GHENT UNIVERSITY

Course Specifications

From the academic year 2019-2020 up to and including the academic year

Causality and Missing Data (C002881)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0 **Study time 150 h** **Contact hrs** 62.5h

Course offerings in academic year 2024-2025

Lecturers in academic year 2024-2025

Offered in the following programmes in 2024-2025	**crdts**	**offering**

Teaching languages

English

Keywords

Causal inference, Experimental studies, Missing data, Observational studies, Selection bias

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.

This course provides an excellent training in scientific thinking.
This course builds on`Analysis of continuous data'

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 techniqies 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, as well as the use of machine learning.
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.
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.

Initial competences

Having successfully completed the course`Analysis of continuous data' or having acquired otherwise the corresponding competences.

Final competences

1 Make the fundamental distinction between association analysis and causal analysis.
2 Recognize the impact of missing data on a statistical analysis.
3 Correct for selection bias and information loss due to missing data in standard statistical analyses.
4 Understand the assumptions underlying different standard methods to correct for selection bias and information loss due to missing data in statistical analyses.
5 Correct for measured (time-dependent) confounders in statistical analyses.
6 Understand the assumptions underlying different standard methods to correct for measured (time-dependent) confounders in statistical analyses.
7 Correct for unmeasured confounders in statistical analyses, when an instrumental variable has been measured.
8 Detect and model interaction terms.

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: coached exercises, Seminar: practical pc room classes

Extra information on the teaching methods

Exercises: written exercises and PC-labs using SAS and R.

Study material

None

References

Little RJ and Rubin DB (1987). Statistical Analysis with Missing Data. New York: Wiley.
Pearl J (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.
van der Laan MJ and Robins JM (2002). Unified Methods for Censored Longitudinal Data and Causality. Springer-Verlag: New-York.

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. Besides the questions that students can ask before, during or after each lecture, there are several possibilities for asking questions: interactive support via Ufora (forum) and, in case of larger numbers of questions, personal coaching after electronic appointment.

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral examination, Open book examination, Written examination with open questions

Examination methods in case of periodic assessment during the second examination period

Oral examination, Open book examination, Written examination with open questions

Examination methods in case of permanent assessment

Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

Theory: oral (open book)
Exercises: written (open book)
Project: written reporting
The entire exam assesses the student's insight into the basic principles of causal inference and incomplete data analysis and his/her ability to actively apply the statistical methods in the course. The exam will consist almost entirely of exercises (e.g. questions to examine whether the student is able to critically read scientific papers that involve causal questions or suffer from missing data, practical insight may also be tested via interpretation of given software-output).

Calculation of the examination mark

Theory: periodic
Exercises: periodic and permanent (project work)
One group project (written reporting) will be assigned.
Calculation of the total score: exam 80%, project 20%. A second examination chance for the project is possible. Non-participation to at least one of the project works implies a maximum score (exam + project) of at most 7/20, regardless of the score obtained on the final exam.