STATS 3001 Statistical Modelling III
Semester 1, 2018
Preliminary examination information

Recommended approaches

- Write out a cheat sheet even though you are **not** allowed one in the exam. This will help you revise the course material and consolidate your knowledge.
- Attempt previous exams (on MyUni) in exam-like conditions.
- Make sure you know assignment, tutorial, practical and lecture content.
- You won't be required to code in R in the exam, but interpreting the output is important.
- Review the definitions, major theorems and proofs from the lectures, some will turn up in the exam.
- Know your formulas obviously. For example, what is $\hat{\boldsymbol{\beta}}$? What is $\text{Var}(A\boldsymbol{Y})$? What is the $100(1-\alpha)\%$ confidence interval and prediction interval of $\boldsymbol{\eta}_0$? etc.
- In the exam, you may separate the appendices from the exam paper so that you don't have to go back-and-forth looking at them and the questions.
- If you don't know an answer, start writing the information you know down. You are likely to be awarded marks for a partial answer and the solution may become apparent after writing down what you know initially.
- The exam paper is worth a total of 70 marks, so you should have plenty of time to complete the 5 questions in 120 minutes. Remember:
  - You don't have to answer questions in the order they appear on the exam paper. For example, if logistic regression is your strong suit, feel free to answer that question first.
  - If a question is taking too much time, leave it and come back to it later - get the easy marks in other questions while you have time.

Exam notes

- The exam will be two (2) hours long, plus 10 minutes reading time.
- You will need a calculator but a scientific calculator is the most advanced calculator allowed. **DO NOT** bring a phone calculator or a graphics calculator as your calculator.
- There will be 5 questions, worth a total of 70 marks.
  - Q1+2+3: multiple regression theory and applications.
  - Q4: short question on geometry of least squares.
  - Q5: logistic regression.
- There will be no questions on Poisson regression in this year's exam - this topic was not included (you can ignore the Poisson regression questions in past exams).
- Ignore the section in the course lecture notes on multistratum experiments, they were not included in the course . Again, ignore any questions on this topic in exams from previous years.

Other notes

- Make sure you keep an eye on your emails/MyUni, as this will be my method of communication.
- Your projects are due Monday 5th June by noon. When these are marked, I will put all projects and uncollected assignments in the Stat Modelling III box on Level 6. I will email you when these are available for collection.
- I will be available for **exam consulting** on
  - **Wednesday** 20^{th} **June noon to 1pm**
  - **Monday** 25^{th} **June 11am to noon**

Checking multiple linear regression assumptions

The assumptions of multiple linear regression are almost sure to come up for easy marks. Make sure you know how to check the assumptions and in this order (you will lose marks if not in this order):

- Linearity
- Homoscedascity (constant variance)
- Normality
- Independence.

Interpreting regression coefficient estimates

Make sure you are able to interpret the parameter estimates from your fitted models correctly. There are plenty of examples given in the assignment, practical and tutorial solutions provided on MyUni.

Lecture summary

This summary is of the lectures actually presented this semester as provided in the Echo 360 videos and given in the annotated lecture pdf files on MyUni. You should use this summary as a guide to the 2018 course content.

(1) Notation, basic distributional results for random vectors.

(2) The multiple linear regression model; least squares estimators and their properties.

(3) Inference for regression; estimable functions; BLUEs.

(4) The Gauss-Markov theorem; inference for regression coefficients; ANOVA.

(5) Prediction; symbolic specification of linear models, model formulæin R.

(6) Factors, factor codings and constraints; model formula operators `+`, `:`, `*`,`^`, `/`, `-`.

(7) The marginality principle, polynomial models, models with factors and covariates.

(8) Regression diagnostics; standardized residuals.

(9) Studentised residuals, leverage, Cook's distance, and influence diagnostics.

(10) Cheese example; Considerations for model building: covariate adjustment for variance reduction.

(11) Covariance adjustment for bias correction and elucidation: observational studies and quasi-experiments.

(12) Model selection algorithms, forward selection, backward elimination, stepwise selection, Mallow's C_p. Box-Cox transformations.

(13) Box-Cox transformations; the profile likelihood.

(14) The profile likelihood (cont.); generalised least squares.

(15) GLS (cont.); weighted least squares; coordinate free approach to least squares: linear subspaces and basic definitions.

(16) Orthogonal projections and least squares.

(17) Reformulation of previous results in the coordinate free setting, ANOVA.

(18) Expected mean squares; sequences of nested hypotheses.

(19) Orthogonality; generalised least squares.

(20) Logistic regression, derivation of the log-likelihood, score vector and Fisher information matrix.

(21) The Fisher scoring algorithm, large sample inference, residual deviance.

(22) Common odds ratios for several 2×2 tables; prospective and retrospective studies.

(23) Assessing model fit; Pearson and Deviance residuals; overdispersion.

May 28, 2018