Subject Information Guide

STAT902: Advanced Data Analysis

Semester 1, 2019

Administration and contact details

Host Department	School of Mathematics and Applied Statistics
Host Institution	University of Wollongong
Name of lecturer	Dr Pavel Krivitsky
Phone number	02 4221 3713
Email Address	pavel_krivitsky@uow.edu.au
Homepage	http://www.krivitsky.net/research
Name of Honours coordinator	
Phone number	
Email Address	

Subject details

Handbook entry URL	https://solss.uow.edu.au/sid/CAL.USER_SUBJECTINFO_SCREEN?p_faccde=24&p_depabb=MAAS&p_subcode=STAT902&p_cal_subject_id=176713&p_year=2019&p_cal_type=P&p_cal_types=UP&p_breadcrumb_type=1&p_menu_type=1&p_cs=5214340372299764567
Subject homepage URL	https://moodle.uowplatform.edu.au/
Honours student hand-out URL	N/A
Start date:	4 March 2019 (first lecture)
End date:	3 June 2019 (last lecture)

Contact hours per week:	2
Lecture day and time:	Mondays 14:30–16:30
Description of electronic access arrangements for students (for example, WebCT)	Resources will be hosted on the University of Wollongong's Moodle e-Learning platform. Arrangements to provide access to the site will be made during first week. It is strongly suggested that the students contact the instructor prior to the first lecture.

Subject content

1. Subject content description

STAT902 introduces a variety of techniques for advanced data analysis, particularly regression, for handling categorical data, dependent data, nonlinear data, and situations where parts of the model are unknown or misspecified. Generalised linear models are considered in detail, as well as other ways of modelling nonlinearity, such as nonlinear models and nonparametric analysis. A variety of ways to model dependent (e.g., repeated measures) data is considered, particularly, linear mixed models, generalised linear mixed models, generalised estimating equations, and latent variable models. Frequentist and Bayesian approaches to inference are considered, including likelihood, quasilikelihood, bootstrap, and sandwich estimation, conjugate priors, Monte-Carlo methods and Markov chain Monte-Carlo, as well as some graphical models, numeric integration, and prior elicitation.

2. Week-by-week topic overview

Note that the following schedule is tentative:
Weeks 1–2: Subject overview; revision of matrix algebra and vector calculus, distributions, and maximum likelihood.
Weeks 3–4: Estimating functions, quasi-likelihood, sandwich estimation, mean-variance misspecification, and bootstrap techniques.
Weeks 4–5: Bayesian inference and computation, prior elicitation, and model selection.
Weeks 6–7: Generalised linear models and nonlinear least-squares.
Weeks 8–9: Linear models for dependent data: linear mixed models and generalised estimating equations.
Weeks 10–11: Nonlinear models for dependent data: generalised linear mixed models and nonlinear mixed models.
Weeks 12–13: Nonparametric methods.

3. Assumed prerequisite knowledge and capabilities

- Equivalent of University of Wollongong's STAT332/STAT921 and STAT333/STAT922: statistical distribution theory, maximum likelihood estimation, fundamentals of statistical inference, multiple regression (linear), logistic and/or Poisson regressions, basic matrix algebra. These topics will be reviewed early in the subject, but not introduced.
- Familiarity with *R* will be very helpful.

4. Learning outcomes and objectives

1) Analyse complex data, particularly data with dependence, repeated measures, categorical response variables, overdispersion, and nonlinearity; and interpret the analyses.
2) Select and apply sophisticated statistical techniques (mostly regression-related) to answer substantive questions. Techniques include generalised linear models, Bayesian inference, mixed models, quasilikelihood, sandwich estimation, bootstrap, generalised estimating equations, model selection, and nonlinear and nonparametric models.
3) Recognise assumptions and limitations of statistical techniques considered, and diagnose their suitability for the data and the research question.
4) Derive expressions for point estimates, variances, and other quantities of interest for the techniques considered.
5) Implement and perform data analyses and diagnostics using R and BUGS/JAGS, including writing independent code in R language.

AQF specific Program Learning Outcomes and Learning Outcome Descriptors (if available):

AQF Program Learning Outcomes addressed in this subject	Associated AQF Learning Outcome Descriptors for this subject
Analyse complex data...	S1, S3, S4, S5, A4
Select and apply sophisticated statistical techniques...	K2, S3, S4, A2
Recognise assumptions and limitations...	K1, K2, S3, A1
Derive expressions for point estimates...	S4, K2, A2
Implement and perform data analyses and diagnostics...	K1, S4

Learning Outcome Descriptors at AQF Level 8

Knowledge

K1: coherent and advanced knowledge of the underlying principles and concepts in one or more disciplines

K2: knowledge of research principles and methods

Skills

S1: cognitive skills to review, analyse, consolidate and synthesise knowledge to identify and provide solutions to complex problem with intellectual independence

S2: cognitive and technical skills to demonstrate a broad understanding of a body of knowledge and theoretical concepts with advanced understanding in some areas

S3: cognitive skills to exercise critical thinking and judgement in developing new understanding

S4: technical skills to design and use in a research project

S5: communication skills to present clear and coherent exposition of knowledge and ideas to a variety of audiences

Application of Knowledge and Skills

A1: with initiative and judgement in professional practice and/or scholarship

A2: to adapt knowledge and skills in diverse contexts

A3: with responsibility and accountability for own learning and practice and in collaboration with others within broad parameters

A4: to plan and execute project work and/or a piece of research and scholarship with some independence

5. Learning resources

Text: Jon Wakefield. *Bayesian and Frequentist Regression Methods*. New York : Springer, **2013**. ISBN: 978-1-4419-0924-4 (Print) 978-1-4419-0925-1 (Online). doi:10.1007/978-1-4419-0925-1

Miscellaneous papers: Linked from the Moodle site.

Lecture notes: Posted on the Moodle site.

R examples: Posted on the Moodle site.

R: Freely available at http://www.r-project.org/; graphical frontends are available as well.

6. Assessment

Exam/assignment/classwork breakdown

Exam	70%	**Assignment**	30%	**Class work**	0%
Assignment due dates		Set Friday of each week	Due Sunday of the subsequent week		
Approximate exam date				15–27 June 2019	

Institution Honours program details

Weight of subject in total honours assessment at host department	1/8
Thesis/subject split at host department	BMath (Hons): Thesis worth 25% BMathAdv (Hons): Thesis worth 37.5%
Honours grade ranges at host department:	
H1	85-100
H2a	75-84
H2b	65-74
H3	50-64