

Contents

Introduction — VII

- 1 Basic statistical concepts — 1**
 - 1.1 Arithmetic mean — 1
 - 1.2 Standard deviation — 1
 - 1.3 Standard error — 2
 - 1.4 Confidence intervals — 3
 - 1.5 Applied example — 4
 - 1.6 P-values — 5
 - 1.7 Testing for normality — 8
 - 1.8 Probability density functions — 10
 - 1.9 Bias and variance — 13
 - 1.10 Normality and confidence intervals revisited — 14

- 2 Bootstrapping — 17**
 - 2.1 Working example — 17
 - 2.2 Requirements — 19
 - 2.3 Diagnostics — 21
 - 2.3.1 Unique values — 22
 - 2.3.2 Distribution of bootstrap-results — 23
 - 2.3.3 Rate of convergence — 25
 - 2.4 The logic of bootstrapping — 26
 - 2.5 Failures of the bootstrap — 29

- 3 Confidence intervals — 32**
 - 3.1 Normal intervals — 32
 - 3.2 Percentile intervals — 33
 - 3.3 Bias corrected intervals (BC) — 34
 - 3.4 Bias corrected and accelerated intervals (BCa) — 37
 - 3.5 Double bootstrap — 39
 - 3.6 Transformations — 42
 - 3.7 Summary and simulation — 44

- 4 Hypothesis testing — 49**
 - 4.1 The logic of hypothesis tests — 49
 - 4.2 Confidence intervals — 51
 - 4.3 Sampling in accordance with the null — 53
 - 4.4 Permutation tests — 54
 - 4.4.1 Paired permutation tests — 59
 - 4.4.2 Boot or permute? — 61

4.5	Power calculations —	63
5	Regression models —	66
5.1	Sampling observations —	66
5.2	Sampling residuals —	70
5.3	Dealing with outliers —	71
5.4	Predicting outcomes and bagging —	75
6	Applied programming examples —	78
6.1	Python —	78
6.1.1	Basics —	78
6.1.2	All confidence intervals —	80
6.1.3	Permutation tests —	85
6.1.4	Multithreading —	89
6.1.5	Importing data —	90
6.1.6	Appendix —	92
6.2	Stata —	94
6.2.1	Descriptive statistics —	95
6.2.2	Estimating models —	97
6.2.3	Writing own programs —	99
6.2.4	Permutation tests —	106
6.2.5	Longitudinal and cluster analyses —	108
6.3	Computational speed —	114
6.4	Further reading —	117
	Bibliography —	119

Introduction

Why bootstrapping?

Few concepts in science come closer to the idea of conjuring something out of thin air than bootstrapping. Although it sounds too good to be true, the concept has been around for a while and has been successfully applied in thousands of studies across all scientific disciplines (Chernick, 2011). The approach was originally developed by Bradley Efron back in the 1970s, when computers easily filled entire rooms and were only accessible to experts. However, technological advances have exponentially increased the amount of computational power available to the average user, making this method much more attractive to a wide audience of people. The main objective of the present book is to convey the central ideas of the technique and demonstrate its flexibility and elegance. Based on this knowledge, each reader should be able to apply bootstrapping to his or her own work and benefit from its many advantages, which can be summarized as follows:

- Fewer assumptions: Many statistical tests make relatively strict assumptions about data, for example, the requirement of normality. If these are not met, which is more often the case than not, conclusions drawn from the analyses might be incorrect. Bootstrapping makes fewer assumptions and often gives better results.
- High flexibility: when there are no analytical formulas available to estimate standard errors for certain statistics, bootstrapping can help out and quantify the uncertainty around a point estimate. Due to its flexibility, bootstrapping computes these estimates for virtually any statistic.
- Accuracy: Bootstrap and related methods, such as permutation tests, are sometimes more accurate than their parametric counterparts and often give highly precise results.

Audience

This book is intended for students and practitioners from the social sciences, especially sociology, psychology, political science, and econometrics. However, the basic ideas and formulas can be applied in all areas of science and are easily generalized to many applications. As bootstrapping is based on a simple statistical concept, this book only requires knowledge about basic statistics (say, Stats I and II). While formulas are often helpful to convey ideas in a precise fashion, the main objective of this book is to be as understandable as possible to a wide audience, so that everyone can profit from the many benefits that bootstrapping delivers.

Structure

Chapter one gives a short summary of central statistical concepts as a foundation (or refreshment) for the topic at hand. Chapter two introduces bootstrapping, when it should be used, why it works, and how its quality can be assessed. Chapter three presents a detailed discussion of bootstrapping confidence intervals, given that they are quite common and are a great option to communicate the uncertainty around a point estimate. Chapter four continues with hypothesis testing, one of the main goals of applied science. Chapter five explains how bootstrapping can be combined with the large family of regression-based approaches, which are one of the most important methods in the social sciences. Finally, chapter six gives concrete examples in Python and Stata, making bootstrapping come alive to help you answer your research questions.

If you want to contact me, you can reach me via email: felix.bittmann@lifbi.de

Acknowledgments

I would like to thank everyone who supported me and this project. Florian Scholze read an early draft and gave valuable comments to improve clarity and precision, especially on statistical concepts. Agata Kraj took care of the language editing and helped convey my ideas in a clear and correct fashion. Stefan Giesen, my editor at De Gruyter Oldenbourg, supported my plans for this book from the very beginning and made the realization possible. Last but not least, my supervisor Steffen Schindler has given me the freedom to pursue side projects and work on various interesting topics. I deeply appreciate your support.