Lecture 10 (of 10) – Wed 25 March 2020

The Assignment

- The assignment consists of two long (20 mark) questions from the 2019 exam.
- **Please submit solutions on Moodle:** under the last lecture there is a tab called "Assignment 2", use this.
- The deadline is on or before 30^{th} April 2020. Late submissions are accepted but will incur a penalty.
- Any submissions via email will be ignored, likewise any "corrections" sent via email. The version you submit on Moodle is the only version that will be considered.
- Either type up using Latex else take a series of high quality pictures. There is an upper limit of 100 MB on your total submission. This should be ample – a decent image will be about 2 MB. Tip: place page flat on table, use a lamp to shine a light directly on top of page.
- Please also submit a cover sheet – this can be handwritten on the first page of your solutions, use the format of the blank cover sheet uploaded on Moodle. Basically I need to see your name, my name, the assignment name, module name, date and a signature. If you are using Latex, do this handwritten.
- I will aim to get marks and solutions to you before the exam.

The Exam

Attached is a slide pack which summarises all the key equations and results from the course. You can't go wrong with working slide by slide and ensuring you understand all the results stated therein. Additionally:

- Make sure you can interpret the output from the R command: summary(model), where model is some glm. In particular, make sure you understand how to use the output to perform goodness of fit testing and model selection. In the exam you are likely to need to use the quantiles from the *Chi-Squared*, *Normal*, *F* and *t* distributions – know how to quickly get these from the Cambridge Statistical Tables.
- A rule of thumb that might come in handy is that if $Z \sim N(0,1)$ then $Pr(Z > 2)$ is approximately 0.025, hence $Pr(-2< Z <2)$ is approximately 0.95. (A more accurate 0.025 upper quantile for the standard normal is 1.96).
- Throughout the second half of the course, you were introduced to some matrix algebra and calculus and also to multivariate probability. There are some standard results that you need to know without hesitation:

1. If x is a random vector and A a matrix: what is E(Ax), Var(Ax)?
2. What is the derivative (w.r.t x) of expressions like x'Ax? Remember that vector differentiation (of a scalar) is merely scalar differentiation (of a scalar), but you then stack the results in a column.
3. We saw the use of matrix square roots (the symmetric version and the unsymmetrical Cholesky version). This required us to recall how transpose and inversion interact – revise these and be comfortable with the manipulations.

- Don't forget the basics from earlier courses – for example, revise your univariate distribution theory: how do you construct a chi-squared random variable (r.v.), t-distributed r.v. and an F-distributed r.v.?
- We used the eigendecomposition (ADA') for real valued, symmetric, positive-definite matrices at least twice. Can you state the key property about the matrix D? Can you state the key property about the side matrix A (it is summed up in the word *orthonormal*)? We

made use of the R function “eigen()” - make sure you understand the output of this function by reading the help file (type “?eigen” into R).

- *Proofs of equivalence*: you may be asked to prove that two statements are equivalent. Make sure you understand the following logic of how to do this. Let A and B be two statements (perhaps collections of infinitely many statements). To say that ***A is equivalent to B*** means both that ***A entails B*** and that ***B entails A***. So to prove that ***A is equivalent to B***, requires two steps:
  1. Assume that A is true, prove that B must also be true. (L2R)
  2. Assume that B is true, prove that A must also be true. (R2L)
- C*ontrolling Error:* we saw many examples of picking a threshold k for a test statistic T in order to control the type 1 error. To do this we needed to figure out the distribution of T under H0. We then said alpha = Pr(error under H0) = Pr(T > k under H0), hence for any given value of alpha (say 0.05) we can solve for the k such that Pr(T > k under H0) = alpha.

Example Q/A

The photos give some example questions from past exams and the solutions. More past exams can be found here: http://www.bbk.ac.uk/library/exam-papers/mathematics-and-statistics. Look at both the undergrad exams “Statistical Modelling” and the MSc exams “Statistical Analysis” regardless of what course you are doing.

Emailing exam questions

In the first instance, if you are stuck on something you should try and discuss it with other people from the course. If after having made at least one serious effort to find the solution, you are still stuck/lost, please send me an email explaining what the problem is, what your serious attempt has been to track down the solution and most importantly, why you think what you tried didn’t work. If you do those things, I will help you.