# Nonparametric Statistics

**Academic Year: ( 2022 / 2023 )** **Review date: 16/05/2022 19:33:30**

**Department assigned to the subject: Statistics Department**

**Coordinating teacher: GARCIA PORTUGUES, EDUARDO**

**Type: Compulsory ECTS Credits : 3.0**

**Year : 1 Semester : 2**

## REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Mathematics for Data Science
Probability
Statistical Inference
Programming in R
Multivariate Analysis
Regression Models
Advanced Programming

## OBJECTIVES

* Basic competences

- CB6: Possess and understand the knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
- CB9: Communicate conclusions, as well as the knowledge and the ultimate reasons that support them, to specialized and non-specialized audiences in a clear and unequivocal manner.
- CB10: Develop the learning skills that enable further study in a manner that is largely self-directed or autonomous.

* General competences

- CG1: Apply the techniques of analysis and representation of information, to adapt it to real problems.
- CG4: Synthesize the conclusions obtained from data analysis and present them clearly and convincingly in a bilingual environment (Spanish and English), both written and oral.
- CG5: Generate new ideas (creativity) and anticipate new situations, in the contexts of data analysis and decision making.
- CG6: Apply social skills for teamwork and to relate with others in an autonomous way.

* Specific competences

- CE1: Apply advanced knowledge of statistical inference in the development of methods for the analysis of real problems.
- CE2: Use free software such as R and Python for the implementation of statistical analysis.
- CE5: Apply advanced statistical fundamentals for the development and analysis of real problems involving the prediction of a variable response.
- CE6: Apply nonparametric models for the interpretation and prediction of random phenomena.
- CE10: Apply statistical modeling in the treatment of relevant problems in the scientific field.

* Learning outcomes

Acquisition of knowledge on: 1) kernel density estimators and their applications; 2) nonparametric regression methods based on smoothing; 3) nonparametric hypothesis testing.

## DESCRIPTION OF CONTENTS: PROGRAMME

This course is designed to give a panoramic view of several tools available for nonparametric statistics, at an intermediate-advanced level. This view covers in-depth the main concepts in the estimation of the density and regression functions through kernel methods (with their corresponding applications), and the description of several popular nonparametric tests. The focus is placed on providing the main insights on the statistical/mathematical foundations of the methods and on showing the effective implementation of the methods through the use of statistical software. This is achieved by a mixture of theory and reproducible code.

1. Introduction
   1.1. Probability review
   1.2. Facts about distributions
   1.3. Stochastic convergence review
   1.4. OP and oP notation
   1.5. Review on basic analytical tools
   1.6. Why nonparametric statistics?

2. Kernel density estimation I
   2.1. Histograms
   2.2. Kernel density estimation
   2.3. Asymptotic properties
   2.4. Bandwidth selection
   2.5. Practical issues
   2.6. Kernel density estimation with ks

3. Kernel density estimation II
   3.1. Multivariate kernel density estimation
   3.2. Asymptotic properties
   3.3. Bandwidth selection
   3.4. Applications of kernel density estimation

4. Kernel regression estimation I
   4.1. Kernel regression estimation
   4.2. Asymptotic properties
   4.3. Bandwidth selection
   4.4. Regressogram
   4.5. Kernel regression estimation with np

5. Kernel regression estimation II
   5.1. Kernel regression with mixed multivariate data
   5.2. Bandwidth selection
   5.3. Prediction and confidence intervals
   5.4. Local likelihood

6. Nonparametric tests
   6.1. Goodness-of-fit tests for distributions
   6.2. Comparison of distributions
   6.3. Independence tests
   6.4. Other tests

The program is subject to modifications due to the course development and/or academic calendar.

## LEARNING ACTIVITIES AND METHODOLOGY

The lessons consist of a mixture of theory (methods description) and practice (implementation and practical usage of methods). The implementation of the methods is done with the statistical language R. Students are expected to bring their own laptops to experience the code during some parts of the lessons.

* Training activities

- AF1: Theoretical lesson.
- AF2: Practical lesson.
- AF5: Tutorials.

- AF6: Group work.
- AF7: Individual work.
- AF8: On-site evaluation tests.

* Teaching methodologies

- MD1: Class lectures by the professor with the support of computer and audiovisual media, in which the main concepts of the subject are developed and the bibliography is provided to complement the students' learning.
- MD3: Resolution of practical cases, problems, etc. posed by the teacher individually or in groups.
- MD4: Presentation and discussion in class, under the moderation of the professor of topics related to the content of the subject, as well as case studies.
- MD5: Preparation of papers and reports individually or in groups.

## ASSESSMENT SYSTEM

| | |
|---|---|
| **% end-of-term-examination/test:** | 0 |
| **% of continuous assessment (assigments, laboratory, practicals…):** | 100 |

The evaluation in the ordinary call is done entirely by continuous evaluation. This is done by a mixture of:

(a) two quizzes about key ideas and theoretical concepts;
(b) a set of practical exercises to be done in groups;
(c) active participation in lessons and voluntary exercises.

The continuous evaluation grade (in the scale 0-10) is

min(0.5 * A + 0.5 * B + 0.10 * C, 10),

where

- A (in the scale 0-10) is the weighted grade of the quizzes;
- B (in the scale 0-10) is the grade of the practical exercises;
- C (in the scale 0-10) is the grade corresponding to (c).

Students who have not followed the continuous evaluation may take a final exam in the ordinary call with a value of 60% of the final grade.

The grade in the extraordinary call is established by a quiz and the delivery of a set of practical exercises.

Further details are provided in Aula Global. The evaluation is subject to modifications due to the course development and/or academic calendar.

## BASIC BIBLIOGRAPHY


- Chacón, J. E. and Duong, T. Multivariate Kernel Smoothing and Its Applications, Chapman and Hall/CRC, 2018

- Fan, J. y Gijbels, I. Local polynomial modelling and its applications, Chapman & Hall, 1996

- Li, Q. and Racine, J. S. Nonparametric Econometrics, Princeton University Press, 2007

- Wand, M. P. and Jones, M. C. Kernel Smoothing, Chapman & Hall, 1995


## ADDITIONAL BIBLIOGRAPHY


- Wasserman, L. All of Nonparametric Statistics, Springer-Verlag, 2006

- Wasserman, L. All of Statistics, Springer-Verlag, 2004