rhap3, rhap4 and rhap5 have been entered here (see Section 3.1).

**data3** the variables *ac* (age class; Section 3.3), *Twlth* and *Thap* (respective totals of the 5 wealth and 5 happiness items; see Section 3.4) have been entered here.

**data4** the variable *mwlth* (the means of the 5 wealth items; see Section 3.5) and the items rwlth1_1 to rwlth5_1 (for the wealth items, the missing values are replaced by a series average; see Section 3.5) are entered here.

**data5** the variable *Twlth* and *Thap* are here the respective sums of rwlth2, wlth4 and wlth5 and hap1, hap3 and hap4 (see Section 4.2).

**family** this is a data file containing the happiness scores of twenty men and their wives, along with the happiness scores from each of their oldest children.

The data files are stored on the website http://www.basisboekstatistiekmetspss.wolters.nl. It would be most helpful if these files were copied to a special 'data' folder/directory on the c-drive (hard drive) of the computer on which you are working.

# Contents

# 1

# How do I prepare myself to work with this *Introduction*?

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

John Lennon & Paul McCartney

*In conducting research, the choice of the appropriate statistical technique to analyze the collected data is an important link in the long chain of decisions that must be taken. Ultimately, the goal of collecting and analysing data is to answer the research question or questions contained in the inquiry.*
*To draw attention to the place of data analysis in the research cycle as a whole, the typical stages of research are listed below. Each stage in the research cycle is stated in the form of a question:*

1 *What is the goal of my research and what is the type of the inquiry?*
2 *How will I acquire information (among other ways, by reviewing the literature)?*
3 *What type of research am I going to conduct?*
4 *What will my research design look like?*
5 *Will the entire population be involved in my research or will I select a sample?*
6 *Which data collection method am I going to use?*
7 *How will I prepare my data for analysis?*
8 *How will I analyze my data?*
9 *How will I report and evaluate my research?*

*This book focuses on stages 7 and 8, along with a portion of 9: preparation, analysis and description of the research data analyzed using SPSS. Referring mostly to a study on the relationship between wealth and happiness (see Figure 1.1), we will investigate step-by-step:*

- *How you should prepare research data for input (Section 1.1 and Chapter 2).*
- *How you can use SPSS to adjust and modify your data. Before you begin the analysis, you often have first to convert (recode) the values of certain data, or combine the values of data to produce new scores (Section 1.1 and Chapter 3), and subsequently check their reliability (Section 1.1 and Chapter 4).*

**Prior knowledge**

- No specific prior knowledge is required.
- Before explaining the uses of SPSS, we will first introduce the practice example: 'Wealth and happiness' (see Figure 1.1).

- *How you choose the appropriate method of analysis. To choose the appropriate method of analysis, you first have to establish the nature of the research question (Section 1.2.1). Does it involve frequencies (Chapter 5), comparisons (Chapter 6) or correlations (Chapter 7). You then must determine the level of measurement for your data (Section 1.2.2). Finally, you have to decide if a sample or a population is involved in your research. You can then use the block diagram* 'How do I analyze my data?' *to ascertain which method of analysis best suits your research question. You will find the diagram on the inside front cover.*
- *How you use SPSS to conduct the analysis and how you should interpret the results. This we indicate for each statistical technique to be discussed. Additionally, we then examine how you can make a report about the results.*

*This chapter fulther deals with a number of important statistical terms, such as normal distribution, significance, probability and standard error (Section 1.3). The final section of the chapter will explain how you can start using SPSS.*

Befor explaining the uses od SPSS, we will first introduce the practice example: 'Wealth and happiness' (see Figure 1.1).
The questionnaire in Figure 1.1 was presented to a representative sample of 500 men and 500 women between the ages of 25 and 55 years old. These age limits were deliberately chosen. Many younger people are still studying and do not, therefore, have any fixed income. For people above the age of 55, there is often the issue of their (partial) withdrawal from the work force, as a result of which they enter another financial category. The data from this research is found on the website http://basisboekstatistiekmetspss.wolters.nl in the file entitled 'data1'.

**Can money buy happiness?**

A researcher would like to know if there is a correlation between wealth and happiness. He asks himself if money can buy happiness. His general research question can then be stated: 'Is there a positive correlation between the amount of wealth at an individual's disposal and the degree of happiness that the individual enjoys?' The concept "wealth" is defined by the researcher as follows: wealth is the quantity of financial means at an individual's disposal. This definition is deliberately broad. In this way, the research not only covers income and assets, but also other financial resources that individuals might have available to them. Happiness is defined as follows: happiness is the degree to which an individual is content about the life that he/she leads. The researcher has operationalized these two concepts in the form of a questionnaire. To measure both the concept of wealth and the concept of happiness, the researcher has devised five statements or items for each.

*Wealth is measured by asking respsondents to agree or disagree with the following statements:*

| | | |
|---|---|---|
| 1 | I own a car | yes/no |
| 2 | I own a home/apartment | yes/no |
| 3 | I own a DVD player | yes/no |
| 4 | I am covered by the Dutch national health plan | yes/no |
| 5 | I receive a rent subsidy | yes/no |

*Happiness is measured by asking repsondents to complete the following multiple choice items:*

1 If I could live my life over again, I would…in the same way

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| certainly not live it | not live it | to some extent (not) live it | live it | definitely live it |

2 Most other people are ... than I am.

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| certainly not better off | not better off | to some extent (not) better off | better off | definitely better off |

3 Things ... as I would like them to be.

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| are certainly not | are not | are to some extent (not) | are | are exactly |

4 Life is …..

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| certainly not difficult | not difficult | to some extent (not) difficult | difficult | extremely difficult |

5 I … lonely.

| ☐ | ☐ | ☐ | ☐ | ☐ |
|---|---|---|---|---|
| certainly do not feel | do not feel | do, to some extent (not), feel | feel | definitely feel |

Since happiness is not only dependent on financial means but also on other factors, the researcher also asks about a number of easily-measured characteristics, such as the sex, age, educational level and marital/family status of the persons being surveyed.

*Sex* ☐ male ☐ female
*Age (in years)* …
*Marital/family statu* ☐ alleen
☐ met partner
☐ met partner en kinderen

*Highest completed educational level*
☐ Lower secondary education
☐ Upper secondary education
☐ Higher education

Figure 1.1 **Model research project 'Wealth and happiness'**

**research questions**

The researcher refined his general research question into a number of specific research questions:

1 How many people own cars, homes, or DVD players?
2 How many people are covered by the national health plan?
3 How many people receive a rent subsidy?
4 How happy are people about their lives?
5 How lonely do people feel?
6 Are there differences between men and women concerning:
   - ownership of a car, home or DVD player
   - national health plan coverage
   - reception of rent subsidies
7 Is there any difference in people's happiness about life when they are living:
   - with or without a partner
   - with or without children

8 Are there any differences in the loneliness felt by men and women?
9 Do men and women experience different degrees of happiness about the life that they are living?
10 Is there a correlation between contentment about life and age?
11 Is there any correlation between wealth and happiness?

## 1.1 How do I prepare the data for analysis?

Before you enter the data, you first have to make, for your own use, a summary of the variables involved in the research and the values that these could have. For example, the variable *sex* appears in our research project and has the possible values 'male' or 'female'. It is useful to convert these values to numerical form, for example '1' for male and '2' for female. You must, of course, appropriately note which values stand for what. You do this by means of a coding scheme or codebook. In Section 2.1, we explain how you can construct such a scheme.

coding scheme

When you have collected the research data and composed a coding scheme, you can enter the data in the computer. You do so by typing your data into the data editor, which appears on the screen when you start SPSS. It is additionally helpful if you immediately assign a name to the variables for which you are entering the scores. Before you begin the analysis of your data, you must check if values or codes are included that do not belong there. For sex, use is most frequently made of the codes '1 = male' and '2 = female'. The entries under sex in your database must, therefore, only be ones and twos. If, for example, a seven appears, you have probably made a mistake when entering the data. You need to correct it before continuing with your analysis.

data editor

Similarly, you have to check if the variables display sufficient variation. If there are very few men in your sample, it will be difficult for you to determine if a difference exists between men and women in, for example, the degree of their happiness. We investigate this further in Section 2.4 ("How do I check if I have made a mistake when entering my data?") and again in Chapter 5, when we deal with the calculation of frequencies.

variation

Sometimes, you have to process the data that you have entered. We are, for example, not only interested in a person's desire to live life over again, or the fact that someone may find life difficult, but above all the extent to which an individual agrees with such 'expressions of (un)happiness' . By adding up the scores from the separate items in the questionnaire, we can arrive at a total score for happiness. The sum, which is accomplished with the help of some SPSS commands, constitutes a new variable. You should have noticed that three of the five statements of (un)happiness have a negative formulation (for example: 'Life is difficult'). It would be helpful if all the scores for the happiness items recode command

pointed in the same direction. In that way, a high score is always an indication of a high degree of happiness. Using a recode command, you can, once the data has been entered, adjust the values of all the variables by means of a single action, so that '5' becomes '1', etc. In Section 3.1, we explain how you can do that.

operationalisation

Sometimes you need to use more than one indicator to measure a concept. We have used five indicators to measure the concept 'wealth'. This translation of an abstract concept, such as wealth, into measurable, concrete characteristics, such as car ownership, is called operationalisation. If you use more than one indicator to measure a concept, such as 'wealth' or 'happiness' and then wish to combine these into one total score, you must invoke some criteria to check if this procedure is allowed. In short, the various operationalisations must all measure the same thing and therefore be homogeneous. With the aid of the homogeneity coefficient alpha (also known as the alpha reliability index), you can see if and, if so, the extent to which they all, in fact, measure the same characteristic. Using an item analysis, you can see if there are questions or items that have a negative effect on the homogeneity. If such is the case, you can decide not to include responses to such questionnaire elements in the total score for a concept. In Chapter 3 ('How do I modify and combine data'), we examine, among other things, how you derive a total score from the scores of the separate items. In Chapter 4, we use an example to provide a detailed answer to the question: How do I check the homogeneity of the collected scores?

item analysis

homogeneity

## 1.2 How do I analyze my data? A data-use key!

The answers to the following questions are important for the choice of method of statistical analysis.

1 Does the general research question involve issues of frequency (how often/to what extent), comparison or correlation? Or does it involve a combination of the above?
2 What is the level of measurement (nominal, ordinal or interval/ratio level) for the data that you have collected?
3 Is a sample or a population involved?

The block diagram 'How do I analyze my data' (to be found on the inside front cover of the book and as a separate insert) has been developed on the basis of these questions. The following sections explore the issue in greater detail.

### 1.2.1 What specific research questions are involved in my general research question?

The nature of the general research question will provide the basis for answering the question about which statistical technique to use. The general research question always contains one or more

specific research questions to which an answer must be given (Introduction).
In general, there are three types of specific research questions that can be distinguished in a research inquiry:
1 Questions concerning how often or to what extent something occurs (*frequency*). An example of this type is: ‘To what extent are people in the Netherlands happy?’ or ‘What percentage of people possess cars?’
2 Questions concerning *a comparison*. Example: ‘Are men happier than women?’
3 Questions concerning *a correlation*. Example: ‘Is there any correlation between wealth and happiness?’

It is clear that research questions 1 to 5 (Introduction) can be characterized as frequency research questions. Research Question 1 is, for example, concerned with the number of people that own such consumer products as DVD players. Chapter 5 presents an example of the analysis used for data involving this type of query. Research questions 6 to 9 are comparative research questions. The analysis of data involving this type of inquiry is dealt with in Chapter 6. Questions 10 and 11 are correlation research questions. In Chapter 7, we examine examples of data analysis for research involving this type of question.

### 1.2.2 At what level can I measure my data?

Once you have determined the type of research question(s) involved in your research (see the first column in the block diagram ‘How do I analyze my data’), you must then decide at which level of measurement the variable(s) has been measured. To do this, see the row below frequency, comparison or correlation in the above-mentioned block diagram.

level of measurement

For each research question, it is necessary to indicate which level of measurement is being used for the variable sex involved. In Section 1.1, the level of measurement for the variable sex in Research Question 9 (on the differences in the extent to which men and women feel themselves to be happy) is different from (and lower than) the level of measurement for the variables in Research Question 10, for example, on the correlation between age and happiness.

For the sex variable, there are only two categories, namely ‘male’ or ‘female’. There is only a straightforward difference involved, no degree of more or less. A man is different from a woman but not more or less than her. The same holds true for marital status; people are married, living together or single. This type of possible answer concerns a nominal level of measurement. You can indicate how many men or women own a car, but not that someone is more ‘male’ or more ‘female’.

nominal level of measurement

Such gradations can indeed be made for data measured on the ordinal, interval or ratio levels.

ordinal level of measurement

Data on the ordinal level of measurement can definitely be described in terms of more or less, but this difference between categories cannot be expressed as a numerical value. For example, level of education is clearly an item describable as a comparison, some having more and others less. The upper high school level is higher than the lower one, but it is not possible to indicate how much higher.

interval and ratio levels of measurement

ratio level

interval level

At the interval and ratio levels of measurement, the difference between the categories conceivable as more or less is also expressible as a number. Temperature provides a good example of this type. The difference between 5 and 10 degrees Celsius is just as large as the difference between 45 and 50 degrees. In contrast to the interval level, the ratio level involves a natural zero-point, such as illustrated by weight or height. The interval level involves comparable intervals, but no natural zero-point. This distinction has consequences for the arithmetic calculation that may be used. With temperature, it is not possible to say that 20 degrees is twice as must as 10 degrees. With weight, it is conversely possible to say that 20 kilos is twice as heavy as 10 kilos. In the case of temperature, 0 degrees does not, after all, constitute the natural zero-point and, consequently, an interval level of measurement is involved. Despite this distinction, SPSS groups the interval and ratio levels of measurement under the heading of Scale. The two other levels of measurement are named Nominal and Ordinal.

Scale
Nominal
Ordinal

continuous variables

discrete variables

We additionally distinguish between continuous and discrete variables. Continuous variables allow you to present a line on which values form a row of connected points: a continuum. No matter how close to each other, any two given points are always separated by still (infinitely more) other possible values. Examples of continuous variables are a person's height, age and intelligence. Variables that can only have whole values are called discrete variables, such as the number of cars that someone can own or the number of children in a family.

### 1.2.3 Is a population or a sample involved?

descriptive statistics
population

There are two types of statistics: descriptive and inferential statistics. Descriptive statistics are used when research is conducted on a population. It is possible to speak of a population when all the units about which you wish to make statements are involved in your research. That is, when you conduct, for example, a survey of all the employees in a company in order to determine their level of job satisfaction.

sample

inductive or inferential statistics

To save costs, you could also solicit responses from just a portion (sample) of the employees that you select at random from the total employee pool. Of course, you still would like to make statements about the entire employee population. In such a case, use must be made of inductive or inferential statistics to allow you, on the basis of a particular case (a sample), to make general statements (about the population).

units

population study

sample study

Before you can begin the analysis of your data, you must therefore pose the question concerning which units (who or what) are the subject of your statements. When these are only the persons or items involved in your research, it is then possible to speak of a population study. If you would also like to make statements about persons or items not involved in the research but, so to speak, represented by the research units that you have selected, you are then undertaking a sample study. In Section 1.4, we briefly deal with a few statistical terms that are continually encountered when testing if the results in a sample are due to chance or if, within a certain margin of uncertainty, it is possible to make generalisations concerning the population from which the sample was drawn.

## 1.3 A few general statistical terms

The purpose of descriptive statistics is to present data in a clear and well- organized manner. If you have determined the job satisfaction of nearly a thousand employees, it makes little sense to present all this data. It is mostly, for example, compiled into a histogram (Section 5.1.2), or formulated as percentages or an average

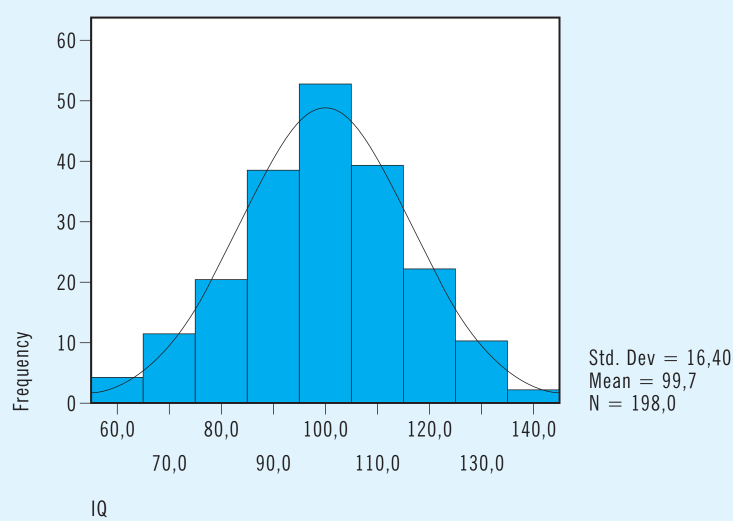



Figure 1.2 **The distribution of the intelligence level**

(Section 5.1.1). You describe your data in a reduced and therefore orderly form. If the data are reported in a graphic manner, the result is quite often the so-called 'normal distribution'. Figure 1.2 provides a (fictive) example of this.

It is the distribution of scores from an intelligence test administered to 198 employees of the 'Labour' company. This distribution closely approximates the form of a normal distribution. By way of comparison, the normal distribution is sketched. Such a distribution is also named the Gauss distribution or the Gauss curve. The characteristic feature of the normal distribution is its bell form. With SPSS, you can verify if the distribution of your data approaches the normal distribution (Section 5.1.2).

normal distribution

Gauss curve

Consider the possibility that 198 employees constitute a random sample from the total employee pool of the 'Labour' company (N = 2213). In such a case, use must be made of inductive or inferential statistics. The question then arises to what extent the mean IQ, found to be 99.7, is representative of the total employee population. In other words, what is the probability of the population having a mean IQ of 99.7 if you could include all of it in the research? This probability is, of course, not so great. For the mean that has been determined for the sample is somewhat dependent on the coincidental composition of the sample group. If we were to select another random sample and still another, etc., it is quite probable that the mean IQ could possibly be somewhat higher or lower. The values will likely deviate a little from each other, but probably not very much. SPSS can be used to calculate this so-called standard error (Section 5.1). The standard error indicates the extent to which the computed sample mean is a good estimate of the population mean. The larger the difference of IQ within the group and the smaller the sample size, the greater is the standard error. The standard error is therefore determined by sample size and sample homogeneity. On the basis of the standard error, you can, for example, state with at least a 95% certainty that the population mean lies between the sample mean minus twice the standard error on the one hand and the same sample mean plus twice the standard error on the other.

inductive or inferential statistics

standard error

sample size

homogeneity

The notion of 'certainty' or probability plays an important role in inductive statistics. Even if you compare the means of two samples, there is some question about the probability that any detected difference in means is representative of the population.

probability

Consider the possibility that the sample of employees from the 'Labour' company consists of 99 women and 99 men. You find that the mean IQ of the women is 101.2 and that of the men is 98.2. Can you then state that the female employees of the 'Labour' company are on average more intelligent than the male employees? You can test if this difference is 'significant'. In Chapter 6, we explain how to do that when comparisons are involved and, in Chapter 7, how you do that when your research involves a corresignificance

one or two- tailed test

one-tailed test

lation. When is it then possible to speak of significance? The general rule is that a finding is significant if the probability of error (p value) is less than 5% or, with larger samples (> 1000) less than 1%. Often, SPSS also indicates if a one or two- tailed test was involved. You conduct a one-tailed test when you have formulated a hypothesis or expectation. If you have a theory on the basis of which you might expect that female employees are more intelligent than their male counterparts, you can then perform a one-tailed test. If you do not, however, have any idea about the possibility of a difference and certainly not about the direction of the difference, your test is then two-tailed.

The determination of significance is based on a few characteristics of the sample, in particular its size and homogeneity. The larger the sample, the smaller is the probability that any finding may be due to chance. The smaller the differences (i.e. range) of a given variable within a group (homogeneous groups), the smaller is the probability that the differences between groups involves a coincidence.

degrees of freedom

df

SPSS output also frequently contains the term degrees of freedom (df). The number of degrees of freedom indicate the extent to which scores can vary. If you know only one of two numbers (i.e. 36) and you know that the mean is 40, the other number must then be 44. This represents 1 degree of freedom. If you in fact know one figure, then you also know the other. For many tests, such as the t test (Section 6.3.1), the number of degrees of freedom is equal to the number of elements in the sample minus 1. For a contingency table (see Section 6.1) the number of degrees of freedom is equal to the number of rows minus 1, multiplied by the number of columns minus 1. For a 2x2 table, the number of degrees of freedom therefore equals 1. If the marginals (row and column totals) of a 2x2 contingency table are known as well as one of the cell frequencies, you can then calculate the other cell frequencies. Degrees of freedom are important when you wish to use a sample in order to estimate a population mean, for example. The significance of an observed difference or correlation in a sample depends on the number of degrees of freedom, which are in turn, except in the case of a contingency table, often dependent on sample size.

When you have established that there is, for example, a significant difference in mean IQ between men and women, you would often also like to know the extent to which the difference can be explained/predicted. To what degree can, for example, difference in mean IQ be attributed to sex? For this purpose there exists, depending on the manner in which the difference is tested, all sorts of measures that are often, however, not included in SPSS. We will explain how you can determine the effect size for each testing method that we discuss.

## 1.4 How does SPSS for Windows work?

Once SPSS has been properly installed on your computer, you should see the following icon on your screen:

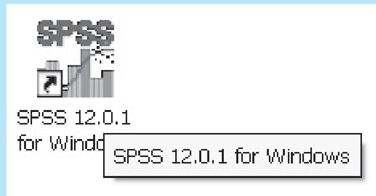



Figure 1.3 **The SPSS-icon**

You open SPSS by double clicking this icon (◎). If you are unable to find the icon, you can always click the start button in the bottom left corner of your screen. You then use the mouse indicator to select “Programs”. You can then see if SPSS appears on the program list displayed on screen. If such is the case, you can start SPSS by double clicking its entry in this list.

> SPSS icon ◎

In SPSS Version 12.0, you then see the welcome screen as shown in Figure 1.4.

**data editor**

This welcome screen offers you various options. If you wish, for example, to enter *new data*, you have to click on “Type in data” and then “OK”. An empty data matrix then appears. In earlier SPSS versions, the first screen that you saw was the empty data matrix. In Chapter 2 it is explained how you can enter data in the now open data editor.

You can also open a *previously-used, existing data file*. If it is a file that you have previously used, you will probably find it under the heading “Open an existing data source”. You then click on the appropriate file followed by “OK”, and the completed data matrix is displayed on screen.

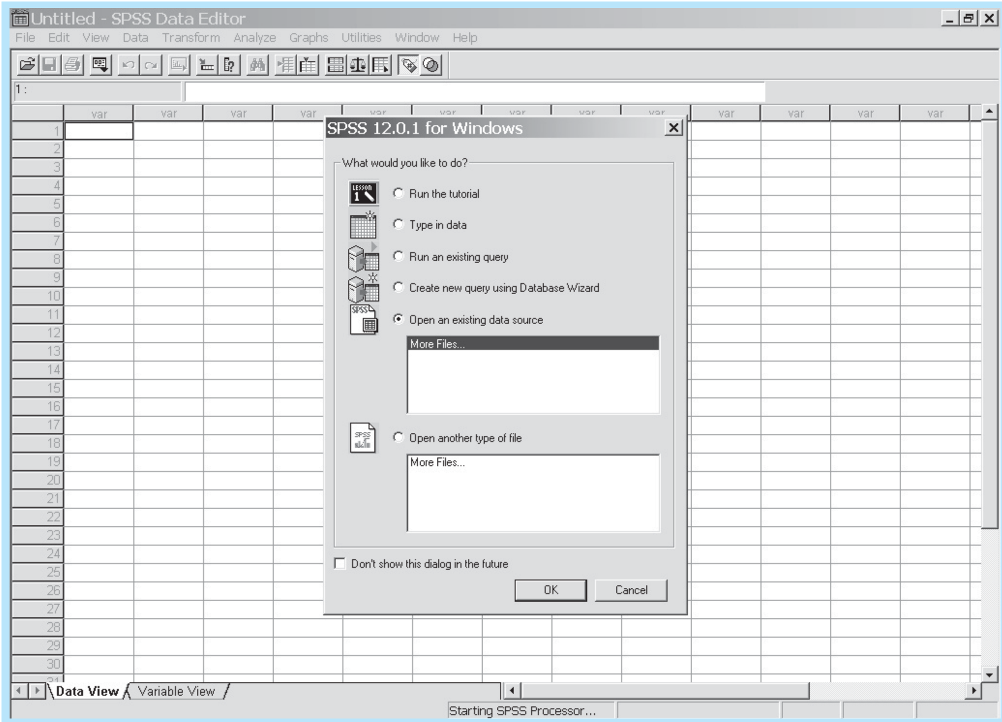


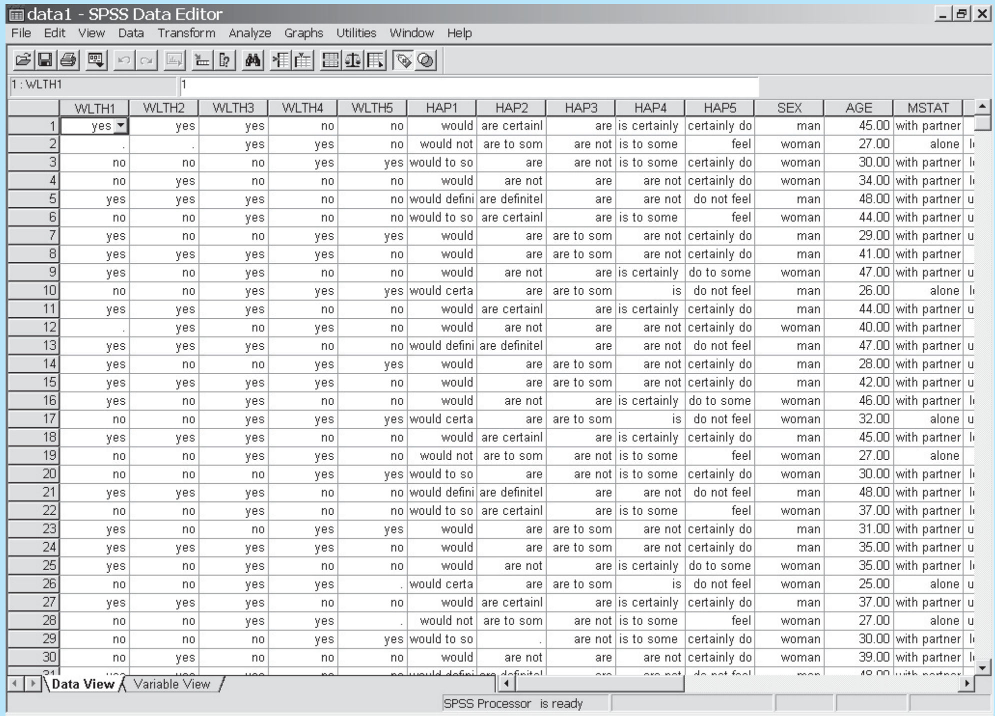

Figure 1.4 **The SPSS welcome screen**

When you wish to access a *still-unused, existing file*, such as the 'data1' file found on the http://www.basisboekstatistiekmet-spss.wolters.nl website, you click "Cancel". The small welcome screen then disappears, while the empty data matrix remains on screen. Now click "File" in the menu bar on the top left corner of the screen, followed by "Open" and "Data". You are then asked to indicate the name and location of the file. Typing in c:\data\data1, for example, and then clicking "Open", causes the sample 'data1' data file from the website to be opened. You will have needed first to have saved the data from the website to the 'data' directory on your c drive as a file named 'data1'.

> Cancel ▼
> File ▼
> Open
> Data ▼
> type (for example) c:\data\data1
> Open ▼

**SPSS data matrix**

The data in the 'data1' data file will then be entered in the SPSS data matrix. You can see this in Figure 1.5:




| | WLTH1 | WLTH2 | WLTH3 | WLTH4 | WLTH5 | HAP1 | HAP2 | HAP3 | HAP4 | HAP5 | SEX | AGE | MSTAT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | yes | yes | yes | no | no | would | are certainl | are | is certainly | certainly do | man | 45.00 | with partner |
| 2 | . | . | yes | yes | no | would not | are to som | are not | is to some | feel | woman | 27.00 | alone |
| 3 | no | no | no | yes | yes | would to so | are | are not | is to some | certainly do | woman | 30.00 | with partner |
| 4 | no | yes | no | no | no | would | are not | are | are not | certainly do | woman | 34.00 | with partner |
| 5 | yes | yes | yes | no | no | would defini | are definitel | are | are not | do not feel | man | 48.00 | with partner |
| 6 | no | no | yes | no | no | would to so | are certainl | are | is to some | feel | woman | 44.00 | with partner |
| 7 | yes | no | no | yes | yes | would | are | are to som | are not | certainly do | man | 29.00 | with partner |
| 8 | yes | yes | yes | yes | no | would | are | are to som | are not | certainly do | man | 41.00 | with partner |
| 9 | yes | no | yes | no | no | would | are not | are | is certainly | do to some | woman | 47.00 | with partner |
| 10 | no | no | yes | yes | yes | would certa | are | are to som | is | do not feel | man | 26.00 | alone |
| 11 | yes | yes | yes | no | no | would | are certainl | are | is certainly | certainly do | man | 44.00 | with partner |
| 12 | . | yes | no | yes | no | would | are not | are | are not | certainly do | woman | 40.00 | with partner |
| 13 | yes | yes | yes | no | no | would defini | are definitel | are | are not | do not feel | man | 47.00 | with partner |
| 14 | yes | no | no | yes | yes | would | are | are to som | are not | certainly do | man | 28.00 | with partner |
| 15 | yes | yes | yes | yes | no | would | are | are to som | are not | certainly do | man | 42.00 | with partner |
| 16 | yes | no | yes | no | no | would | are not | are | is certainly | do to some | woman | 46.00 | with partner |
| 17 | no | no | yes | yes | yes | would certa | are | are to som | is | do not feel | woman | 32.00 | alone |
| 18 | yes | yes | yes | no | no | would | are certainl | are | is certainly | certainly do | man | 45.00 | with partner |
| 19 | no | no | yes | yes | no | would not | are to som | are not | is to some | feel | woman | 27.00 | alone |
| 20 | no | no | no | yes | yes | would to so | are | are not | is to some | certainly do | woman | 30.00 | with partner |
| 21 | yes | yes | yes | no | no | would defini | are definitel | are | are not | do not feel | man | 48.00 | with partner |
| 22 | no | no | yes | no | no | would to so | are certainl | are | is to some | feel | woman | 37.00 | with partner |
| 23 | yes | no | no | yes | yes | would | are | are to som | are not | certainly do | man | 31.00 | with partner |
| 24 | yes | yes | yes | yes | no | would | are | are to som | are not | certainly do | man | 35.00 | with partner |
| 25 | yes | no | yes | no | no | would | are not | are | is certainly | do to some | woman | 35.00 | with partner |
| 26 | no | no | yes | yes | . | would certa | are | are to som | is | do not feel | woman | 25.00 | alone |
| 27 | yes | yes | yes | no | no | would | are certainl | are | is certainly | certainly do | man | 37.00 | with partner |
| 28 | no | no | yes | yes | . | would not | are to som | are not | is to some | feel | woman | 27.00 | alone |
| 29 | no | no | no | yes | yes | would to so | . | are not | is to some | certainly do | woman | 30.00 | with partner |
| 30 | no | yes | no | no | no | would | are not | are | are not | certainly do | woman | 39.00 | with partner |

Figure 1.5 **The screen that appears ...**

**Analyze**
**Graphs**

You can now do a number of things. You can begin processing the data. You can also perform analyses by means of "Analyze", or compose graphs using "Graphs". In subsequent chapters, we will use concrete examples to elaborate these possibilities.

Questions that undoubtedly occur to you are: 'How can I shut the SPSS program off again?' and 'How can I obtain help when working with SPSS?'

*You shut the SPSS program down* by clicking on the small black cross in the top right corner of the screen. When the program is shutting down, you are asked if you wish to save any changes that you (may) have made to the file. If you indicate that you will *save* the modified file, you are requested to give the file a name. If you type in a new name, the modified file will be stored in addition to the unmodified version. If you leave the name unaltered, the modified file will be 'written' over the earlier version of the file.

When working with SPSS, you can receive help by clicking "Help" in the menu bar displayed on screen. In the roll-down menu that appears, you subsequently select "Topics". You then encounter the help menu, which is illustrated in Figure 1.6. If, for example, you no longer remember how you saved your data, you can type in 'save'. By clicking on "data files" and "Display", you obtain information about the procedure for saving your data. There are, additionally, help buttons for providing information about specific procedures involved in the various SPSS tasks themselves.

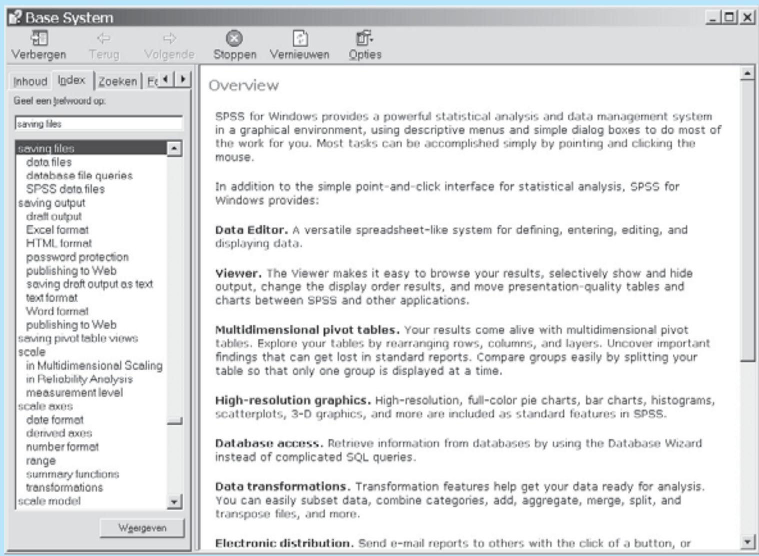



Figure 1.6 **The help menu**

Finally, we draw your attention to *all the various tutorials* included in SPSS. When you open SPSS, you see, among the series of options, that there is one entitled "Run tutorial" (see Figure 1.4). Clicking this option gains you access to a list of tutorials. In addition, SPSS has its own *website* containing all types of information about the program: www.spss.com.

**Key words**

Introduction

- research cycle
- research questions

**1.1**

- data preparation
- coding scheme
- codebook
- data editor
- variation
- recoding
- operationalisation
- item analysis
- homogeneity

1.2.1

- inquiry
- frequency research question
- comparative research question
- correlation research question

1.2.2

- level of measurement
- nominal level (in SPSS: Nominal)
- ordinal level (in SPSS: Ordinal)
- interval level (in SPSS: Scale)
- ratio level (in SPSS: Scale)
- continuous variables
- discrete variables

1.2.3

- population
- descriptive statistics
- sample
- inductive statistics
- inferential statistics
- units

**1.3**

- normal distribution
- Gauss curve
- standard error
- sample size
- homogeneity
- significance
- probability
- probability of error
- p value
- alpha ($\alpha$)
- significant
- statistical testing
- one or two-tailed testing
- degrees of freedom (df)
- effect size

**1.4**

- data file
- data matrix
- data editor
- analyze
- graphs
- help function
- tutorials