Ve406 Applied Regression Analysis using R

1 Introduction

1.1 Course Profile

1.1.1 Course Information

- Course Description:

  This course provides an introduction to the process and procedures of statistical modelling. We will explore real data sets, examine various models for the data, assess the validity of their assumptions, and determine which conclusions we can make, if any. In this course you will learn how to program in R and how to use R for effective data analysis.

- Learning Outcomes:

  After successful completion of this course, you should be able to
  1. Explore data graphically.
  2. Select appropriate models.
  3. Implement those models in R.
  4. Examine those models critically.
  5. Interpret the results of those models to non-statisticians.

- Who should take this class?

  The prerequisite for this class is computer/programming knowledge at the level of Vg101 (or above), and statistics knowledge at the level of Ve401 (or above). Both undergraduates and graduate students are welcome to take the course.

1.1.2 Contact Information

- Instructor:

  Tong Zhu

- Lectures:

  Tuesday (06:20pm – 08:55pm) in E3-102
  Thursday (06:20pm – 08:00pm) in E3-102

- Office Hours:

  See Canvas

- Email:

  See Canvas

- Teaching Assistant(s):

  See Canvas

1.1.3 Grading Policy

- Assignment:
  15% There will be 5 assignments in the form of problem sets.
- Lab/Workshop:
  10% There will be 4 labs/Workshops.
- Project:
  25% There will be a project in the form of a challenge.
- Exam:
  50% There will be two exams:

Midterm	Final
25%	25%

- Quiz (Optional):
  15% Quizzes will be given frequently in class.
- For those who attempt all quizzes, their grade is whichever is the higher of:
  0. 25% ALW + 25% Proj + 0% Quiz + 50% Exam
  1. 25% ALW + 10% Proj + 15% Quiz + 50% Exam
  2. 25% ALW + 40% Proj + 15% Quiz + 20% Exam
- For this course, the grade will be curved to achieve a median grade of "B".

1.1.4 Project

- Each of you need to be in one and only one 3-member team for the project.
- The project will be graded according to the following three aspects:
  1. Oral Presentation of your model
  2. Poster Presentation of your model
  3. Prediction Accuracy of your model

  each of those three aspects has an equal weight.
- Each member of the same team will receive the same project mark.
- You will be working on real data (TESLA), some of part 1 is here

  but part 2 will not materialise before the due date.

1.1.5 Honour Code

- Honesty and trust are important. Students are responsible for familiarising themselves with what is considered as a violation of honour code.
- Assignments/projects are to be solved by each student individually. You are encouraged to discuss problems with other students, but you are advised not to show your written work to others. Copying someone else's work is a very serious violation of the honour code.

- Students may read resources on the Internet, such as articles on Wikipedia, Wolfram MathWorld or any other forums, but you are not allowed to post the original assignment question online and ask for answers. It is regarded as a violation of the honour code.
- Since it is impossible to list all conceivable instance of honour code violations, the students has the responsibility to always act in a professional manner and to seek clarification from appropriate sources if their or another student's conduct is suspected to be in conflict with the intended spirit of the honour code.

1.1.6 Teaching Schedule

Week	Topics	Others
1	Introduction Simple Linear Regression Least-Squares and Maximum Likelihood	 A1R
2	Diagnostics Transformation	 W1R
3	Workshop 1 Multiple Linear Regression Diagnostics	 A1D A2R
4	**National Holiday**	
5	Inference Categorical regressors Collinearity	W1D W2R
6	Variable Selections Influential points and Outliers	A2D A3R
7	Workshop 2 Heteroskedasticity Correlated Noise	W2D W3R
8	Nonlinear Regression **Midterm Exam**	A3D A4R
9	Workshop 3 Logistic Regression Poisson Regression	W3D W4R
10	Nonparametric Regression Generalized Linear Model	A4D A5R
11	Workshop 4 Estimating Equations Generalised Additive Models	W4D W5R
12	Principal Component Analysis Factor Analysis	A5D Poster
13	Mixture Models Survival Analysis (Optional) Project Presentation	
14	**Final Exam**	

1.1.7 Textbook

- Simon J. Sheather (2010)
  A Modern Approach to Regression with R

- Grolemund and Wickham (2016)
  R for Data Science
- Myers et al. (2010)
  Generalized Linear Models with Applications in Engineering and the Sciences
- Fox. (2015)
  Applied Regression Analysis and Generalized Linear Models