Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Coordinating unit: 270 - FIB - Barcelona School of Informatics

Teaching unit: 749 - MAT - Department of Mathematics

Academic year: 2019

Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Teaching unit Compulsory)

ECTS credits: 7,5 Teaching languages: Catalan, Spanish

Prior skills

The contents of the previous subjects in the Degree.

Degree competences to which the subject contributes

Basic:

CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.

CB3. That students have the ability to gather and interpret relevant data (usually within their area of ??study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy

Specific:

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

Generical:

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

Transversal:

CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Teaching methodology

Theory:
Lectures develop the theory and include illustrative examples.

Problems:
The students have in advance the list of problems relevant to the topic being developed in theory. They had the opportunity to try to solve problems before the problems class. They require the teacher's help in the points where they have encountered difficulties. The teacher solves these questions in the blackboard and develops the full solution of some problems that he or she considers that are especially challenging.

Laboratory:
The teacher introduces the R language during the course, with special emphasis on random variables simulation tools, descriptive statistics and univariate statistical inference.

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Learning objectives of the subject

1.At the end of the course, students will know the definition of probability and their properties, and will apply them to solve probability calculation problems.
2.At the end of the course students will know how to use the concept of random variable to formalize and solve probability calculation problems.
3.At the end of the course students will know how to simulate complex random phenomena with the computer and deduce approximate values of amounts of interest (probabilities, characteristics of random variables) that are difficult to calculate analytically.
4.At the end of the course students will know the most common probabilistic distributions and will be able to recognize situations where they are used to model real phenomena.
5.At the end of the course, students will know how to calculate distributions and expected expectations and use them in prediction.
6.At the end of the course, students will know whether two random variables are independent, and if they are not, they will be able to measure the linear correlation coefficient.
7.At the end of the course, students will know the Law of the Great Names and the Central Limit Theorem.
8.At the end of the course students will know the concepts of population, sample, parameter and estimator, and know the basic properties.
9.At the end of the course students will know the basic tools of descriptive statistics and will know how to apply them.
10.At the end of the course, students will know the basics of statistical inference (timely estimation, confidence intervals and hypothesis tests) and will know how to calculate them in real situations

Study load

Total learning time: 187h 30m	Hours large group:	45h	24.00%
	Hours small group:	30h	16.00%
	Guided activities:	0h	0.00%
	Self study:	112h 30m	60.00%

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Content

Probability spaces and random variables

Degree competences to which the content contributes:

Description:
Random experiences. Algebra of events. Probability space. Conditional probability. Independence of events. Bayes' Theorem. Simulation of random experiments.

Random variables

Degree competences to which the content contributes:

Description:
Definition of random variable. Probability distribution function. Discrete random variables (probability function) and continuous variables (probability density function). Expectation and moments. Models of usual distributions. Simulation of random variables.

Random vectors

Degree competences to which the content contributes:

Description:
Multidimensional distributions. Independence. Conditioned distributions. Covariance and correlation. Expectation and covariance matrix. Conditioned expectation. Multinomial distribution. Multivariate normal distribution.

Sum of random variables

Degree competences to which the content contributes:

Description:
Distribution of the sum. Markov's, Chebyshev's and Chernoff's Inequalities. Law of Large Numbers. Central Limit Theorem.

Population and sample

Degree competences to which the content contributes:

Description:
Random sample. Empirical distribution function. Parametric statistical models. Parameters and estimators. Descriptive statistics.

Point estimation

Degree competences to which the content contributes:

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Description:
Method of moments. Maximum likelihood. Properties of the estimators (bias, variance, mean square error, consistency, adequacy, efficiency).

Confidence Intervals

Degree competences to which the content contributes:

Description:
Confidence Intervals

Hypothesis testing

Degree competences to which the content contributes:

Description:
Type I and Type II Errors. Power. Relation with confidence intervals. The likelihood ratio test.

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Planning of activities

Developing the Topic "Probability and random variables"	Hours: 17h Theory classes: 6h Practical classes: 2h Laboratory classes: 2h Guided activities: 0h Self study: 7h
Description: Developing the Topic "Probability and random variables" Specific objectives: 1, 3	

Developing the Chapter "Random variables"	Hours: 17h Theory classes: 6h Practical classes: 2h Laboratory classes: 2h Guided activities: 0h Self study: 7h
Description: Developing the Chapter "Random variables" Specific objectives: 2, 3, 4, 9	

Developing the Chapter "Random vectors"	Hours: 18h 48m Theory classes: 7h 30m Practical classes: 3h Laboratory classes: 2h Guided activities: 0h Self study: 6h 18m
Description: Developing the Chapter "Random vectors" Specific objectives: 3, 5, 6, 9	

Developing the Chapter "Sum of random variables"	Hours: 12h 48m Theory classes: 4h 30m Practical classes: 1h 30m Laboratory classes: 1h 30m Guided activities: 0h Self study: 5h 18m

270206 - PIE1 - Probability and Statistics 1

Description:
Developing the Chapter "Sum of random variables"

Specific objectives:
3, 5, 7

Developing the Chapter "Population and sample"	Hours: 12h 42m Theory classes: 4h 30m Practical classes: 1h 30m Laboratory classes: 1h 30m Guided activities: 0h Self study: 5h 12m

Description:
Developing the Chapter "Population and sample"

Specific objectives:
3, 8, 9

Developing the Chapter "Point estimation"	Hours: 18h 42m Theory classes: 7h 30m Practical classes: 2h Laboratory classes: 3h Guided activities: 0h Self study: 6h 12m

Description:
Developing the Chapter "Point estimation"

Specific objectives:
10

Developing the Chapter "Confidence intervals"	Hours: 7h 30m Theory classes: 3h Practical classes: 1h Laboratory classes: 1h Guided activities: 0h Self study: 2h 30m

Description:
Developing the Chapter "Confidence intervals"

Specific objectives:
10

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Developing the Chapter "Hypothesis testing"	Hours: 15h 30m Theory classes: 6h Practical classes: 2h Laboratory classes: 2h Guided activities: 0h Self study: 5h 30m
Description: Developing the Chapter "Hypothesis testing"	

Final examination	Hours: 33h Guided activities: 3h Self study: 30h
Description: Final examination Specific objectives: 1, 2, 4, 5, 6, 7, 8, 10	

Mid-term examination	Hours: 12h Guided activities: 2h Self study: 10h
Description: Mid-term examination Specific objectives: 1, 2, 4, 5, 6	

Mid-term lab examination	Hours: 11h Guided activities: 1h Self study: 10h
Description: Mid-term lab examination Specific objectives: 3, 9	

Final lab examination	Hours: 11h 30m Guided activities: 1h 30m Self study: 10h
Description: Final lab examination	

Last update: 19-07-2019

270206 - PIE1 - Probability and Statistics 1

Specific objectives:
3, 9, 10

Qualification system

A partial exam (EP) and a final exam (EF). Each of them has a part of theory and problems, and another part of laboratory.

During the course (approximately biweekly), short activities (ACT) will be proposed (a problem to solve analytically or with the computer).

The final grade (NF) is computed as

NF = 0.65 * EF + 0.25* max(EP,EF) + 0.10*max(ACT,EF)

The re-evaluation exam grade (ER) replaces the 100% of the final exam grade. So the final grade after re-evaluation (NFreav) will be

NFreav = 0.65 * ER + 0.25* max(EP,ER) + 0.10*max(ACT,ER)

Bibliography

Basic:

Baumer, B..S.; Kaplan, D.T.; Horton, N.J. Modern data science with R. Boca Raton: Taylor & Francis CRC Press, 2017. ISBN 9781498724487.

DeGroot, M.H.; Schervish, M.J. Probability and statistics. 4th ed. Boston: Pearson, 2012. ISBN 9780321709707.

Baron, M. Probability and statistics for computer scientists. 3rd ed. Boca Raton, FL: CRC Press, 2019. ISBN 9781138044487.

Evans, M.J.; Rosenthal, J.S. Probability and statistics: the science of uncertainty. 2nd ed. New York: W.H. Freeman and Company, 2010. ISBN 9781429224628.

Bertsekas, D.P.; Tsitsiklis, J.N. Introduction to probability. 2nd ed. Belmont, Massachussets: Athena Scientific, 2008. ISBN 9781886529236.

Complementary:

Pitman, J. Probability. New York: Springer, 1993. ISBN 0387979743.

Mitzenmacher, M.; Upfal, E. Probability and computing: randomized algorithms and probabilistic analysis. Cambridge: Cambridge University Press, 2005. ISBN 0521835402.

Grinstead, C.M.; Snell, J.L. Introduction to probability. 2nd rev. ed. Providence (R.I.): American Mathematical Society, 1997. ISBN 0821807498.

Others resources:

Hyperlink

https://math.dartmouth.edu/~prob/prob/prob.pdf