STAT378

Statistical Computing

S2 Day 2017

Dept of Statistics

Contents

General Information

Unit convenor and teaching staff Lecturer in charge Jun Ma jun.ma@mq.edu.au Contact via 98508548 Room 526, 12 Wally's Walk (E7A) Lecturer Hassan Doosti hassan.doosti@mq.edu.au Contact via 98504884 Room 534, 12 Wally's Walk (E7A)
Credit points 3
Prerequisites 6cp at 200 level including (STAT272 or STAT273 or STAT278)
Corequisites
Co-badged status
Unit description This unit develops basic computer-intensive statistical methods. Many of them find applications in scientific research and industry. Topics include: Monte-Carlo simulation; bootstrapping; regression computations which include collinearity diagnostic and models selection using cross-validation; alternatives to least squares; ridge regression, weighted least squares and logistic regression; maximum likelihood computations using iterative methods, such as Newton-Raphson and Fisher scoring; and applications of the maximum likelihood method.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at https://www.mq.edu.au/study/calendar-of-dates

Learning Outcomes

On successful completion of this unit, you will be able to:

By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment Tasks

Name	Weighting	Hurdle	Due
Assignment 1	13%	No	Week 6 Tue (5/9/2017)
Assignment 2	13%	No	Week 12 Tue (31/10/2017)
Group presentation	14%	No	Week 13 lecture
Final exam	60%	No	University Exam Period

Assignment 1

Due: **Week 6 Tue (5/9/2017)**
Weighting: **13%**

Assignment 1 will be available on the unit webpage in week 3 and due in week 6. Assignments may be handwritten or word-processed, and submitted in person to Dr Hassan Doosti during the week 6 lecture on Tuesday. Late submissions without approval will be penalized at a rate of 5% of earned mark per day, up to maximum of 50%. This implies that submissions will not be accepted more than 10 days late. Assignments must be each student’s own work.

This Assessment Task relates to the following Learning Outcomes:

- Understand the details of linear regression computations.
- Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity.
- Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods.
- Be able to write programs using MATLAB.

On successful completion you will be able to:

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assignment 2

Due: **Week 12 Tue (31/10/2017)**
Weighting: **13%**

Assignment 2 will be available on the unit webpage in week 9 and due in the week 12 lecture. Assignments may be handwritten or word-processed, and submitted in person to A/Prof Jun Ma during the 12 lecture. For policy on later submission and other issues please see the Assignment 1 description.

This Assessment Task relates to the following Learning Outcomes:

- Understand Monte-Carlo simulation and know how to apply it to solve problems.
- Understand the bootstrapping method and its applications.
- Understand nonlinear regression and neuron networks.
- Understand Bayesian computations and be familiar with Winbugs.
- Know how to perform maximum likelihood computations using Newton or scoring method.
- Be able to write programs using MATLAB.

On successful completion you will be able to:

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Group presentation

Due: **Week 13 lecture**
Weighting: **14%**

There will be group presentations during the week 13 lecture. Details on topics and groups, as well as marking criteria, will be announced later. Basically, each group will be given a paper to read (for 3 weeks) and present it to other students and lecturers. Other groups and lecturers will mark the presenting group, and an weighted average mark will be the final mark for this group students.

On successful completion you will be able to:

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Final exam

Due: **University Exam Period**
Weighting: **60%**

This is a written exam and it is to be scheduled in the university exam period. This examination involves conceptual questions, simple calculation questions and simple programming questions. For example, it may ask students to write a bootstrapping program to compute bootstrap confidence intervals. For this exam, students are allowed to bring into the exam room TWO A4 paper notes written/typed on both sides; photocopies are not allowed. Only non-programmable calculators that do not have text retrieval capacity are allowed. Note that students who apply for Supplementary Exams must make themselves available over the supplementary exam period (11/12/17 -15/12/17).

This Assessment Task relates to the following Learning Outcomes:

1. Understand Monte-Carlo simulation and know how to apply it to solve problems.
2. Understand the bootstrapping method and its applications.
3. Understand the details of linear regression computations.
4. Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity.

5. Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods.
6. Understand nonlinear regression and neuron networks.
7. Understand Bayesian computations and be familiar with Winbugs.
8. Know how to perform maximum likelihood computations using Newton or scoring method.

On successful completion you will be able to:

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Delivery and Resources

Classes

You are required to attend a 3-hour lecture each week at Tuesday 10 – 1pm in E4B 208 (FBE Computer Lab)

You are also required to attend an 1-hour tutorial each week at Tuesday 4 - 5pm in E4B 214 (FBE Computer Lab).

Prescribed texts

There is no prescribed textbook for this unit. Students should obtain lecture overheads from iLearn prior to the lecture. The following are recommended reading books for this unit

- Wendy L. Martinez, Angel R. Martinez, "Computational Statistics Handbook with MATLAB", Third Edition (2015 by Chapman and Hall/CRC).
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, "An Introduction to Statistical Learning with Applications in R", (2013 by Springer).
- Robert I. Jennrich, "An Introduction to Computational Statistics – regression analysis", (QA278.2.J46 1995).
- J.H. Maindonald, "Statistical Computation", (QA276.4.M25)

- James E. Gentle, "Elements of Computational Statistics", (QA276.4.G455)
- Kleijnen, Jack. Simulation: a statistical perspective. (QA76.9.C65.K5913)
- Draper and Smith. Applied Regression Analysis. 2nd edition. (QA278.2.D7)

Computing packages used

We primarily use the software package MATLAB in this Unit. MATLAB is becoming increasingly important for training students in scientific computations. More information about MATLAB can be found at the web site "https://au.mathworks.com/products/matlab-online.html" and students are entitled to one year free license.

We use Winbugs when teaching Bayesian methods.

Unit webpage

Unit webpage is located on Moodle at https://ilearn.mq.edu.au. You can access the material on Moodle only if you are enrolled in the unit. All lecturing materials are available at this webpage.

Teaching and Learning Strategy

The unit is taught in traditional mode:

- Students are required to attend weekly lectures and tutorial classes.
- Students are expected to have read through the material, provided on iLearn, to be covered in each week.
- Submit assignments. ·
- Contact the unit convenor in advance if for any reason, you cannot hand in your assessment tasks on time.
- Collect their marked assessment from the lecturer during the lecture

Examination

If you notify the University of your disruption to studies for your final examination, you must make yourself available for the supplementary examination. If you are not available at that time, there is no guarantee an additional examination time will be offered.

Software

We are using MATLAB (or R) and WinBUGS in teaching this unit. R and WinBUGS are free software and are widely used nowadays by statisticians. More information about R can be found at http://www.r-project.org/, and WinBUGS at “http://www.mrc-bsu.cam.ac.uk/bugs/”.

Unit Schedule

Week	Topic

1	**Introduction and Review** 1. Introduction to statistical computing. 2. Some common probability distributions and their random number generation. 3. Quick introduction to MATLAB.
2	**Regression Computation: Linear regression** 1. Linear regression model. 2. Least-squares criterion and normal equations. 3. Matrices expression of linear regression.
3	**Regression Computation: Linear regression** 1. Solving linear system of equations. 2. Properties of the estimator. 3. Residual analysis. 4. Regression Diagnostics.
4	**Regression Computation: Linear regression** 1. Transformation and variance stabilizing. 2. Box-Cox transformation 3. Weighted regression.
5	**Regression Computation: Linear regression** 1. Ridge regression. 2. Computer assisted model building: (1) Stepwise. (2) Best subset. (3) Cross-validation. (4) Cp and PRESS.
6	**Regression Computation: Linear regression** 1. Model selection using Lasso. 2. Examples.
7&8	**Monte Carlo Simulation** 1. Introduction to Monte Carlo simulations. 2. Examples **Bootstrapping** 1. Introduction to Bootstrapping method. 2. Examples. 3. Further discussions.

8&9	**Nonlinear regression and neural networks** 1. Nonlinear regression. 2. Least squares estimation. 3. Gauss-Newton algorithm. 4. Neural network modelling
10&11	**Maximum Likelihood (ML) Estimation** 1. Introduction to maximum likelihood estimation. 2. An example in medical imaging. 3. Algorithms for ML computing.
11&12	**Bayesian methods** 1. Introduction. 2. Prior and posterior distributions. 3. Generate random number from the posterior distributions. 4. Bayesian computations using WinBugs.
13	Students group presentation

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central. Students should be aware of the following policies in particular with regard to Learning and Teaching:

Academic Honesty Policy http://mq.edu.au/policy/docs/academic_honesty/policy.html

Assessment Policy http://mq.edu.au/policy/docs/assessment/policy_2016.html

Grade Appeal Policy http://mq.edu.au/policy/docs/gradeappeal/policy.html

Complaint Management Procedure for Students and Members of the Public http://www.mq.edu.au/policy/docs/complaint_management/procedure.html

Disruption to Studies Policy (in effect until Dec 4th, 2017): http://www.mq.edu.au/policy/docs/disruption_studies/policy.html

Special Consideration Policy (in effect from Dec 4th, 2017): https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policies/special-consideration

In addition, a number of other policies can be found in the Learning and Teaching Category of Policy Central.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/support/student_conduct/

Results

Results shown in *iLearn*, or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in eStudent. For more information visit ask.mq.edu.au.

Student Support

Macquarie University provides a range of support services for students. For details, visit http://students.mq.edu.au/support/

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to improve your marks and take control of your study.

- Workshops
- StudyWise
- Academic Integrity Module for Students
- Ask a Learning Adviser

Student Services and Support

Students with a disability are encouraged to contact the Disability Service who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

IT Help

For help with University computer systems and technology, visit http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/.

When using the University's IT, you must adhere to the Acceptable Use of IT Resources Policy. The policy applies to all who connect to the MQ network including students.

Graduate Capabilities

Creative and Innovative

Our graduates will also be capable of creative thinking and of creating knowledge. They will be imaginative and open to experience and capable of innovation at work and in the community. We want them to be engaged in applying their critical, creative thinking.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment task

- Group presentation

Commitment to Continuous Learning

Our graduates will have enquiring minds and a literate curiosity which will lead them to pursue knowledge for its own sake. They will continue to pursue learning in their careers and as they participate in the world. They will be capable of reflecting on their experiences and relationships with others and the environment, learning from them, and growing - personally, professionally and socially.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Discipline Specific Knowledge and Skills

Our graduates will take with them the intellectual development, depth and breadth of knowledge, scholarly understanding, and specific subject content in their chosen fields to make them competent and confident in their subject or profession. They will be able to demonstrate, where relevant, professional technical competence and meet professional standards. They will be able to articulate the structure of knowledge of their discipline, be able to adapt discipline-specific knowledge to novel situations, and be able to contribute from their discipline to inter-disciplinary solutions to problems.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment tasks

- Assignment 1
- Assignment 2
- Group presentation
- Final exam

Critical, Analytical and Integrative Thinking

We want our graduates to be capable of reasoning, questioning and analysing, and to integrate and synthesise learning and knowledge from a range of sources and environments; to be able to critique constraints, assumptions and limitations; to be able to think independently and systemically in relation to scholarly activity, in the workplace, and in the world. We want them to have a level of scientific and information technology literacy.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment tasks

- Assignment 1
- Assignment 2
- Group presentation
- Final exam

Problem Solving and Research Capability

Our graduates should be capable of researching; of analysing, and interpreting and assessing data and information in various forms; of drawing connections across fields of knowledge; and they should be able to relate their knowledge to complex situations at work or in the world, in order to diagnose and solve problems. We want them to have the confidence to take the initiative in doing so, within an awareness of their own limitations.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment tasks

- Assignment 1
- Assignment 2
- Group presentation
- Final exam

Effective Communication

We want to develop in our students the ability to communicate and convey their views in forms effective with different audiences. We want our graduates to take with them the capability to read, listen, question, gather and evaluate information resources in a variety of formats, assess, write clearly, speak effectively, and to use visual communication and communication technologies as appropriate.

This graduate capability is supported by:

Learning outcome

- By the end of this unit students should be able to: (1) Understand Monte-Carlo simulation and know how to apply it to solve problems. (2) Understand the bootstrapping method and its applications. (3) Understand the details of linear regression computations. (4) Know how to perform regression diagnostics, particularly checking model assumptions, detecting outliers and diagnosing collinearity. (5) Know model selection techniques, including Stepwise, best subsets, CV, Cp, PRESS and LASSO methods. (6) Understand nonlinear regression and neuron networks. (7) Understand Bayesian computations and be familiar with Winbugs. (8) Know how to perform maximum likelihood computations using Newton or scoring method. (9) Be able to program uins MATLAB

Assessment tasks

- Assignment 1
- Assignment 2
- Group presentation