

OIECOX09 (S1) &
OIECOY09 (S2)- Data
Analysis

Professor: Alexandra SCHAFFAR
Contact information:
Cellular: 0628684398
e-mail: schaffar@univ-tln.fr
e-mail:
alexandra.dimou@dauphine.psl.eu
Semester: 1 &2

Course level: Bachelor (L)
Domain: Business, Economics
Language: English
Number of hours per semester: 33H
ECTS: 6

Course description and objectives

This course introduces students to fundamental methods in data analysis using the statistical software package R. It aims to teach students how to effectively present, analyze, and interpret data. Students will gain familiarity with the R ecosystem and learn how to use R for various data analysis projects. Key topics include numerical and graphical summaries of data, univariate and bivariate analysis, analysis of variance, regression, principal components, factor analysis, and cluster analysis. Practical application is emphasized through hands-on sessions in a computer lab.

Prerequisites

Basic knowledge of probability and statistics is required. Familiarity with any programming language is beneficial but not mandatory.

Learning outcomes

By the end of the course, students will be able to:

- Describe and present data using appropriate statistical and graphical methods.
- Summarize different types of variables and their distributions.
- Analyze relationships between variables through bivariate analysis.
- Conduct regression analysis and make predictions.
- Perform multivariate analyses, including principal component analysis and factor analysis.
- Execute cluster analysis to identify patterns within data.
- Confidently navigate and utilize the R environment for data analysis tasks.
- Apply advanced R features and packages for data visualization and manipulation.
- Work effectively in teams and communicate data analysis results clearly.

Assignments and grading

Active participation is crucial for this course. Students will engage in individual and group work on real datasets using R. Assessment consists of:

- Class Participation and Homework (40%): Regular attendance, active engagement in lab sessions, and completion of homework assignments.
 - Class Participation (10%): Attendance and participation in discussions and lab sessions.
 - Homework (30%): Regular assignments based on lecture and lab exercises.
- Weekly Quizzes (10%): Short quizzes to reinforce key concepts.
- Group Project (20%): A comprehensive project involving data analysis and presentation.
- Final Exam (30%): A comprehensive, practical exam conducted on the computer, assessing students' ability to apply course concepts using R.

Course structure

Each session consists of a theoretical part covering the statistical concepts needed for analysis and a practical part using R software to work with examples.

<i>Session</i>	<i>Topic</i>
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

- Wickham, H., & Golemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage Publications.
- Kabacoff, R. I. (2015). *R in Action: Data Analysis and Graphics with R*. Manning

Publications.

- Peng, R. D. (2016). *R Programming for Data Science*. Leanpub.
- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press.
- Nina Zumel and John Mount (2019), *Practical Data Science with R*
- 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

R Documentation and Tutorials:

- [The R Project for Statistical Computing](#)
- RStudio Education Resources
- CRAN R Documentation

Moodle

This course is on Moodle: yes

Academic integrity

Be aware of the rules in University 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.