Dauphine | PSL★
UNIVERSITÉ PARIS

0IECOY09 – Data Analysis

Professor: Alexandra SCHAFFAR
Contact information: schaffar@univ-tln.fr

Department: International affairs
Semester: 2

Course level: L3 Undergraduate
Domain: Business / Economics
Teaching language: English
Number of in-class hours: 33
Number of course sessions: 10 + Exam
ECTS: 6

Course description and objectives

The aim of this course is to introduce the students to the fundamental methods in Data analysis. This course aims to teach students how to present, analyze and interpret data by using the statistical analysis software package R. Following the course will help students to get familiar to the R ecosystem and learn how to use R for the most common data analysis's projects. Topics include numerical and graphical summaries of data, qualitative and quantitative univariate analysis, bivariate analysis with the study of the links between two variables and more, analysis of variance, regression, principal components, factor analysis and cluster analysis. The course focuses on simple predictive analysis (linear regression or multidimensional analysis, factor approach, principal components approach). The courses take place in the computer lab in order to emphasize on practical aspects of data analysis.

Prerequisites

Basic knowledge on probabilities and statistics.

Learning outcomes

At the end of the course, students are able to describe and present data, to summarize different types of variables, to analyze the relation between these variables, to practice regression and prediction, to cluster and compare different groups of observations. At the end of the course all students are quite familiar with the R environment.

Assignments and grading

Active class participation is a main feature of the course's organization. Students work independently on real datasets with data analysis software R. A final exam evaluates the students. The final exam is also on computer. The final exam counts for 50% of the note. Participation and homework counts for another 50% of the note.

The numerical grade distribution will dictate the final grade. The passing grade for a course is 10/20.

Class participation: Active class participation – this is what makes classes lively and instructive. Come on time and prepared. Class participation is based on quality of comments, not quantity.
Exam policy: In the exam, students will not be allowed to bring any document (except if allowed by the lecturer). Unexcused absences from exams or failure to submit cases will result in zero grades in the calculation of numerical averages. Exams are collected at the end of examination periods.

Course structure

Session	Topic

1	Introduction to R-software – Descriptive statistics
2	Sampling and statistical inference
3	Analyzing relationships among two categorical variables
4	t.test and ANOVA
5	Correlation and Simple regression
6	Multiple regression
7	Logistic regression
8	Principal components analysis
9	Correspondence analysis
10	Clustering
11	**Final Exam**

Bibliography

- Heumann (2016), Introduction to Statistics and Data Analysis, Springer, 455 pages
- J.L. Devore (2011), Introduction to Statistics and Data Analysis, 4th Edition, 944 pages
- C. Judd (2017), Data Analysis, New Edition, 366 pages
- D.S Moore (2009), Introduction to the Practice of Statistics, Freeman, 690 pages
- Daniel J. Denis (2020), Univariate, Bivariate, and Multivariate Statistics Using R, Wiley, 384 pages
- Mustapha Abiodun Akinkunmi (2019), Business Statistics with Solutions in R, De Gruyter, 276 pages
- Christian Heumann, Michael Schomaker, et al. (2017), Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R, Springer

Lecturer's biography

Moodle

This course is on Moodle: **No**

Academic integrity

Be aware of the rules in Université Paris Dauphine about plagiarism and cheating during exams. All work turned in for this course must be your own work, or that of your own group. Working as part of a group implies that you are an active participant and fully contributed to the output produced by that group.