

Module 0: Intro and Course Overview

Dr Ian Hunt

¹Statistician/Data Scientist
Senior Lecturer, TIA (Launceston), UTAS
ian.hunt@utas.edu.au

May 2, 2022

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping

- 1 Be realistic, but optimistic, about using statistics and probability reasoning.
- 2 Use a computer to properly analyse scientific data (which entails going beyond Microsoft Excel).
- 3 Make justifiable and reproducible statistical inferences using real data.
- 4 Plan for the coherent analysis of data in your future studies, careers and life.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping

Be realistic

- *“Methodology, like sex, is better demonstrated than discussed, though often better anticipated than experienced.”*
Leamer (1983, Lets take the con out of econometrics, *AER*, page 40)
- Analysing a real data set for research can take **hundreds of hours** i.e. weeks of work.
- Critiquing and understanding others' research is difficult.
- This course is the *start of a journey*.

Philosophy of Applied Statistics?



Figure 1: Bradley Efron.

- For over 50 years, Bradley Efron has been a leading researcher in the famous Stanford University Statistics department.
- Bootstrapping, named from Efron's favourite Baron Munchausen story, was initially treated with scepticism. We now take it for granted, though he probably doubts that we should.

"I have a feeling that statisticians are cynics, because you realise how much of the stuff that you are told is true in the world is actually just that month's accident that worked out, or that month's disaster that happened. Appreciating how much randomness there is in everyday experience helps a lot." Efron (2010)

Who am I?

- My name is Ian Hunt and I am the resident statistician at the Tasmanian Institute of Agriculture (TIA).
 - I get to poke my nose into everyone's research and data here at TIA. Vicariously, so do you!
 - I have been a statistical consultant for 20 years, and I am a Chartered Statistician (CStat) and Fellow of the Royal Statistical Society. I hold post-graduate degrees, with honours, from Université Catholique de Lille (France), the London School of Economics (London), City University (London) and Otago (New Zealand). I also have an online post-grad certification in data science from MIT (USA).
 - In 2015 I began lecturing in France (MSc level statistics). For the past few years I have been a full-time academic in Australia.
- My accent in English, French and Spanish is terrible.

I try, but I am not perfect

Some past course feedback.

- *"Almost all other courses on stats at Monash are a total waste of time. But this stats course was really good—and I'm really looking forward to the next one. Ian has a great teaching style directed at major take away issues and then encourages attendees to take the next step through further reading."*
- *"Ian clearly knew exactly what he was talking about and this was the most interest I have ever had in statistics - it has always been a guilty gap in my knowledge that I've avoided but Ian presented it in a lucid and understandable way that intrigued me rather than scared me. However, I don't think that these sessions were anywhere near enough to consider that I can now 'do' stats. It was a good introduction, but this is an entire discipline and I need clear guidance to wade through it."*
- *"Further to Ian's question about whether there were any students with maths-anxiety (and poor schooling) present – I'm one! As such, I thought it might be useful to know that I found the sessions really informative and the readings very enlightening (though they were very slow going due to my lack of familiarity with the concepts). In particular the p-value paper really got me understanding that it isn't as hard as it sounds. However, because the ideas were so unfamiliar, it took me close to the whole day to read ... I thought it might be uncommon for students to admit to how challenging some of this material can be – or maybe to the level of deficit some of us have."*
- *"While the lecturer was very knowledgeable, for someone who has no foundation in statistics, this introductory course was a mess, with the lecturer assuming that the students have a lot of assumed knowledge and understandings all his technical terms. In additions, the lecturer spoke way too fast, was mumbling, had confusing slides, and did not explain any of the terms that he used in a way that actually made sense for someone taking a 101 course!"*

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping

- There are two one hour lectures per week, which are online.
- But three of the lectures are designated as “keynote”. These will hopefully be in-person and I will invite some of TIAs researchers to present their statistics.
- There is one two-hour tutorial per week. These are hopefully all in-person, depending on COVID 19 restrictions.

The lectures are based on a book, which I expect you to purchase.

- In particular, we are reading this book in class: Spiegelhalter (2019),



- On average, you will need to read one chapter of the book each week.
- Buy the book online as soon as possible. It costs less than AUD\$20. There is an ebook available on Amazon or a paperback.

- You must read the book, attend the lectures and participate in classroom discussions.
- Each week *during class* I will compile a **short summary** of what we have learned from the book and how this relates specifically to agrifood research.
- In the first hour of lecture time each week we will go over the chapter contents for that week.
- The second hour of lecture time will be largely devoted to interactively building the summary.
- It is going to be particularly important for the quality of the weekly summaries, on which a substantial portion of the exam will be about, that there is a detailed **class discussion and interaction** as we sum up. **You should take your own notes too.**

The tutorials (or so-called “practicals”) will be based on a **set of notes** written by me.

- There is an entire section on coding. This uses a comprehensive set of notes on using R.
- There are three other “modules”.
- The idea is that we will collectively work through all the notes in class, with our computers open. There will be much code writing, data manipulating and real-time result interpretation.
- Keeping up is going to be a challenge, especially if we are forced online.

My philosophical approach to teaching

There's no question that literacy and pedagogy must evolve alongside technology. It's possible to recognise this while also defending an educational humanism – with a small 'h' – that values the time and space given to a young person to mess around, try things out, make mistakes, have a say, and not immediately find out what score they've got as a result. It has become clear, as we witness the advance of Panopto, Class Dojo and the rest of the EdTech industry, that one of the great things about an old-fashioned classroom is the facilitation of unrecorded, unaudited speech, and of uninterrupted reading and writing.

How many words does it take to make a mistake? William Davies on the mechanisation of learning. *London Review of Books*. Vol. 44, No. 4, 24 February 2022.

There are several assignments and an exam.

- The first assignment will require computer programming with R, alongside Microsoft Excel.
- The second assignment will be on statistical inference, and will require the use of R.
- The assignments may be broken into several pieces, to be handed in separately.
- The exam will cover the material from both the lectures and tutorials.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research**
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping

- There will be a gentle introduction to R and RStudio, using a tried-and-tested set of notes with examples to type in.
- The notes are suitable for complete beginners and those with a background in coding other languages.
- *Using* R will be essential to learning the statistical concepts in this course.
- Our R coding will not be done in one big block — we will be using the R notes as a reference the whole time, as we work through the notes from other modules.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics**
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping

Computing and Statistics module outcomes

- Understand the basics of “statistical computing”.
- Select the right software or language for your research and data.
- Avoid time-wasting, needless complexity and dirty habits.
- Recognise some good ideas.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance**
- 6 Module 3: Data Analysis
- 7 House keeping

- Understand the **foundations** of statistics and sampling variance.
- Run **useful hypothesis tests**.
- **Understand** p-values, test size, test power and confidence intervals.
- Use “statistical analysis” for **research and critical thinking**.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis**
- 7 House keeping

Data Analysis module outcomes

- Recognise some good ideas (including those related to statistical summaries and charts).
- Summarise and explore continuous data.
- Summarise and explore categorical data.
- Assess interesting estimates, models and hypotheses.

Table of Contents

- 1 Introduction and reality check
- 2 Course structure
- 3 Using R for statistical research
- 4 Module 1: Computing and Statistics
- 5 Module 2: Estimation and Sampling Variance
- 6 Module 3: Data Analysis
- 7 House keeping**

Official “intended learning outcomes” ...

- The official intended learning outcomes are those from 2021.
- The assessment task dates are not set-in-stone.
- Whilst it was too late to change everything when I took over for 2022, course outcomes and assessments are essentially the same as explained in the official “course outline”.

Overall course outcomes ...

- 1 Be realistic, but optimistic, about using statistics and probability reasoning.
- 2 Use a computer to properly analyse scientific data (which entails going beyond Microsoft Excel).
- 3 Make justifiable and reproducible statistical inferences using real data.
- 4 Plan for the coherent analysis of data in your future studies, careers and life.

Efron, B. (2010). Editorial interview. *Significance (Royal Statistical Society)*.

Leamer, E. (1983). Let's take the con out of econometrics. *The American Economic Review* 73, 31–40.

Spiegelhalter, D. (2019). *The Art of Statistics: Learning from Data*. Penguin UK.