



Causale analyse en ontbrekende gegevens (C002674)

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

Studiepunten 6.0 Studietijd 165 u Contacturen 45.0 u

Aanbodssessies en werkvormen in academiejaar 2019-2020

A (semester 2)	Engels	werkcollege: geleide oefeningen	2.5 u
		werkcollege: PC-klasoefeningen	20.0 u
		hoorcollege	22.5 u

Lesgevers in academiejaar 2019-2020

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

Aangeboden in onderstaande opleidingen in 2019-2020

	stptn	aanbodssessie
Educatieve Master of Science in de wetenschappen en technologie (afstudeerrichting wiskunde)	6	A
Master of Science in de wiskunde	6	A
Uitwisselingsprogramma wiskunde (niveau master)	6	A

Onderwijstalen

Engels

Trefwoorden

Causale analyse, Data analyse, Experimentele studies, Observationele studies, Ontbrekende gegevens, Selectie-bias, Statistiek

Situering

De student in staat stellen om

- diverse vormen van vertekening, ten gevolge van ontbrekende gegevens, confounding en selectie-bias, te herkennen in statistische analyses,
- het design van een studie anders op te zetten teneinde dergelijke vormen van vertekening te vermijden,
- voor dergelijke vertekening te corrigeren door de statistische data-analyse.

Inhoud

Deze cursus biedt een diepgaand onderzoek van statistische methoden om causale besluiten te trekken uit experimentele en observationele gegevens. Deze methodologie is rijkkelijk toepasbaar in de epidemiologie, klinische studies, volksgezondheid, landbouwkunde, sociologie, opvoedingskunde, demografie, economie...

In het eerste deel van deze cursus wordt een algemene causale theorie ingevoerd die zal toelaten om belangrijke types vertekening in de statistische analyse systematisch te bestuderen. Epidemiologische concepten zoals rechtstreeks causaal effect, onrechtstreeks causaal effect, confounding, selectie-bias en intermediaire variabelen worden formeel gedefinieerd aan de hand van potentiële uitkomsten en intuïtief inzichtelijk gemaakt aan de hand van causale diagrammen. Vertekeningen ten gevolge van ontbrekende gegevens en van on gepaste correctie voor intermediaire variabelen en tijdsafhankelijke confounders worden bestudeerd als bijzondere gevallen van deze algemene causale theorie, eerder dan als alleenstaande gevallen.

Vervolgens wordt ingegaan op diverse technieken voor confounding correctie bij het inschatten van het effect van een enkele blootstelling op een uitkomst. Naast standaard regressie en standardisatie wordt ingegaan propensity score gebaseerde technieken die gebruik maken van subclassificatie, matching, regressie of inverse wegen, alsook op het gebruik van machine learning.

Het derde deel van de cursus behandelt mediatie-analyse. Het behandelt controlled direct effects, natural direct and indirect effects, de mediatieformule, natural effect

models en tot technieken om met tijdsafhankelijke confounding om te gaan: inverse wegen en G-estimation.

In het vierde deel van de cursus worden marginale structurele modellen ingevoerd voor de analyse van tijdsafhankelijke blootstellingen in de aanwezigheid van tijdsafhankelijke confounders.

Tot slot wordt concreet het probleem van ontbrekende gegevens in meer detail bestudeerd. De klassieke missing data taxonomie (missing (completely) at random, missing not at random) wordt beschreven en haar plausibiliteit geëvalueerd in diverse contexten. Multipale imputatietechnieken worden in detail besproken; EM-algoritme en inverse probability weighted schatters worden slechts kort vermeld. Het belang van sensitiviteitsanalyse wordt onderstreept.

Indien de tijd het toelaat, worden instrumentele variabele methoden besproken en toegelicht in de context van imperfecte gerandomiseerde studies (bijvoorbeeld, gerandomiseerde klinische studies met therapie-ontrouw).

De verschillende concepten en technieken worden telkens geïllustreerd aan de hand van actuele voorbeelden uit de epidemiologie, sociologie en economie. Als ondersteuning bij deze cursus wordt tevens een beknopte inleiding gegeven tot logistische regressie.

Begincompetenties

Basiskennis statistiek en lineaire modellen (zoals in de cursus 'Statistiek III: regressie-analyse').

Eindcompetenties

- 1 De studenten zullen het fundamentele onderscheid kunnen maken tussen associatie-analyse en causale analyse.
- 2 Zij zullen het belang van missing data erkennen, specifieke methoden kennen die dit kunnen minimaliseren op niveau van design en in rekening kunnen brengen op het niveau van de analyse.
- 3 Zij zullen voor (tijdsafhankelijke) confounders kunnen corrigeren en interactietermen kunnen detecteren en modelleren.
- 4 Zij zullen de gevoeligheid van conclusies aan assumpties kunnen evalueren.

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

Hoorcollege, werkcollege: geleide oefeningen, werkcollege: PC-klasoefeningen

Leermateriaal

Engelstalige syllabus en wetenschappelijke artikels. Geraamde totaalprijs: 5 EUR

Referenties

- 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.

Vakinhoudelijke studiebegeleiding

De studenten zullen de concepten en methoden die tijdens de hoorcolleges worden aangereikt veelvuldig inoefenen door realistische datasets te analyseren tijdens de practica, waar de studenten nauwgezet worden opgevolgd, en tijdens het maken van hun projectwerk. Bijkomende uitleg kan steeds bekomen worden. Er is ook interactieve ondersteuning via Ufora (forum, e-mail).

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijk examen met open vragen, openboekexamen, mondeling examen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijk examen met open vragen, openboekexamen, mondeling examen

Evaluatievormen bij niet-periodegebonden evaluatie

Verslag

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

Theorie: periodegebonden; mondeling (open boek)

Oefeningen: periodegebonden en permanent; schriftelijk (open boek) en project: schriftelijke rapportering.

Eindscoreberekening

Project: 4/20

Schriftelijk en mondeling examen: 16/20