UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Causal Inference

2526-1-F8205B018

Obiettivi formativi

1. **Knowledge and Understanding:**

The course aims to introduce principles of causal inference in observational studies to be able to define (1) the causal effect, (2) the difference between association and causation, (3) the concept of confounding and selection bias, (4) direct acyclic graphs, (5) estimation of causal effects, (6) statistical models for causal effects (7) mediation (8) target emulation trials

2. **Applying knowledge and understanding:**

Based on the concepts and methods presented in class, students will be able to draw cuasal graphs and assess how to run statistical models to estimate causal effects using both Stata (StataCorporation), widely used in the field of epidemiology and biostatistics and R.

3. **Making judgements:**

Having learnt methods and applications, students will show which methods and techniques are to be used given the study and the data at hand, by critically assessing the statistical and clinical implications of the models being used.

4. **Communication skills:**

The student will be able to explain how the methods can be applied and interpret the results in a clear and simple way.

5. **Learning skills:**

The students will learn how to analyse basic and more advanced statistical methods to fit causal models, by understanding the methodology and by using Stata.

Contenuti sintetici

Causal inference from observational data is a key task of biostatistics and of allied sciences such as sociology, education, behavioral sciences, demography, economics, health services research, etc.

These disciplines share a methodological framework for causal inference that has beendeveloped over the last decades.

This course presents this unifying causal theory and shows how biostatistical concepts and methods can be understood within this general framework. The course emphasizes conceptualization but also introduces statistical models and methods for causal inference.

Specifically, the content is:
a) formally define causal concepts such as causal effect and confounding
b) identify the conditions required to estimate causal effects
c) use analytical methods that, under those conditions, provide estimates that can be endowed with a causal interpretation.

The (causal) methods can be used under less restrictive conditions than the traditional statistical methods. For example, instrumental v ariable methods allow one toestimate the causal effect of an exposure in the presence of unmeasured confounders of the exposure and outcome.

Programma esteso

Section (I): Causal inference without models:
1 A definition of causal effect
1.1 Individual causal effects
1.2 Average causal effects
1.3 Measures of causal effect
1.4 Random variability
1.5 Causation versus association

2 Randomized experiments
2.1 Randomization
2.2 Conditional randomization
2.3 Standardization
2.4 Inverse probability weighting
3 Observational studies
3.1 Identifiability conditions
3.2 Exchangeability
3.3 Positivity
3.4 Consistency: First, define the counterfactual outcome
3.5 Consistency: Second, link counterfactuals to the observed data
3.6 The target trial

Graphical representation of causal effects
6.1 Causal diagrams
6.2 Causal diagrams and marginal independence
6.3 Causal diagrams and conditional independence

6.4 Positivity and consistency in causal diagrams
6.5 A structural classification of bias
6.6 The structure of effect modification
7 Confounding
7.1 The structure of confounding
7.2 Confounding and exchangeability
7.3 Confounding and the backdoor criterion
7.4 Confounding and confounders
7.5 Single-world intervention graphs
7.6 Confounding adjustment

8 Selection bias
8.1 The structure of selection bias
8.2 Examples of selection bias
8.3 Selection bias and confounding
8.4 Selection bias and censoring
8.5 How to adjust for selection bias
8.6 Selection without bias

**II Causal inference with models **

11 Why model?
11.1 Data cannot speak for themselves
11.2 Parametric estimators of the conditional mean
11.3 Nonparametric estimators of the conditional mean
11.4 Smoothing
11.5 The bias-variance trade-off

12 IP weighting and marginal structural models
12.1 The causal question
12.2 Estimating IP weights via modeling
12.3 Stabilized IP weights
12.4 Marginal structural models
12.5 Effect modification and marginal structural models
12.6 Censoring and missing data
13 Standardization and the parametric g-formula
13.1 Standardization as an alternative to IP weighting
13.2 Estimating the mean outcome via modeling
13.3 Standardizing the mean outcome to the confounder distribution
13.4 IP weighting or standardization?
13.5 How seriously do we take our estimates?
14 G-estimation of structural nested models
14.1 The causal question revisited
14.2 Exchangeability revisited
14.3 Structural nested mean models
14.4 Rank preservation
14.5 G-estimation
14.6 Structural nested models with two or more parameters
15 Outcome regression and propensity scores
15.1 Outcome regression
15.2 Propensity scores
15.3 Propensity stratification and standardization
15.4 Propensity matching
15.5 Propensity models, structural models, predictive models
16 Instrumental variable estimation

16.1 The three instrumental conditions
16.2 The usual IV estimand
16.3 A fourth identifying condition: homogeneity
16.4 An alternative fourth condition: monotonicity
16.5 The three instrumental conditions revisited
16.6 Instrumental variable estimation versus other methods

**III Causal inference for time-varying treatments **

19 Time-varying treatments
19.1 The causal effect of time-varying treatments
19.2 Treatment strategies
19.3 Sequentially randomized experiments
19.4 Sequential exchangeability
19.5 Identifiability under some but not all treatment strategies
19.6 Time-varying confounding and time-varying confounders
20 Treatment-confounder feedback
20.1 The elements of treatment-confounder feedback
20.2 The bias of traditional methods
20.3 Why traditional methods fail
20.4 Why traditional methods cannot be fixed
20.5 Adjusting for past treatment
21 G-methods for time-varying treatments
21.1 The g-formula for time-varying treatments
21.2 IP weighting for time-varying treatments
21.3 A doubly robust estimator for time-varying treatments
21.4 G-estimation for time-varying treatments
21.5 Censoring is a time-varying treatment
21.6 The big g-formula
22 Target trial emulation
22.1 Intention-to-treat effect and per-protocol effect
22.2 A target trial with sustained treatment strategies
22.3 Emulating a target trial with sustained strategies
22.4 Time zero . . . . . .
22.5 A unified approach to answer What If questions with data
23 Causal mediation
23.1 Mediation analysis under attack
23.2 A defense of mediation analysis
23.3 Empirically verifiable mediation
23.4 An interventionist theory of mediation

Prerequisiti

The course is offered in english, fluency in the english language is an important condition for successfully attending the lectures (listening comprehension and reading ability) and reading the additional material.

Metodi didattici

Blended/elearning: frontal lectures, online lectures, class work, seminars, use of Stata and Dagitty. Online teaching

will be used to allow students (not coming to class) to follow some of the lectures of the course. Lectures will be prepared and uploaded during the course. Moreover, homeworks will be discussed and possible solutions proposed once the homeworks are corrected.

Modalità di verifica dell'apprendimento

1. Class Project (group): (80 %)
2. Final Written exam (20 %)

Both call project based on team work and the final exam will help to evaluate the student knowledge necessary to assess the goal of the study being proposed to be analyzed using dagitty and Stata

Testi di riferimento

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman &Hall/CRC. http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/

Periodo di erogazione dell'insegnamento

II Semester, IV term

Lingua di insegnamento

English

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ