Statistical Methods
MT2300
2019/2020

Introductory Material

Lecturer: Dr Alexey Koloydenko
Office Hours: TBA, TBA, TBA
Room 0-29
e-mail: alexey.koloydenko@rhul.ac.uk
Workshop tutors: Dr Alexey Koloydenko, TBA

Aims

To study important aspects of statistical modelling in an integrated way and develop some expertise both in the theory and application of linear models.

Objectives: On completion of the course, students should be able to

- demonstrate familiarity with the main methods based on linear models
- to apply these methods to analyse data and interpret the results from such analysis
- use R effectively in the analysis of relevant data.

Content

- **Principles of statistical modelling and terminology:** Systematic and random components; types of variables.
- **Simple and multiple linear regression:** Design matrix notation; fitting the model, ordinary least squares; inferences about regression parameters; prediction; assessing the regression.
- **Some special cases:** Polynomial Models, models that incorporate factors.
- **Model building:** Testing significance of specified subsets of variables; examining all subsets; sequential methods.
- **Model validation and comparison of regressions:** Examination of residuals; influential observations; some possible problems and remedial actions; dummy variables.
- **Qualitative explanatory variables - Analysis of variance:** One-way and two way ANOVA; point estimation; linear contrasts; a general approach via multiple regression.
- **Some non-parametric methods (if time permits):** The sign test, the Wilcoxon test, the Kolmogorov-Smirnov goodness-of-fit test.

Indicative Texts

- *Linear Models with R* - J.J. Faraway (CRC Press 2014), Library - ebook.
- *Introduction to Statistical Modelling* - W.J. Krzanowski (Arnold) Library Ref. 518.3 KRZ
- *Applied Regression Analysis and Other Multivariable Methods* - D.G. Kleinbaum, L.L. Kupper, K.E. Muller, A. Nizam. 3rd Edition (Duxbury Press 1998) Library Ref. 518.3 KLE
- *Data Analysis Using Regression Models* - E. W. Frees (Prentice Hall, 1996). Library Ref. 300.184 FRE
- *A second Course in Statistics: Regression Analysis* - W. Mendenhall, T. Sincich. Fifth Edition (Prentice Hall, 1996). Library Ref. 518.3 MEN
- *Applied linear regression* - S. Weisberg. Third Edition (Wiley series in probability and statistics, 2005) Library Ref:
  eBook `https://www-dawsonera-com.ezproxy01.rhul.ac.uk/abstract/9780471704089`, and 519.536 WEI.
- *Applied linear statistical models* - M. H. Kutner et. al. Fifth Edition (McGraw-Hill, 2005). Library Ref: 519.5 APP
- *Applied linear statistical models : regression, analysis of variance, and experimental designs* - M. H. Kutner, J. Neter, W. Wasserman. Third Edition (Irwin , 1990) Library Ref: 518.3 NET
- *Applied regression analysis* - N. R. Draper. Second Edition (Wiley, 1981) Library Ref: 518.3 DRA
- *Regression. Linear Models in Statistics* N.H. Bingham, John M. Fry. Springer Undergraduate Mathematics Series. (Spinger 2010). Library Ref: 519.536 BIN

Pre-requisites

A knowledge of elementary probability and the basic theory of statistics is required (MT1300). In particular, the following concepts are assumed to be familiar to the students: a real valued random variable X; the cumulative distribution function F and probability density function f of a continuously distributed random variable X, and their basic properties; the expectation $E[X]$ and the variance $Var[X]$ of X; point and interval estimation, and statistical hypotheses concerning means and variances; standard probability distributions (Normal, i.e. Gaussian, Student t, etc); further details of confidence intervals and hypothesis tests, such as the types of error, the p-value, power and significance level.

You will also need to be competent in the use of statistical tables and have some familiarity with using a statistical package.

To revise the above, you may want to consult any standard introductory text such as:

- *A Basic Course in Statistics* (4th or later edition) - G.M. Clarke and D. Cooke (Arnold 1998). Library Ref. 518.3 CLA
- *Statistics for Business and Economics* (4th edition) - P. Newbold (Prentice-Hall 1991). Library Ref. 338.754033 NEW

Course organization

The course consists of **25 lectures** *and* **10 workshops** *in Term 2.* **In the last teaching week of Term 1 (December)**, there will be an introductory R session, with some light exercises assigned for the holiday period. In **teaching weeks 1, 2 and 3** there will be three lectures per week, R workshops begin in the teaching week 2, and starting from the teaching week 4, **there will be two lectures per week (one lecture slot will be replaced by the workshop.** You will be told which workshop group you have been assigned to by the tenth teaching week of term 1.

The lecture notes will be available on Moodle. There will be weekly problem sheets, eight of which will be marked. The problem sheets will appear on Moodle at the beginning of each week. You should hand in your solutions on xxx (TBA) in the following week. Please hand in your work at the end of your xxx (TBA) lecture. Model solutions will appear on Moodle on xxx (TBA). Unless accompanied by an approved notification of absence form, late submissions cannot be marked. More detailed feedback can be given in person during the office hours.

The workshop sheets will be made available in Moodle in due course (weeks 4 to 11). Information and data for the ***examination project*** *will appear in Moodle by the middle of the term.*

Assessment

There will be (1) a two-hour written examination (70%), weekly homeworks (10%), and a R examination project (20%); the project will be due in the last week of term.