STAT 3012/3912 Applied Linear Models – Semester 1, 2018

Unit Information Sheet

Websites: It is important that you check the STAT3012 website regularly which may be accessed through the canvas website `https://canvas.sydney.edu.au/courses/6134`.

Unit lecturers:

- STAT3012 week 1-10:
  Michael Stewart (office 818, tel. 02 9351 5765, `michael.stewart@sydney.edu.au`)
- STAT3012 week 11-13:
  Emi Tanaka (office 827, tel. 02 9351 3039, `emi.tanaka@sydney.edu.au`)
- STAT3912 week 1, 3-13:
  Michael Stewart (office 818, tel. 02 9351 5765, `michael.stewart@sydney.edu.au`)

Contact hours:

- Michael, TBC in Carslaw 818 for Week 2-10
- Emi, Wed 10-11am in Carslaw 829 for Week 11-13

Timetable: Note the following adjustments to the normal weekly timetable

- Week 01: for STAT3**0**12 students only, no tutorial
- Week 02: for STAT3**9**12 (Advanced) students only, no advanced lecture
- Week 04: Wed, 28/03/18 a quiz in place of the lecture
- Week 04: Fri, 14/04/17 no classes because of Good Friday
- Week 10: Fri, 18/05/17 a quiz in place of the lecture
- Week 13: for STAT3**9**12 (Advanced) students only, an advanced quiz in place of the advanced lecture
- Week 13: no computer lab

Tutorials and labs: Labs (one per week) start in week 1 and tutorials (one per week for STAT3012 only) start in week 2. You should attend the tutorial given on your personal timetable. Attendance at labs and tutorials will be recorded. Because there is no tutorial for students in the advanced stream, their Friday lab will be a mix between tutorial and computer problems.

Prerequisites: (STAT2012 or STAT2912 or STAT2004) and (MATH1002 or MATH1902)

Prohibitions: STAT3002, STAT3902, STAT3004, STAT3904, STAT3022, STAT3922

Suggested reading: There are no required texts for this unit. The following text is recommended as a reference book and is available as a 2 hour loan from the library.

- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. (2005). Applied Linear Statistical Models (5th Ed.), Mc Graw-Hill/Irwin.

Further, the books and WWW resources listed below may prove helpful.

- Mason, R.L., Gunst, R.F., Hess, J.L. (2003). Statistical Design & Analysis of Experiments (2nd edition). Wiley.
- Mendenhall, W. and Sincich, T. A. (2003). Second Course in Statistics: Regression Analysis (6th edition). Pearson Education Inc.: New Jersey.
- Montgomery, D.C. (2004). Design and Analysis of Experiments (6th edition). Wiley.
- Faraway, J.J. (2002). Practical Regression and ANOVA using R. `http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf`
- Bailey, R. (2008) Design of Comparative Experiments. Cambridge University Press.
- R Development Core Team: An Introduction to R. `http://cran.r-project.org/doc/manuals/R-intro.pdf`

In addition, The Comprehensive R Archive Network site at `http://cran.r-project.org/` contains lots of useful material. In particular, you can download your own free copy of the R package, and browse through the FAQs (Frequently Asked Questions).

Computing & data sets: The computer package R with the RStudio interface will be used during this course. R can be freely downloaded from the CRAN site given above and RStudio is available from `https://www.rstudio.com/`. All data sets used in lectures, tutorials and computer practicals will be made available from the course webpage.

Assessment: Your final raw mark for this unit will be calculated as follows:

method of assessment	tentative timing	weighting	
		STAT 3012	STAT 3912
assignment	week 7 & 13	15%	10%
quiz	week 4 & 10	30%	25%
advanced quiz	week 13	–	10%
2h written examination	June examination period	55%	55%

Both assignments have to be handed in via Turnitin by the due date.

High Distinction (HD), 85-100: representing complete or close to complete mastery of the material; **Distinction (D), 75-84**: representing excellence, but substantially less than complete mastery; **Credit (CR), 65-74**: representing a creditable performance that goes beyond routine knowledge and understanding, but less than excellence; **Pass (P), 50-64**: representing at least routine knowledge and understanding over a spectrum of topics and important ideas and concepts in the course.

Objectives: The main objective of this course is to introduce the fundamental concepts of analysis of data from both observational studies and experimental designs using classical linear methods, together with the teaching of concepts of collection of data and design of experiments. Additional objectives are to gain competency in the application and understanding of linear models and regression methods with diagnostics for checking appropriateness of models; to be introduced to robust regression methods; to be introduced to the design and analysis of experiments and to further understand the notions of replication, randomisation and ideas of factorial designs; to enhance proficiency in the use of the R statistical package to give analyses and graphical displays.

Outcomes: Students who successfully complete this unit will be able to demonstrate

- proficiency in the use of the general F-test as the main tool to choose between two nested regression models
- proficiency in assessing model assumptions and outlier detection in regression models through standard diagnostic plots (box plot, scatterplot, Q-Q-plot, Cook's distance plot, leverage vs residual plot), through influence measures (leverage values, Cook's distance) and through tests (Bartlett test against homoscedasticiy and normality tests)
- proficiency in the understanding and application of multiple linear regression and in the understanding of R^2 and the adjusted R^2
- proficiency in the understanding and application of 1-way ANOVA models of type I and II, including finding an interpretation of the TSS term through using the concept of orthogonal contrasts and making inference on all parameters
- proficiency in the understanding and application of 2-way ANOVA models of type I and making inference on all parameters
- proficiency in the calculation and decomposition of sum of squares terms in multi-way ANOVA for orthogonal designs
- competency in correcting multiple pairwise comparisons by applying the Tukey, Scheffé and Bonferroni correction
- competency in deriving the least-squares estimator in linear regression
- competency in the calculation and interpretation of confidence intervals for all parameters in linear regression
- competency in the understanding of the difference between confidence intervals and prediction intervals
- competency in model selection through using the F-test, t-test, AIC or BIC through full searches or by using step-wise procedures (backward, forward, stepwise)
- competency of polynomial regression models and their selection through using orthogonal polynomials
- competency in using the R function `lmer` for the fitting of mixed models and a basic understanding of these complicated models
- competency in reducing a nominal factor in a multi-way ANOVA to a continuous variable through using linear contrast coefficients
- competency in calculating the distribution for contrasts and using this to calculate confidence intervals for contrasts
- competency in the design of an appropriate scheme for treatment allocation and data collection as well as the correct analysis for complete randomised designs (CBD), randomised CBD (RCBD), Latin square designs (LSD), incomplete block designs (IBD) and balanced IBD (BIBD), ANCOVAs, and nested designs
- competency in the understanding of blocks, nested factors, interactions terms and confounding in experimental designs
- competency in using R to compute estimates and standard errors for regression parameters without built-in functions such as `lm` and `aov`, for generating treatment allocation lists for the CBD, RCBD and LSD.
- basic understanding of L1 regression, M regression and MM regression

- advanced stream students will additionally have competency in theoretical aspects of regression methods, in particular the Gauss-Markov theorem and appreciation of the F-test; if time permits, partial correlation coefficients will be taught and in that case a level of competency should be reached

Intended topics outline:

Week	Topics
1	Experimental designs, observational studies, software R, simple linear regression
2	Model diagnostics, inference for linear regression, fitting multiple linear regression models
3	Inference for multiple regression models, multiple correlation coefficients, Leverage and Cook's distance, the general F-test
4	Subset selection using stepwise procedures and AIC, Cp and BIC
5	Polynomial regression, orthogonal polynomials, Robust regression, 1-way ANOVA
6	Simultaneous CIs, decomposing sums of squares
7	Quantitative factors, 2-way ANOVA, interactions
8	2-way ANOVA with interactions, Normality tests, experimental designs
9	Randomized complete block designs, Latin square designs, incomplete block designs
10	Analysis of covariance, nested factors
11	Revisiting Experimental Design, nested designs, random effect model
12	Variance component estimation, mixed effects models, longitudinal data
13	Agricultural data, hierarchical data, revision

---

last adjustments: February 27, 2018