University of Rome Tor Vergata

Academic year 2021/2022

Bachelor degree in Business Administration and Economics

Quantitative Methods III
(English)

6 CFU

SSD: SECS-S/03

Rosario Barone
rosario.barone@uniroma2.it

April 13th - May 20th

1 Aims and course description

The course *Quantitative Methods III* is focused on statistical methods for categorical data. The R software for statistical computing will be introduced and used throughout. Since the aim of the course is that of giving students the ability to perform data prepation and analysis, we will start with classes dedicated to an introduction of R software. We will then introduce Generalized Linear Models (GLM) for binary and count outcomes. GLM may be

interpreted as a modification of linear regression for the case of categorical outcomes. For example, in econometrics such models are used to predict the risk that a firm is about to bankrupt: suppose we have a sample of firms and to observe several features, such as status (response variable), number of employees, liquidity ratio, profitability ratio, leverage ratio, etc. By defining a GLM, we may estimate the effect of the selected covariates on the default (status) for the observed firms and then predict bankrupt for new firms. Methods more tailored for prediction will be discussed, also with multi-category outcomes. Lastly, the course will provide tools for dealing with clustered data, arising when units are dependent in groups or when there are repeated measurements over the same units (panel data). In summary, the main goal is to provide to the students theretical and practical skills, such that they are able to select the technique needed to answer specific questions (based on data) in order to perform data analysis appropriately.

2 Prerequisites

Attending students from the B.D. in business administration and economics must have passed Quantitative Methods I and Quantitative Methods II (or similar courses). A good understanding of multiple linear regression models is mandatory. More in detail, all attending students must be familiar with the following topics (which will not be reviewed in class):

- *Basic Probability theory*: Events. Marginal and conditional probabilities. Random variables. Probability mass function. Cumulative distribution function. Density function. Quantile function. Expected values. Main probability distributions.

- *Descriptive Statistics*: Frequencies, bar plots, mean, median, standard deviation, inter quartile range, histograms, boxplots. Scatterplots, Pearson correlation, Spearman correlation.

- *Statistical Inference*: Theory of point estimation: consistency, unbiasedness. Standard error. Confidence intervals for means and probabilities. Hypothesis testing: one-sample t-test, two-sample t-test, paired t-test, chi-square test.

- *Linear regression*: Simple linear regression, multiple linear regression, collinearity. Testing regression coefficients.

Taking the course without knowledge of these topics is discouraged. Students having attended introductory statistics and econometrics courses (covering also probability theory), but unfamiliar with any of the topics above, are strongly encouraged to study them on their own before attending classes.

3 Teaching Method

Some classes will be held in the computer lab (depending on the number of attendant students). Each topic will be described in theory and, wheter is possible, will be implemented in R. Focus will be on the theoretical definition and practical implementation of each technique, as well as on the interpretation of results.

4 Schedule of topics

- Introduction to the R software

- Categorical data: multi-way tables, distributions, the generalized linear model

- Principles of causal analysis: selection, bias, confounding. Confounders, colliders, mediators. Simpson's Paradox.

- Logistic regression, Poisson regression

- Analysis of clustered and panel data: mixed models

- Principles of supervised learning. Classification trees and random forests

5 Textbook

- Cameron, A. C. and Trivedi, P. K. (2005) *Microeconometrics: Methods and Applications.* Cambridge University Press. (More technical)

- Alan Agresti (2018) *An Introduction to Categorical Data Analysis*, 3rd Edition. Wiley (More introductive)

- Everitt, B. S. and Hothorn, T. (2006) *A Handbook of Statistical Analyses Using R*. CRC Press. Available for free at: http://www.ecostat.unical.it/tarsitano/Didattica/LabStat2/Everitt.pdf (For practicing and learning a bit more about R)

6 Assessment

The course attendance is not mandatory (but is recomanded). The final evaluation is based on a mandatory group assignment (also for non attending students) and a written exam. An oral proof may be required from the students. There will be no mid-term exam. The group assignment must be completed and sent to the my e-mail address at least one week prior to the written exam attended by any member of the group. In order to enter the group assignment phase students must send an email with an Excel file containing the student number, last and first name of all members of the group. No additions to the group are allowed after the assignment is sent, while drop-out of members is allowed. Groups are made by a minimum of one to a maximum of six students. Students will submit a text file containing the code used to undertake the assignment, and a brief report (no more than 8 pages including tables and graphics). The assignment will be judged on the basis of (I) correctness, (ii) ability to correctly interpret the results, (iii) ability to draw grounded conclusions, (iv) clarity of the presentation of the report, (v) clarity of the R code. The group assignment will contribute to 50% of the final mark, all students in the group will receive the same mark.

The written exam will cover the entire course's program. It will include open questions, multiple choice questions, and commentary on outputs from R aimed at evaluating the knowledge and understanding of the theoretical as well as applied notions illustrated during the course. Textbooks, notes or other material are not allowed, and their use by the students will invalid the exam. Students must book for the exam. Students not booked will not be allowed to take the exam.

Results are published on my personal webpage. Within two days from the publication of the written exam results, students may require an oral proof in order to increase the marks (by sending and e-mail).

The oral proof will inlcude (few) questions on the entire course's program and/or on the group assignment. Final marks are uploaded on the Delphi system so to be individually received by email by candidates.

7 Other info

Office hours: upon appointment by e-mail sent at least 24 hours before.

Calendar available on: https://economia.uniroma2.it/ba/business-administration-economics/corso/lezioni/2153/

Exams date:

- May 27th, 09:00 am;

- June 15th, 10:00 am;

- July 8th, 10:00 am;

- September 7th, 10:00 am.