

Statistics III: Regression Analysis (C003560)

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

Credits 6.0

Study time 165 h

Course offerings and teaching methods in academic year 2024-2025

A (semester 1)

Dutch, English

Gent

lecture

seminar

Lecturers in academic year 2024-2025

Vansteelandt, Stijn

WE02

lecturer-in-charge

Offered in the following programmes in 2024-2025

[Bachelor of Science in Mathematics](#)

crdts

6

offering

A

[Preparatory Course Master of Science in Mathematics](#)

6

A

Teaching languages

English, Dutch

Keywords

Multivariate linear models, statistical data analysis

Position of the course

To provide the student with a basis for statistical modelling. Allow her/him to solve a range of practical problems through a relatively simple (multivariate linear) statistical model. Give him/her the tools and skills to fit the model to a given dataset and to draw appropriate conclusion based on the output. To introduce the basic concepts of design, analysis, confounding, interaction, information, efficiency, multi-collinearity, significance, power, influential observations, goodness-of-fit, robustness and model building in the context of the linear model. Derive theoretical properties of estimators and predictions following the multivariate linear model. Understanding how to align analysis with prediction versus quantifying associations and effects. Understanding the dangers of data-adaptive analyses and being able to use advanced solutions. To be able to quantify the importance of predictors in terms of the strength of their residual association with an outcome.

Contents

Introduction to modeling

- The curse of dimensionality
- Prediction, association, causation
- The simple and multiple linear model
- Confounding and interaction
- Regression to the mean

Linear regression for prediction

- The least squares principle
- Maximum likelihood theory for linear models
- Prediction intervals based on a pre-selected model
- Graphical control of assumptions made based on residues
- Detection of influential measures via leverage and Cook's distance
- Criteria for modeling aimed at prediction: Mallow's Cp, AIC, cross-validation
- Model building by means of best subset selection, step selection, ridge regression, lasso
- Splines

Linear regression for associations and effects

- Counterfactuals and Causal Effects
- Analysis of randomized experiments
 - Decision-making based on a pre-selected model

- Balancing bias and efficiency in randomized studies
 - Sandwich estimators
 - Analysis of observational studies
 - The impact of model misspecification on bias and precision; extrapolation and multicollinearity
 - Causal diagrams, confounding and selection bias
 - Model building for associations and effects; Occam's dilemma; Post-selection inference
 - Double selection
 - Direct standardization
 - Impact of measurement errors
- Linear regression to quantify the importance of predictors
- Multiple correlation coefficient

Initial competences

Final competences of the course Statistics II.

Final competences

- 1 The student can recognize practical problems which can be solved by means of the (multivariate) linear model.
- 2 She/he knows the distinction between association, prediction and causation, and is able to place the notion of confounding in this context.
- 3 She/he can interpret the linear model correctly, can fit it to a data set and draw justified conclusions in the theoretical as well as the practical sense.
- 4 She/he knows and can derive properties of the estimators, paying attention to the distinction between estimation and prediction.
- 5 She/he can link regression analysis with analysis of variance.
- 6 She/he also recognizes the limitations of the linear model, can suggest appropriate extensions and develop the corresponding methods via the maximum likelihood approach.
- 7 She/he can balance efficiency and bias when designing a statistical study.
- 8 She/he must be able to clearly and correctly report the method and results of an applied linear regression analysis, both orally and in writing. He/she should also be able to work in team on a solution of a practical problem situated in this context.
- 9 He/she must be able to perform regression analysis with high dimensional predictors.

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

Seminar, Lecture

Extra information on the teaching methods

Exercises: work sessions and PC practicals partly through independent work followed by question and answer sessions, and partly on campus under supervision. Project work with implementation in a team and also independent reporting of a data analysis.

Study material

Type: Syllabus

Name: Statistics III: Regression Analysis

Indicative price: € 15

Optional: no

Available on Ufora : Yes

Online Available : Yes

Available in the Library : No

Available through Student Association : Yes

References

Wooldridge, J. (2012). Introductory Econometrics: A Modern Approach, 5th Edition.

Course content-related study coaching

Students are supervised in class during exercise sessions and computer practicals. They can exchange questions and answers outside lecture hours among themselves and with lecturers. A project will provide the students with practical experience in data analysis.

Assessment moments

end-of-term and continuous assessment

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

Oral assessment, Written assessment with open-ended questions

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

Oral assessment, Written assessment with open-ended questions

Examination methods in case of permanent assessment

Skills test, Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

The knowledge and problem solving skills of the students are tested by means of a written exam and a data analytic project.

Calculation of the examination mark

Permanent evaluation (20%) + periodic evaluation (80%).