Statistics for Non-Statisticians

Birger Stjernholm Madsen

Statistics for Non-Statisticians

Second Edition

Springer

Birger Stjernholm Madsen
Novozymes A/S
Bagsvaerd, Denmark

The Danish edition was published in 2012 as “Statistik for ikke-statistikere” by Samfundslitteratur, Frederiksberg, Denmark.

ISBN 978-3-662-49348-9 ISBN 978-3-662-49349-6 (eBook)
DOI 10.1007/978-3-662-49349-6

Library of Congress Control Number: 2016940499

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

Preface

Never have so many organizational decisions been taken based on statistics as today! Everything is supported by numbers. This applies to marketing, economics, social sciences, natural sciences, industry, and administrative work within organizations, businesses, and institutions. It is, therefore, important to have insight into basic statistical concepts, when assessing statistical data material, as well as when preparing an investigation, so that it produces useful statistical results.

There are several books on elementary statistics. So why write another? The simple answer is: because it is needed! This book fills a gap in the existing literature about statistics. Most existing short introductions to statistics take one of the following approaches:

- An approach based primarily on descriptive statistics (charts, tables, etc.)
- A purely verbal approach without any mathematical formula, but also without practical guidelines
- An approach based primarily on probability theory

In contrast, this book is intended to be a "first course for the practitioner," giving a lot of useful details for, e.g., planning of surveys. Comparing this book to standard 500–600 pages textbooks on statistics, you will actually find a lot of practical information in this book that is not available in the standard textbooks!

I have for some decades taught statistics at all levels. It is my experience that the most important concepts of statistics can be explained, so that "ordinary" people can understand it. I have experienced this through hundreds of courses for many different audiences. Now, I have put my words on paper!

Who Is this Book Written for?

The book is written for those who need to know how to collect, analyze, and present data. You may be working with administrative data, financial data, or data from the social sciences or natural sciences. Maybe you plan to collect data through sample surveys, such as customer surveys or similar.

You do not know much about statistics. Maybe you have learned a little about the topic earlier, but forgotten most of it again. Maybe you never learned anything about the topic, but you are curious!

Although the book does not require knowledge of statistics, I assume that you are not totally unfamiliar with numbers! You are able to perform simple calculations with a calculator. And you don't panic, when you see a simple formula containing a square root! Don't worry: This book is not loaded with mathematical formulas. But it is unfortunately impossible to introduce statistical concepts without a minimum of mathematical calculations.

It is an advantage, if you have a basic understanding of spreadsheets. This book is not a course in the use of spreadsheets—the easiest way to learn spreadsheets is by reading a computer booklet or taking a course!

Neither is this a "How to do statistics with Excel" book—you can use the references in the literature list, if you need this. There are many books of this kind, often occupying hundreds of pages. . .

Yet it may be useful to know how the most important statistical calculations can be performed using the features of a spreadsheet. Spreadsheets have nowadays made numbers and graphs accessible to most people. This also applies to statistical calculations!

If you do not have access to a spreadsheet, I can recommend the spreadsheet Calc from Open Office (a free Office suite). See links to software in Appendix. Virtually everything discussed in this book can be done with this spreadsheet!

I recommend that, while reading the book, you work with some simple data, which you enter in a spreadsheet. It is easier to learn statistics if you work a little with the substance!

The beginner may be satisfied with a spreadsheet as a tool for statistical analysis. In professional work with statistics, however, you will very quickly discover the limitations of a spreadsheet. Then it is time to consider a better tool for the purpose! Therefore, the Appendix presents some of the major programs for statistics as well as links, where you can find more about them.

It is my hope that the book can be used for private study and as supplementary reading at business colleges, technical schools, high schools, and the initial training of statistics at business schools and social sciences at a university level. The book is not written for any specific education.

After reading this book, you should have the ability to dig further into some of the many other books on statistics available on the market. It is my hope that this book can ease the transition to the reading of the (many) more advanced books on

the subject. The number of books on statistics grows dramatically as the professional level rises!

The mathematically oriented reader has to accept that the book does not achieve 100 % mathematical precision everywhere. Focus is on an understandable rather than mathematical precise presentation.

Some topics in the book are a bit more "technical" than the rest of the book. These issues can be skipped, without thereby losing coherence. Some of these items are placed in a text frame and entitled "Technical note: . . ." Some topics provide a clear indication that they may be skipped. They are often put at the end of a chapter.

There are also many examples of using spreadsheets. If you do not use spreadsheets, you can just read the examples, without bothering about how the results were obtained in the spreadsheet.

Structure of the Book

The book is structured in such a way that what you learn in one chapter is used in the following chapters. This means that you should read it from the beginning, at least up to and including Chap. 5.

Chapters 1 and 2 are about the collection and presentation of data. These are crucial issues for most people working with statistics.

Chapters 3–5 are the core of the book. They introduce the basic statistical concepts, including descriptive statistics, the normal distribution, and statistical tests.

When you have read Chap. 5, Chaps. 6–8 can be read independently of each other.

Chapter 6 supplements Chap. 1; it is about the planning of sample surveys and experiments.

Chapters 7 and 8 supplement Chap. 4 on the normal distribution. Chapter 8 is probably the "heaviest" material of the book and appropriately placed at the end!

The Appendices of the book contain a lot of hopefully useful information: review of probability theory, bibliography, glossary of statistical terms, list of statistical functions in spreadsheets, list of statistical software, useful links, and various useful tables. All words in this book, which are marked with an *asterisk* (*), are explained in the glossary.

At the publisher's website, you will find additional material for the book: useful worksheets, further explanation, examples, etc. Of course, there is also a spreadsheet with the example dataset "Fitness Club," which is used as recurrent example.

I wish you a pleasant reading!

Copenhagen, Denmark Birger Stjernholm Madsen

Preface to the 2nd edition

During the years since the first edition of this book, I have received many comments. Some pointed out that a few useful sections could be added in order to make the book available to a wider audience, including people working in industry. The main enhancements are as follows:

1. New sections in Chap. 4 about the lognormal distribution, control charts and process capability have been added.
2. A new section in Chap. 8 about ANOVA (with 1 factor) has been added.

The appendices about references, links and software have also been updated.

It is my hope that these changes and additions have improved the book substantially. I urge the readers to further comment on the book!

Copenhagen, Denmark — Birger Stjernholm Madsen

Acknowledgements

I send my warm thanks to my editors at Springer, Barbara Fess and Johannes Gläser, who came with a lot of good suggestions.

I also want to thank my editor on the Danish editions, Peter Byriel from Samfundslitteratur. He gave a lot of constructive criticism along the way.

In addition, I want to thank the following Danish statisticians for valuable comments: Leif Albert Jørgensen, Niels Landvad and Anders Milhøj.

Finally, I thank my wife, Yrsa, for being extremely patient with me in the busy periods!

Contents

Abbreviations

$\bar{x}$	Sample average
ANOVA	Analysis of Variance
B(n,p)	Binomial distribution (*n* observations, probability *p*)
Cp	Process Capability Index
Cpk	Minimum Process Capability Index
CV	Coefficient of Variation
DF	Degrees of Freedom
DOE	Design of Experiments
E(X)	Mean of *X*. E =“Expectation”
H_0	Null hypothesis
H_1	Alternative hypothesis
LCL	Lower Control Limit
LSL	Lower Specification Limit
N(0,1)	Standardized normal distribution (mean 0, variance 1)
$N(\mu,\sigma^2)$	Normal distribution with mean μ and variance σ^2
R	Sample range
s	Sample standard deviation
s^2	Sample variance
SPC	Statistical Process Control
UCL	Upper Control Limit
USL	Upper Specification Limit
V(X)	Variance of *X*
μ	Population mean
σ	Population standard deviation
σ^2	Population variance
Σ	Sum
χ^2	Chi-squared (distribution or test)

About the Author

Birger Stjernholm Madsen has a Master of Science in Statistics and Mathematics.

He has several years of experience as a statistician at major Danish companies within several industries as well as within national statistics.

He has also taught statistics for several years at University of Copenhagen and has held hundreds of statistics courses for various audiences throughout decades. This book is a direct result of his teaching.