

Course guide

270009 - PE - Probability and Statistics

Last modified: 30/01/2024

Unit in charge: Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.

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

Academic year: 2023 **ECTS Credits:** 6.0 **Languages:** Catalan, Spanish, English

LECTURER

Coordinating lecturer: JOSÉ ANTONIO GONZÁLEZ ALASTRUE

Others:

Primer quadrimestre:

NIHAN ACAR DENIZLI - 12
ERIK COBO VALERI - 31
JORDI CORTÉS MARTÍNEZ - 32
JOSÉ ANTONIO GONZÁLEZ ALASTRUE - 11
KLAUS GERHARD LANGOHR - 22
MIREIA LOPEZ BELTRAN - 41
BHUMIKA ASHVINBHAI PATEL - 13
NURIA PEREZ ALVAREZ - 23
ROSER RIUS CARRASCO - 21, 51

Segon quadrimestre:

ERIK COBO VALERI - 11
JOSÉ ANTONIO GONZÁLEZ ALASTRUE - 11
MIREIA LOPEZ BELTRAN - 12
ROSER RIUS CARRASCO - 41

PRIOR SKILLS

Students need to be sufficiently knowledgeable about algebra and mathematical analysis to be able to assimilate concepts related to the algebra of sets, numerical series, functions of real variables of one or more dimensions, differentiation and integration. They should also be able to understand technical English.

REQUIREMENTS

- Prerequisite M1
- Prerequisite M2

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

CT1.2A. To interpret, select and value concepts, theories, uses and technological developments related to computer science and its application derived from the needed fundamentals of mathematics, statistics and physics. Capacity to solve the mathematical problems presented in engineering. Talent to apply the knowledge about: algebra, differential and integral calculus and numeric methods; statistics and optimization.

CT8.3. To demonstrate knowledge and be able to apply appropriate techniques for modelling and analysing different kinds of decisions.

Generical:

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

TEACHING METHODOLOGY

The subject is based on the active learning of the student, guided and guided by the teacher with the help of e-status (interactive platform that, with data individualized by the exercises, allows to evaluate and learn thanks to a feedback immediate).

The teaching scheme of the 4 specific blocks consists of the repetition of cycles based on: exposition of theoretical concepts, numerical resolution of exercises, support for exercises with R (using laptops in the classroom, or in the laboratory), follow-up tests by of the teacher, and autonomous practice of exercises.

The application block develops the transversal competence with the application in group to a specific case contributed by the students, analyzed with R, under the direction of the teacher.

LEARNING OBJECTIVES OF THE SUBJECT

- 1.2. Define and calculate probabilities for a random experience.
2. Calculate the conditional and joint probabilities and detect whether there is (in)dependence for a random experience with two variables and apply Bayes' theorem to locating the conditional probabilities for the other variable.
3. Graphically represent a random experience.
4. Calculate mean and variance for given probability and distribution functions for a discrete random variable.
5. Identify the most appropriate theoretical model to represent a given random variable from among the following: Bernoulli, binomial, Poisson, Geometric, Normal, uniform and exponential.
6. Calculate cumulative probabilities for certain values from the parameter for theoretical models with the help of tables or R; conversely, locate the random variable values from the desired cumulative probabilities.
7. Calculate and interpret the covariance and correlation values for two random variables.
8. Calculate, using sample data, statistics that reflect central tendency (mean) and dispersion (variance and standard deviation).
9. From sample indicators, obtained from a s.r.s., he/she will compute confidence intervals for certain parameters. For example: from the mean, the standard deviation and the sample size of a variable with Normal distribution, the student will calculate the CI95%.
10. Based on a hypothesis and the sample mean and standard deviation for a normally distributed variable, calculate the P-value and justify the evidence against the hypothesis.
11. From the data of a comparative test (e.g., performance of two computer products), the student will use the confidence interval to obtain a wide range of possible values of the difference in the outcome.
12. Using the summary of the model, obtain and interpret the estimators of the model, compute and interpret the R-squared coefficient, obtain the estimators of the uncertainty of the estimate and build a CI for the population values.
13. Make predictions and assess their degree of uncertainty using summary data from the adjusted model.
14. Based on the graphs of the adjusted model, analyze the premises of the model and, if necessary, propose transformations of the variables.
15. Design a prediction study, collect data and analyse and interpret results.
16. Identify, for a deterministic process, variability sources and magnitudes.
18. Design a comparative test of computer products, collect data and analyse and interpret results.

STUDY LOAD

Type	Hours	Percentage
Guided activities	6,0	4.00
Hours large group	15,0	10.00
Hours small group	30,0	20.00
Self study	84,0	56.00
Hours medium group	15,0	10.00

Total learning time: 150 h

CONTENTS

Block A. Probability and random variables

Description:

Random experiment. Probability, conditional probability, joint probability. Definition of random variable and types. Probability function, probability density function and probability distribution function. Joint probability function. Indicators: expectation, variance, standard deviation, covariance, correlation. Independence between two random variables.

Block B. Probabilistic models

Description:

Parameterised theoretical models of random variables. Direct and inverse probabilities computation, with R. Introduction to simulation. Sample mean distribution. Central Limit Theorem, Normal approximations.

Block C. Basis of statistics

Description:

Population and sample. Parameter, statistic and estimator. Bias of an estimator. Confidence interval for a parameter, and for the difference of two parameters. Hypothesis test

Block D. Statistical models and forecasting

Description:

Comparació de dos grups, disseny aparellat, mostres independents. Model lineal. Indicadors de la qualitat de l'ajustament. Validació de les premisses. Introducció a la ciència de dades. Estudis experimentals i observacionals. Ètica de la ciència, research waste.

Block T. Application.

Description:

Identifying sources of variability in computer processes. Design of a study with planning of the goal, data collection, statistical analysis with R and results interpretation.

ACTIVITIES

Block A activities. Probability and random variables

Description:

Locate probability and statistics, especially in the IT field. Provide a grounding in probability. Be able to calculate and analyze joint and conditional probabilities. Analyze whether there is independence or not. Define random variable (RV), discrete and continuous RV. Define probability function, cumulative probability function and joint probability function. Relate RV indicators to sample indicators.

Specific objectives:

1, 2, 3, 4, 7, 16

Related competencies :

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Full-or-part-time: 27h

Theory classes: 6h

Practical classes: 6h

Self study: 15h

Block B activities. Probabilistic models

Description:

Define the theoretical, discrete and continuous models typically used in the IT field and their characteristics and parameters.

Specific objectives:

5, 6

Related competencies :

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Full-or-part-time: 27h

Theory classes: 6h

Practical classes: 6h

Self study: 15h

Mid-semester exam 1

Description:

Mid-semester exam consisting of problems corresponding to topics 1 to 3 (learning objectives 1 to 8).

Specific objectives:

1, 2, 3, 4, 5, 6, 7, 16

Related competencies :

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Full-or-part-time: 8h

Guided activities: 2h

Self study: 6h

Block C activities. Basis of statistics

Description:

Basic population, sampling, parameter and estimator concepts. Introduction to statistics; definition and linking of confidence intervals (CI) and hypothesis testing (HT).

Specific objectives:

8, 9, 10, 11

Full-or-part-time: 27h

Theory classes: 6h

Practical classes: 6h

Self study: 15h

Block D activities. Statistical models and forecasting

Description:

Definition of statistical models. Analysis of variability. Paired design and independent samples. Linear model. Validation of premises, possible transformations, predictions. Some models for data science. Implications of research.

Specific objectives:

12, 13, 14

Full-or-part-time: 27h

Theory classes: 6h

Practical classes: 6h

Self study: 15h

Application activities

Description:

Identify problems in the IT field for a probability or statistical study. Design a study, collect data and analyse and interpret results. Summarise conclusions critically.

Specific objectives:

15, 16, 18

Related competencies :

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Full-or-part-time: 24h

Theory classes: 6h

Practical classes: 6h

Self study: 12h



Mid-semester exam 2

Description:

Mid-semester exam consisting of problems corresponding to topics 4 to 6 (learning objectives 9 to 17).

Specific objectives:

8, 9, 10, 11, 12, 13, 14, 15, 18

Related competencies :

G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Full-or-part-time: 8h

Guided activities: 2h

Self study: 6h

Final Exam

Description:

Covers all the topics.

Full-or-part-time: 2h

Guided activities: 2h

GRADING SYSTEM

The qualification of the subject is obtained by means of the continuous evaluation (AC) during the 15 weeks of class or with the final exam (EF).

PE is divided into 5 topics or blocks: 4 specific (A, B, C, D) and one cross-disciplinary applications topic of statistical application (T).

Each block results in a Block Note (NB.i, i = A,B,C,D,T). The following formula is applied in the AC:

$$AC = [3 \text{ NB.A} + 3 \text{ NB.B} + 3 \text{ NB.C} + 3 \text{ NB.D} + 2 \text{ NB.T}] / 14$$

If $AC \geq 5$, the student can be released from the final exam.

Please note that the EF may consider the grade for the transversal competence:

$$EF = \max\{ ef, (12 ef + 2 \text{ NB.T}) / 14 \}$$

where "ef" is the actual grade for the final exam.

The course grade of the subject PE is $\max(AC, EF)$.

The qualification of the transversal competence is:

A and $\text{NB.T} \geq 8.5$; B for 6.5

Calculating NB.i grades:

- the first 4 have an assessment based on a Block Problem (PB.i, i = A,B,C,D) in a mid-term exam out of class hours. Usually there are 2 tests that give rise to the grades for the 4 blocks.

In addition, a Block Monitoring factor (SB.i, i = A,B,C,D) is obtained for each of the four theoretical blocks, based on 3 tests: 2 written tests done in the classroom, and a mark for problems solved outside the classroom. The SB.i factor increases the grade for the corresponding Block Problem (PB.i) to obtain the Block Grade according to:

$$\text{NB.i} = \min(10, \text{PB.i} * \text{SB.i}) \text{ for } i = A,B,C,D$$

(SB.i factor is $1 + \sum p_j$, where p_j is a number between 0 and 0.05, coming from the different block monitoring tests; the exact number of tests may be less than 3 if there are unforeseen changes to the school calendar, with consequent loss of classes).

- The T-Block grade (NB.T) is calculated on the basis of two reports and a final presentation.



BIBLIOGRAPHY

Basic:

- González, J.A.; Cobo, E.; Muñoz, P.; Martí-Recober, M. Estadística per a enginyers informàtics. Edicions UPC, 2008. ISBN 9788483019535.
- Baron, M. Probability and statistics for computer scientists. Third edition. CRC Press, 2019. ISBN 9781138044487.
- Wonnacott, T.H.; Wonnacott, R.J. Introducción a la estadística. 2a ed. Limusa, 1997. ISBN 9681845099.

Complementary:

- Horgan, J.M. Probability with R: an introduction with computer science applications. Wiley, 2009. ISBN 9780470280737.
- Peña, D. Estadística: modelos y métodos: vol.1: fundamentos. 2a ed rev. Alianza, 1986. ISBN 8420689931.
- Ross, S.M. Probability models for computer science. Harcourt/Academic Press, 2002. ISBN 9780125980517.
- Matloff, N.S. From algorithms to Z-Scores: probabilistic and statistical modeling in computer science. University Press of Florida, 2009. ISBN 9781616100360.

RESOURCES

Hyperlink:

- <https://estatus.upc.edu>- <https://www-eio.upc.edu/teaching/pe/>- https://www-eio.upc.edu/~josean/shinyweb/jag_shiny.php