Statistical Methods for Business and Economics

Statistical Methods for Business and Economics

Gert Nieuwenhuis
Tilburg University
The Netherlands

The McGraw·Hill Companies

London Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco
St. Louis Bangkok Bogotá Caracas Kuala Lumpur Lisbon Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

Statistical Methods for Business and Economics
Gert Nieuwenhuis
ISBN-13 9780077109875
ISBN-10 0077109872

McGraw-Hill Higher Education

Published by McGraw-Hill Education
Shoppenhangers Road
Maidenhead
Berkshire
SL6 2QL
Telephone: 44 (0) 1628 502 500
Fax: 44 (0) 1628 770 224
Website: ***www.mcgraw-hill.co.uk***

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

Library of Congress Cataloging in Publication Data
The Library of Congress data for this book has been applied for from the Library of Congress

Acquisitions Editor: Rachel Gear
Head of Development: Caroline Prodger
Marketing Manager: Mark Barratt
Production Editor: James Bishop

Text design by Hard Lines
Cover design by Paul Fielding
Printed and bound in Italy by Rotolito, Lombarda

Published by McGraw-Hill Education (UK) Limited, an imprint of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020.

ISBN-13 9780077109875
ISBN-10 0077109872

The McGraw-Hill Companies

Brief Table of Contents

Detailed Table of Contents

Preface

Statistics has to do with variation, variability. The gross national product changes from year to year; people differ in opinion; sales on the market vary daily. Therefore the main theme of this book is **variation**. Statistics tries to describe and analyse variation, and above all, to explain it. Variation is the reason for statistics.

Why I wrote this book

During the past two decades, new directions in (international) economics came into existence. The growing importance of the European market and the accompanying internationalization of many organizations caused a serious need for research and knowledge about internationally oriented economics and business. The increased competition gave rise to quantification: to measure the quality of products, to explore the risks of new investments, to learn about the market and the competitors, to learn about other countries and their possibilities for investments.

At the economic faculties of universities, the process sketched above, the disappearance of the boundaries between the EU member states and the introduction of the euro stimulated the creation of new study opportunities: international business, international economics, international finance, business studies, etc. Many universities in Europe opened their doors to students from abroad, while domestic students are encouraged to do a part of their study at other universities in Europe. These developments have several consequences for the courses offered to students in economics and business. New courses on international competition prepare the students for the new situation in the European market. Other courses are adapted to include new ideas and results. Students are challenged and encouraged to widen their horizons.

Apart from the use of the computer in textbooks, introductory statistics courses for students in business and economics have hardly changed during the past thirty years. Although the growing international character should stimulate students to learn as much as possible about new ideas and methods, the courses in elementary statistics remained more or less the same. The introduction of the computer even had a serious negative side effect: statistics partly degenerated into a push-the-button science. Students learn to do the trick, but they are not encouraged to learn why this trick is a good one. It would appear that computers are so impressive that calculation is more important than understanding. Furthermore, the (often American) textbooks do not counterbalance this development. Although the need for critical and creative quantitatively oriented economists is great, students are hardly encouraged to understand the things they are doing. Books on introductory statistics do not offer a step-by-step path that students can follow to learn what statistical procedures are and how they can be used to solve problems in business and economics. Practice is that most students just use the formulae and often apply them without any understanding.

In this book I have tried to stop, and partly reverse, this process. Of course, the computer is very important for an economist and it really is indispensable for this book too. But a computer is only a powerful calculator, and a statistical computer package is no statistician. It is primarily the **understanding** of the statistical procedures that statistics in economics and business has to be about. The technical knowledge about how to perform the statistical methods with a computer is also important, but very much secondary. Students have to be challenged to understand these methods, to stimulate their creativity. It is not enough that they know the buttons to be pushed; they also have to know why. They have to be challenged to reach as high as they can. The present competitive situation in Europe demands creative and motivated economists and managers.

What distinguishes this book from others

In this book, students are challenged to understand the **statistical thinking** behind the **methods**. To accomplish this, the following guidelines are used:

- There is no reluctance to express methods as formulae.
- However, only the formulae that really increase understanding are presented.
- New methods are analysed thoroughly, until complete understanding is achieved.
- To increase understanding, emphasis is on the common elements of many seemingly different methods.
- Basic statistical methods, such as hypothesis tests, are presented as step procedures.
- Many examples are used to increase understanding of the statistical methods.

Indeed, formulae are slightly more important than in many other introductory books on statistics. But on the other hand, much more effort than usual is made to teach the ability to read the formulae and to emphasize that a formula is shorthand notation for an idea that can be expressed in words as well. The underlying aim is to explain why a formula looks as it does, to avoid the 'learning it by heart' and 'treating it as a black box'.

Much understanding can also be gained by emphasizing the common form and common ingredients of many statistical methods. To start with, many formulae about population variables in descriptive statistics and random variables in probability are basically identical; it is a waste and a shame not to point out and make use of these similarities. As a second example, the test statistics of many hypothesis tests have a common basic form. By emphasizing this underlying common structure, many formulae turn out to be similar. To stress the common features of many basic statistical methods, some of them are presented as multiple-step procedures. For instance, a hypothesis test is presented as a five-step procedure.

Many examples and exercises are about European circumstances, about EU countries or enterprises in the EU. Many of the datasets originally come from institutions such as Eurostat, OECD, World Bank and the European Central Bank. However, examples about non-economic topics, for example games and sports, can also be very stimulating. The book also contains examples using data from Statistics Netherlands, from other international statistical agencies and from my private archives. Such examples are usually European in nature: similar data might have been obtained in other countries as well.

Traditionally, introductory books on statistics offer introductions to the four sub-fields of descriptive statistics, probability theory, sampling theory and inferential statistics, treated in this order. This book also has this useful subdivision. Part 1 'Descriptive Statistics' discusses how to summarize a dataset by way of tables, graphs and statistics. If the dataset consists only of measurements on a part (sample) of the population (i.e. all objects of interest), the descriptive findings of this sample dataset are used in inferential statistics (the subject of Part 4) to draw conclusions about the whole population. It is important to note that these general conclusions are valid only if the sample is obtained in a very precise way. The sub-field of sampling theory (Part 3) discusses sampling procedures that allow such general conclusions. As usual in introductory texts on statistics, only random sampling is treated here in detail. The sub-field of probability theory (Part 2) is partly independent, but it also has to build a bridge between descriptive statistics and inferential statistics: based on the sample information and the sampling procedure it shows how to draw valid conclusions and to ascertain the precision of these conclusions.

When compared with other introductory books, this book pays more attention to the sub-fields of descriptive statistics and probability theory. Furthermore, the links between the four sub-fields and their main similarities – such as their joint purpose to describe variation of variables – are emphasized.

Introductory descriptive statistics is traditionally the least challenging part of statistics. It is heavily based on computer work and hence the underlying intentions easily get lost in viewing so many data. To overcome this, its preparatory role with respect to inferential statistics is emphasized.

For instance, in Chapter 5 the basic idea behind regression analysis – the wish to understand why a variable shows variation – is considered (and partly worked out).

Indeed, probability theory is an independent science and offers elegant, stimulating examples. But its role as intermediary between descriptive statistics and inferential statistics must also be emphasized. In many introductory books on statistics, this role does not become clear; the emerging difficulties are avoided. Discussion of probability theory often constitutes an island in isolation. In the present book, I have tried to demystify the role of probability. On the one hand, this is done by looking back to descriptive statistics and putting emphasis on the experiment 'random observation'. On the other hand, the gap with inferential statistics is bridged by looking forward and by considering probability results that are basic for inferential statistical methods. Any emerging theoretical difficulties are tackled by carefully explaining all steps and by giving examples. Some of the basic probabilistic results that underlie the theory of confidence intervals and hypothesis testing are treated in the parts of the book that deal with the sub-fields of probability and sampling. This is done to make the intermediate roles of these sub-fields more transparent and to facilitate the introduction of the statistical procedures in inferential statistics.

As mentioned at the beginning, the book concentrates on variation. This concept is crucial for economists and managers since it is often the variation of datasets and variables that is of interest. In studies regarding incomes or GDPs, measures of variation give information about income inequality. In research on product satisfaction (as in marketing) or on political opinions, little variation refers to consensus. In studies regarding investment, variation is often related to risk. The underlying purpose of many papers in economics and business is to detect the factors that, at least partially, cause the variation of the variable of interest. That is why it is extremely important to have a good understanding of the concept 'variation' and its complicated measures (such as variance, standard deviation, standard error), and of their importance for inferential statistics. In my opinion, it is not possible to inform students about similarities and differences between the many related concepts on variation without occasionally being a bit formal.

In brief, the objectives of this book are:

- to stimulate the students to reach as high as they can;
- to challenge, to increase the understanding, to make the learning by heart unnecessary;
- to demonstrate the coherence of the four sub-fields of statistics;
- to demonstrate the importance of the concept 'variation';
- to illustrate the methods with European examples.

Special notes for students and instructors

Computer packages

Most of the graphs and printouts in the book are created with Excel or SPSS. However, within the text, examples and exercises, references to these computer packages are omitted. This is done to make it possible to use the book with other computer packages as well.

For students and instructors who do prefer to use Excel and/or SPSS, the explanations of techniques are placed in Appendix A1 and put on the internet. In this appendix, the subdivision into sections is such that, for instance, A1.8 is about Excel and SPSS techniques for Chapter 8 of the book. Among Sections A1.1–A1.25, the package Excel is most important in the first sections and SPSS in the last. The reasons for putting emphasis on Excel in the first half of Appendix A1 are:

- Excel is more accessible than SPSS;
- many students have already used Excel at school or college;
- Excel is less a 'black box' than SPSS and hence fits better with the objectives of this book;
- Excel has nice options that allow data manipulations (such as the Fill Handle, which enables data to be filled into adjacent cells).

The reasons for increasing the role of SPSS throughout Appendix A1 are:

- SPSS has standard (built-in) statistical procedures;
- SPSS is especially suitable for inferential statistics.

But again, it is possible to use these packages otherwise and even to use other packages.

Traditionally, probabilities for distributions are determined with tables. I believe that tables are incomplete and outdated, and that their use has to be discouraged. However, in tutorials not all students have access to a computer, while graphical calculators can usually only deal with the normal distribution. That is why I have decided to include some tables in the book and to put other tables on the internet. However, in the text of the book, probabilities are calculated with a computer.

Sometimes a probability can be calculated just by using common sense. But in other cases the computer is needed to calculate probabilities that come from special families of distributions. In this book I have used the icon (*) to indicate that a computer is used in the calculation of a probability.

Exercises

Each of the 25 chapters ends with an exercise section: some simple exercises to practise the mechanics and to better understand the theory, some exercises to apply the theory, some more advanced exercises to challenge the reader.

Some exercises are based on datasets, others are not. For some exercises a computer is necessary to summarize the data; these exercises are marked '(computer)'. In other exercises the underlying dataset is added but not really needed to answer the questions since the data are already summarized in the text of the exercise. If wanted, such exercises can also be used on a computer practical by inviting the students first to check the summarized results.

Internet

For students, written solutions of the odd-numbered exercises and of most case studies are available on the internet. For the instructors, all solutions are available. All datasets are placed on the internet. In the datasets the decimal point is used; not the decimal comma.

Also PowerPoint files are available on the internet, one file for each of the 25 chapters. These ppt files summarize the chapters and can be used by instructors.

Although I did my utmost to avoid them, the book will probably contain errors and mistakes. I invite students and instructors to mail all errors as soon as they are detected. A file will be posted on the internet that contains the list of errors found so far. If necessary, it will be regularly updated. Of course I am also interested in general opinions about the book. Please contact me for discussion.

Cases

The book contains many cases, one at the start of each chapter (except Chapter 1) and usually one or even more at the end. They are meant to motivate and illustrate the contents of the chapters and can be used by instructors during their lectures. In each chapter, the solution of the initial case is given in the course of the chapter; the solutions of other cases are available on the internet.

Special notes for students

From the many years of my experience I know that a considerable number of students try to learn statistics by doing only the exercises. This approach will not work! The text (theory) is an essential part of the book since it explains the methods. If only the exercises are done, students will get lost in the seemingly enormous number of formulae and tricks; they will have a horrible time. But if the text is read before the exercises are attempted, the methods of the exercises are revealed and become easy to remember.

The book makes use of many symbols and letters, including Greek letters. A list of those used in the book is given in Appendix A3.

Special notes for instructors

I have tried to follow international notations as much as possible. However, I noticed that common notations are not always consistent. Since I believe that students have to learn right from the start to distinguish between the methods and the realizations that are the results of applications of the methods, I have decided to be slightly more consistent than the authors of many other books. In this book, random variables and test statistics are usually denoted by capitals (X, Y, T, G) and their realizations by small letters (x, y, t, g). Furthermore, population statistics (parameters) are usually denoted by Greek letters; sample statistics by suitable Latin letters. However, I have decided not to be too provocative and to write p for a population proportion (although π would have been more consistent). For the random sample proportion and its realization, I use the respective notations $\hat{P}$ and $\hat{p}$.

There is one concept for which I have introduced a private naming: the number that in a sense lies between the null hypothesis and the alternative hypothesis, the number that SPSS calls the test value. Since I do not know of another common name for it and since 'test value' is not suitable since it is often confused with value of the test statistic or critical value, I have called it 'hinge'.

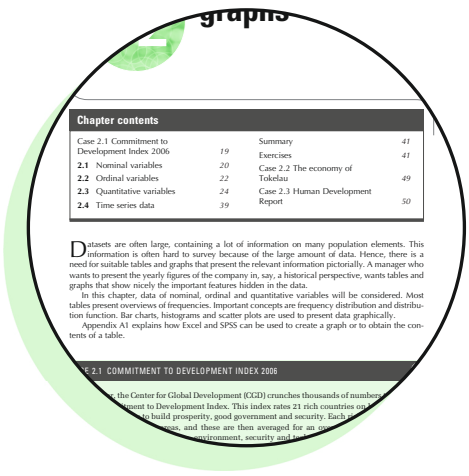
The level of mathematics needed to read this book is the ordinary level of those who finished secondary school with the intention to do a further university education in business or economics. In Chapter 8 (on probability distributions, expectations, variances), the mathematical topic differentiation is cautiously used. Integration is also used, but only for those who are familiar with it. In my experience, students learned about the summation operator at secondary school but many of them forgot about it. That is why this topic is intensively (but separately) considered in Appendix A2.

The book has 25 chapters, slightly more than most other books. Some of the chapters are small but others are rather large. If wanted, some chapters can be combined and treated in one lecture, for instance Chapters 6–7 and Chapters 12–14. I have decided to place the definitions of probability and the probability rules in different chapters (6 and 7). The main reason is that Chapter 6 is rather philosophical and, being not too large, offers the opportunity to recover from being confronted with so many descriptive statistics in Chapters 1–5.

Some sections and subsections are optional, for instance Sections 9.3 (Poisson distributions) and 10.2 (exponential distributions). If wanted, Sections 22.5 (instrumental variables) and 22.6 (logit model) can be omitted too. Even the whole of Chapter 22 (model violations for regression) can, if wanted, be omitted, since elementary residual analysis is also part of Chapters 19 (simple linear regression) and 21 (multiple linear regression: extension).

The order of the chapters is not always strict. For instance, it is possible to treat Chapters 24 and 25 immediately after Chapter 18.

Guided Tour



graphs

Chapter contents

Datasets are often large, containing a lot of information on many population elements. This information is often hard to survey because of the large amount of data. Hence, there is a need for suitable tables and graphs that present the relevant information pictorially. A manager who wants to present the yearly figures of the company in, say, a historical perspective, wants tables and graphs that show nicely the important features hidden in the data.

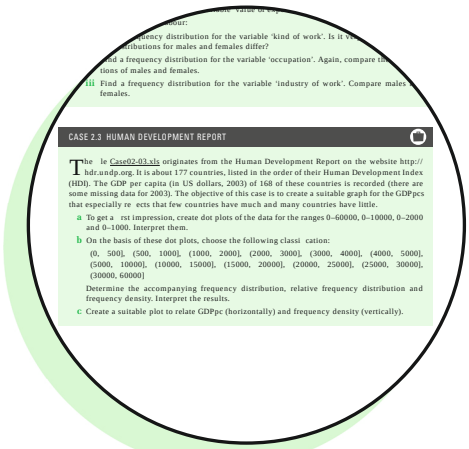
In this chapter, data of nominal, ordinal and quantitative variables will be considered. Most tables present overviews of frequencies. Important concepts are frequency distribution and distribution function. Bar charts, histograms and scatter plots are used to present data graphically.

Appendix A1 explains how Excel and SPSS can be used to create a graph or to obtain the contents of a table.

E 2.1 COMMITMENT TO DEVELOPMENT INDEX 2006

Introduction

Each chapter opens with an outline of the main techniques and methods covered in the chapter, summarizing what knowledge, skills or understanding readers should acquire once they have read it.



CASE 2.3 HUMAN DEVELOPMENT REPORT

The le Case02-03.xls originates from the Human Development Report on the website http://hdr.undp.org. It is about 177 countries, listed in the order of their Human Development Index (HDI). The GDP per capita (in US dollars, 2003) of 168 of these countries is recorded (there are some missing data for 2003). The objective of this case is to create a suitable graph for the GDPpcs that especially re ects that few countries have much and many countries have little.

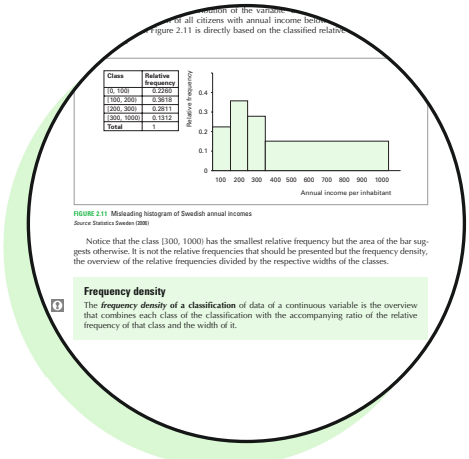
a To get a rst impression, create dot plots of the data for the ranges 0–60000, 0–10000, 0–2000 and 0–1000. Interpret them.

b On the basis of these dot plots, choose the following classi cation:
(0, 500], (500, 1000], (1000, 2000], (2000, 3000], (3000, 4000], (4000, 5000], (5000, 10000], (10000, 15000], (15000, 20000], (20000, 25000], (25000, 30000], (30000, 60000]
Determine the accompanying frequency distribution, relative frequency distribution and frequency density. Interpret the results.

c Create a suitable plot to relate GDPpc (horizontally) and frequency density (vertically).

Real-life case studies to apply statistics to business

The book includes chapter case studies designed to test how well you can apply the main techniques learned. The initial case study is revisited within the chapter so that you can see how to arrive at solving the problems. There is also a selection of longer cases at the end of most chapters for extra examples.



Class	Relative frequency
[0, 100)	0.2269
[100, 200)	0.3616
[200, 300)	0.2811
[300, 1000)	0.1312
Total	1

FIGURE 2.11 Misleading histogram of Swedish annual incomes
Source: Statistics Sweden (2006)

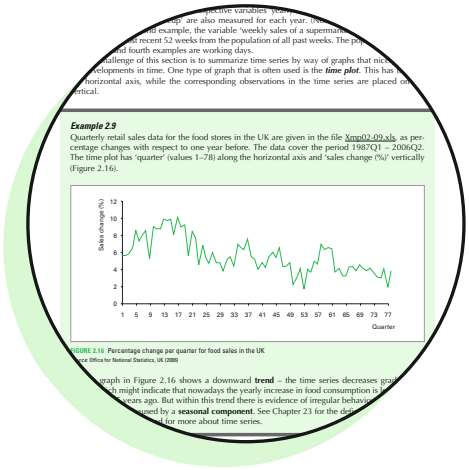
Notice that the class [300, 1000) has the smallest relative frequency but the area of the bar suggests otherwise. It is not the relative frequencies that should be presented but the frequency density, the overview of the relative frequencies divided by the respective widths of the classes.

Frequency density

The ***frequency density* of a classification** of data of a continuous variable is the overview that combines each class of the classification with the accompanying ratio of the relative frequency of that class and the width of it.

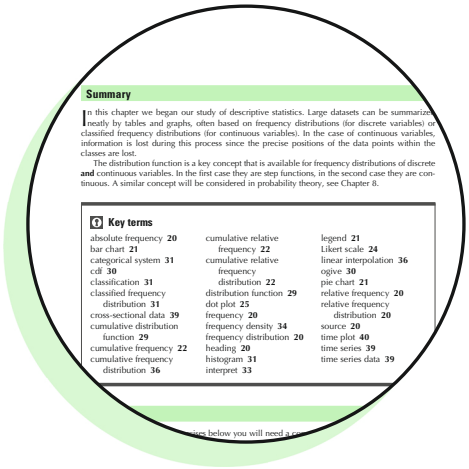
Key terms and key equations – highlighting what you need to know

Key terms are highlighted throughout the chapter in bold italic, with page number references at the end of each chapter so they can be found quickly and easily. Key equations and formulae are also highlighted in the book, and symbols listed at the end of each chapter too. An ideal tool for last minute revision or to check key formulae as you read.



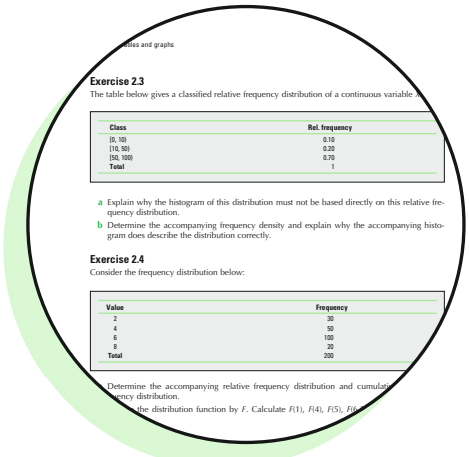
Packed with examples

Each chapter includes lots of short examples. They aim to show how a particular concept or statistical technique is used in practice, by providing data and examples showing how statistics can be applied in a business or economics context.



A useful chapter summary

This briefly reviews and reinforces the main topics you will have covered in each chapter to ensure you have acquired a solid understanding of the key topics. Use it as a quick reference to check you've understood the chapter. Each summary also includes a list of key terms in statistics.



Plenty of exercises

These questions encourage you to review and apply the knowledge you have acquired from each chapter. They are a useful revision tool to check that you have mastered statistical techniques; they can also be used by your lecturer as assignments or practice exam questions.

Technology to enhance learning and teaching

*Visit **www.mcgraw-hill.co.uk/textbooks/nieuwenhuis** today*

Online Learning Centre (OLC)

After completing each chapter, log on to the supporting Online Learning Centre website. Take advantage of the study tools offered to reinforce the material you have read in the text, and to develop your knowledge in a fun and effective way.

Resources for students include:

- Solutions to the odd-numbered exercises, to allow students to check their progress as they work through the exercises
- Solutions to selected case study problems
- Datasets from the text

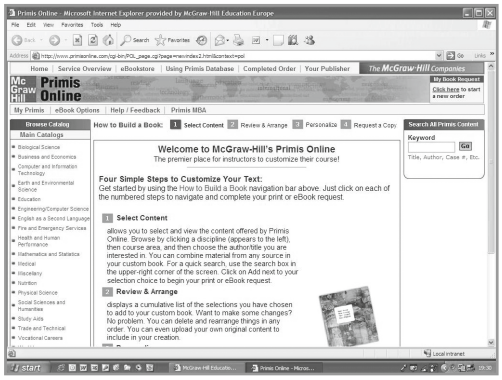
Also available for lecturers:

- Chapter by chapter PowerPoint for use in presentations or as handouts
- All solutions to the exercises
- Other additional material and updates

Custom Publishing Solutions: Let us help make our **content** your **solution**

At McGraw-Hill Education our aim is to help lecturers to find the most suitable content for their needs delivered to their students in the most appropriate way. Our **custom publishing solutions** offer the ideal combination of content delivered in the way which best suits lecturer and students.

Our custom publishing programme offers lecturers the opportunity to select just the chapters or sections of material they wish to deliver to their students from a database called Primis at ***www.primisonline.com***



Primis contains over two million pages of content from:

- textbooks
- professional books
- case books – Harvard Articles, Insead, Ivey, Darden, Thunderbird and BusinessWeek
- Taking Sides – debate materials

Across the following imprints:

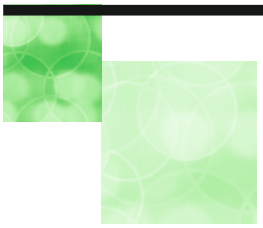
- McGraw-Hill Education
- Open University Press
- Harvard Business School Press
- US and European material

There is also the option to include additional material authored by lecturers in the custom product – this does not necessarily have to be in English.

We will take care of everything from start to finish in the process of developing and delivering a custom product to ensure that lecturers and students receive exactly the material needed in the most suitable way.

With a Custom Publishing Solution, students enjoy the best selection of material deemed to be the most suitable for learning everything they need for their courses – something of real value to support their learning. Teachers are able to use exactly the material they want, in the way they want, to support their teaching on the course.

Please contact *your local McGraw-Hill representative* with any questions or alternatively contact Warren Eels **e:** *warren_eels@mcgraw-hill.com*.



About the author

About the Author

Gert Nieuwenhuis is associate professor of probability and statistics at Tilburg University. He works at the Faculty of Economics and Business Administration, at the department of Econometrics and Operations Research. He has more than 30 years experience of teaching basic probability and statistics, regression analysis, time series forecasting, actuarial sciences, risk theory and basic econometrics to both undergraduate and graduate business and economics students. Together with Hans Moors and Maarten Janssens he has also written a series of four books, *Statistics for Economics* (in Dutch). In his spare time Professor Nieuwenhuis enjoys reading and listening to rock music, and likes to run and cycle through the holms of the river Maas and the hills of Nijmegen.

About Tilburg University

Tilburg University is a compact institution for higher education, specialized in human and social sciences and located in the southern part of the Netherlands. It has an outstanding international track record for teaching and research excellence. Its business and economics institute GentER is a world-class research institute.

Acknowledgements

Many people have contributed to the realization of this book. I want to thank all colleague lecturers and all students of Tilburg University who gave me their fruitful comments. Also many referees have given me useful comments; I want to thank them all. Many of their suggestions are incorporated in the final text: In particular, I want to thank Noud van Giersbergen for his detailed comments.

I also want to thank McGraw-Hill for giving me the opportunity to publish my educational ideas about statistics and its relation to business and economics.

I want especially to thank my colleagues and friends Hans Moors and Gé Groenewegen. Cooperation with Hans in a former book project, where we were co-authors, was very stimulating for me and helped me to develop my ideas. I also want to thank Hans for reading a part of the manuscript and for permitting me the use of previously published material and ideas. Gé was my anchor during the often exhausting process of writing the book. Apart from his critical reading of parts of the manuscript, he handed me many datasets and ideas for examples and exercises. I really want to thank him for that.

I also want to thank our children Gijs, Nienke, Martijn, Bas and Lonneke. I want to thank them for accepting that I was not always accessible, not even in the few cases when I was physically present.

But most of all I want to thank Ineke. She really was great, although it sometimes must have been a hard job to find the real Gert in the abstract world of statistics. She always remained sweet, careful, enthusiastic and supporting, even when confronted with so much physical and spiritual absence. I love you and I promise to do better.

Gert Nieuwenhuis
g.nieuwenhuis@uvt.nl
Malden
September 2008