

# Statistical Techniques for Biological and Environmental Sciences

Brad Duthie

2022-07-12



# Contents

<b>Preface</b>	<b>11</b>
0.1 What is statistics? . . . . .	11
0.2 Why this module is important . . . . .	11
0.3 Teaching overview . . . . .	11
0.4 Assessment overview . . . . .	11
0.5 Jamovi statistical software . . . . .	12
0.6 Textbooks . . . . .	12
0.7 Canvas . . . . .	12
0.8 Timetable . . . . .	12
<b>I Background mathematics and data organisation</b>	<b>13</b>
<b>1 Background mathematics</b>	<b>17</b>
1.1 Numbers and operations . . . . .	17
1.2 Order of operations . . . . .	17
<b>2 Data organisation</b>	<b>19</b>
2.1 Tidy data . . . . .	19
2.2 Data files . . . . .	19
<b>3 <i>Practical:</i> Preparing real datasets</b>	<b>21</b>
3.1 LibreOffice Calc . . . . .	21
3.2 Exercise Organising data 1 . . . . .	21

3.3	Exercise Organising data 2 . . . . .	21
3.4	Exercise Organising data 3 . . . . .	21
3.5	Summary of exercises and why they are useful. . . . .	22
<b>II</b>	<b>Statistical concepts</b>	<b>23</b>
4	Recap of some statistical concepts	27
5	Why study statistics?	29
6	Populations and samples	31
7	Types of variables	33
8	Units, precision, and accuracy	35
9	Uncertainty propogation	37
10	<i>Practical.</i> Introduction to Jamovi	39
10.1	Exercise for summary statistics . . . . .	39
10.2	Exercise to compute variable . . . . .	39
10.3	Exercise on transforming variables . . . . .	39
<b>III</b>	<b>Summary statistics</b>	<b>41</b>
11	Decimal places and significant figures	45
12	The mean	47
13	The mode	49
14	The median and quantiles	51
15	Mean, mode, median, and resistance	53

<i>CONTENTS</i>	5
<b>16 Plots</b>	<b>55</b>
16.1 General principles . . . . .	55
16.2 Histograms . . . . .	55
16.3 Box-whisker plots . . . . .	55
<b>17 <i>Practical.</i> Real data with Jamovi</b>	<b>57</b>
17.1 Some biological example . . . . .	57
17.2 Some environmental example . . . . .	57
17.3 Summary of exercises and why they are useful. . . . .	57
 <b>IV Probability models and the Central Limit Theorem</b>	 <b>59</b>
<b>18 Introduction to probability models</b>	<b>63</b>
18.1 A practical example . . . . .	63
18.2 Probability distributions . . . . .	63
<b>19 The Central Limit Theorem (CLT)</b>	<b>65</b>
19.1 Examples of the CLT in action . . . . .	65
19.2 The standard normal distribution . . . . .	65
19.3 What are z-scores? . . . . .	65
<b>20 <i>Practical.</i> Probability and simulation</b>	<b>67</b>
20.1 Calculating probability exercise 1 . . . . .	67
20.2 Calculating probability exercise 2 . . . . .	67
20.3 Calculating probability from normal distribution . . . . .	67
20.4 Normal distribution and sample size . . . . .	67
20.5 Simulating the central limit theorem . . . . .	68
 <b>V Statistical inference</b>	 <b>69</b>
<b>21 Sample statistics and population parameters</b>	<b>73</b>
<b>22 Standard Normal Distribution</b>	<b>75</b>

<b>23 Confidence intervals</b>	<b>77</b>
<b>24 The t-interval</b>	<b>79</b>
<b>25 <i>Practical.</i> z- and t- intervals</b>	<b>81</b>
25.1 Example constructing confidence intervals . . . . .	81
25.2 Confidence interval for different levels (t- and z-) . . . . .	81
25.3 Proportion confidence intervals . . . . .	81
25.4 Another confidence interval example? . . . . .	81
<b>VI Review of parts I-V</b>	<b>83</b>
<b>VII Hypothesis testing</b>	<b>87</b>
<b>26 What is hypothesis testing?</b>	<b>91</b>
<b>27 Making and using hypotheses and types of tests</b>	<b>93</b>
<b>28 An example of hypothesis testing</b>	<b>95</b>
<b>29 Hypothesis testing and confidence intervals</b>	<b>97</b>
<b>30 Student t-distribution and one sample t-test</b>	<b>99</b>
<b>31 Another example of a one sample t-test</b>	<b>101</b>
<b>32 Independent t-test</b>	<b>103</b>
<b>33 Paired sample t-test</b>	<b>105</b>
<b>34 Violations of assumptions</b>	<b>107</b>
<b>35 Non-parametric tests, and what these are.</b>	<b>109</b>

<i>CONTENTS</i>	7
<b>36 <i>Practical.</i> Hypothesis testing and t-tests</b>	<b>111</b>
36.1 Exercise on a simple one sample t-test . . . . .	111
36.2 Exercise on an independent sample t-test . . . . .	111
36.3 Exercise involving multiple comparisons . . . . .	111
36.4 Exercise with non-parametric . . . . .	111
36.5 Another exercise with non-parametric . . . . .	111
 <b>VIII Analysis of Variance (ANOVA)</b>	 <b>113</b>
<b>37 What is ANOVA?</b>	<b>117</b>
<b>38 One-way ANOVA</b>	<b>119</b>
<b>39 Two-way ANOVA</b>	<b>121</b>
<b>40 Kruskal-Wallis H test</b>	<b>123</b>
<b>41 <i>Practical.</i> ANOVA and associated tests</b>	<b>125</b>
41.1 ANOVA Exercise 1 . . . . .	125
41.2 ANOVA Exercise 2 . . . . .	125
41.3 ANOVA Exercise 3 . . . . .	125
41.4 ANOVA Exercise 4 . . . . .	125
 <b>IX Counts and Correlation</b>	 <b>127</b>
<b>42 Frequency and count data</b>	<b>131</b>
<b>43 Chi-squared goodness of fit</b>	<b>133</b>
<b>44 Chi-squared test of association</b>	<b>135</b>
<b>45 Correlation key concepts</b>	<b>137</b>
<b>46 Correlation mathematics</b>	<b>139</b>

<b>47 Correlation hypothesis testing</b>	<b>141</b>
<b>48 <i>Practical.</i> Analysis of count data, correlation, and regression</b>	<b>143</b>
48.1 Chi-Square Exercise 1 . . . . .	143
48.2 Chi-Square association Exercise 2 . . . . .	143
48.3 Correlation Exercise 3 . . . . .	143
48.4 Correlation Exercise 4 . . . . .	143
 <b>X Linear Regression</b>	 <b>145</b>
<b>49 Regression key concepts</b>	<b>149</b>
<b>50 Regression validity</b>	<b>151</b>
<b>51 Introduction to multiple regression</b>	<b>153</b>
<b>52 Model selection (maybe remove this?)</b>	<b>155</b>
<b>53 <i>Practical.</i> Using regression</b>	<b>157</b>
53.1 Regression Exercise 1 . . . . .	157
53.2 Regression Exercise 2 . . . . .	157
53.3 Regression Exercise 3 . . . . .	157
53.4 Regression Exercise 4 . . . . .	157
 <b>XI Randomisation approaches</b>	 <b>159</b>
<b>54 Introduction to randomisation</b>	<b>163</b>
<b>55 Assumptions of randomisation</b>	<b>165</b>
<b>56 Bootstrapping</b>	<b>167</b>
<b>57 Monte Carlo</b>	<b>169</b>



<i>CONTENTS</i>	9
<b>58 <i>Practical.</i> Using R</b>	<b>171</b>
58.1 R Exercise 1 . . . . .	171
58.2 R Exercise 2 . . . . .	171
58.3 R Exercise 3 . . . . .	171
 <b>XII Statistical Reporting</b>	 <b>173</b>
<b>59 Reporting statistics</b>	<b>177</b>
<b>60 More introduction to R</b>	<b>179</b>
<b>61 More getting started with R</b>	<b>181</b>
<b>62 <i>Practical.</i> Using R</b>	<b>183</b>
62.1 R Exercise 1 . . . . .	183
62.2 R Exercise 2 . . . . .	183
62.3 R Exercise 3 . . . . .	183
 <b>XIII Review of parts (VII-XII)</b>	 <b>185</b>
<b>A Statistical tables</b>	<b>189</b>



# Preface

Welcome to the module. This workbook will be used throughout the semester and contain all of the information that you need for the statistical techniques (SCIU4T4) module.

## **0.1 What is statistics?**

An explanation of the material, and what will be taught.

## **0.2 Why this module is important**

Some discussion of module importance

## **0.3 Teaching overview**

Here is how you will be taught, with online lectures, reading assignments, and face-to-face practicals.

## **0.4 Assessment overview**

You will have one formative test and two summative tests. You will also have one mock exam and one exam exam.

### **0.4.1 Test 1F**

Information about Test 1F

**0.4.2 Test 1S**

Information about Test 1S

**0.4.3 Test 2S**

Information about Test 1S

**0.4.4 Mock Exam**

Information about the mock exam

**0.4.5 Exam**

Information about the exam

**0.5 Jamovi statistical software**

Introduction to Jamovi, and why we are using it instead of other software.

**0.6 Textbooks**

Introduction to the primary textbook Learning statistics with jamovi, and a mention of other sources.

**0.7 Canvas**

How we will use Canvas, and how this book relates to it (Learning and Teaching content, where lectures, assessments, and discussions can be found).

**0.8 Timetable**

## Part I

# Background mathematics and data organisation



In week 1, we will focus on a refresher of some necessary background mathematics for this module. We will then turn to the topic of how to organise data sets. We will then practice organising datasets and saving them in a usable format.

Week: 1 Dates: Suggested Readings: Textbook intro, Hadley's paper Assessments: Practice quiz Practical: Quick summary of topics covered





# Chapter 1

## Background mathematics

Some of this will be review, but it is important. Suggested reading for this (some mathematics text).

### 1.1 Numbers and operations

A very broad reminder of mathematics, which you will need for this module

### 1.2 Order of operations

This is easy to forget



## Chapter 2

# Data organisation

It is important to organise data properly so that statistical analysis can be done. Here I explain the tidy approach to data. Suggested reading Hadley Wickam's paper.

### 2.1 Tidy data

### 2.2 Data files



## Chapter 3

# *Practical:* Preparing real datasets

In this practical, we will use a spreadsheet to organise datastes.

### 3.1 LibreOffice Calc

LibreOffice Calc is a free and open source spreadsheet program. The instructions for this section will be identical to the more popular commercial Microsoft Excel.

### 3.2 Exercise Organising data 1

Walks through Exercise 1.3.2

### 3.3 Exercise Organising data 2

Walks through Exercise 1.3.3

### 3.4 Exercise Organising data 3

Walks through Exercise 1.3.4, saving all of these as CSV files

### 3.5 Summary of exercises and why they are useful.

## Part II

# Statistical concepts





Overview of what this week will include.

Week: 2 Dates: Suggested Readings: Textbook intro to Jamovi Assessments:  
Practice quiz



## Chapter 4

# Recap of some statistical concepts

Some introduction



## Chapter 5

# Why study statistics?

General discussion



## Chapter 6

# Populations and samples

Explanation of the mode





## Chapter 7

# Types of variables

Categorical, ordinal, continuous, etc.



## Chapter 8

# Units, precision, and accuracy

Resistance in statistics



## Chapter 9

# Uncertainty propogation

Some simple equations



## Chapter 10

# *Practical.* Introduction to Jamovi

Some introductory text to Jamovi. Particular attention, and maybe an example, on different data types and how to find them in Jamovi.

### 10.1 Exercise for summary statistics

### 10.2 Exercise to compute variable

### 10.3 Exercise on transforming variables





## Part III

# Summary statistics



Overview of what this week will include.

Week: 3 Dates: Suggested Readings: Textbook intro to Jamovi Assessments:  
Practice quiz



## Chapter 11

# Decimal places and significant figures

It is important to know how to write a number you have calculated to the appropriate number of digits, typically either defined as a number of decimal places or as a number of 'significant figures'. This is especially important in any module (for example, this one) where you are required to put numeric answers into a test or exam, as the computer will only recognise the answer as being correct if it is expressed in the way stated in the question. If you are unfamiliar with how to work out the right number of decimal places or significant figures, then these guides should be useful, although there are many other web-sources that would explain the concepts too.



## Chapter 12

# The mean

Explanation of the mean





## Chapter 13

# The mode

Explanation of the mode



## Chapter 14

# The median and quantiles

Notes on the median and quantiles



## Chapter 15

# Mean, mode, median, and resistance

Resistance in statistics



## Chapter 16

# Plots

Graphics are critical for visualising data, which is always important

### 16.1 General principles

Some points about plots

### 16.2 Histograms

Histograms are special, and introduce the concept of a distribution.

### 16.3 Box-whisker plots





## Chapter 17

# *Practical.* Real data with Jamovi

Using some real datasets in Jamovi

### 17.1 Some biological example

Remember to first put it in a tidy format. Get summary statistics here too.

### 17.2 Some environmental example

Transform and compute a new variable, plotting in Jamovi.

### 17.3 Summary of exercises and why they are useful.

Useful for scientific publications, reading, and dissertation work.



## Part IV

# Probability models and the Central Limit Theorem



General overview of what will be the focus of this week.

Week: 4 Dates: Suggested Readings: Textbook intro to probability Assessments: Practice quiz Practical:



## Chapter 18

# Introduction to probability models

Some background

### 18.1 A practical example

How to think about probability

### 18.2 Probability distributions

Some more useful examples

#### 18.2.1 Binomial distribution

Explanation, fairly straightforward

#### 18.2.2 Poisson distribution

Another example

#### 18.2.3 Normal distribution

Why this is so important





## Chapter 19

# The Central Limit Theorem (CLT)

General overview

19.1 Examples of the CLT in action

19.2 The standard normal distribution

19.3 What are z-scores?



## Chapter 20

# *Practical.* Probability and simulation

Some of these examples will be similar to what will be on the assessment

### 20.1 Calculating probability exercise 1

Example exercise 1 with some simple probability calculations

### 20.2 Calculating probability exercise 2

Example exercise 2 with some simple probability calculations

### 20.3 Calculating probability from normal distribution

Example exercise for getting a the probability of some value sampled above, below, or between some threshold in Jamovi.

### 20.4 Normal distribution and sample size

Showing how we get closer to the normal distribution as sample size increases in Jamovi.

## **20.5   Simulating the central limit theorem**

Doing the example from a uniform distribution in Jamovi.

## Part V

# Statistical inference



General overview of what will be the focus of this week.

Week: 5 Dates: Suggested Readings: Textbook intro to probability Assessments: Practice quiz Practical:





## Chapter 21

# Sample statistics and population parameters

An explanation of this



## Chapter 22

# Standard Normal Distribution

What this means, and why it is important.



## Chapter 23

# Confidence intervals

How these are calculated, and how to interpret them



## Chapter 24

# The t-interval

What this is and how it relates to the normal distribution, and why it is important.





## Chapter 25

### *Practical.* z- and t-intervals

25.1 Example constructing confidence intervals

25.2 Confidence interval for different levels (t- and z-)

25.3 Proportion confidence intervals

25.4 Another confidence interval example?



## Part VI

# Review of parts I-V



This is a special chapter for week 6, which is a reading week, and it will function as a very brief pause for review. It will also ensure that the numbers of chapters will correspond to weeks.



## Part VII

# Hypothesis testing





General overview of what will be the focus of this week.

Week: 7 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical:



## Chapter 26

# What is hypothesis testing?

An explanation of this, and that we are starting to get into some of the more interesting bits of inferential statistics.



## Chapter 27

# Making and using hypotheses and types of tests

What this means, and why it is important.



## Chapter 28

# An example of hypothesis testing

Errors





## Chapter 29

# Hypothesis testing and confidence intervals

Relationship between these two.



## Chapter 30

# Student t-distribution and one sample t-test

What this is and how to do it in Jamovi.



## Chapter 31

# Another example of a one sample t-test

From the lectures



## Chapter 32

# Independent t-test

What this is and how to use it in Jamovi.





## Chapter 33

# Paired sample t-test

Another explanation, example, and how to do it in Jamovi.



## Chapter 34

# Violations of assumptions

What to do in this case



## Chapter 35

# Non-parametric tests, and what these are.

Explanation of how to do them in Jamovi.



## Chapter 36

### *Practical.* Hypothesis testing and t-tests

- 36.1 Exercise on a simple one sample t-test
- 36.2 Exercise on an independent sample t-test
- 36.3 Exercise involving multiple comparisons
- 36.4 Exercise with non-parametric
- 36.5 Another exercise with non-parametric





## Part VIII

# Analysis of Variance (ANOVA)



General overview of what will be the focus of this week.

Week: 8 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical:



## Chapter 37

# What is ANOVA?

General explanation



## Chapter 38

# One-way ANOVA

Explain what this is.





## Chapter 39

# Two-way ANOVA

More explanation



## Chapter 40

# Kruskal-Wallis H test

Non-parametric explanation



## Chapter 41

### *Practical.* ANOVA and associated tests

41.1 ANOVA Exercise 1

41.2 ANOVA Exercise 2

41.3 ANOVA Exercise 3

41.4 ANOVA Exercise 4



## Part IX

# Counts and Correlation





General overview of what will be the focus of this week.

Week: 9 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical:



## Chapter 42

# Frequency and count data

General explanation



## Chapter 43

# Chi-squared goodness of fit

Explain what this is.



## Chapter 44

# Chi-squared test of association

More explanation





## Chapter 45

# Correlation key concepts



## Chapter 46

# Correlation mathematics



## Chapter 47

# Correlation hypothesis testing



## Chapter 48

# *Practical.* Analysis of count data, correlation, and regression

48.1 Chi-Square Exercise 1

48.2 Chi-Square association Exercise 2

48.3 Correlation Exercise 3

48.4 Correlation Exercise 4





## Part X

# Linear Regression



General overview of what will be the focus of this week.

Week: 10 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical:



## Chapter 49

# Regression key concepts



## Chapter 50

# Regression validity





## Chapter 51

# Introduction to multiple regression

General explanation



## Chapter 52

# Model selection (maybe remove this?)

Seriously consider moving the regression into this week. and ease the amount of material in previous weeks.



## Chapter 53

### *Practical.* Using regression

53.1 Regression Exercise 1

53.2 Regression Exercise 2

53.3 Regression Exercise 3

53.4 Regression Exercise 4



## Part XI

# Randomisation approaches





The aim of this lecture is to introduce the randomisation approach to statistical hypothesis testing. We will first introduce the general idea of what randomisation is and how it relates to the hypothesis testing that we have been doing since week five. We will then consider an instructive example in which a randomisation approach is used in place of a traditional t-test to test whether or not the mean values of two different groups are identical. We will then compare the assumptions underlying randomisation and how they differ slightly from the assumptions of traditional hypothesis testing. We will then look at how randomisation can be used to build confidence intervals and test hypotheses that would be difficult to test with other approaches. In learning about randomisation approaches, we will also review some key concepts from earlier in the module. The aim is not to understand all of the nuances of randomisation, but to understand, conceptually, what is going on in the methods described below.

Week: 11 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical: R starts creeping in now?



## Chapter 54

# Introduction to randomisation

General explanation



## Chapter 55

# Assumptions of randomisation

How these differ



## Chapter 56

# Bootstrapping

What this is and why we use it.





## Chapter 57

# Monte Carlo



## Chapter 58

### *Practical.* Using R

58.1 R Exercise 1

58.2 R Exercise 2

58.3 R Exercise 3



## Part XII

# Statistical Reporting



Week: 12 Dates: Suggested Readings: Textbook Assessments: Practice quiz  
Practical: R starts creeping in now?





## Chapter 59

# Reporting statistics

General explanation



## Chapter 60

# More introduction to R

How these differ



## Chapter 61

# More getting started with R

Just more to do.



## Chapter 62

### *Practical.* Using R

62.1 R Exercise 1

62.2 R Exercise 2

62.3 R Exercise 3





## Part XIII

### Review of parts (VII-XII)



This chapter will be specifically to prepare for exam.



## Appendix A

### Statistical tables