

DMA-MBD-513 Fundamentals of Mathematical Data Analytics

SEMESTER: Fall

CREDITS: 30 hours

LANGUAGE: Spanish

DEGREES: Master in Big Data Technologies and Advanced Analytics

Course overview

This course is an introduction to statistics with a special emphasis on its mathematical foundations. The subject aims to introduce the student to basic, but fundamental, concepts such as distributions, probability or inference. The subject also provides a first course of basic R by combining exposure of the main concepts in statistics and tutorial R sessions.

Prerequisites

Basic knowledge of Calculus and Algebra is required (understand and manipulate equations, manipulate exponents and logarithms using their basic rules, full understanding of functions and inverse functions, understand limits, derivatives and integrals, know rules for product and summation, etc.) Basic knowledge of Statistics (descriptive statistics, discrete and continuous probability distribution models, sampling and basics of statistical inference) is highly recommended but not required.

Basic knowledge of Programming languages is required, ideally in R or Python.

Course contents

Theory:

1. Introduction: Types of variables. Levels of measurement. Frequency tables.
2. Graphics: For categorical data. For quantitative data. Relationship between variables.
3. Distributions: Shape of distributions. Measures of centrality and dispersion. Summaries and boxplots.
4. Probability I: Contingency tables. The rules of probability. Conditional probability. Relationship between categorical variables.

5. Probability II: Normal. Properties of the normal distribution. Poisson. Weibull. Exponential. Sampling. Confidence Intervals. Central limit theorem.
6. Linear Regression: Least squares. Coefficient of correlation and interpretation. Regression analysis.
7. Logistic regression: Conditions of applicability. Interpretation.
8. Introduction to Bayesian Statistics: Bayes Theorem. Monte Carlo. Bayesian software.

Textbooks

While we will not follow a textbook, we find the following books quite remarkable in their central topics (R, regression and Bayesian statistics, respectively). The first one in particular is very close to the goals of this course.

- **Grolemund, G. and Wickham, H.** (2017). R for Data Science. O'Reilly. Disponible online en <https://r4ds.had.co.nz>
- **Peng, R.** (2016) R Programming for Data Science. Disponible online en <https://bookdown.org/rdpeng/rprogdatascience/>
- **Kabacoff, R.,** (2011). *R in Action*. 1st Edition. Manning Publications.
- **Gelman, A. and Hill, J.,** (2006), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st Edition. Cambridge University Press.
- **Kruschke, J.,** (2014), *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*, 2nd Edition. Academic Press.

Grading

The following conditions must be accomplished to pass the course:

- A minimum overall grade of at least 5 over 10.
- A minimum grade in the final exam of 4 over 10.

The overall grade is obtained as follows:

- Final exam accounts for 60% of the final grade if the grade in this exam is at least 4. In other case, final exam accounts for 100 % of the overall grade.
- Laboratory sessions work (in class and homework) accounts for 40% of the final grade.