Course Specifications
Valid as from the academic year 2015-2016

Causality and Missing Data (C002674)

Course
Specifications

Validity

Course size

| Credits | Study time 165 h | Contact hrs 45.0 h |

Course offerings and teaching methods in academic year 2016-2017

A (semester 2) English seminar: practical 20.0 h seminar: coached 2.5 h lecture 22.5 h

Lecturers in academic year 2016-2017

Vansteelandt, Stijn WE02 lecturer-in-charge

Offered in the following programmes in 2016-2017

| Master of Science in Mathematics (main subject Applied Mathematics) | 6 | A |
| Master of Science in Mathematics (main subject Mathematical Physics and Astronomy) | 6 | A |
| Master of Science in Mathematics (main subject Pure Mathematics) | 6 | A |
| Master of Science in Mathematics | 6 | A |
| Exchange Programme in Mathematics (master's level) | 6 | A |

Teaching languages

English

Keywords

Causal inference, Data analysis, Experimental studies, Missing data, Observational studies, Selection bias, Statistics

Position of the course

To enable the master to
• recognize diverse forms of bias, due to missing data, confounding and selection bias, in statistical analyses
• avoid such biases through study design
• correct for such biases through statistical data analysis.

Contents

This course offers a thorough investigation of statistical methods for causal inference from experimental and observational data. This methodology has wide applications in epidemiology, clinical studies, public health, agriculture, sociology, pedagogy, demography, economics...

In the first part of this course, a general causal theory will be introduced which will enable a systematic study of different important types of bias in the statistical analysis of experimental and observational data. Epidemiological concepts such as direct causal effect, indirect causal effect, confounding, selection bias and intermediate variables will be formally defined using potential outcomes and made intuitive using causal diagrams. Biases due to missing data in empirical studies and due to inappropriate adjustment for intermediate variables and time-dependent confounders will be studied as special cases of this general causal theory rather than as separate cases.

Next, several techniques for confounding adjustment in point treatment studies are discussed: standard regression adjustment, standardisation and propensity score based methods that use subclassification, matching, regression or inverse weighting. The third part of the course focuses on mediation analysis. It discusses controlled direct effects, natural direct and indirect effects, the mediation formula, natural effect models and finally techniques for handling time-varying confounding: inverse weighting and G-estimation.

(Approved)
In the fourth part of this course, we will introduce marginal structural models for the analysis of time-dependent exposures in the presence of time-dependent confounders. Finally, the problem of incomplete data will be studied in more detail. The classical missing data taxonomy (missing (completely) at random, missing not at random) will be described and its plausibility evaluated in different settings. Imputation methods are discussed in detail; the EM-algorithm and inverse probability weighted estimators are only briefly mentioned. The importance of sensitivity analyses will be stressed. If time allows, instrumental variable methods will be discussed and illustrated in the context of imperfect randomized studies (e.g. randomized clinical trials with noncompliance).

The different concepts and techniques will be illustrated using real data sets from epidemiology, sociology and economics. To support this course, a brief introduction to logistic regression will be given.

Initial competences
Basic knowledge of statistics and linear models (as in the course `Statistic Models and Data Analysis`).

Final competences
1 The students will be able to make the fundamental distinction between association analysis and causal analysis.
2 They will recognize the importance of missing data, know specific methods to minimize missing data at the design stage and have the capability to correct for them through the statistical analysis.
3 They will be able to correct for (time-dependent) confounders and have the capability to detect and model interaction terms.
4 They will be able to evaluate the sensitivity of conclusions to the underlying assumptions.

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
Lecture, seminar: coached exercises, seminar: practical PC room classes

Learning materials and price
English syllabus and scientific articles. Cost: 5 EUR

References

Course content-related study coaching
The students will frequently exercise the concepts and methods explained during the lectures, by analyzing realistic data sets during the practical sessions, where students will be closely supervised, and while making their project work. The lecturer is available for supplementary help. There is also interactive support through Minerva (forum, e-mail).

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, oral examination

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

Examination methods in case of permanent evaluation
Report

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

(Approved)
Extra information on the examination methods

- Theory: by period; oral (open book).
- Exercises: by period and permanent; written (open book) and project: written reporting.

Calculation of the examination mark

- Project 1: 3/20
- Project 2: 3/20
- Written and oral examination: 14/20