Getting Started

Human–Computer Interaction (HCI) is concerned with understanding the relationship between people and interactive digital systems. It draws on diverse disciplines, in particular, psychology, computer science, social sciences and, increasingly, design, creative arts and critical theory. There are therefore occasional existential crises in HCI as to whether it is a science, a design discipline, a technical discipline, a social science or all of these or none of them (Reeves, 2015). Regardless of what HCI is or should be, it is clear that experiments, quantitative data and, therefore, statistics are regularly used to advance our understanding of how people experience, use and are used by digital systems.

However, many of us who set out to do HCI did not also set out to become statisticians. Typically, we have arrived at statistics, and the experimental methods that go with them, as the tools we need to tackle the research questions that interest us. Fortunately, there are many good textbooks and courses, particularly coming from psychology, that are well suited to supporting such HCI research, my personal favourites being Howell (2016) (currently in its 9th edition) and Pagano (2012). The problem is that textbook descriptions and examples are not really like the messy business of acquiring quantitative data and then trying to work out what to do with it. The fledging researcher very soon meets the reality of data that is not normal or has unexpected outliers and a plethora of tests that might (or might not) help. The Internet seems to help with useful comforting advice like t-tests are robust to deviations from assumptions and other such folklore. But then the Internet also suggests that left-handed haggis are dying out so you cannot always believe what you read on the Internet.

At this point, many researchers need a more experienced practitioner to talk through the specific questions they have about their data or the data they would like to collect. For many years, I have been teaching statistics and so I have ended up as just such an experienced person who people turn to. On the whole, the questions that people ask boil down to seeking reassurance and the general question of 'Am I doing this right?' Unfortunately, the answer is more often than not 'It depends'. Which of course begs the question of what it depends on. The purpose of this book is to provide a resource for researchers that hopefully addresses some of their questions about what might be the right way to gather and analyse data. What I am very keen to avoid, though, is simply providing a new source of folklore to compete with Internet 'wisdom'. For this reason, I have deliberately taken a quite scholarly approach of digging into the questions and trying to show where any existing folklore comes from, what the debate around current thinking is and, where possible, propose some possible answers or at least ways forward. What I rarely offer is definitive answers.

At the end of the day, when it comes to data collected with people and the uncertainty inherent in such data, it is very hard to have clear answers to statistical questions that provide the reassurance that people seek. The goal therefore is not to provide a definitive statement about what is the right way to analyse data but instead to help guide the reader to do a bit better and to be, at least, explicit about the compromises that have to be made.

Who Should Read This Book?

This book is intended for anyone who is about to collect or about to analyse quantitative data for an HCI study and is worried about whether they are doing it right. Understandably, in such data, you might have a whole host of questions and this book could not pretend to answer them all. Instead, I have tried to address the questions that I am asked most often or, in some cases, ones that I think HCI researchers should be asking more often.

As such, the reader is expected to be in a position to do HCI research but this could be anyone from a first-year undergraduate doing an evaluation to an experienced researcher meeting a new problem in their data. However, if you are undertaking a research project then I would hope you have read at least one book about planning quantitative studies, such as Harris (2008) and another about the standard statistical methods, such as Howell (2016). There are other excellent textbooks like these. You may also have had a course in experimental methods and statistics, which would be another very good place to start.

What I am not expecting is that you necessarily have, what might be called, advanced knowledge. On the whole, as discussed with reference to simplicity (Chapter 5), I do not think more sophisticated statistics are the solution to most (or possibly even any) of the problems that people ask me about. For this reason, the textbooks that I recommend are more or less accessible to anyone entering higher education. That is, your level of background knowledge should not be a bar to getting something out of any of the chapters in this book though you may have to do some homework to really get to grips with some parts.

Hopefully, right now, you will only be at the planning stage of your study. Some of the questions in this book, for instance about how many options a Likert item should have (Chapter 15), will help you to design your study well. In reality, I know that some of you will be reading this having designed an experiment, gathered the data and now are stuck trying to analyse it. I cannot promise that these essays will help you retrieve something useful from your hard work but they may help you navigate some of the most common problems such as the accusation of fishing (Chapter 13) or how to deal with outliers (Chapter 8). But don't do it again! And read the chapter about planning your analysis (Chapter 5).

This book could also be of use to researchers outside of HCI. However, the examples and practical concerns are from the context of my own research field and, in many cases, my own research experience. At the same time, the questions addressed could be relevant to any researcher considering the quantitative analysis of data arising from human participants. Thus, this book should be of use to you if you come from the psychological and behavioural sciences though some of the examples might not be so illuminating.

The Structure of the Book

As will become clear from this book, I do not see statistics as an end in themselves. Statistics are one part of a larger toolset that helps us to answer research questions. Statistics are inherently interwoven with the studies that generate the data to be analysed. The studies in turn are deeply connected with the new knowledge we seek. How knowledge, studies and statistics work together to form science is a matter of philosophy. While it might be simpler to say that such philosophy is irrelevant to practical statistics, I think one of the key things I am trying to say throughout this book is that you cannot really understand why statistics is done the way it is without thinking about why studies are done the way they are. For this reason, Part I of this book offers the philosophical backdrop to the more practical answers offered in Part II. I do not frequently encounter the questions that the chapters in Part I address but I do find that, regularly, the answers that I give about other questions need these more philosophical explanations. One of the best examples of this is the question 'Why can't I do all the statistical tests on my data that I like? After all, it does not change the data.' This seems reasonable but it flies in the face of how statistical tests work in concert with the experiment that generated the data being tested (Chapter 1). A knowledgeable person might then come back at me and say that that is not a problem for Bayesian statistics, it is just a flaw in how we traditionally do statistics. But that's not true! Unfortunately, not many explanations of Bayesian statistics actually address this issue properly (Chapter 3).

It is not essential that every reader read every chapter of Part I before tackling any chapter in Part II. In fact, you could probably safely launch straight into any of the chapters in Part II. But it just might be that to really answer the question you have, you will need to go back to the fundamentals in Part I of why we do statistical tests the way we do. Where I have anticipated this, I have indicated it in the chapters.

Each chapter in Part II is intended to stand on its own and is motivated by particular practical questions or issues relevant to gathering or analysing quantitative data in HCI. As such, each of these chapters addresses a set of questions around a topic, such as what do if data are not normal, and these questions are listed at the start of each chapter. If your question is listed then hopefully the chapter will help you to think about possible solutions to your problems. Ideally, you would not need to read several chapters at once: you should get some useful answers to the typical questions without reading further. In addition, each chapter is deliberately intended to be a short essay that targets the motivating questions so that even if you do not find the answers you seek, you will not have spent too long finding that out.

What you will not find are definitive answers. In some cases, there are answers that are best practice but even they are not necessarily the last word. I think the key to doing statistics well is to keep learning, to recognise the limitations of what we know and what we can do and to try to work out how we might do things a bit better. Definitive answers tend to give the impression that there is nothing more to learn so I do not try to give any.

Of course, in trying to answer some questions, we are led to further questions, and discussions of one issue may simultaneously address other issues. Thus, you will find in each chapter links to the topics and issues addressed in other chapters. Some chapters even thread together to make coherent units that would provide a broader background to more general topics in statistical methods. For example, some topic threads are:

1. Using modern robust statistics: Chapters 4, 5, 7, 10 and 11.
2. Problems with distributions or the shape of data: Chapters 7, 8 and 14.
3. Developing questionnaires: Chapters 15, 16 and 17.
4. Choosing suitable tests: Chapters 4, 5, 6 and 18.

I should emphasise that I chose the topics of the chapters to address the questions that I regularly get asked. As such, there is a degree of selectivity and possibly even idiosyncrasy in the topics of these essays. I do not aim, like the father of essays, Montaigne (1958), to be the 'matter of my book' but at the same time I recognise that the topics covered and the approach to those topics are based on my own particular experiences. For that reason, I do present my opinions but I will try to clearly indicate wherever it is my opinion in the absence of more authoritative viewpoints and also I am happy to write in the first person. These are my essays: my attempts to communicate statistics my way. The final test of the success of this approach is with you the reader. If you like this approach, let me know. If you do not, also let me know! And if you have other questions that this book does not answer, then I would be happy to hear from you and to see if I can help.

Statistics Software

Though occasionally it may be informative to perform the calculations of a statistical test with pencil and paper, in practice all statistical testing is done using statistics software in one form or another. Indeed, it should be this way: the most careful hand calculation can still have errors in it even with quite small datasets. Statistics software will not help you to do the right calculation but once you are doing the right calculation, the software will help you to keep doing it right.

Unfortunately, there is a plethora of software options for the practising researcher to use. These include:

- Well-established commercial packages such as SPSS, SAS and MiniTab
- Open source options like R and the statistics.py library for Python
- Function libraries as part of other data and mathematical software packages such as MatLab, Mathematica and Excel
- Bespoke online web apps, some of essentially unknown provenance and quality.

Because this book is not intended to be a textbook of statistics, I had hoped to side-step the issue of which package is best and in doing so avoid any ideological battles about what makes a good package. Part of my reasoning was also that it is a topic that I have found hard to care about. It seemed to me, as an HCI researcher, that your choice of statistical package boiled down to choosing the one that annoyed you least. They will all annoy you to some extent, through a diverse range of usability problems, but you will be able to tolerate some better than others depending on your particular experience and disposition. However, in writing this book and researching some of the most modern approaches, it is clear that R is becoming an important package that is starting to dominate statistical software.

There may be interesting cultural and disciplinary reasons for this. For instance, R is open source, which makes it easy for people to obtain and work with, unlike commercial software, which may not only require a licence but one that requires regular renewal. My own story is that originally I used SPSS because it was commonly used and available in the psychology department where I started teaching statistics. However, with each SPSS version that came out it seemed to get slower and slower and at the same time introduce more and more interface and interaction inconsistencies and problems. Even now I am still not sure what I will find under the SPSS menu option 'General Linear Model' as opposed to the 'Generalized Linear Model' and reliably finding a Mann–Whitney test is a thing of the past.

In recent years, I have therefore looked for other options and R has been useful, allowing a level of control and consistency that SPSS did not offer and at the same time ensuring that my students could do at home exactly what they had seen done in classes or labs. Moreover, it was free!

But these are not my reasons for recommending R. My main reason is that it became clear in my reading for this book that statisticians are using R as a way to embody their statistical knowledge. Whereas previously a statistician might outline a calculation or function in a paper and the reader would have to work out how to perform the calculation in practice, that outline is now backed up by R code, both as a listing and a downloadable package. The result is that you can apply it almost immediately if you are using R. A particularly important example of this is the `WRS2` package developed by Mair and Wilcox (Available from: cran.r-project.org). This provides many of the statistical analysis functions that Wilcox promotes as best practice in modern statistics (Wilcox, 2017), including robust versions of the *t*-test and ANOVA that I discuss in Chapters 11 and 12.

This is not to say R is perfect. Far from it. Its language is a computer scientist's nightmare, being a bastardisation of list-processing, weakly-typed, object-oriented and procedural paradigms with unflagged case-sensitivity thrown in for good measure. As I said, all statistics packages are annoying. Nonetheless, it is the environment where cutting-edge, best-practice statistics are being made available.

I would therefore strongly recommend that you get familiar with R so that you are in a position to use the best techniques and methods, not only those that are currently out there but those that may be developed in future. I tend to use R through the RStudio interface because that just helps to organise scripts, output and workspace. There are good online resources and textbooks to help the new user. I've found Quick R (www.statmethods.net) useful because that assumes you have good experience of statistics and just want to transfer that over to R. I have also found Matloff (2011) useful to help make sense of the almost arbitrary semantics of the R language. And if you are coming to R and statistics afresh then I can also recommend the R version of Andy Field's book (Field, Miles, and Field, 2012) and if you like your statistics textbooks to have a healthy dose of steampunk you might even like Field (2016). Learning R may not be straightforward but it is worth the effort.

References

Field, Andy (2016). *An Adventure in Statistics: The Reality Enigma*. Sage.

Field, Andy, Jeremy Miles and Zoe Field (2012). *Discovering Statistics Using R*. Sage.

Harris, Peter (2008). *Designing and Reporting Experiments in Psychology*. McGraw-Hill Education.

Howell, David C. (2016). *Fundamental Statistics for the Behavioral Sciences*. Nelson Education.

Matloff, Norman (2011). *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.

Montaigne, Michel de (1958). *Complete Essays of Montaigne*; Trans. by Donald M. Frame. Stanford University Press.

Pagano, Robert R. (2012). *Understanding Statistics in the Behavioral Sciences*. Cengage Learning.

Reeves, Stuart (2015). 'Human-computer interaction as science'. *Proceedings of the Fifth Decennial Aarhus Conference on Critical Alternatives*. Aarhus University Press, pp. 73–84.

Wilcox, Rand R. (2017). *Introduction to Robust Estimation and Hypothesis Testing*. 4th edn. Academic Press.