

Department of Statistics

STATS 760: A Survey of Modern Applied Statistics

Course Information 2017

Lecturer:

Alan Lee

Department of Statistics

City Office: Room 367, Mathematics and Physics Building

Telephone : 373 7599 extension 88749

Fax: 373 7000

DD: 923 8749

email: lee@stat.auckland.ac.nz

Course web page: www.stat.auckland.ac.nz/~lee/760

Introduction:

The aim of this course is twofold: to introduce you to some new statistical techniques, and give you some experience in finding out details of statistical techniques that are unfamiliar to you. If you are working in statistics after graduation, you will have to teach yourself a lot of new techniques. This course is designed to give you practice in doing this.

Thus, the course is rather different from most postgraduate courses, in that there will be fewer lectures than usual, and some of these will be given by you!

The course is based on the books “Statistical Models in S” by Chambers and Hastie, “Modern Applied Statistics with S-Plus” by Venables and Ripley, “The Elements of Statistical Learning” by Hastie, Tibshirani and Friedman, and An Introduction to Statistical Learning by James, Witten, Hastie and Tibshirani. All four books are available in the University library, the last three as e-books. I will provide photocopies of the relevant sections of C&H.

You will work through selected chapters of the books, at your own pace. Each student will be assigned 5 topics, three of these will be compulsory, and you can choose two others.

The compulsory ones (together with the relevant books) are

Linear and generalized linear models: (C&H, V&R) The emphasis will be on understanding how R fits linear models.

Modern regression and classification techniques (V&R, JWHT, HT&F) Gams, regression trees, neural networks, smoothing, nearest neighbours, boosting and bagging, random forests, support vector machines.

Unsupervised Learning and Visualization (V&R, JWHT , HT&F)., K means, hierarchical clustering, Self-organizing maps.

The remaining two topics will be chosen from

Plotting and graphical displays, including animation
Mixed models
Multivariate analysis
Survival analysis
Time series analysis
Spatial statistics

The primary resource for these will be V&R.

I will negotiate the choice of the last two topics in the first week. As a general rule, you should not do topics you are already familiar with. Thus, each of you will have a different study program.

Course organization:

We will divide into ~6 small groups, and I will have a meeting with each group once a week, at a time and place to be arranged. This should last about one hour. At the meeting, you will discuss how you are progressing with your mastering of the assigned chapters. You will be expected to research each topic, and gain the following

- An appreciation of why the technique is important
- The kinds of applied problems that the technique can solve
- The types of data that call for the technique
- The R software that you need to implement the technique
- How to interpret the output from the computer runs
- Any diagnostic techniques that are important
- How the techniques relate to others (e.g. how to glms relate to linear models, GAMS to glms)

I will assign some problems for you to work, which I will mark and return.

In addition, each group will choose one of the 5 topics for more formal study. For this topic, I will require the group to write a short paper (around 10-20 pages) describing the

technique and covering the bullet points above. In addition, each group will deliver a 20 minute presentation to the class describing your findings.

I will also require you to keep a journal, in which you will make notes recording what you are learning. You will also include computer output in your journal. Discussion of your journal entries and computer runs will form the basis of our meetings. Journals are not formal documents – they are a record of your discovery process, and can be handwritten. However, I will have to be able to read them!!

Assessment:

This will be on the basis of your journal, your paper and presentation, and also on the basis of a mid-semester test and an oral exam at our final meeting when we will review what you have learnt. The breakdown will be

Journals/meetings	20%
Oral exam	20%
Paper/presentation	20%
Test	20%
Assignments	20%

NB: There is no final exam.

Getting started - the first week:

At the first lecture (in Rm 310, building 303, at 12 noon on Friday March 10) I will discuss the assignment of topics, based on your course selections to date. The topics will be subject to negotiation, but in general I will expect the material covered to be new to you. We will also discuss how you can pair up for the paper and the presentation.

In addition, I will discuss the resources you can use to research your topics. The primary resources will be the two books mentioned above, but there are many excellent texts you can use for supplementary reading. There is also the R and S-Plus online help. The internet (particularly using Google) can also be very helpful.

We will also begin a discussion of linear models.

In weeks 2-10, I will give the occasional “overview” lecture on the topics that are unfamiliar to most students. These will cover mostly the three compulsory topics, but maybe some of the other topics. The lectures will be introductory and will be designed to get you started on these topics – dates to be arranged. The rest of the time will be taken up with self-study, our weekly meetings, and preparation of your papers and presentations. The presentations will be held in the last week of term. I expect all students to attend all these lectures and be prepared to make constructive comments during question time. Papers will be due on the last Friday of the semester (Friday 9th June).

Collaboration:

It is my view that collaboration is an important part of the learning process and I encourage you to discuss the material you are studying with each other (and me!) However, you must not copy another person's journal.

Course Planner:

In the diagram below, 0 is the first week, discussed above. The other figures denote "periods" and are a guide to what topics you should be working on in any particular week. You will be assigned one topic for each period.

Week beginning					
	Mon	Tue	Wed	Thu	Fri
6/03/2016	0	0	0	0	Lecture
13/03/2016	1	1	1	1	1
20/03/2016	1	1	1	1	1
27/03/2016	2	2	2	2	2
3/04/2016	2	2	2	2	2
10/04/2016	3	3	3	3	Easter
17/04/2016	mid-semester break				
1/05/2016	3	3	3	3	3
8/05/2016	4	4	4	4	4
15/05/2016	4	4	4	4	4
22/05/2016	5	5	5	5	5
29/05/2016	5	5	5	5	5
5/06/2013	Presentations		Paper due		

Web page: The course web page is at www.stat.auckland.ac.nz/~lee/760 and I will post various resources on this page, plus administrative notices. Please check it regularly.