STATS 101: a modern introduction

Dr Ian Hunt[1]

[1]Manager, Statistical Consulting Service
Data Science and AI Platform
Monash University
ihunt@bunhill.co.uk; ian.hunt@monash.edu; statisticalconsulting@monash.edu

November 13, 2020

1 Introduction and reality check

2 Session 1: Estimation and Sampling Variance

3 Session 2: Inference

4 Session 3: Data Analysis

Course outcomes

1. Be realistic.

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 ahead for the statistical analysis of data for journal articles.

Table of Contents

Recognise some good ideas

- Data summaries come before complex models.

- Lump and split data groups dynamically.

- Be flexible with "data types".

- Keep naming conventions simple and recognisable.

- Comment on any code and analysis steps that you take.

- Re-use your code and functions.

- Start with simple analysis and *then* add complexity.

- Test code to ensure you get the right answers.
  - Compare your R code with what you get in SPSS, for example.
  - Use fake data or small subsets for testing.

Table of Contents

Topics and resources (Session 1)

- Gentle introduction to R and RStudio
  - *R for Statistics* notes (especially chapter 2).

- Overview of hypothesis testing
  - *Statistical Inference for Research* notes (chapter 2).

- Random variables and probability.
  - R code developed in class.

- Simulation and resampling.
  - R code developed in class and *Statistical Inference for Research* notes (chapter 1).

- Means, the central limit theorem and normal distributions.
  - R code developed in class and the example from **?**, chapter 7.

Table of Contents

Topics and resources (Session 2)

- Theoretical sampling variance.
    - *Statistical Inference for Research* (chapter 1).

- Bootstrap sampling variance
    - *Statistical Inference for Research* (chapter 1) and R code from this class.

- t-tests, p-values and confidence intervals.
    - *Statistical Inference for Research* (chapter 3) and R code from class.

- Size and Power.
    - *Statistical Inference for Research* (chapter 3) and R code from class.

- Multiple hypothesis testing.
    - *Statistical Inference for Research* (chapter 4) and R code from class.

Table of Contents

Topics and resources (Session 3)

In this session we will use the notes called *Straight-bat data analysis*.

- Data summaries.

- Histograms.

- QQ-plots.

- Boxplots and "error bars".

- Correlation measures, outliers and ranks.

- Mann-Whitney/Wilcoxon tests.

Course outcomes

1. Be realistic.

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 ahead for the statistical analysis of data for journal articles.

Bibliography I