200635 - PDE - Statistical Data Protection

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2015
Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 5
Teaching languages: Catalan

Teaching staff
Coordinator: JORDI CASTRO PÉREZ
Others: JORDI CASTRO PÉREZ - A

Prior skills

Basic concepts of Statistics and Operations Research.

Degree competences to which the subject contributes

Specific:

6. CE-2. Ability to master the proper terminology in a field that is necessary to apply statistical or operations research models and methods to solve real problems.
7. CE-3. Ability to formulate, analyze and validate models applicable to practical problems. Ability to select the method and / or statistical or operations research technique more appropriate to apply this model to the situation or problem.
8. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.

9. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.
10. CE-7. Ability to understand statistical and operations research papers of an advanced level. Know the research procedures for both the production of new knowledge and its transmission.
11. CE-8. Ability to discuss the validity, scope and relevance of these solutions and be able to present and defend their conclusions.
12. CE-9. Ability to implement statistical and operations research algorithms.

Transversal:

1. ENTREPRENEURSHIP AND INNOVATION: Being aware of and understanding how companies are organised and the principles that govern their activity, and being able to understand employment regulations and the relationships between planning, industrial and commercial strategies, quality and profit.
2. SUSTAINABILITY AND SOCIAL COMMITMENT: Being aware of and understanding the complexity of the economic and social phenomena typical of a welfare society, and being able to relate social welfare to globalisation and sustainability and to use technique, technology, economics and sustainability in a balanced and compatible manner.
3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.
5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English.
The goal of the course is to introduce the student to the field of Statistical Disclosure Control or statistical secret. This discipline proposes a joining of methods in order to guarantee individual data confidentiality when disseminating statistical data, whether it be microdata or data added in tabular form. This problem is of great importance for the National Institutes of Statistics and, in general, any other private entity or official organism that needs to divulge data. Upon completion of the course, the student should know how to apply the technical principles of microdata protection and tabular data, as well as be familiar with the software applied to these methods. The last part of the course presents the solution of some statistical problems by optimization techniques (orthogonal latin squares, clustering problems, etc.)

Abilities to Acquire:
* Knowledge of what is the field of Statistical Disclosure Control or Statistical Data Protection.
* Knowledge of the technical principals of Microdata Protection and Aggregate Data Protection.
* Knowledge of data protection software.
* The ability to protect data by using some existing technique.
* Familiarization with recent literature regarding this field.

Learning objectives of the subject

Teaching methodology

Theory:
The contents of the subject are presented and discussed with a combination of explanations on the board and with transparencies.
Training:
Laboratory sessions which demonstrate the use of data protection software.

Language:
The course can be imparted in either English, Catalan or Spanish.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group: 30h</th>
<th>24.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group: 15h</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 80h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Content

### Introduction to Statistical Disclosure Control.

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

**Description:**

### Methods for microdata.

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

**Description:**
Perturbation Methods: Microaggregation, Noise Adding, Rank-Swapping; Non-Perturbation Methods: Recodification.

### Methods for tabular data.

**Learning time:** 30h
- Theory classes: 24h
- Laboratory classes: 6h

**Description:**
Determination of Sensitive Cells. Non-Perturbation Methods: Cell Suppression Problem; Exact and Heuristic Methods. Perturbation Methods: Controlled Rounding; Minimum Distance Controlled Adjustment.

### Optimization in statistical problems

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

**Description:**

## Qualification system

Completion of exercises and class work.

## Bibliography

**Basic:**

- Articles en revistes d'estadística i investigació operativa dels darrers 15 anys.