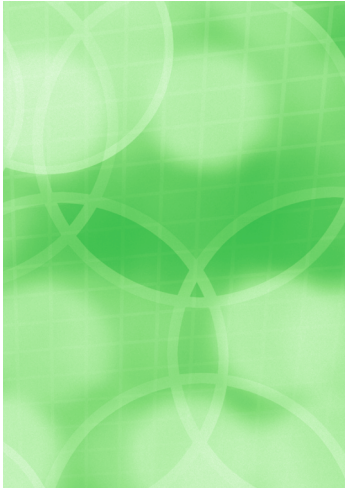


# **Statistical Methods for Business and Economics**







# 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



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](http://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. Copyright © 2009 by McGraw-Hill Education (UK) Limited. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Fictitious names of companies, products, people, characters and/or data that may be used herein (in case studies or in examples) are not intended to represent any real individual, company, product or event.

ISBN-13 9780077109875  
ISBN-10 0077109872

© 2009. Exclusive rights by The McGraw-Hill Companies, Inc. for manufacture and export. This book cannot be re-exported from the country to which it is sold by McGraw-Hill.

The McGraw-Hill Companies



# Brief Table of Contents

Preface	<i>xi</i>
Guided tour	<i>xvi</i>
Technology to enhance learning and teaching	<i>xviii</i>
About the author	<i>xxi</i>
Acknowledgements	<i>xxii</i>
<b>1</b> Introduction and basic concepts	<b>1</b>
<b>PART 1: Descriptive Statistics</b>	<b>17</b>
<b>2</b> Tables and graphs	<b>19</b>
<b>3</b> Measures of location	<b>51</b>
<b>4</b> Measures of variation	<b>87</b>
<b>5</b> Pairs of variables	<b>125</b>
<b>PART 2: Probability Theory</b>	<b>171</b>
<b>6</b> Definitions of probability	<b>173</b>
<b>7</b> Calculation of probabilities	<b>199</b>
<b>8</b> Probability distribution, expectation and variance	<b>233</b>
<b>9</b> Families of discrete distributions	<b>287</b>
<b>10</b> Families of continuous distributions	<b>309</b>
<b>11</b> Joint probability distributions	<b>338</b>
<b>PART 3: Sampling Theory</b>	<b>371</b>
<b>12</b> Random samples	<b>373</b>
<b>13</b> The sample mean	<b>391</b>
<b>14</b> Sample proportion and other sample statistics	<b>415</b>
<b>PART 4: Inferential Statistics</b>	<b>433</b>
<b>15</b> Interval estimation and hypothesis testing: a general introduction	<b>435</b>
<b>16</b> Confidence intervals and tests for $\mu$ and $p$	<b>472</b>
<b>17</b> Statistical inference about $\sigma^2$	<b>498</b>
<b>18</b> Confidence intervals and tests to compare two parameters	<b>520</b>
<b>19</b> Simple linear regression	<b>559</b>
<b>20</b> Multiple linear regression: introduction	<b>616</b>
<b>21</b> Multiple linear regression: extension	<b>648</b>
<b>22</b> Multiple linear regression: model violations	<b>694</b>
<b>23</b> Time series and forecasting	<b>733</b>
<b>24</b> Chi-square tests	<b>769</b>
<b>25</b> Non-parametric statistics	<b>792</b>
<b>Appendices</b>	<b>813</b>
<b>A1</b> Excel and SPSS (Internet)	<b>813</b>
<b>A2</b> Summation operator $\Sigma$	<b>814</b>
<b>A3</b> Greek letters	<b>819</b>
<b>A4</b> Tables	<b>820</b>
<b>A5</b> Some numeric answers to exercises	<b>845</b>
<b>Index</b>	<b>857</b>





# Detailed Table of Contents

Preface	<i>xi</i>	4.2 Measures based on deviations from the mean	96
Guided tour	<i>xvi</i>	4.3 Interpretation of the standard deviation	103
Technology to enhance learning and teaching	<i>xviii</i>	4.4 z-Scores	105
About the author	<i>xxi</i>	4.5 The variation of 0-1 data	106
Acknowledgements	<i>xxii</i>	4.6 The variance of a frequency distribution	107
<b>1 Introduction and basic concepts</b>	<b>1</b>	<i>Summary</i>	109
1.1 What is statistics?	1	<i>Exercises</i>	111
1.2 Subdivision of statistics	4	<i>Case 4.2 The reigns of British kings and queens</i>	123
1.3 Variables	5	<i>Case 4.3 Food insecurity in the world</i>	123
1.4 Populations versus samples	9	<b>5 Pairs of variables</b>	125
<i>Summary</i>	12	<i>Case 5.1 Women's world records approach men's world records</i>	126
<i>Exercises</i>	13	5.1 Scatter plot, covariance and correlation	126
<i>Case 1.1 Trading partners of the EU25</i>	16	5.2 Regression line	141
<b>PART 1: Descriptive Statistics</b>	<b>17</b>	5.3 Linear transformations	150
<b>2 Tables and graphs</b>	<b>19</b>	5.4 Relationship between two qualitative variables	154
<i>Case 2.1 Commitment to Development Index 2006</i>	19	<i>Summary</i>	157
2.1 Nominal variables	20	<i>Exercises</i>	158
2.2 Ordinal variables	22	<i>Case 5.2 Mercer Quality of Living Survey</i>	168
2.3 Quantitative variables	24	<i>Case 5.3 Anscombe's Quartet</i>	169
2.4 Time series data	39	<b>PART 2: Probability Theory</b>	<b>171</b>
<i>Summary</i>	41	<b>6 Definitions of probability</b>	173
<i>Exercises</i>	41	<i>Case 6.1 Chances of positive or negative returns on a portfolio</i>	173
<i>Case 2.2 The economy of Tokelau</i>	49	6.1 Random experiments	174
<i>Case 2.3 Human Development Report</i>	50	6.2 Rules for sets	178
<b>3 Measures of location</b>	<b>51</b>	6.3 Historical definitions of probability	182
<i>Case 3.1 The gender gap in employment rates</i>	52	6.4 General definition of Kolmogorov	187
3.1 Nominal variables	52	<i>Summary</i>	190
3.2 Ordinal variables	55	<i>Exercises</i>	190
3.3 Quantitative variables	55	<i>Case 6.2 The mysterious dice</i>	197
3.4 Relationship between mean / median / mode and skewness	70	<i>Case 6.3 A three-question quiz about risk (part I)</i>	198
<i>Summary</i>	72	<b>7 Calculation of probabilities</b>	199
<i>Exercises</i>	73	<i>Case 7.1 Internet connection problems and the LinkNet Router</i>	199
<i>Case 3.2 The paradox of means (Simpson's paradox)</i>	84	7.1 Basic properties	200
<i>Case 3.3 Did the euro cause price increases?</i>	85	7.2 Rules for counting	202
<b>4 Measures of variation</b>	<b>87</b>	7.3 Random drawing and random sampling	207
<i>Case 4.1 Ericsson shares versus Carlsberg shares</i>	88		
4.1 Measures based on quartiles and percentiles	89		

7.4 Conditional probabilities and independence	210	11.2 Covariance and correlation	342
7.5 Bayes' rule	216	11.3 Conditional probabilities and independence of random variables	346
Summary	218	11.4 Linear combinations of random variables	352
Exercises	220	Summary	359
Case 7.2 <i>The market for wool detergent</i>	231	Exercises	361
Case 7.3 <i>Russian roulette</i>	232	Case 11.2 <i>Portfolios of the stocks Philips and Ahold</i>	368
Case 7.4 <i>Was the draw for the UEFA Euro 2004 play-offs fair?</i>	232	Case 11.3 <i>Introduction to Markowitz's portfolio theory (part II)</i>	369
<b>8 Probability distribution, expectation and variance</b>	233	<b>PART 3: Sampling theory</b>	<b>371</b>
Case 8.1 <i>Expected return and risk of Carlsberg Breweries stock</i>	233	<b>12 Random samples</b>	373
8.1 Random variables	234	Case 12.1 <i>Number of defects on electronic circuit boards</i>	373
8.2 Probability distributions	236	12.1 Sampling methods	373
8.3 Functions of random variables	250	12.2 Random samples with replacement ( <i>iid</i> samples)	376
8.4 Expectation, variance and standard deviation	252	12.3 Random samples without replacement	378
8.5 Rules for expectation and variance	260	12.4 Sample statistics and estimators	380
8.6 Random observations	267	Summary	383
8.7 Other statistics of probability distributions	271	Exercises	384
Summary	274	Case 12.2 <i>Households statistics (part I)</i>	389
Exercises	275	<b>13 The sample mean</b>	391
Case 8.2 <i>A three-question quiz about risk (part II)</i>	284	Case 13.1 <i>The ruin probability of insurance company Lowlands</i>	391
Case 8.3 <i>Introduction to Markowitz's portfolio theory (part I)</i>	284	13.1 Expectation, variance and Chebyshev's rule	392
<b>9 Families of discrete distributions</b>	287	13.2 Concerning the exact probability distribution of the sample mean	396
Case 9.1 <i>Defective computer chips</i>	287	13.3 The central limit theorem	399
9.1 Bernoulli distributions	287	13.4 Consequences of the CLT	403
9.2 Binomial and hypergeometric distributions	289	Summary	406
9.3 Poisson distributions	299	Exercises	408
Summary	302	Case 13.2 <i>Households statistics (part II)</i>	413
Exercises	303	<b>14 Sample proportion and other sample statistics</b>	415
Case 9.2 <i>The non-business mobile phone market</i>	307	Case 14.1 <i>Approval probabilities in quality control</i>	416
<b>10 Families of continuous distributions</b>	309	14.1 Properties of the sample proportion	416
Case 10.1 <i>EU limit for carbon dioxide emissions</i>	309	14.2 Properties of other sample statistics	422
10.1 Uniform distributions	309	14.3 Standard errors	424
10.2 Exponential distributions	313	Summary	425
10.3 Normal distributions	315	Exercises	426
Summary	328	Case 14.2 <i>Households statistics (part III)</i>	431
Exercises	329		
Case 10.2 <i>The green and red people</i>	336		
<b>11 Joint probability distributions</b>	338		
Case 11.1 <i>Insurance against bicycle theft</i>	338		
11.1 Discrete joint probability density function	339		



<b>PART 4: Inferential Statistics</b>	433		
<b>15 Interval estimation and hypothesis testing: a general introduction</b>	435		
Case 15.1 <i>Should industrial activities be moved to another country?</i>	435		
15.1 Initial approach to statistical procedures	436		
15.2 Point and interval estimation	440		
15.3 Hypothesis testing	444		
Summary	460		
Exercises	462		
Case 15.2 <i>Personality traits of graduates (part I)</i>	470		
<b>16 Confidence intervals and tests for <math>\mu</math> and <math>p</math></b>	472		
Case 16.1 <i>If it says McDonald's, then it must be good</i>	472		
16.1 Standardized sample mean and $t$ -distribution	473		
16.2 Confidence intervals and tests for $\mu$	477		
16.3 Confidence intervals and tests for $p$ : large sample approach	483		
16.4 Common formats so far	488		
Summary	489		
Exercises	490		
Case 16.2 <i>The effects of an increase in the minimum wage (part I)</i>	497		
<b>17 Statistical inference about <math>\sigma^2</math></b>	498		
Case 17.1 <i>Standard deviation as a measure of disunity</i>	498		
17.1 Recap and introduction	499		
17.2 A property of the estimator	502		
17.3 Confidence intervals	506		
17.4 Tests	509		
Summary	514		
Exercises	515		
Case 17.2 <i>FTSE 100 and the collapse of the US housing market</i>	519		
<b>18 Confidence intervals and tests to compare two parameters</b>	520		
Case 18.1 <i>Did the changes to Statistics 2 increase the grades?</i>	520		
18.1 Some problems with two parameters	521		
18.2 The difference between two population means	521		
18.3 The ratio of two population variances	534		
18.4 The difference between two population proportions	542		
Summary	546		
Exercises	547		
Case 18.2 <i>Fund ABN AMRO AEX FD9 and beating the market</i>	557		
		Case 18.3 <i>Ambitiousness of students</i>	557
		Case 18.4 <i>The effects of an increase in the minimum wage (part II)</i>	558
		<b>19 Simple linear regression</b>	559
		Case 19.1 <i>Wage differentials between men and women (part I)</i>	560
		19.1 Relating a variable to other variables	561
		19.2 The simple linear regression model	563
		19.3 Point estimators of $\beta_0$ , $\beta_1$ and $\sigma_\varepsilon^2$	570
		19.4 Properties of the estimators	575
		19.5 Inference about the parameter $\beta_1$	578
		19.6 ANOVA table and degree of usefulness	585
		19.7 Conclusions about $Y$ and $E(Y)$	592
		19.8 Residual analysis	597
		Summary	602
		Exercises	604
		Case 19.2 <i>Profits of top corporations in the USA (part I)</i>	613
		Case 19.3 <i>Income and education level of identical twins (part I)</i>	614
		<b>20 Multiple linear regression: introduction</b>	616
		Case 20.1 <i>Pollution due to traffic (part I)</i>	617
		20.1 The multiple linear regression model	617
		20.2 Properties of the point estimators	624
		20.3 ANOVA table	625
		20.4 Usefulness of the model	627
		20.5 Inference about the individual regression coefficients	632
		20.6 Conclusions about $Y$ and $E(Y)$	636
		20.7 Residual analysis	637
		Summary	639
		Exercises	640
		Case 20.2 <i>Income and education level of identical twins (part II)</i>	647
		<b>21 Multiple linear regression: extension</b>	648
		Case 21.1 <i>Pricing diamond stones</i>	649
		21.1 Usefulness of portions of a model	649
		21.2 Collinearity	653
		21.3 Higher-order terms and interaction terms	655
		21.4 Logarithmic transformations	660
		21.5 Analysis of variance by way of dummy variables	662
		21.6 Model building	673
		Summary	679
		Exercises	679
		Case 21.2 <i>Profits of top corporations in the USA (part II)</i>	693

Case 21.3 <i>Personality traits of graduates (part II)</i>	693	24.1 Introduction	770
Case 21.4 <i>Pollution due to traffic (part II)</i>	693	24.2 Goodness of fit tests	771
<b>22 Multiple linear regression: model violations</b>	694	24.3 Tests for independence and homogeneity	779
Case 22.1 <i>Income and education level of identical twins (part III)</i>	694	Summary	785
22.1 Collinearity	695	Exercises	785
22.2 Heteroskedasticity	699	Case 24.2 <i>Different views in the EU about illegal activity</i>	791
22.3 Non-linearity and non-normality	704	<b>25 Non-parametric statistics</b>	792
22.4 Dependence of the error terms	708	Case 25.1 <i>Business start-ups and lack of capital</i>	792
22.5 Instrumental variables	714	25.1 Introduction	793
22.6 Introduction to binary choice models and the logit model	717	25.2 Two independent samples	794
Summary	722	25.3 Two matched samples	799
Exercises	723	25.4 Two or more independent samples	803
Case 22.2 <i>Profits of top corporations in the USA (part III)</i>	731	Summary	806
Case 22.3 <i>A 'final' model for the wage differentials case</i>	731	Exercises	807
<b>23 Time series and forecasting</b>	733	<b>Appendices</b>	813
Case 23.1 <i>Forecasting the price of Microsoft stock</i>	734	A1 Excel and SPSS (Internet)	813
23.1 Introduction	734	A2 Summation operator $\Sigma$	814
23.2 Components of time series	735	A3 Greek letters	819
23.3 Smoothing techniques: moving averages, exponential smoothing	737	A4 Tables	820
23.4 Exponential smoothing and forecasting	741	Table 1. Binomial distributions	820
23.5 Linear regression and forecasting	742	Table 2. Poisson distributions	827
23.6 Autoregressive model and forecasting	749	Table 3. Distribution function of the standard normal distribution	829
Summary	753	Table 4. Quantiles of $t$ -distributions	830
Exercises	754	Table 5. $\chi^2$ -distributions	831
Case 23.2 <i>Persistence of the capital market rate (worked out)</i>	766	Table 6. $F$ -distributions	832
<b>24 Chi-square tests</b>	769	Table 7. Durbin-Watson bounds	842
Case 24.1 <i>Kicks from the penalty mark in soccer</i>	769	Table 8. Critical values for the Wilcoxon rank sum test	843
		Table 9. Critical values for the Wilcoxon signed rank sum test	844
		A5 Some numeric answers to exercises	845
		<b>Index</b>	857



# 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  $\pi$  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’.

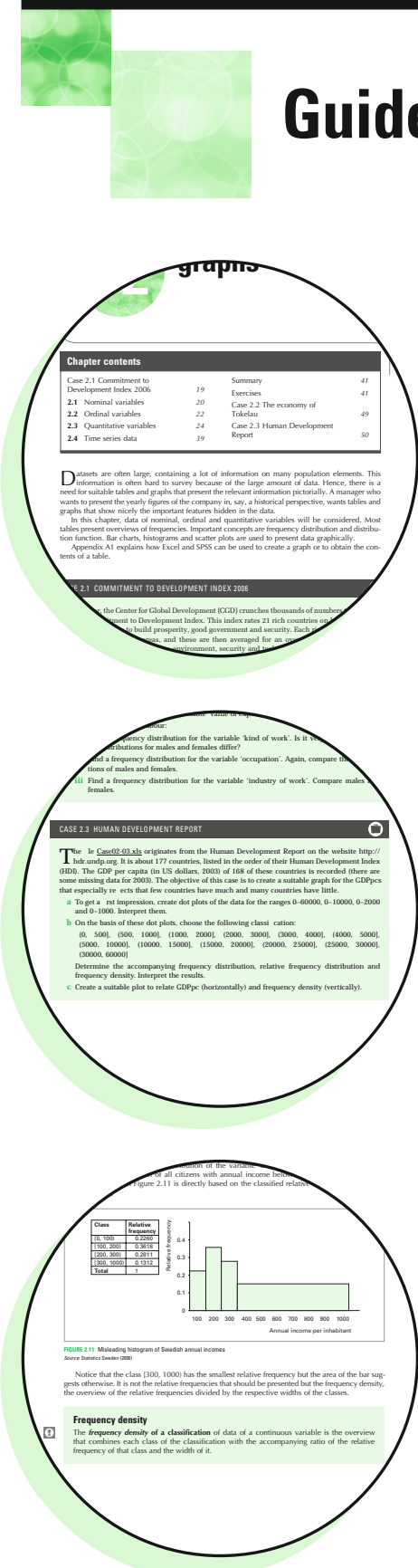
The level of mathematics needed to read this book is the ordinary level of those who finished secondary school with the intention to do a further university education in business or economics. In Chapter 8 (on probability distributions, expectations, variances), the mathematical topic differentiation 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



## 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.

## 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.

## 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.



...variables (time series) are also measured for each year. For example, the variable weekly sales of a supermarket is measured for the recent 52 weeks from the population of all past weeks. The past four examples are working days.

The purpose of this section is to summarise time series by way of graphs that show developments in time. One type of graph that is often used is the **time plot**. This has a horizontal axis, while the corresponding observations in the time series are placed vertically.

**Example 2.9**  
Quarterly retail sales data for the food stores in the UK are given in the file `Temp02-09.xls`, as percentage changes with respect to one year before. The data cover the period 1987Q1 – 2006Q2. The time plot has ‘quarter’ (values 1–78) along the horizontal axis and ‘sales change (%)’ vertically (Figure 2.16).

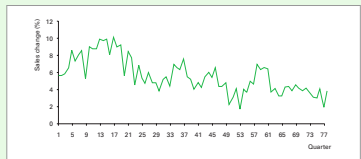


FIGURE 2.16 Percentage change per quarter for food sales in the UK over the two-decade timespan, 1987–2006.

The graph in Figure 2.16 shows a **downward trend** – the time series decreases gradually. It might indicate that nowadays the yearly increase in food consumption is less than 10 years ago. But within this trend there is evidence of irregular behaviour, which is caused by a **seasonal component**. See Chapter 23 for the details. For more about time series, see Chapter 23.

## 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.

### Summary

In this chapter we began our study of descriptive statistics. Large datasets can be summarized neatly by tables and graphs, often based on frequency distributions (for discrete variables) or classified frequency distributions (for continuous variables). In the case of continuous variables, information is lost during this process since the precise positions of the data points within the classes are lost.

The distribution function is a key concept that is available for frequency distributions of discrete and continuous variables. In the first case they are step functions, in the second case they are continuous. A similar concept will be considered in probability theory, see Chapter 8.

#### Key terms

absolute frequency 20	cumulative relative frequency 22	legend 21
bar chart 21	cumulative relative frequency 22	Likert scale 24
categorical system 31	frequency 20	linear interpolation 36
cd 30	frequency distribution 22	ogive 30
classification 31	distribution function 29	pie chart 21
classified frequency distribution 31	dot plot 25	relative frequency 20
cross-sectional data 39	frequency 20	relative frequency distribution 20
cumulative distribution function 29	frequency density 34	source 20
cumulative frequency 22	frequency distribution 20	time plot 40
cumulative frequency distribution 36	heading 20	time series 39
	histogram 31	time series data 39
	interpret 33	

## 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.

### Exercise 2.3

The table below gives a classified relative frequency distribution of a continuous variable.

Class	Rel. frequency
[0, 10)	0.10
[10, 50)	0.20
[50, 100)	0.70
Total	1

- Explain why the histogram of this distribution must not be based directly on this relative frequency distribution.
- Determine the accompanying frequency density and explain why the accompanying histogram does describe the distribution correctly.

### Exercise 2.4

Consider the frequency distribution below:

Value	Frequency
2	30
4	50
6	100
8	20
Total	200

- Determine the accompanying relative frequency distribution and cumulative frequency distribution.
- Determine the distribution function  $F$ . Calculate  $F(1)$ ,  $F(4)$ ,  $F(5)$ ,  $F(6)$ .

## 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](http://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

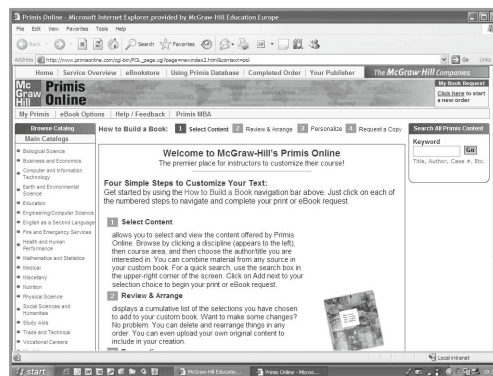
### Also available for lecturers:

- 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](http://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:

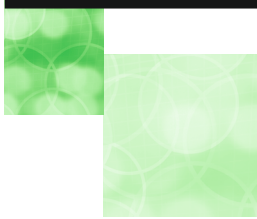
- McGraw-Hill Education
- Open University Press
- Harvard Business School Press
- US and European material

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.

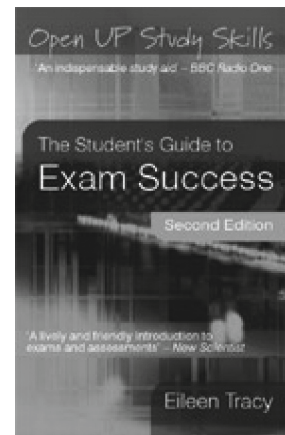
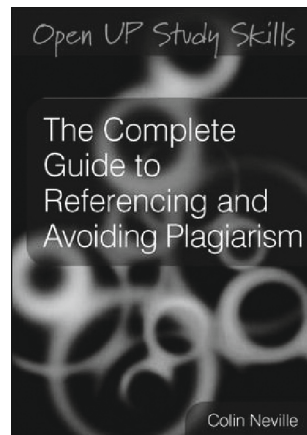
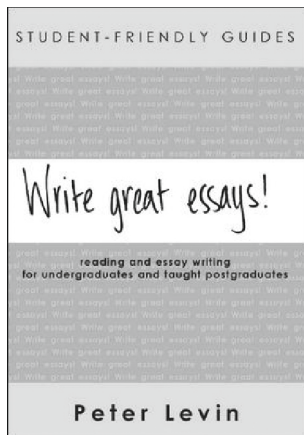
We will take care of everything from start to finish in the process of developing and delivering a custom product to ensure that lecturers and students receive exactly the material needed in the most suitable way.

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](mailto:warren_eels@mcgraw-hill.com).



## Make the grade!



# 30% Off any Study Skills book!

Our Study Skills books are packed with practical advice and tips that are easy to put into practice and will really improve the way you study. Topics include:

- techniques to help you pass exams
- advice to improve your essay writing
- help in putting together the perfect seminar presentation
- tips on how to balance studying and your personal life

[www.study-skills.co.uk](http://www.study-skills.co.uk)

Visit our website to read helpful hints about essays, exams, dissertations and much more.

**Special offer!** As a valued customer, buy online and receive 30% off any of our Study Skills books by entering the promo code **getahead**



---

# 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

## *A note from the Publisher*

Every effort has been made to trace and acknowledge ownership of copyright and to clear permission for material reproduced in this book. The publishers will be pleased to make suitable arrangements to clear permission with any copyright holders whom it has not been possible to contact.