Course guides
250334 - PROBESTAD - Probability and Statistics

Last modified: 06/10/2020

Unit in charge: Barcelona School of Civil Engineering
Teaching unit: 751 - DECA - Department of Civil and Environmental Engineering.

Degree: BACHELOR'S DEGREE IN GEOLOGICAL ENGINEERING (Syllabus 2010). (Compulsory subject).

Academic year: 2020 **ECTS Credits:** 4.5 **Languages:** Spanish

LECTURER

Coordinating lecturer: MARIA CAMINO TEOFILA BALBUENA MARTINEZ

Others: MARIA CAMINO TEOFILA BALBUENA MARTINEZ, MARÍA ISABEL ORTEGO MARTÍNEZ

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
4055. Understanding of the concepts of randomness of physical, social and economic phenomena, and uncertainty

Transversal:
592. EFFICIENT ORAL AND WRITTEN COMMUNICATION - Level 2. Using strategies for preparing and giving oral presentations. Writing texts and documents whose content is coherent, well structured and free of spelling and grammatical errors.
595. TEAMWORK - Level 2. Contributing to the consolidation of a team by planning targets and working efficiently to favor communication, task assignment and cohesion.
599. EFFECTIVE USE OF INFORMATI0N RESOURCES - Level 3. Planning and using the information necessary for an academic assignment (a final thesis, for example) based on a critical appraisal of the information resources used.
602. SELF-DIRECTED LEARNING - Level 3. Applying the knowledge gained in completing a task according to its relevance and importance. Deciding how to carry out a task, the amount of time to be devoted to it and the most suitable information sources.
584. THIRD LANGUAGE. Learning a third language, preferably English, to a degree of oral and written fluency that fits in with the future needs of the graduates of each course.

TEACHING METHODOLOGY

The course consists of 2 hours per week of classroom activity (large size group) and 1 hour weekly with half the students (medium size group).

The 2 hours in the large size groups are devoted to theoretical lectures, in which the teacher presents the basic concepts and topics of the subject, shows examples and solves exercises.

The 1 hour in the medium size groups is devoted to solving by using computer practical problems and to laboratory practice with greater interaction with the students. The objective of these practical exercises is to consolidate the acquired knowledge both general and specific.
.

Support material in the form of a detailed teaching plan is provided using the virtual campus ATENEA: content, program of learning and assessment activities conducted and literature.

LEARNING OBJECTIVES OF THE SUBJECT

Students will learn to analyse, represent and process data, solve probability and statistical problems and learn how these techniques are used in science, technology and applied technological problems.

Upon completion of the course, students will be able to: 1. Use specific software and the techniques studied during the course to analyse the data from a geological engineering problem;
2. Use specific software to carry out multiple linear regression analysis;
3. Carry out data simulations and transformations of random variables and study the resulting distributions.

Data analysis; Regression models, parameter estimation; Probability and uncertainty; Definition and interpretation of random variables and operations between random variables; Probabilistic models: Bernoulli distributions, Poisson distributions and other distributions; Asymptotic probabilistic models: Normal distribution to transformed distributions; Return period estimation; Parameter estimation, maximum likelihood method, interval estimation; Hypothesis testing and goodness-of-fit testing; Bayesian estimation and statistical evaluation of regression models

STUDY LOAD

Type	Hours	Percentage
Hours small group	14,0	12.44
Hours medium group	11,0	9.78
Hours large group	20,0	17.78
Guided activities	4,5	4.00
Self study	63,0	56.00

Total learning time: 112.5 h

CONTENTS

Item 1. Probability

Description:
Introduction. Basic concepts and properties. Independence
events. Conditional probability. Multiplication rule.
Theorem of total probability and Bayes. Exercises.
Problems of probability, independence, total probability, Bayes. Combinatorial review.

Full-or-part-time: 8h 24m
Theory classes: 2h
Practical classes: 1h 30m
Self study : 4h 54m

Practice1: Data Analysis.

Description:
Introduction to the software: Excel, R, SPSS.
Definition of variables and data entry. Classification of variables. Frequency tables, histograms and bar graphs. Transform procedure: Recode. Main statistics. Procedure explore: Boxplot Graphs; outliers. Scale problems

Full-or-part-time: 2h 24m
Laboratory classes: 1h
Self study : 1h 24m

Item 2. Random variables

Description:
Distribution functions. Summary measures
probability distribution associated to a random variable: Expected value, median, variance, standard deviation, Chebychev inequality.
In an Excel spreadsheet we are working the issue of release of 2 dice: Possible cases. Associated random variables, corresponding distribution functions, associated measures.

Full-or-part-time: 7h 11m
Theory classes: 2h
Laboratory classes: 1h
Self study : 4h 11m

Item 3. Some discrete probability models.

Description:
Discrete random variables. Common discrete probability models: binomial distribution. Poisson distribution. Binomial and Poisson approximation.

Using R: Charts binomial and Poisson probability for different parameters. Solving exercises.

Full-or-part-time: 9h 36m
Theory classes: 2h 30m
Practical classes: 1h 30m
Self study : 5h 36m

Item 4. Some continuous probability models.

Description:
Continuous Random Variables: Density function.
Exponential and Gamma distribution; return periods. Normal distribution. Lognormal distribution.
Continuous distributions with Excel. Solving problems with exponential variables (durations, average life, etc) and gamma variables.
Continuous distributions with R. Solving problems at exponential and normal scenarios.

Full-or-part-time: 13h 12m
Theory classes: 3h
Practical classes: 2h 30m
Self study : 7h 42m

Practice 2: Description of models

Description:
To describe in detail in an excel spreadsheet the experiment of tossing a coin 10, 100, 1000 times. Approximating a binomial and Poisson distribution through Normal distibution.

Full-or-part-time: 2h 24m
Laboratory classes: 1h
Self study : 1h 24m

Controls

Full-or-part-time: 9h 36m
Laboratory classes: 4h
Self study : 5h 36m

Tema 5: Sampling and sampling distribution.

Description:
Random sample. Distribution of the mean and sample variance: Central limit theorem.
Weak law of large numbers. Chi-quadrat Distribution. Order statistics. Introduction of extreme models Weibull, Gumbel and Frechet.
In an excel spreadsheet exercises are solved raised about the mean and sample variance.
In an Excel spreadsheet models are plotted ends of the three types. Parameters are verified and checked what kind of problems should be used each of the three models.

Full-or-part-time: 12h
Theory classes: 3h
Practical classes: 2h
Self study : 7h

Item 6. Parametric point estimation

Description:
Methods of finding estimators: Method of moments. Maximum likelihood method. Mean square error.

Are solved exercises need to know the estimation methods studied in theory: Method of moments, maximum likelihood method
In an Excel spreadsheet 100 or more simulated data. Then create a macro to generate a button to automatically select different samples of size 30. Average variable is built and verified to follow a normal distribution. This illustrates the central limit theorem.
Eventually performed the same exercise with R and compared the two methodologies.

Full-or-part-time: 14h 23m
Theory classes: 2h
Practical classes: 2h
Laboratory classes: 2h
Self study : 8h 23m

Item 7. Confidence intervals and hypothesis test

Description:
1. Concept of confidence interval. 2. Pivots in Normal data: The Case of a sample and two independent samples. 3. Confidence Intervals for normal populations.
4. Other confidence intervals. 5. Minimum sample size.
Parametric hypothesis testing. Errors of types I and II. Significance of a contrast. Power of a contrast. Normal populations: contrasts on the mean and variance. Comparison of means of independent normal populations. Some additional considerations: Election of the null hypothesis, choice of level
significance, the p-value or significance level of
sample. Frequently Hypothesis.
Suggests exercises to calculate confidence intervals on parameters. They relate to the contrast raised on parametric assumptions. Is seen as a pivot in a statistical transforna when using the information given by the null hypothesis, or (primary)
Graphics P-P and Q-Q. The Kolmogorov-Smirnov. Contrast
Pearson Chi-square, contingency tables.

Full-or-part-time: 14h 23m
Theory classes: 3h 30m
Practical classes: 1h 30m
Laboratory classes: 1h
Self study : 8h 23m

Lab 3: Linear model.

Description:
Scatterplot. Bivariate correlation. Correlation
part. Simple linear regression. Multiple linear regression.
Analysis of variance.
1. Choose a response variable and once as explanatory
and graphically generate the regression line. 2. Run
the simple linear regression procedure and interpret
results. Analysis of variance and hypothesis.
3. Variables to generate predicted values and residual values.
4. Conduct analysis of the waste.

Full-or-part-time: 14h 23m
Theory classes: 2h
Laboratory classes: 4h
Self study : 8h 23m

GRADING SYSTEM

The course rating is derived from scores obtained from three controls and other activities related to evaluation.

The first two controls are multiple choice tests that aim to revise and consolidate the knowledge acquired. The third control will be carried out at the end of the semester and will consist of individual resolution of one or more problems in the classroom. Other evaluations may be conducted both individually and in groups. These will include the drafting of a statistical report on a data set.

The final mark is derived from a weighted average of the three controls and directed activities.

Criteria for re-evaluation qualification and eligibility: Students that failed the ordinary evaluation and have regularly attended all evaluation tests will have the opportunity of carrying out a re-evaluation test during the period specified in the academic calendar. Students who have already passed the test or were qualified as non-attending will not be admitted to the re-evaluation test. The maximum mark for the re-evaluation exam will be five over ten (5.0). The non-attendance of a student to the re-evaluation test, in the date specified will not grant access to further re-evaluation tests. Students unable to attend any of the continuous assessment tests due to certifiable force majeure will be ensured extraordinary evaluation periods.

These tests must be authorized by the corresponding Head of Studies, at the request of the professor responsible for the course, and will be carried out within the corresponding academic period.

EXAMINATION RULES.

Failure to perform some of the three controls in the scheduled period will result in a mark of zero in that activity.

BIBLIOGRAPHY

Basic:

- Canavos, G.C. Probabilidad y estadística: aplicaciones y métodos. México: McGraw Hill, 1999. ISBN 9684518560.
- Horra, J. de la. Estadística aplicada. 3a ed. Madrid: Díaz de Santos, 2003. ISBN 8479785543.
- Peña, D. Estadística : modelos y métodos : vol. 1 : fundamentos. 2a ed. rev. Madrid: Alianza, 1986-1991. ISBN 8420681091 (VOL.1).

Complementary:

- Mood, A. M; Graybill, F.A; Boes, D.C. Introduction to the theory of statistics. 3rd. New York: McGraw-Hill, 1974. ISBN 0070428646.