Universidad Zaragoza

Academic Year/course: 2022/23

29507 - Probability and Statistics

Syllabus Information

Academic Year: 2022/23
Subject: 29507 - Probability and Statistics
Faculty / School: 175 - Escuela Universitaria Politécnica de La Almunia
Degree: 625 - Bachelor's Degree in Industrial Processes' Data Engineering
ECTS: 6.0
Year: 2
Semester: First semester
Subject Type: Basic Education
Module:

1. General information

1.1. Aims of the course

In this subjet the student is introduced to the second large block of statistics, statistical inference based on the calculation of probabilities.

Starting from the possibility of occurrence of an event, the concept of probability associated with experiments in which there are some uncertainties about what will happen. This is the starting point for the study of probability, which provides solutions to different problems. Probability Theory is the basis for the study of Statistical Inference, in which mathematical models will be provided to help us know the different random variables from the data of a sample.

The final objective is that the student integrates the basic knowledge that, together with the skills with the tools used in the subject, will be able to make decisions, as well as the elaboration of reports for the professional development as a data engineer.

These approaches and objectives are aligned with the following Sustainable Development Goals (SDGs) of the United Nations Agenda 2030 (https://www.un.org/sustainabledevelopment/es/) so that the acquisition of the learning outcomes of the subject provides training and competence to contribute to some extent to its achievement: Goal 4 Quality education.

1.2. Context and importance of this course in the degree

The subjet is compulsory and is part of the basic training for students of Data Engineering in Industrial Processes. This course is studied in the first semester of the second year of the curriculum, which assumes that the student has already acquired training in the learning outcomes of Fundamentals of Mathematics I and II.

The study of Probability and Statistics provides skills in tools that will be very useful in different assignments in later courses. They could be used for the global understanding and decision making in situations of uncertainty in subjects, such as digital signal processing, quality control and improvement, etc. For all these reasons, this course is a basic tool in the training of a data engineer.

1.3. Recommendations to take this course

It is recommended that the student has basic knowledge of differential and integral calculuations. It is also advisable to have some familiarity with symbolic and numerical calcululations programs.

In order to get the most out of the course, regular attendance to both classes, theoretical and practical, besides participation in them is important. This is how the students will be able to acquire, in a sequential way, the theoretical knowledge and skills with the computer tool used.

2. Learning goals

2.1. Competences

Core Competencies:
CB2 - That the students know how to apply their knowledge to their work or vocation in a professional manner and possess

the competencies, that are usually demonstrated through the elaboration and defense of arguments and the resolution of problems within their area of study.
CB4 - Students will be able to convey information, ideas, problems and solutions to both, specialized and non-specialized, audiences.
CB5 - That students have developed those learning skills necessary to undertake further studies with a high degree of autonomy further studies with a high degree of autonomy.

General competences:
CG03 - Apply techniques for data acquisition, management and processing in Engineering.
CG06 - Build solutions derived from data analysis that optimize production processes in industry.

Transversal competences:
CT03 - Search, select and responsibly manage information and knowledge.
CT04 - Develop critical thinking and reasoning.
CT05 - Communicate results effectively.
CT07 - Analyze and solve problems autonomously, adapt to unforeseen situations and make decisions.

Specific competences:
CE04 - Solve mathematical problems that may arise in engineering.

2.2. Learning goals

The student, in order to pass this course, must demonstrate the following results....

To know the reflexive use of symbolic and numerical calculation tools.
To master the modeling of engineering environments under stochastic nature by means of random variables, as well as the performance of calculations in situations of uncertainty.
Possess skills of scientific-mathematical thinking, which allow them to ask and answer certain mathematical questions.
Have the ability to handle mathematical language; in particular, symbolic and formal language.

2.3. Importance of learning goals

In the subjet of Probability and Statistics, students will learn and practice statistical support in decision making about a population based on a certain number of sample observations.

Students will develop the skills to work with different data sets, to recognize models on which to rely in order to reach efficient conclusions in the face of uncertainty.

In the professional work of a data engineer, he/she will have to report on a data set in which he/she will need a descriptive study such as the inference techniques that will be covered in this course.

3. Assessment (1st and 2nd call)

3.1. Assessment tasks (description of tasks, marking system and assessment criteria)

Continuous assessment system:

Exams: During the course two main exams will be conducted. They will focus on theoretical and / or practical aspects of the subject:

- Written test 1: Week 8 will be held and will focus on the subject in the first 8 weeks of the course. Its weight in the final grade will be 30%.
- Written test 2: Week 15 will be made and will focus on the subject in the second half of the course. Its weight in the final grade will be 30%.

Participatory controls: Throughout the course, students will perform six participatory controls together valued at 20% of the final grade, which consist of conducting practical exercises.

Applied work: Throughout the course, students will perform two works applied to matters of the subject, its valuation is 20% of the final grade.

Overall Assessment: Students who have not passed the subject with the system of continuous assessment, have to pass a global exam whose weight in the final grade will be 80%. Also, they must submit the two applied work required during the course.

Evaluation criteria

In the written tests, controls and work participation will be evaluated: Practical exercises must be properly raised. If a computer program is used in solving exercises, the code used and in any case the results are clearly explained be detailed. The probability distribution assigned to each random variable must be duly justified, identifying the value or values ??of the model parameters. Hypothesis testing will arise clear and defined manner.

4. Methodology, learning tasks, syllabus and resources

4.1. Methodological overview

The proposed methodology seeks to promote student work and continued focus on the more practical aspects of statistics: working with real data.

In order to achieve this goal all practical classes (2 hours per week) will be held in the computer room, using R programming language. The theoretical explanations of the concepts of the subject (2 hours weekly) will be reinforced by examples or case studies analyzed with the computer

The approach, methodology and assessment of this guide are intended to be the same for any teaching scenarios. They will be adapted to the social-health situation at any particular time, as well as to the instructions given by the authorities concerned.

4.2. Learning tasks

The course includes the following learning tasks:

The course is organized with 4 hours of class a week for the 15 weeks of the semester. Some of these hours are taught in the computer room, and in them, the teacher explains the more practical aspects of the subject, which are reinforced with practical work by using statistical analysis programs.

Tutored self-employment: 2 hours per week for 15 weeks where the student works autonomously in the computer room in performing work.

Personal work: 60 hours

4.3. Syllabus

The course will address the following topics:

- **Probability**: Elements of probability: Event. Probability. Probabilistic space. Conditional probability. Total probability theorem. Bayes theorem. Random variables: Discrete: mass function, distribution function. continuous: density function, distribution function. Expected value: mean, variance. Discrete distributions: Bernoulli trials, binomial, Poisson distribution. Continuous distributions: normal, exponential, uniform distribution, beta, gamma distribution. Multivariate random Variables: probability function, expected value, covariance, independence V.A., distribution chi-square, Student's t, F Snedecor.
- **Introduction to reliability theory**: Quality and reliability function, reliability and risk function. Exponential distribution, Weibull distribution.
- **Inference**: Parameter estimation: Population and sample. Random, stratified, cluster and systematic sampling. Statistics and Estimator. Simulation. Method of moments, maximum likelihood method. Desirable properties of estimators: bias, efficiency, consistency. Point estimation and interval. Fisher theorem. Central limit theorem. Confidence intervals.
- **Hypothesis testing**: null and alternative hypotheses. Error type I and II, significance level, power of contrast. unilateral and bilateral tests. P-value.
- **The goodness of fit**: Kolmogorov-Smirnov test. Multivariate linear regression model: Parameter Estimation. stepwise procedures: backward and forward. Akaike index. Residue analysis.
- **Time series**: Components. Combination of components. Trend analysis. Seasonal variations.

4.4. Course planning and calendar

Since the subject consists of 6 ECTS credits, and each consists of 25 hours divided into 10 hours of supervised work and 15 hours of autonomous work, activities of classroom learning (lectures, practical classes and seminars) and activities continuous assessment (participatory controls and written tests) will occupy 60 hours during the semester.

Other classroom activities as personal and tutorials non-contact as the study for the assimilation of concepts and techniques, practice for familiarization with computer tools, problem-solving and test preparation, will require 90 hours of independent student work.

All these activities should add the 150 hours required to achieve learning outcomes pursued the subject. The concrete and comprehensive planning of the course it will be informed to students at the beginning of the course. Also from the beginning of the course, it will be set the dates of the official announcements from the school management.

4.5. Bibliography and recommended resources

http://psfunizar10.unizar.es/br13/egAsignaturas.php?codigo=29507