Categorical Data Analysis (C002880)

Course Specifications
Valid as from the academic year 2018-2019

Course

Lecturers in academic year 2018-2019
De Neve, Jan PP01 lecturer-in-charge
Loeys, Tom PP01 co-lecturer

Offered in the following programmes in 2018-2019
Master of Science in Bioinformatics (main subject Systems Biology) 5 B

Teaching languages
English

Keywords
Categorical data analysis, generalized linear models

Position of the course
This course builds on `Principles of Statistical Data Analysis` to enable students to understand and apply the most frequently used methods of categorical data analysis. It is in turn one of the courses that prepares for the course on data mining, epidemiology, and survival analysis.
This fits in with the following learning outcomes of the master program: M.1.2., M.1.3., M.1.4., M.1.5., M.1.6, M.1.7., M.2.1., M.2.4., M.2.5., M.3.1., M.4.3.

Contents
• Distribution and inference for categorical data
• Analysis of contingency tables with approximate and exact methods
• Generalised linear models
• Logistic regression
• Ordinal logistic regression, Proportional odds models
• Multinomial logistic regression
• Poisson regression, negative binomial regression, zero-inflated models
• Loglinear models for (paired) tables
• Generalized additive models

Initial competences
Basic knowledge of probability and mathematical statistics. Basic material from the course `Analysis of continuous data`.

Final competences
1 The student can report accurately the results and limitations of a categorical data analysis.
2 The student has theoretical knowledge about the most frequently used methods of categorical data analysis.

Course size (nominal values; actual values may depend on programme)
Credits 5.0
Study time 150 h
Contact hrs 62.5 h

Course offerings and teaching methods in academic year 2018-2019
A (semester 1)
English
lecture 22.5 h
group work 32.5 h
seminar: practical PC room classes 7.5 h

B (semester 1)
lecture 22.5 h
group work 32.5 h
seminar: practical PC room classes 7.5 h

Course size (nominal values; actual values may depend on programme)
Credits 5.0
Study time 150 h
Contact hrs 62.5 h

Teaching languages
English

Keywords
Categorical data analysis, generalized linear models

Position of the course
This course builds on `Principles of Statistical Data Analysis` to enable students to understand and apply the most frequently used methods of categorical data analysis. It is in turn one of the courses that prepares for the course on data mining, epidemiology, and survival analysis.
This fits in with the following learning outcomes of the master program: M.1.2., M.1.3., M.1.4., M.1.5., M.1.6, M.1.7., M.2.1., M.2.4., M.2.5., M.3.1., M.4.3.

Contents
• Distribution and inference for categorical data
• Analysis of contingency tables with approximate and exact methods
• Generalised linear models
• Logistic regression
• Ordinal logistic regression, Proportional odds models
• Multinomial logistic regression
• Poisson regression, negative binomial regression, zero-inflated models
• Loglinear models for (paired) tables
• Generalized additive models

Initial competences
Basic knowledge of probability and mathematical statistics. Basic material from the course `Analysis of continuous data`.

Final competences
1 The student can report accurately the results and limitations of a categorical data analysis.
2 The student has theoretical knowledge about the most frequently used methods of categorical data analysis.
3 The student can correctly interpret and critically assess the results of a categorical data analysis.
4 The student can select appropriate statistical methods for categorical data analysis.
5 The student participates actively and works constructively in a group context to solve problems involving categorical data analysis.
6 The student has practical skills to analyze real-world datasets with categorical responses.

Conditions for credit contract
Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract
This course unit cannot be taken via an exam contract

Teaching methods
Group work, lecture, seminar: practical PC room classes

Extra information on the teaching methods
Exercises: PC-labs (using R) with written tutorials and feedback material.

Learning materials and price
A pdf version of the course notes is freely available via Minerva. Estimated cost of the printed version: 10 EUR

References

Course content-related study coaching
PC-projects and independent work are supported by (written) R tutorials with feedback during the lectures.

Evaluation methods
end-of-term evaluation and continuous assessment

Examination methods in case of periodic evaluation during the first examination period
Written examination with open questions, open book examination

Examination methods in case of periodic evaluation during the second examination period
Written examination with open questions, open book examination

Examination methods in case of permanent evaluation
Assignment

Possibilities of retake in case of permanent evaluation
examination during the second examination period is possible

Extra information on the examination methods
Periodical evaluation: written (open book)
Permanent evaluation: written (open book)
The periodical exam aims to assess if the student understands the basic theory of categorical data analysis. A part of the exam consists of exercises that involve the interpretation of given software output. The take home independent work takes the form of small-scale data-analytic projects, intermixed with small computational problems.

Calculation of the examination mark
Permanent evaluation (25% of the final score) takes the form of three take home problem sets for group work. In addition, there is an exam during the exam period at the end of the semester (75% of the final score). When the student scores less than 10/20 for one of the components, he/she can no longer pass the entire course unit. If the total score is a mark of ten or more out of twenty, then this is reduced to the highest failing mark (9/20). For failed students who want a retake in the second examination period, the permanent evaluation will take the form of a data-analytic project.