Preface

For those of us who are passionately interested in research, the methods we use are at least as important as our findings. We need to have confidence that our quantitative methods give us more than just an illusion of rigor; they should provide genuine insight into the problems which interest us. This book is a tool to assist researchers in making sense of their quantitative data without confusion, bias, or arbitrary conventions.

A few years ago we wrote a critical examination of statistical practice in Human–Computer Interaction which was presented at CHI 2012. Due to the lively discussions that surfaced at the conference, we thought there was more to "statistical methods in HCI" than could be conveyed through a conference paper. Initially, we set out to work on a journal paper which would both examine current practice, as well as introduce a number of statistical methods that are not covered in introductory research methods and statistics lectures but could, in our view, strengthen the field. The article then became lengthy, and we really wanted it to be hands-on. It spiraled out of control: we started involving both experts in different methods, as well as users of "a-typical" methods in HCI, and discussed their possible contributions to the article. Hence, mid 2014, we decided, in collaboration with Springer, to turn our article into a book. And here it is.

About the Authors

We confess to the reader at the outset that our own statistical practices are not perfect. In fact, over the years, both of us have committed about all of the potential errors identified in this volume (and they are there in the literature for you to sorrow over). These less than perfect analyses stem from ignorance of the pitfalls of null hypothesis significance testing—honestly!—rather than an intention to deceive the reading public. On occasion however, we have consciously used "traditional" methods rather than their more recent counterparts in order to tailor the analyses to

the reviewers' expectations. It would be better for our field if authors didn't have to do this. You can help by leaving your copy of this book on coffee tables in HCI conference venues, with Part V of Chap. 14 helpfully highlighted.

Maurits

I am a social scientist and researcher primarily interested in persuasion, quantitative research methods, and optimal design. After doing my master's in (economic-) Psychology at the University of Tilburg, and doing a post-master program in User–System Interaction at the Eindhoven University of Technology, I received my Ph.D. with honors from the Eindhoven University of Technology, Eindhoven, the Netherlands. Next, I worked as a postdoctoral researcher at the Aalto school of Economics, Aalto, Finland. Afterwards, I worked for 2 years as assistant professor of Statistics and Research Methods at the University of Tilburg. During my Ph.D. I also worked as a research scientist at Philips Research, Eindhoven, the Netherlands and as a distinguished visiting scholar at the CHIMe lab of Stanford University, Stanford CA, USA.

Currently I am assistant professor in Artificial Intelligence (AI) at the Radboud University Nijmegen. Also, I am the track leader of a master track called "Web and Language." I (amongst other courses) teach a course on AI techniques on the web called "AI at the Webscale." You can find the website of the research lab that I run right here: www.nth-iteration.com.

At the end of 2012 my first "popular" (Dutch) book called "Digitale Verleiding" which, since 2015, is also available in English under the name "Persuasion Profiling" was released. I am also a founder of PersuasionAPI, "pioneers in persuasion profiling" (see www.sciencerockstars.com). The company is now owned by Webpower b.v.

My prime research interest are:

- Persuasive technologies. I focus on the real-time adaptation of the use of distinct persuasive principles in interactive technologies.
- Research methods. I study both parametric and non-parametric statistical methods, hierarchical models, and time-series.
- Online/streaming learning. I work quite a bit on how to fit hierarchical models online.
- Bandit problems. I have worked on policies for multi-armed bandit problems.
- Dynamic Adaptation. I have been involved in several attempts to model, in real-time, consumer behavior and adapt e-commerce attempts accordingly.

Obviously, my interest in research methods drove me to start editing this book. I have, throughout my studies and work, been interested in quantitative methods in diverse fields, ranging from social science, to computer science, to physics and

engineering. And, I think the social sciences can broaden their views on their methods by looking around, and by keeping a close eye to modern developments. In this book we are trying to alert readers to methods that they might not have covered in their introductory stats course, but which are of use for their research and practice.

Judy

I am a computer scientist by training, although I am now a professor in the School of Education at the University of Edinburgh. I managed to graduate from two university degrees without ever taking a course in statistics; everything I know about statistics is self-taught. This is why I ended wandering down the path of editing this book—I kept innocently searching for answers to questions which should have been simple but weren't (such as "how do you analyse Likert data?"). It's quite easy to fall down a statistical rabbit hole if you don't start with the "traditional" knowledge about which statistical techniques to use in particular cases. I first became aware that null hypothesis significance testing (NHST) was a crumbling edifice when I listened to a podcast interview with Eric-Jan Wagenmakers who used Bayesian methods to demonstrate why experimental results from a series of PSI studies were flawed. While mildly disappointed that PSI doesn't really exist, I had a new rabbit hole to explore. But although I understood the criticisms of NHST and was interested to try new analytic techniques, I couldn't find many examples of such analyses in the HCI literature. I teamed up with Maurits, and we set out to draw together a book that would help other HCI researchers like me to apply appropriate statistical techniques to the sort of research problems we encounter daily. This is the book I wish I had beside me when struggling with countless analyses. I can attest to the fact it is useful: since getting initial chapter drafts from the authors, I have referred to them on many occasions already. I hope it is useful to you too.

Who Is This Book For?

This book is for Human−Computer Interaction (HCI) researchers who want to get better at making sense of quantitative data in a rigorous and thoughtful way. It aims both to critically reflect on current statistical methods used in HCI, and to introduce a number of novel methods to the HCI audience. Throughout the book HCI examples are used. However, we sincerely hope the book will be of use to a wider audience: this book, as far as we know, is one of the first attempts to bundle together non-introductory-course methods in a critical yet usable fashion.

It is not an introductory textbook—we assume you have a basic grasp of probability and commonly used NHST techniques such as t-tests, analysis of variance, and regression. We also assume that you have enough of a background in programing (or motivation to learn it) not to be intimidated by [R] code.

The Structure of the Book

The book opens with our own introduction which explains why the statistical methods we choose matter so much, and how researchers in other fields have moved away from blind reliance on the familiar null hypothesis significance testing (NHST) framework. It identifies some of the most common misunderstandings about NHST and often encountered misapplications of these methods. These themes are picked up again in the discussion Chaps. 13 and 14 at the end of the book.

In Part I (*Getting Started With Data Analysis*), the authors introduce the [R] environment (Chap. 2), explain how to visualize data to gain an intuitive understanding of your dataset before embarking on further analysis, and to illustrate your argument clearly in publications (Chap. 3). In the initial stages of the analysis, it is sensible to consider what to do about missing data—Chap. 4 explains how. Hint: you will always have more missing data than you expect and the solution is not to sweep it under the carpet.

Part II (*Classical Null Hypothesis Significance Testing Done Properly*) takes the pragmatic view that as many researchers will continue to use NHST, we might as well focus on applying such methods correctly and them interpreting them meaningfully. The issues of effect size and power are discussed in Chap. 5, highlighting the inconvenient truth that you generally need more participants than you want to find and that the size of an effect is more useful for interpreting the real-world significance of your results than a gloriously small p-value. Chapter 6 covers techniques for handling time with repeated measures analysis of variance and event history analysis. The importance of checking the assumptions underlying the tests you employ is emphasized in Chap. 7, along with a useful guide to the appropriate use of non-parametric tests.

We venture into the world of *Bayesian Inference* in Part III. Chapter 8 introduces the concepts of Bayesian reasoning and how they can be usefully applied to help us interpret data, while Chap. 9 illustrates how to quantify the strength of evidence for multiple competing hypotheses. This will equip you with the tools to weigh up the merits of alternative substantive hypotheses rather than examining the rather weak and watery hypothesis that the means of two groups are identical.

In Part IV we address some techniques for *Advanced Modeling in HCI*. Chapter 10 introduces latent variable analysis and structural equation modeling which can be used to infer properties of variables which cannot be directly measured. Chapter 11 discusses generalized linear mixed models, link functions and how to deal with data with a nested structure. In Chap. 12, techniques for latent class analysis and mixture models are explained with running examples.

Part V is a reflection on *Improving Statistical Practice in HCI*. Chapter 13 is a clear and well-argued call to use estimation methods to support fair statistical communication. The author offers many tips that researchers can follow to improve the clarity of their reporting. The concluding chapter by the editors draws together the chapters in the book and how the techniques presented here—and other novel

techniques which we did not have the space to cover—could address some of the problems pervading the use of statistical methods in social sciences. To allay any doubts of those who believe that HCI is immune to the methodological failings of other fields, we critically analyze the methods used in eight case studies of top cited quantitative studies in our field. We consider changing attitudes to quantitative methods in HCI and conclude with some recommendations for authors, reviewers, and journal editors which we hope will help in improving the clarity and fairness of statistical reporting in our field.

The Sample Dataset: The Mango Watch

We decided that it would be helpful to have running examples relating to HCI, as presently this is hard to find in the literature. A shared dataset would provide some continuity between chapters and a link to topics of interest to HCI researchers. We were unable to find an open-source dataset which suited this purpose, so we provide a simulated dataset from the following hypothetical scenario:

> *A company has recently invested in smart watches (known as the Mango) for all sales executives. They have tasked the world famous UX expert Professor Houdini and his team to evaluate the impact of the Mango on productivity among the sales staff. They firmly believe that having email on one's wrist*[1] *will cause faster response times to customers, and therefore more sales. The professor, at long last unfettered by constraints on resources imposed by measly academic research grants, has come up with the following data collection plan.*
>
> *Data will be collected for Sales Team A over a period of 3 months after the introduction of Mangos. Control data will be collected from members of Sales Team B over the same period—this group will not yet be equipped with Mangos. However, they will receive their Mangos in month 4, giving the researchers both between subjects (A compared to B) and within subjects (pre-Mango B compared to post-Mango B) data.*
>
> The measures used are:
>
> - A 7 point SUS-like usability scale about experience of using Mango (ordinal).
> - Average response time to answer customer emails (ratio/interval).
> - Sales Team A and B staff are distributed over several geographical regions:
> - Sales Team A: Aachen, Abik, Aberdeen,
> - Sales Team B: Babol, Bakersfield, Barrie.

More details of the [R] code which generated the dataset can be found in the supplementary materials to Chap. 2. You will find that the chapter authors have extended and changed the scenario to better illustrate their techniques. For most of the chapters online supplements are available (http://extras.springer.com). If the authors deviate from, or extend the initial Mango watch dataset they either use [R] code—presented in the chapter text—to generate their data, or their datasets are

[1] It is company policy that Mangos must be worn at all times, even in bed.

available in the supplementary materials. As the code generates the data randomly, the data sets used by different authors will differ.[2] For this reason, we don't recommend trying to make sense of the experimental findings across chapters. Sadly, this book will tell you nothing of value about the impact of fruit-based smart watches on the performance of sales executives. However, the Mango watch scenario does offer a context that is shared between all chapters which saves the authors the effort (and space) to introduce novel settings. Needless to say, the editors remain open to funding offers from such companies who do wish to investigate fruit-based smart watch performance.

Our Approach

We chose to use [R] as the language for conducting the analyses in the book because it a widely used, free, open-source platform which evolves fast enough to keep up with new techniques as they emerge. We realize that [R] will introduce a learning curve for some readers (but in Judy's recent personal experience it is worth investing the time to make the switch from a graphical package). We have asked the chapter authors to explain the statistical concepts in plain language, and to introduce the supporting mathematics where necessary. You will find the finer mathematical details in footnotes from time to time, in places where Maurits' mathematical conscience has been sufficiently troubled. Above all, the authors have made a real effort throughout to help the reader to interpret the numerical results in the context of a research problem, as you will need to do yourself when making sense of your data for publication.

This book might be disappointing for some of you: in many ways the book will lead to more questions than answers. However, this is inherent in our subject: there is no single best method to use to analyze your data. Many methods could be considered, tried, and possibly combined. You may notice that the authors contradict each other from time to time, and that different analytical techniques lead to quite different interpretations of the data. This is a wider point of the book: we must choose the techniques we use wisely, and report it wisely so that other researchers in the future are aware of our assumptions and are informed to make their own interpretations.

It will become obvious as you read through the volume that the method of analysis really does matter. The decisions made by the analyst make a difference to the results and how they are interpreted. Interpreting quantitative data is more prone to subjectivity than you might think. It is better to acknowledge this fact and deal with it sensibly, than to hide it in a clutter of procedural use of "accepted" statistical methods.

[2]Some authors explicitly mention their seeds for reproducibility. Others provide their dataset in the supplementary materials.

Acknowledgements

We sincerely thank the chapter authors for all their patience, effort, and generosity in making this book work. We would also like to thank Shumin Zhai for his comments on our earlier writings and suggesting that we should edit a collection on the topic.

http://www.springer.com/978-3-319-26631-2

Modern Statistical Methods for HCI
Robertson, J.; Kaptein, M. (Eds.)
2016, XX, 348 p. 77 illus., Hardcover
ISBN: 978-3-319-26631-2