

Statistiek III: regressieanalyse (C003560)

Cursusomvang *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

Studiepunten 6.0 **Studietijd 165 u**

Aanbodsessies en werkvormen in academiejaar 2024-2025

A (semester 1)	Engels, Nederlands	Gent	hoorcollege werkcollege
----------------	--------------------	------	----------------------------

Lesgevers in academiejaar 2024-2025

Vansteelandt, Stijn	WE02	Verantwoordelijk lesgever
---------------------	------	---------------------------

Aangeboden in onderstaande opleidingen in 2024-2025

	stptn	aanbodsessie
Bachelor of Science in de wiskunde	6	A
Vorbereidingsprogramma tot Master of Science in de wiskunde	6	A

Onderwijsstalen

Engels, Nederlands

Trefwoorden

Multivariate lineaire modellen, statistische data-analyse

Situering

De student een basis geven voor statistische modelbouw. Haar/hem toelaten om voor een waaier aan praktische problemen via een relatief eenvoudig (multivariaat lineair) statistisch model een oplossing te vinden. Met behulp van software het lineaire model kunnen aanpassen aan een dataset en er de gepaste conclusies bij kunnen trekken. In de context van het lineaire model de basisbegrippen van design, analyse, confounding, interactie, informatie, efficiëntie, multi-collineariteit, significantie, kracht, goodness-of-fit, invloedrijke observaties, robuustheid en modelbouw bijbrengen. De theoretische en praktische eigenschappen kunnen afleiden van de schatters en predicties die voortvloeien uit het multivariaat lineair model. Begrijpen hoe de analyse af te stemmen op predictie, versus het kwantificeren van associaties en effecten. De gevaren van data-adaptieve analyses begrijpen, en van geavanceerde oplossingen gebruik kunnen maken. Het belang van predictoren, in termen van de sterkte van hun residuele associatie met een uitkomst, kunnen kwantificeren.

Inhoud

Inleiding tot modelleren

- De nood aan modelleren
- Predictie, associatie, causatie
- Het enkelvoudige en meervoudige lineaire model
- Confounding en interactie
- Regression to the mean

Lineaire regressie voor predictie

- Het kleinste kwadratenprincipe
- Maximum likelihood theorie voor lineaire modellen
- Predictie-intervallen op basis van een vooraf gekozen model
- Grafische controle van gemaakte vooronderstellingen op basis van residuen
- Detectie van invloedrijke metingen via leverage en Cook's afstand
- Criteria voor modelbouw gericht op predictie: Mallow's Cp, AIC, cross-validatie
- Modelbouw d.m.v. best subset selectie, stapsgewijze selectie, ridge regression, lasso
- Splines

Lineaire regressie voor associaties en effecten

- Counterfactuals en causale effecten
- Analyse van gerandomiseerde experimenten

- Besluitvorming op basis van een vooraf gekozen model
- Het afwegen van vertekening en efficiëntie in gerandomiseerde studies
- Sandwich schatters
- Analyse van observationele studies
 - De impact van model misspecificatie op vertekening en precisie; extrapolatie en multicollineariteit
 - Causal diagrammen, confounding en selectiebias
 - Modelbouw voor associaties en effecten; Occam's dilemma; Post-selection inference
 - Double selection
 - Directe standaardisatie
- Impact van meetfouten

Lineaire regressie om het belang van predictoren te kwantificeren

- Meervoudige correlatiecoëfficiënt

Begincompetenties

Eindcompetenties van het vak Statistiek II.

Eindcompetenties

- 1 De student moet in staat zijn praktische problemen te herkennen die kunnen opgelost worden met behulp van een (multivariaat) lineair model.
- 2 Hij/zij moet het onderscheid kunnen maken tussen associatie, predictie en causatie en het begrip confounding in die context kunnen plaatsen.
- 3 Hij/zij moet het lineaire model op correcte wijze kunnen interpreteren, het kunnen aanpassen aan een gegeven dataset en er verantwoordde conclusies kunnen uit trekken, zowel in formele als in toegepaste zin.
- 4 Hij/zij moet de eigenschappen van de schatters kennen en kunnen afleiden, met aandacht voor het onderscheid tussen een betrouwbaarheidsinterval en een predictie-interval.
- 5 Hij/zij moet het verband kennen tussen lineaire regressie en variantie-analyse.
- 6 Hij/zij moet ook de beperkingen van het lineaire model onderkennen, in een concrete context de nodige uitbreidingen kunnen suggereren en die met de maximum likelihood methode kunnen uitwerken.
- 7 Hij/zij moet een goede balans tussen efficiëntie en vertekening kunnen nastreven bij het ontwerpen van een statistische studie.
- 8 Hij/zij moet de methode en resultaten van een lineaire regressie analyse zowel schriftelijk als mondeling correct en helder kunnen rapporteren. Hij/zij moet in deze context ook in team kunnen werken aan een oplossing voor een praktisch probleem.
- 9 Hij/zij moet regressie-analyse met hoog-dimensionale predictoren kunnen uitwerken.

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Werkcollege, Hoorcollege

Toelichtingen bij de didactische werkvormen

Oefeningen: werksessies en PC-practica gedeeltelijk via zelfstandig werk gevolgd door vraag- en antwoordsessies, en deels on campus onder begeleiding. Projectwerk met uitvoering in team en ook zelfstandig rapporteren van een data-analyse.

Studiemateriaal

Type: Syllabus

- Naam: Statistiek III: Regressie Analyse
- Richtprijs: € 15
- Optioneel: nee
- Beschikbaar op Ufora : Ja
- Online beschikbaar : Ja
- Beschikbaar in de bibliotheek : Nee
- Beschikbaar via studentenvereniging : Ja

Referenties

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

Vakinhoudelijke studiebegeleiding

De studenten worden begeleid door een assistent bij praktische oefeningensessies en tijdens

(Goedgekeurd)

PC practica. Buiten de lesuren kunnen zij met de lesgevers vragen en antwoorden uitwisselen. Een project zal hen in staat stellen data-analytische vaardigheden daadwerkelijk in de praktijk te brengen.

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Mondelinge evaluatie, Schriftelijke evaluatie met open vragen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Mondelinge evaluatie, Schriftelijke evaluatie met open vragen

Evaluatievormen bij niet-periodegebonden evaluatie

Vaardigheidstest, Werkstuk

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

Periodegebonden evaluatie: de kennis en het probleemoplossend vermogen van de student worden schriftelijk getoetst.

Niet-periodegebonden evaluatie: een project waarin een dataset wordt geanalyseerd.

Eindscoreberekening

Niet-periodegebonden evaluatie (20%) + periodegebonden evaluatie (80%).