270009 - PE - Probability and Statistics

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2019
Degree: BACHELOR'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2010). (Teaching unit Compulsory)
ECTS credits: 6
Teaching languages: Catalan, Spanish

Teaching staff

Coordinator:
- José Antonio González Alaстрue (jose.a.gonzalez@upc.edu)

Others:
- Erik Cobo Valeri (erik.cobo@upc.edu)
- Joan Garcia Subirana (joan.garcia-subirana@upc.edu)
- Jordi Cortés Martinez (jordi.cortes-martinez@upc.edu)
- Jordi Escayola Mansilla (jordi.escayola@upc.edu)
- Klaus Gerhard Langohr (klaus.langohr@upc.edu)
- Mireia Lopez Beltran (mireia.lopez.beltran@upc.edu)
- Miriam Mota Foix (miriam.mota@upc.edu)
- Nuria Perez Alvarez (nuria.perez@upc.edu)
- Roser Rius Carrasco (roser.rius@upc.edu)

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 students' active learning, guided and directed by the lecturer with the help of e-status (an interactive instant-feedback platform with individual exercise data).

The teaching method based on six specific topics consists of repeating cycles based on: theoretical explanations, numerical solutions for exercises, guidance in laboratory classes, follow-up tests by the group teacher and independent practice of exercises.

The applications topic develops transferable competencies through group work on specific cases put forward by students under the lecturer's guidance.

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 and 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. Construct a confidence interval for the mean of a normally distributed variable from the sample mean and standard deviation.
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. Quantify both the performance difference and the imprecision of random sampling using comparative performance test data for two computer products and report the value of the difference if the test has covered all possible situations of interest.
12. Design a comparative test of two computer products, collect data and analyse and interpret results.
13. Using the summary data for two variables, obtain and interpret the estimators for the regression line variables, compute and interpret the R-squared coefficient, obtain the estimators of the uncertainty of the estimate and build a CI for the population values.
14. Make predictions and assess their degree of uncertainty using summary data for two variables and the adjusted model.
15. Analyse the model premises and, if necessary, propose variable transformations for the adjusted model of the graphs for two variables.
16. Design a prediction study, collect data and analyse and interpret results.
17. Identify, for a deterministic process, variability sources and magnitudes.
# Study load

<table>
<thead>
<tr>
<th></th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total learning time:</td>
<td>150h</td>
<td></td>
</tr>
<tr>
<td>Hours large group:</td>
<td>15h</td>
<td>10.00%</td>
</tr>
<tr>
<td>Hours medium group:</td>
<td>15h</td>
<td>10.00%</td>
</tr>
<tr>
<td>Hours small group:</td>
<td>30h</td>
<td>20.00%</td>
</tr>
<tr>
<td>Guided activities:</td>
<td>6h</td>
<td>4.00%</td>
</tr>
<tr>
<td>Self study:</td>
<td>84h</td>
<td>56.00%</td>
</tr>
</tbody>
</table>
Block 1. Probability calculations

Degree competences to which the content contributes:
Description:

Block 2. Random variables

Degree competences to which the content contributes:
Description:

Block 3. Random variable models

Degree competences to which the content contributes:
Description:
Parameterised theoretical models of discrete and continuous random variables. Direct and inverse probabilities computation, with statistical tables and R. Sample mean distribution. Central Limit Theorem, Normal approximations.

Block 4. Evidence: principles of inference

Degree competences to which the content contributes:
Description:

Block 5. Experiment design

Degree competences to which the content contributes:
Description:

Block 6. Statistical models and forecasting

Degree competences to which the content contributes:
270009 - PE - Probability and Statistics

**Description:**
Graphical fitting of the relationship between two numerical variables. Estimation of a linear model. Indicators of the fit quality. Validation of the premises and transformations. Predictions for an individual value and the average.

**Application.**

**Degree competences to which the content contributes:**

**Description:**
Identifying sources of variability in computer processes. Design a study with planning of the goal, data collection, statistical analysis and results interpretation.
## Planning of activities

| Topic 1 activities. Probability calculations | Hours: 18h  
Theory classes: 2h  
Practical classes: 2h  
Laboratory classes: 4h  
Guided activities: 0h  
Self study: 10h |
| Description:  
Situate probability and statistics, especially in the IT setting. Basic probability concepts. Calculating and analysing conditional and joint probabilities. Analysing for (in)dependence.  
Specific objectives:  
1, 2, 3, 17 |

| Topic 2 activities. Random variables. | Hours: 18h  
Theory classes: 2h  
Practical classes: 2h  
Laboratory classes: 4h  
Guided activities: 0h  
Self study: 10h |
| Description:  
Define random variable, discrete random variable and continuous random variable. Define the probability function, the probability distribution function and the joint probability function. Link random variable indicators with sampling indicators.  
Specific objectives:  
4, 6, 7 |

| Topic 3 activities. Random variable models | Hours: 18h  
Theory classes: 2h  
Practical classes: 2h  
Laboratory classes: 4h  
Guided activities: 0h  
Self study: 10h |
| Description:  
Define the theoretical, discrete and continuous models typically used in the IT field and their characteristics and parameters.  
Specific objectives:  
5, 6 |

| Mid-semester exam 1 | Hours: 8h  
Guided activities: 2h  
Self study: 6h |
## 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, 17

### Topic 4 activities. Evidence: principles of inference

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 10h</td>
</tr>
</tbody>
</table>

**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

### Topic 5 activities. Experiment design

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 10h</td>
</tr>
</tbody>
</table>

**Description:**
Define tests with independent and paired samples. Situate and specify the comparison of two means (using Student-t, CIs and HTs in independent paired samples) and the comparison of two variances (in independent samples and suitable transformations).

**Specific objectives:**
11, 12

### Topic 6 activities. Statistical models and forecasting

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 10h</td>
</tr>
</tbody>
</table>

**Description:**
Define a relational model between two variables, analyse the variability, validate the premises, consider possible transformations and make predictions.

**Specific objectives:**
13, 14, 15
## Application activities

<table>
<thead>
<tr>
<th>Hours: 24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 3h</td>
</tr>
<tr>
<td>Practical classes: 3h</td>
</tr>
<tr>
<td>Laboratory classes: 6h</td>
</tr>
<tr>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td>Self study: 12h</td>
</tr>
</tbody>
</table>

**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:**
12, 16, 17

## Mid-semester exam 2

<table>
<thead>
<tr>
<th>Hours: 8h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guided activities: 2h</td>
</tr>
<tr>
<td>Self study: 6h</td>
</tr>
</tbody>
</table>

**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, 16

## Final Exam

<table>
<thead>
<tr>
<th>Hours: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guided activities: 2h</td>
</tr>
<tr>
<td>Self study: 0h</td>
</tr>
</tbody>
</table>

**Description:** Covers all the topics.
The subject is divided into seven topics: six specific topics and one cross-disciplinary applications topic.

For the first six topics, two mid-semester exams (depending on the calendar for each semester) result in six marks (PBi, i = 1 ... 6). For the same six topics, an assessment mark is calculated based on marks for two written exercises completed in the classroom and a mark for problems solved outside class time. The marks are transformed in a Follow-up Factor (SBi, i = 1 .. 6), which can increase the corresponding mark PBi, in order to obtain the block mark:

\[ NB_i = \min(10, PBi \cdot SBi) \quad \text{for } i=1..6 \]

(factor SBi is \(1 + \sum p_j\), where \(p_j\) is a mark between 0 and 0.05, coming from each of the block tests; unexpected events causing loss of classes might decrease the number of marks for a given block)

Topic 7 has no mid-semester exam; hence NB7 will be calculated on the basis of the final report and the presentation.

Given the cumulative nature of the material, the topics will have the following continuous assessment (AC) weighting:

\[ AC = \left[ \frac{10 \cdot NB1 + 11 \cdot NB2 + 12 \cdot NB3 + 13 \cdot NB4 + 14 \cdot NB5 + 15 \cdot NB6 + 10 \cdot NB7}{85} \right] \]

Student who receive AC >= 5 do not have to do the final exam EF.

Keep in mind that EF can consider the transferable competency mark:

\[ EF = \max \{ef, (75 \cdot ef + 10 \cdot NB7)/85\} \]

where "ef" is the proper note of the final exam.

The course mark for the subject is max(AC,EF).

The transferable competency is graded as follows:
A if NB7 >= 8.5; B if 6.5 <= NB7 < 8.5; C if 5 <= NB7 < 6.5; and D if NB7 < 5.
270009 - PE - Probability and Statistics

Bibliography

Basic:


Complementary:


Others resources:

Hyperlink

http://wise.cgu.edu/

http://onlinestatbook.com/stat_sim/

https://lstat.kuleuven.be/newjava/vestac/

http://www.janehorgan.com/

http://heather.cs.ucdavis.edu/probstatbook