# Preface {-}

## History and License {.unnumbered .unlisted}

This book is an adaptation of DJ Navarro (2018). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6). [https://learningstatisticswithr.com/](https://learningstatisticswithr.com/).

The book is released under a [creative commons CC BY-SA 4.0 licence](https://creativecommons.org/licenses/by-sa/4.0/). This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license - CC BY-SA.

## Preface to Version 0.75 {.unnumbered .unlisted}

In this version we have updated the figures, images and text to maintain

compatibility with latest versions of jamovi (2.2); many thanks to Peter

Fisk for his help with this. Also tweaked and corrected are a few

sections where improvements have been suggested by readers. This has

mainly included fixing typos but also in places correcting conceptual

detail, for example we have updated the information on kurtosis to

reflect that it isn’t really about distribution “pointiness” and instead

kurtosis is about whether data distributions have thin or fat tails.

Thanks to all the readers who made suggestions, either through

contacting me by email, or raising an issue on github.

\*David Foxcroft

February 9th, 2022\*

## Preface to Version 0.70 {.unnumbered .unlisted}

This update from version 0.65 introduces some new analyses. In the ANOVA

chapters we have added sections on repeated measures ANOVA and analysis

of covariance (ANCOVA). In a new chapter we have introduced Factor

Analysis and related techniques. Hopefully the style of this new

material is consistent with the rest of the book, though eagle-eyed

readers might spot a bit more of an emphasis on conceptual and practical

explanations, and a bit less algebra. I’m not sure this is a good thing,

and might add the algebra in a bit later. But it reflects both my

approach to understanding and teaching statistics, and also some

feedback I have received from students on a course I teach. In line with

this, I have also been through the rest of the book and tried to

separate out some of the algebra by putting it into a box or frame. It’s

not that this stuff is not important or useful, but for some students

they may wish to skip over it and therefore the boxing of these parts

should help some readers.

With this version I am very grateful to comments and feedback received

from my students and colleagues, notably Wakefield Morys-Carter, and

also to numerous people all over the world who have sent in small

suggestions and corrections - much appreciated, and keep them coming!

One pretty neat new feature is that the example data files for the book

can now be loaded into jamovi as an add-on module - thanks to Jonathon

Love for helping with that.

\*David Foxcroft

February 1st, 2019\*

## Preface to Version 0.65 {.unnumbered .unlisted}

In this adaptation of the excellent ‘Learning statistics with R’, by

Danielle Navarro, we have replaced the statistical software used for the

analyses and examples with jamovi. Although R is a powerful statistical

programming language, it is not the first choice for every instructor

and student at the beginning of their statistical learning. Some

instructors and students tend to prefer the point-and-click style of

software, and that’s where jamovi comes in. jamovi is software that aims

to simplify two aspects of using R. It offers a point-and-click

graphical user interface (GUI), and it also provides functions that

combine the capabilities of many others, bringing a more SPSS- or

SAS-like method of programming to R. Importantly, jamovi will always be

free and open - that’s one of its core values - because jamovi is made

by the scientific community, for the scientific community.

With this version I am very grateful for the help of others who have

read through drafts and provided excellent suggestions and corrections,

particularly Dr David Emery and Kirsty Walter.

\*David Foxcroft

July 1st, 2018\*

## Preface to Version 0.6 {.unnumbered .unlisted}

The book hasn’t changed much since 2015 when I released Version 0.5 –

it’s probably fair to say that I’ve changed more than it has. I moved

from Adelaide to Sydney in 2016 and my teaching profile at UNSW is

different to what it was at Adelaide, and I haven’t really had a chance

to work on it since arriving here! It’s a little strange looking back at

this actually. A few quick comments…

- Weirdly, the book consistently misgenders me, but I suppose I have

only myself to blame for that one :-) There’s now a brief footnote

on page 12 that mentions this issue; in real life I’ve been working

through a gender affirmation process for the last two years and

mostly go by she/her pronouns. I am, however, just as lazy as I ever

was so I haven’t bothered updating the text in the book.

- For Version 0.6 I haven’t changed much I’ve made a few minor changes

when people have pointed out typos or other errors. In particular

it’s worth noting the issue associated with the etaSquared function

in the lsr package (which isn’t really being maintained any more) in

Section 14.4. The function works fine for the simple examples in the

book, but there are definitely bugs in there that I haven’t found

time to check! So please take care with that one.

- The biggest change really is the licensing! I’ve released it under a

Creative Commons licence (CC BY-SA 4.0, specifically), and placed

all the source files to the associated GitHub repository, if anyone

wants to adapt it.

Maybe someone would like to write a version that makes use of the

tidyverse… I hear that’s become rather important to R these days :-)

Best,

\*Danielle Navarro\*

## Preface to Version 0.5 {.unnumbered .unlisted}

Another year, another update. This time around, the update has focused

almost entirely on the theory sections of the book. Chapters 9, 10 and

11 have been rewritten, hopefully for the better. Along the same lines,

Chapter 17 is entirely new, and focuses on Bayesian statistics. I think

the changes have improved the book a great deal. I’ve always felt

uncomfortable about the fact that all the inferential statistics in the

book are presented from an orthodox perspective, even though I almost

always present Bayesian data analyses in my own work. Now that I’ve

managed to squeeze Bayesian methods into the book somewhere, I’m

starting to feel better about the book as a whole. I wanted to get a few

other things done in this update, but as usual I’m running into teaching

deadlines, so the update has to go out the way it is!

\*Danielle Navarro

February 16, 2015\*

## Preface to Version 0.4 {.unnumbered .unlisted}

A year has gone by since I wrote the last preface. The book has changed

in a few important ways: Chapters 3 and 4 do a better job of documenting

some of the time saving features of Rstudio, Chapters 12 and 13 now make

use of new functions in the lsr package for running chi-square tests and

t tests, and the discussion of correlations has been adapted to refer to

the new functions in the lsr package. The soft copy of 0.4 now has

better internal referencing (i.e., actual hyperlinks between sections),

though that was introduced in 0.3.1. There’s a few tweaks here and

there, and many typo corrections (thank you to everyone who pointed out

typos!), but overall 0.4 isn’t massively different from 0.3.

I wish I’d had more time over the last 12 months to add more content.

The absence of any discussion of repeated measures ANOVA and mixed

models more generally really does annoy me. My excuse for this lack of

progress is that my second child was born at the start of 2013, and so I

spent most of last year just trying to keep my head above water. As a

consequence, unpaid side projects like this book got sidelined in favour

of things that actually pay my salary! Things are a little calmer now,

so with any luck version 0.5 will be a bigger step forward.

One thing that has surprised me is the number of downloads the book

gets. I finally got some basic tracking information from the website a

couple of months ago, and (after excluding obvious robots) the book has

been averaging about 90 downloads per day. That’s encouraging: there’s

at least a few people who find the book useful!

\*Danielle Navarro

February 4, 2014\*

## Preface to Version 0.3 {.unnumbered .unlisted}

There’s a part of me that really doesn’t want to publish this book. It’s

not finished.

And when I say that, I mean it. The referencing is spotty at best, the

chapter summaries are just lists of section titles, there’s no index,

there are no exercises for the reader, the organisation is suboptimal,

and the coverage of topics is just not comprehensive enough for my

liking. Additionally, there are sections with content that I’m not happy

with, figures that really need to be redrawn, and I’ve had almost no

time to hunt down inconsistencies, typos, or errors. In other words,

this book is not finished. If I didn’t have a looming teaching deadline

and a baby due in a few weeks, I really wouldn’t be making this

available at all.

What this means is that if you are an academic looking for teaching

materials, a Ph.D. student looking to learn R, or just a member of the

general public interested in statistics, I would advise you to be

cautious. What you’re looking at is a first draft, and it may not serve

your purposes. If we were living in the days when publishing was

expensive and the internet wasn’t around, I would never consider

releasing a book in this form. The thought of someone shelling out $80

for this (which is what a commercial publisher told me it would retail

for when they offered to distribute it) makes me feel more than a little

uncomfortable. However, it’s the 21st century, so I can post the pdf on

my website for free, and I can distribute hard copies via a

print-on-demand service for less than half what a textbook publisher

would charge. And so my guilt is assuaged, and I’m willing to share!

With that in mind, you can obtain free soft copies and cheap hard copies

online, from the following webpages:

Soft copy:

[www.compcogscisydney.com/learning-statistics-with-r.html](https://www.compcogscisydney.com/learning-statistics-with-r.html)

Hard copy:

[www.lulu.com/content/13570633](https://www.lulu.com/content/13570633)

[\*\*\*Ed: these links are defunct, try this instead:

[learningstatisticswithr.com](https://learningstatisticswithr.com)\*\*\*]

Even so, the warning still stands: what you are looking at is Version

0.3 of a work in progress. If and when it hits Version 1.0, I would be

willing to stand behind the work and say, yes, this is a textbook that I

would encourage other people to use. At that point, I’ll probably start

shamelessly flogging the thing on the internet and generally acting like

a tool. But until that day comes, I’d like it to be made clear that I’m

really ambivalent about the work as it stands.

All of the above being said, there is one group of people that I can

enthusiastically endorse this book to: the psychology students taking

our undergraduate research methods classes (DRIP and DRIP:A) in 2013.

For you, this book is ideal, because it was written to accompany your

stats lectures. If a problem arises due to a shortcoming of these notes,

I can and will adapt content on the fly to fix that problem.

Effectively, you’ve got a textbook written specifically for your

classes, distributed for free (electronic copy) or at near-cost prices

(hard copy). Better yet, the notes have been tested: Version 0.1 of

these notes was used in the 2011 class, Version 0.2 was used in the 2012

class, and now you’re looking at the new and improved Version 0.3. I’m

not saying these notes are titanium plated awesomeness on a stick –

though if you wanted to say so on the student evaluation forms, then

you’re totally welcome to – because they’re not. But I am saying that

they’ve been tried out in previous years and they seem to work okay.

Besides, there’s a group of us around to troubleshoot if any problems

come up, and you can guarantee that at least one of your lecturers has

read the whole thing cover to cover!

Okay, with all that out of the way, I should say something about what

the book aims to be. At its core, it is an introductory statistics

textbook pitched primarily at psychology students. As such, it covers

the standard topics that you’d expect of such a book: study design,

descriptive statistics, the theory of hypothesis testing, t-tests, χ 2

tests, ANOVA and regression. However, there are also several chapters

devoted to the R statistical package, including a chapter on data

manipulation and another one on scripts and programming. Moreover, when

you look at the content presented in the book, you’ll notice a lot of

topics that are traditionally swept under the carpet when teaching

statistics to psychology students. The Bayesian/frequentist divide is

openly disussed in the probability chapter, and the disagreement between

Neyman and Fisher about hypothesis testing makes an appearance. The

difference between probability and density is discussed. A detailed

treatment of Type I, II and III sums of squares for unbalanced factorial

ANOVA is provided. And if you have a look in the Epilogue, it should be

clear that my intention is to add a lot more advanced content.

My reasons for pursuing this approach are pretty simple: the students

can handle it, and they even seem to enjoy it. Over the last few years

I’ve been pleasantly surprised at just how little difficulty I’ve had in

getting undergraduate psych students to learn R. It’s certainly not easy

for them, and I’ve found I need to be a little charitable in setting

marking standards, but they do eventually get there. Similarly, they

don’t seem to have a lot of problems tolerating ambiguity and complexity

in presentation of statistical ideas, as long as they are assured that

the assessment standards will be set in a fashion that is appropriate

for them. So if the students can handle it, why not teach it? The

potential gains are pretty enticing. If they learn R, the students get

access to CRAN, which is perhaps the largest and most comprehensive

library of statistical tools in existence. And if they learn about

probability theory in detail, it’s easier for them to switch from

orthodox null hypothesis testing to Bayesian methods if they want to.

Better yet, they learn data analysis skills that they can take to an

employer without being dependent on expensive and proprietary software.

Sadly, this book isn’t the silver bullet that makes all this possible.

It’s a work in progress, and maybe when it is finished it will be a

useful tool. One among many, I would think. There are a number of other

books that try to provide a basic introduction to statistics using R,

and I’m not arrogant enough to believe that mine is better. Still, I

rather like the book, and maybe other people will find it useful,

incomplete though it is.

\*Danielle Navarro

January 13, 2013\*