



MT3508: Applied Statistics (2020/2021)

26 January 2021

Contents

Contact Information	3
Module Description	4
Intended Learning Outcomes	4
Timetable	6
Lectures	6
Practical Sessions	7
Office Hours	7
Assessment	8
R-Studio Cloud	9
What is R-Studio Cloud?	9
Why are we using R-Studio Cloud?	9
How will we use R-Studio Cloud?	10
What about R-Studio desktop	11
Reference Books	12
eBooks	12
Print	13

Contact Information

Important: Please check [Moodle](#) and Teams regularly for announcements.

Module Co-ordinator: Dr Hannah Worthington

Lecturers: Dr Hannah Worthington and Dr Michail Papathomas

Contact Information:

Please use email or Teams to contact us.

- Hannah: [hw233@](#)
- Michail: [M.Papathomas@](#)

Please begin the email subject with *MT3508*.

Module Description

Together with MT3507, this module provides a bridge between second year and Honours modules in Statistics. It deals with the application of statistical methods to test hypotheses and draw inferences from data. This includes a number of nonparametric methods and statistical tests (goodness-of-fit tests and tests of independence). Inference methods include model fitting by least squares and maximum likelihood, and variance estimation by means of the information matrix and the bootstrap. The framework of the generalised linear model is presented covering parameter estimation, deviance, model selection and diagnostics. Further applications include multiple regression, analysis of variance and the (normal) linear model.

Pre-requisites: Before taking this module you must pass MT2508.

Intended Learning Outcomes

- Understand inference by maximum likelihood sufficiently well to conduct maximum likelihood inference on unseen problems using the statistical software R, and to draw appropriate conclusions.
- Be able to construct appropriate likelihood functions from non-mathematical problem descriptions, for problems involving uncertainty, and in which observations are independent.
- Be able to use R to implement likelihood functions, to maximise them with respect to unknown parameters, and obtain confidence intervals for model parameters and functions of parameters.
- Understand the relationships between ANOVA models, linear regression models, generalised linear models, and more general statistical models that do not fall into any of these categories.
- Be able to conduct appropriate model selection and diagnostic tests for these models, to assess model adequacy.
- Be able to obtain Wald confidence intervals, profile likelihood confidence intervals, and bootstrap confidence intervals for parameters and functions

of parameters.

Timetable

The following information may change as the semester develops. Whilst I shall endeavour to keep this document up-to-date, please regularly check Moodle and Teams for announcements of any changes. Get in touch if anything is unclear.

I'll introduce the module during a live Teams session on ***Tuesday 26th January 12pm (Week 1)***. Come along with suggestions or questions about the semester.

Lectures

Lectures will take place over Teams on Mondays (even weeks only), Tuesdays and Thursday at 12pm Noon. The sessions will be recorded and available through the module [Moodle page](#) (hopefully later the same day).

Below is a *rough* schedule indicating which parts of the lecture notes are to be covered each week. Please see the ***Lecture Plan*** document for information on individual sessions.

- ***Week 1*** - Introduction to the module and *Section 1.1* introducing some examples that will carry through the module
- ***Week 2*** - Finish *Chapter 1* and start *Chapter 2* up to discussing *optim*
- ***Week 3*** - Finish *Chapter 2*
- ***Week 4*** - *Chapter 3* and start *Chapter 4* up to beginning to discuss goodness-of-fit tests
- ***Week 5*** - Finish *Chapter 4*
- ***Week 6*** - *Chapter 5* up to starting to look at examples
- ***Week 7*** - Finish *Chapter 5* and start *Chapter 6* up to beginning to look at one-way ANOVA
- ***Week 8*** - *Chapter 6* possibly finish two-way ANOVA

Spring Break

- ***Week 9*** - *Chapter 6* regression, looking at examples and diagnostics
- ***Week 10/11*** - Finish *Chapter 6*

Practical Sessions

This module is very hands-on in using R. Starting in Week 2 there will be weekly practical sessions during which you can discuss and get help on the question sheets.

Practical sessions will take place over Teams and we will be using R-Studio Cloud as a means to share code. We'd like to make use of the breakout rooms in teams to see some small group discussions taking place amongst the class, so please do your best to attend a session each week.

For Weeks 2-11 the weekly sessions will take place on:

Thursday 10am and Thursday 3pm

You'll be able to sign up for a group on MySaint, but you are welcome to attend both if you wish!

Office Hours

If you have questions about the module materials you are welcome to get in touch over email or Teams at any point. If you have queries you want to discuss in real time (i.e. live face-to-face over teams) then you can make use of the following office hour:

Wednesday 12pm Weeks 1-11.

Assessment

Moodle Quizzes - $4 \times 5\%$

Quiz 1: Due *Wednesday February 10th 1pm (Week 3)*

Quiz 2: Due *Wednesday February 24th 1pm (Week 5)*

Quiz 3: Due *Wednesday March 17th 1pm (Week 8)*

Quiz 4: Due *Wednesday April 14th 1pm (Week 10)*

There will also be a practice quiz in Week 2 so that you can see the potential format and some of the different styles of questions that may be asked.

Final Exam - 80%

You must complete at least 90% of assessment in a module to have the opportunity to gain credit in the module. If you do not complete at least 90% of assessment you will be given a FINAL academic alert in the module and receive a grade of 0X.

R-Studio Cloud

Warning: I'm also rather new to using R-Studio Cloud. (In fact some of you may have already logged more hours if you have met it in other modules...) So this is a learning experience for everyone.

What is R-Studio Cloud?

This is a version of R-Studio that runs in your browser as a cloud-based solution for remote collaboration, code sharing and teaching.

You can access R-Studio Cloud [here](#). When you enter your university email address an option will appear to allow you to log-in through the University. Access is still being arranged so some of you may not be able to do so just yet, but it should be possible very shortly.

Why are we using R-Studio Cloud?

You will be getting a fair bit of programming practice in this module. Ordinarily we would be teaching in an environment where tutors can simply look over your shoulder to help out or offer suggestions.

This is rather tricky at the moment.

If/when issues arise, the most efficient way to offer help is to be able to see *exactly* the problems you are having.

It's my experience that screen sharing/emailing code/screen grab/copying the error message, only offers a quick fix when the issue is trivial. It is a frequent problem that the errors can not be recreated or there are version issues/updates/missing packages/missing information/other conflicts that cause more problems.

R-Studio Cloud gets around this by allowing tutors to see your work exactly as you are seeing it, error messages and all. We hope (and we are taking a bit of a leap of faith here) that the R-Studio Cloud is a straightforward way to offer help on coding in a way that is efficient, but still familiar, for everyone involved.

How will we use R-Studio Cloud?

Each practical sheet will have it's own project area in R-Studio Cloud. These are projects that I will create and then share with you, from which you can duplicate a copy and then work on just as you would in R-Studio normally.

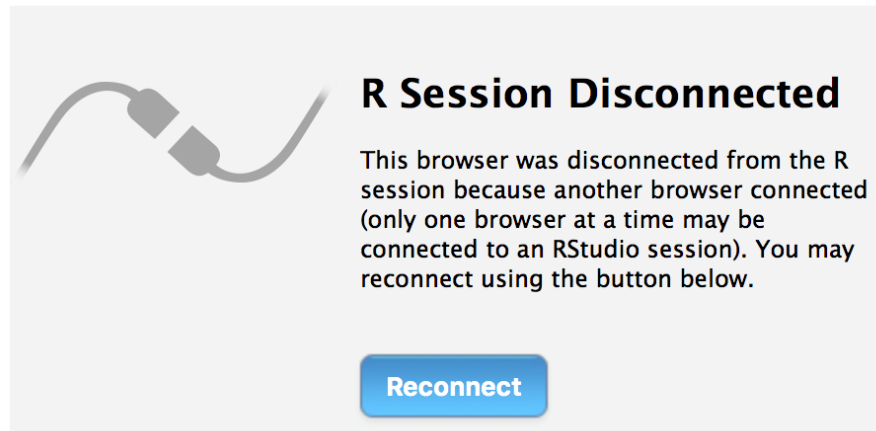
The idea is, if you run into problems you can let a tutor know and they can jump into your session to see what is happening.

Pros:

- You can continue working on your code whilst waiting for help.
- When a tutor takes a look they are seeing the most up-to-date version.
- The tutor can see all of your code. This is particularly helpful since error messages are often associated with something earlier on going astray and only being flagged later on.
- You can work asynchronously to the tutor but only have one version on the go.
- It avoids AppsAnywhere!

Cons:

- The system isn't perfect. You will temporarily lose access to your work whilst a tutor is taking a look. If you see the following pop-up



it's probably an indication that a tutor has jumped into your area and is taking a look at your work.

- We will need to figure out a workflow for requesting help etc.
- You will lose access when term ends (I don't know the exact date).

At any point you can download a copy of your work. I will send out reminders to do so as we complete the practical sheets so that you can retain copies of your work.

What about R-Studio desktop

If you would prefer to work on your local version of R-Studio, offline, then you are welcome to do so. But R-Studio Cloud will be the mechanism through which we will be offering coding and teaching support (so you may just need to move your work round a little).

Reference Books

There is no single text that covers everything that we will do in MT3508. The notes and practicals are comprehensive for our purposes. However, if you would like some supplementary reading, the following books cover various bits of the module.

eBooks

The following are available as eBooks through the library.

Wood, S. N. (2017). *Generalized additive models: An introduction with R (2nd edition)*. CRC Press

While we do not cover generalised additive models (GAMs) in MT3508, this book contains a lot of stuff that we do cover in its lead-up to GAMs.

- Appendix A contains some general maximum likelihood results.
- Chapter 3 contains a lot of good material on GLMs - treated with more theoretical rigour than we do in MT3508, but also some examples of applications of GLMs using R.
- Chapter 1 contains a lot of good material on linear models (LMs: linear regression and ANOVA) - again treated with more theoretical rigour than we do in MT3508, but also some examples of applications of LMs.

Wood, S. N. (2015). *Core statistics*. Cambridge University Press.

This is an excellent reference text at a fairly advanced level. The Preface says “*The aim is to offer a concise coverage of the core knowledge needed to understand and use parametric statistical methods and to build new methods for analysing data.*” It does this very well, but much of its contents is beyond the scope of MT3508. The most useful chapter to use are probably:

- Chapter 3, which covers basic R programming and concepts;
- Chapter 4, which covers maximum likelihood theory (often at a more advanced level than MT3508 does); and
- Chapter 7, which covers LMs (often at a more advanced level than MT3508 does).

Print

The following is *not* available as an eBook but you may be able to request a scan of appropriate sections from the library.

Miller, R. B. (2011). *Maximum likelihood estimation and inference, with examples in R, SAS and ADMB.* Wiley

This covers a lot more than we do, and is more advanced in many places.

- Chapters 1-6 overlap a good deal with what we do in MT3508.
- The first part of Chapter 7 covers some material on generalised linear models (GLMs).
- Chapters 11 and 12 cover the theory underlying maximum likelihood methods (some of which we use explicitly in MT3508 but none of which we derive; these Chapters contain derivations).