Exploratory data analysis

Academic Year: (2021 / 2022) **Review date: 08/06/2021 10:43:17**

Department assigned to the subject: Statistics Department

Coordinating teacher: AUSIN OLIVERA, MARIA CONCEPCION

Type: Basic Core ECTS Credits : 6.0

Year : 1 Semester : 1

Branch of knowledge: Social Sciences and Law

OBJECTIVES

The aim of this course is that students learn how to organize, represent, analyze and summarize the information contained in a dataset by the use of the appropriate graphical, tabular and numerical methods according to the type of data and variables observed.

SPECIFIC COMPETENCES
1. Distinguish different types of variables and data.
2. Synthesize tabular, numeric and graphical statistical information.
3. Propose and validate the simple linear regression model as a model for the relationship between two continuous variables.

TRANSVERSAL COMPETENCES:
1. Capacity of analysis and synthesis of information.
2. Setting up and solving practical problems.
3. Written and verbal communication.

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction
1.1. What is Statistics. Definition.
1.2. General concepts.
1.3. Sample methods.

2. Descriptive statistics for a single variable.
2.1 Frequency distribution. Grouping by classes.
2.2. Frequency distribution. Grouping by class intervals.
2.3. Graphical displays.
2.4. Numerical measures for a univariate distribution.

3. Transformations.
3.1. Linear transformations.
3.2. Non linear transformations.

4. Joint description of various variables.
4.1. Two-way tables. Joint frequency distribution.
4.2. Graphical displays.
4.3. Marginal frequency distributions. Conditional frequency distributions.
4.4. Numerical measures for linear association. Pearson's correlation coefficient.
4.5. Spearman's correlation coefficient.
4.6. Association measures for contingency tables.

5. Relations between variables.
5.1. Simple linear regression. The least squares method.

1. Introduction
1.1. What is Statistics. Definition.
1.2. General concepts.
1.3. Sample methods.

2. Descriptive statistics for a single variable.
2.1 Frequency distribution. Grouping by classes.
2.2. Frequency distribution. Grouping by class intervals.
2.3. Graphical displays.
2.4. Numerical measures for a univariate distribution.

3. Transformations.
3.1. Linear transformations.
3.2. Non linear transformations.

4. Joint description of various variables.
4.1. Two-way tables. Joint frequency distribution.
4.2. Graphical displays.
4.3. Marginal frequency distributions. Conditional frequency distributions.
4.4. Numerical measures for linear association. Pearson's correlation coefficient.
4.5. Spearman's correlation coefficient.
4.6. Association measures for contingency tables.

5. Relations between variables.
5.1. Simple linear regression. The least squares method.

LEARNING ACTIVITIES AND METHODOLOGY

Theory classes with support material available on the web, problem solving classes, practical classes with statistical computing packages in computing labs.

ASSESSMENT SYSTEM

% end-of-term-examination/test:	60
% of continuous assessment (assigments, laboratory, practicals…):	40

Final exam (60%), two midterm exams (15%+20%), computational exercises to handle (5%). It is requiered to obtain 4 points out of 10 in the final exam.

BASIC BIBLIOGRAPHY

- A. Agresti Categorical Data Analysis, Wiley, 2002

- J. Tukey Exploratory Data Analysis, Addison-Wesley, 1977