International Programme of Advanced Epidemiology and Statistics (IPAES)

ESCOLA SALUT PÚBLICA
ESCUELA SALUD PÚBLICA
LLATZERET - MENORCA

Causal inference

September 17-21, 2018

Welcome

Welcome to the 2nd meeting of the International Programme of Advanced Epidemiology and Statistics. This 2nd meeting features two course sequences, run in parallel. **Causal Inference Sequence I** includes two modules: *Causal inference with directed graphs* and *Applied causal mediation analysis*. **Causal Inference Sequence II** includes three modules: *Propensity score and other matched-sampling strategies for observational studies*, *Methods for attrition and missing data*, and *Instrumental variables analysis*. The goal of these courses is to empower epidemiologists, biomedical scientists, public health professionals, research clinicians, social scientists, and statisticians to apply causal inference with confidence. The courses are taught in English by renowned international lecturers. The programme is held in the *Llazeret* of Maó within the Public Health School of Menorca. Participants will enjoy an intense learning experience in a unique historical environment. The *Llatzeret* is a small island at the port of Maó wherein all boats entering Spain during the 19th century had to stop for quarantine to get a certificate of being free of infection diseases.

Mònica Guxens
Programme director

Causal Inference Sequence I

Causal Inference with Directed Graphs - *Felix Elwert*

This 2.5-day module offers an applied introduction to directed acyclic graphs (DAGs) for causal inference from observational data. DAGs are rigorous and accessible tools for understanding and solving complicated causal problems. All causal inference relies on causal assumptions, and DAGs are a graphical notation for these causal assumptions. Analysts can use DAGs to derive the statistical implications of their causal assumptions and determine which statistical associations equal ("identify") causal effects. We will use DAGs to study the identification of causal effects, with particular emphasis on identification by adjustment, which underlies the use of regression and matching techniques for causal inference. We will animate the material with numerous examples from the social and biomedical sciences.

Outline:

1. Counterfactual causality
2. Directed Acyclic Graphs (DAGs)
3. Statistical implications of causal assumptions
4. Causal identification analysis
5. Selection bias

Applied Causal Mediation Analysis – *Theis Lange*

This 2.5-day module consists in one part on the theoretical foundation for causal mediation analysis and another one hands-on conduction of causal mediation analyses. Mediation analysis concerns assessing the mechanisms and pathways by which causal effects operate. Discussion will be given as to when the standard approaches to mediation analysis are or are not valid. The no-confounding assumptions needed for these techniques will be described. The use and implementation of sensitivity analysis techniques to assess the how sensitive conclusions are to violations of assumptions will be covered, as will extensions to multiple mediators. We will also discuss interaction on additive and multiplicative scales, and their relation to statistical models. The practical conduction of mediation analysis will be based on natural effects model and conduction in different software packages (R, SAS and Stata) will be thoroughly addressed.

Outline:

1. Concepts and methods on causal mediation
2. Natural effect models definitions and use
3. Sensitivity analysis and multiple mediators
4. Conduction in multiple software packages
5. Case studies

Causal Inference Sequence II

Propensity Score and other Matched-Sampling Strategies for Observational Studies - *Marie-Abèle Bind*

This 1.5 day offers an overview of matching methods that facilitate the validity of estimated causal effects of a non-randomized exposure on an outcome. Matched-sampling strategies, including propensity score analysis, help the reconstruction of hypothetical randomized experiments before any outcome data are observed, in which the statistical analysis can be easily performed by comparing the outcome in the exposed and non-exposed units. We will illustrate some matching approaches in a specific observational study and contrast the methods to standard regression-based adjustments.

Outline:

1. Reconstruction of hypothetical experiments
2. Propensity score and other matched-sampling strategies (e.g., optimal, pair matching)
3. Analysis of matched datasets
4. Application to a real data set

Methods for Attrition and Missing data - *Xavier Basagaña*

This 1-day module will provide methodological tools to overcome potential selection biases induced by attrition, i.e. selected individuals not participating in the study or being lost to follow up, or by having missing information in some of the key study variables. Incomplete data are ubiquitous in observational and experimental research. An inappropriate analysis of a study with incomplete data can lead to incorrect inferences, both in terms of bias and in the quantification of uncertainty. We will discuss the potential problems induced by incomplete data and will provide an overview of two main methods that can deal with such problems, namely inverse probability weighting and multiple imputation. The module will give special focus on the practical implementation of these techniques in realistic settings.

Outline:

1. Implications of attrition and missing data for causality
2. Overview of methods and assumptions
3. Multiple imputations
4. Inverse probability weighting
5. Combination of multiple imputation and inverse probability weighting

Instrumental Variables Analysis - *Felix Elwert*

This 2.5-day module covers instrumental variables (IV) analysis for applied researchers. Instrumental variables estimation is a quasi-experimental approach for identifying causal effects when ordinary regression fails. In observational studies, instrumental variables estimation recovers causal effects in the presence of unmeasured confounding and measurement error. In randomized experiments, instrumental variables analysis corrects for non-compliance. In epidemiology, “Mendelian Randomization” has become an important application of instrumental variables analysis, which uses genes as instrumental variables. Building on a thorough understanding of the basics, advanced topics will include non-linear instrumental variables models (including survival models) and instrumental variables estimation for heterogeneous treatment effects (i.e. “LATE” effects). Throughout, the course will focus on building substantive intuitions with several real running examples, including Mendelian randomization. The course aims to empower participants to adjudicate whether instrumental variables estimation solves their specific applied problems.

Outline:

1. Rules for working with linear models
2. Instrumental variables: a solution for omitted variables bias
   2.1. Some math and a lot of substance
   2.2. Just-identified models (IV)
   2.3. Over-identified models (multiple IVs)
3. Understanding exclusion violations: rules and applied heuristics
4. Estimation options—IV, 2SLS, GMM, LIML
   4.1. Getting standard errors right
5. Weak and many instruments bias
6. IV as a solution for measurement error
7. Mendelian randomization
8. Non-linear IV
   8.1. Heterogeneous effects and LATE

Faculty

Felix Elwert

Dr. Elwert is Romnes Professor of Sociology and affiliated Professor of Biostatistics and Population Health Sciences at the University of Wisconsin–Madison. He holds graduate degrees in sociology and in statistics from Harvard University. Methodologically, he works on problems of causal inference in observational settings. Substantively, he conducts observational and field-experimental research in social demography, social inequality, and social networks. Dr. Elwert was the first winner of the *Causality in Statistics Education Award* from the American Statistical Association in 2013.

Theis Lange

Dr. Lange is Associate Professor of at the Section of Biostatistics of the Department of Public Health of the University of Copenhagen. He holds a degree in mathematical statistics from the University of Copenhagen. His primary fields of research are developing methods for causal inference and mediation, analyzing causal pathways involved in creating social inequality in cancer, and comparing different measures of mediation in a survival analysis context.

Marie-Abèle Bind

Dr. Bind is a John Harvard Distinguished Science Fellow and Research Associate at the Department of Statistics at Harvard University. She received a dual Doctor of Science degree in Environmental Health and Biostatistics from the Harvard School of Public Health. Her research interests focus on defining causal questions being asked by describing real or hypothetical multifactorial interventions, and developing new statistical methods for quantifying the causal effects of environmental exposure on health outcomes.

Xavier Basagaña

Dr. Basagaña is Associate Research Professor in biostatistics at the Barcelona Institute for Global Health. He holds a graduate degree in biostatistics from Harvard University. His research concerns methodology on statistical methods to produce valid interferences in observational studies (e.g. on measurement error and missing data), relationship between extreme temperatures and health, and relationship between air pollution and health.

Who should attend?

The programme is aimed at applied researchers with an interest in causal inference from observational data, in particular for epidemiologists, biomedical scientists, public health professionals, research clinicians, social scientists, and statisticians. Participants should have a good working knowledge of applied regression analysis and an intermediate knowledge level of epidemiology.

Materials

Participants will receive a bound manual containing detailed lecture notes, reading materials, and many other useful features. In the "Causal inference and directed acyclic graphs" module, which aims to strengthen your ability to *think through* causal problems, we will work through numerous pencil-and-paper exercises. In the rest of modules participants need to bring their laptop computers to participate in the software-based exercises. Software code in SAS, STATA, and R will be provided.

Registration

The registration fee of each course sequence is 1,200€ and includes all course materials, lunch and coffees from Monday September 17 to Friday September 21 2018, and the Course Dinner on Tuesday September 18. Participants have to book their accommodation themselves. We will provide a bus service for the transfer between the hotels located in Maó and the *Llatzeret*. Further details for registration will be announced shortly. For further information, please contact us at ipaes.emps@gmail.com or visit our webpage www.emsp.cime.es.

With the collaboration of ISGlobal