200609 - ATV - Lifetime Data Analysis

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
1004 - UB - (ENG)Universitat de Barcelona

Academic year: 2018
Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 5
Teaching languages: English

Teaching staff
Coordinator: GUADALUPE GÓMEZ MELIS
Others: Primer quadrimestre:
GUADALUPE GÓMEZ MELIS - A
KLAUS GERHARD LANGOHR - A

Prior skills
In order to follow the course successfully the student has to be familiar with the following concepts: estimation theory and confidence intervals, likelihood function, maximum likelihood estimation, regression models, hypothesis tests. The student will have to use the R software for homework and data analysis. Chapters 1 through 3 of the book "Principles of Statistical Inference" Cox, Cambridge University Press (2006) should be mastered.

Degree competences to which the subject contributes

Specific:
3. 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.
4. 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.
5. 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.

Translate to english
6. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.

Transversal:
2. 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.
200609 - ATV - Lifetime Data Analysis

Teaching methodology

Lectures:
One hour and a half sessions in which the main concepts and topics are introduced. The lecturer will use a computer to introduce the course content. Emphasis is put on ideas and intuition. Topics are discussed from the point of view of real situations concerning clinical trials or epidemiological studies.

Problem-solving sessions:
Incorporated into the practical sessions.

Laboratory sessions:
One hour and a half sessions held in the computer lab in which theoretical problems are tackled and exercises are carried out using computers.

Learning objectives of the subject

Survival analysis is employed in many fields to analyze data representing the duration or elapsed time between two events. It is also known as event history analysis, lifetime data analysis, reliability analysis and time to event analysis. A key characteristic that distinguishes survival analysis from other areas of statistics is that survival data are usually censored, sometimes truncated and the normality hypothesis is inadequate. Censoring occurs when the information for some individuals is incomplete, what may happen for different reasons discussed in class.

The course Lifetime Data Analysis covers a series of procedures and techniques for analyzing censored and/or truncated data. While the course is focused on medical applications in public health and in epidemiology, it also has direct applications to other disciplines such as economics, actuarial sciences, engineering and demography.

The aim of the course is to develop the core of survival analysis and to put into practice the knowledge acquired by means of the statistical software package R.

Abilities to be acquired:
* Identification of those situations or studies in which it is necessary to use Survival Analysis methodology. The ability to define the events and times relevant to each situation.
* Identification and knowledge of the different types of censoring and truncation. The ability to construct the likelihood in each case.
* Knowledge on the most common parametric models: Exponential, Weibull, Gamma, Gompertz, Lognormal and Log-logistic. The ability to evaluate the most adequate model in a concrete example.
* The ability to obtain and interpret the Kaplan-Meier estimator, to know its most important properties and how to calculate estimators for the cumulative risk functions.
* Knowledge on how to present different hypothesis tests in order to compare two or more survival curves. The ability to select the most appropriate test according to the type of alternative hypothesis.
* Knowledge on how to use accelerated lifetime regression models: the Weibull and the log-logistic model. Knowledge of their relationships and differences.
* The ability to set out and interpret a proportional hazard model, as well as checking the goodness-of-fit by means of studying different residuals.
## Study load

<table>
<thead>
<tr>
<th></th>
<th>Hours large group:</th>
<th>Hours medium group:</th>
<th>Hours small group:</th>
<th>Guided activities:</th>
<th>Self study:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total learning time:</strong></td>
<td>30h</td>
<td>0h</td>
<td>15h</td>
<td>0h</td>
<td>80h</td>
</tr>
<tr>
<td></td>
<td>24.00%</td>
<td>0.00%</td>
<td>12.00%</td>
<td>0.00%</td>
<td>64.00%</td>
</tr>
<tr>
<td>Content</td>
<td>Learning time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Basic concepts and parametric models</strong></td>
<td>6h 30m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival function. Hazard function.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean and median life</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal parametric models.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Censoring and truncation</strong></td>
<td>5h 30m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different types of right censoring.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left and interval censoring.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building the likelihood function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left truncation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>One sample non-parametric inference</strong></td>
<td>9h 30m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier estimator for the survival function.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson-Aalen estimator for the cumulative risk function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic Properties.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence intervals and confidence bands.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two sample comparison</strong></td>
<td>8h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two sample comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The (weighted) log-rank test.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleming-Harrington tests family.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stratified tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Parametric regression

**Description:**
- Accelerated life models.
- Log-linear, proportional hazards and proportional odds models.
- Weibull regression model.
- Log-logistic model.

**Learning time:** 7h 30m
- Theory classes: 4h 30m
- Laboratory classes: 3h

---

### Semi-parametric regression: Cox Model

**Description:**
- Cox’s regression model.
- Partial likelihood.
- Inference within Cox model.
- Different type of residuals for Cox model.
- Validation of the Cox model.

**Learning time:** 8h
- Theory classes: 6h
- Laboratory classes: 2h

---

### Qualification system

Assessment is based on the following:
* Problems solved and handed in throughout the course (3 sets) (25%)
* Case study with real data (25%)
* Final exam (50%)

### Regulations for carrying out activities

The student will be informed at the beginning of the course on the dates of each deliverable.
Bibliography

Basic:

Complementary: