STAT210: Applied Statistics

General Course Information

This document provides a general description of STAT210, and should be treated as a broad overview and subject to minor change from one year to another. A more detailed Course Information document is provided each semester.

Lectures

There are three lectures per week. Lectures are normally delivered in person but may be delivered online as needed, e.g. if an instructor is isolating. We normally record lectures as well in order to support revision and students who may not be able to attend, e.g. due to illness or an allowed clash. From time to time a lecture may not record correctly, and they do not capture everything that occurs in class, so should not be treated as a normal replacement for attending lectures.

Tutorials/Practicals

There are multiple tutorial streams (which we also call practicals). These are a key way to develop R programming skills and are very helpful for assignments. They are also a great place to ask conceptual questions. Each students is assigned to a stream, but we are usually able to accommodate students who wish to attend more than one stream.

Website

Copies of the lecture slides and assignments will be available via a course website, along with other resources such as announcements, discussion boards, and recorded lectures.

Prerequisites

The prerequisite for the course is either STAT110 or STAT115.

Content Overview

Statistics is a critically important part of scientific research. Beyond science, it is used in nearly every aspect of life, including public policy, marketing, insurance, economics, manufacturing, and social media. Statistics is concerned with understanding how data can inform us about the world in which we live. It involves study design, data collection, analysis, and communication. During this course, we will look at a range of commonly used statistical techniques, using a broad range of research questions to motivate the methodology. We will encounter linear regression, logistic regression, and multivariate methods, while learning about the importance of sampling and study design to understanding.

Learning Outcomes

At the end of STAT210 a student should be able to independently design studies and analyze various types of data. They should be able to use R to input data and fit basic statistical models. They should be able to understand the assumptions underlying a range of models and identify when a more complex approach is required.

Modules

The topics covered in STAT210 include:

- Linear regression

- Models for binomial and count data

- Principal components analysis

- Clustering and discrimination

- Study design

Outline

Week	Topics
1	Sampling and inference
2	Models and estimation 1
3	Linear regression
4	Linear regression
5	Properties of linear regression
6	Remedial measures
7	Study design
8	Logistic regression
9	Logistic regression
10	Principal components analysis
11	Discrimination
12	Clustering
13	Course review

Software

Throughout this course we will be using the software RStudio and Rmarkdown. This software is freely available. It is recommended that you download the required software onto a laptop or home computer. While most students choose this, we provide alternative computing support via two options:

- The software can be accessed online. A video will be posted to the resources page that shows how to access the server, upload files to the online version, and subsequently download files back onto your home computer.

- Use the computing labs available in the Department of Mathematics and Statistics.

More information about RStudio and Rmarkdown, including how they can be installed, is available in documents on the STAT 210 resource page.

Grading

The course has three component: assignments (A), mid-term test (T), and exam (E). The final mark (F) is

$$F = 0.24A + 0.16T + 0.60E.$$

Terms

STAT210 has two terms requirements, both of which must be met to pass the paper:

- A minimum mark of 40% on the Assignments

- A minimum mark of 40% on the Exam

Assignments

There will be **ten assignments** throughout the semester, approximately every week. To accommodate most circumstances that may crop up during a semester, e.g. work commitments, illness, family events, the assignment mark will be based on the **best eight** submitted assignments.

- The assignments will be submitted electronically (through your personal resources page on the STAT210 resources website). You will be required to write your assignments using R Markdown, compiled into Word (both the .Rmd and .doc file will need to be uploaded). No other format will be accepted.

- Late assignments will not be marked unless prior arrangements have been made with the lecturer, or unless told otherwise in the detailed Course Information document provided each semester.

- If a student has a valid reason to miss more than two assignments (e.g. serious and lengthy illness), they should contact the Course Coordinator to discuss options. Appropriate documentation would be required.

Mid-term Test

The mid-term test is 50 minutes in length, normally held close to the mid-semester break.

Final Exam

The final exam is comprehensive and will be 3 hours in duration.