Causal Inference

Andrea Rotnitzky
Universidad Di Tella and Harvard University

This course discusses statistical theory and methodology aimed at addressing causal inquiries from observational data and complex randomized designs.
The students will learn to:

1) Formulate causal contrasts of interest for addressing specific scientific inquiries.
2) Derive graphical models for investigating the conditions under which the causal contrasts of interest are identified from data collected under specific study designs.
3) Formulate adequate structural models for making inference about the causal contrasts of interest
4) Derive efficient, double robust, estimators for the causal contrasts of interest under the postulated structural models.

Two key modeling approaches are studied: the directed acyclic graphs (DAG) models and models for counterfactual variables (structural models).

The theory of DAGS is used to:

i) formally define and connect key causal concepts including confounding, exchangeability, overall effects, direct and indirect effects, intermediate variables, instrumental variables, non-compliance and selection bias; and
ii) derive conditions for identifiability of causal contrasts.

DAGS will also used to demonstrate that parameters of standard regression models do not have interpretation as causal contrasts for the effects of time-varying exposures in the presence of time dependent covariates that are simultaneously confounders and intermediate variables.

Two broad classes of models for counterfactual variables will then be introduced whose parameters have causal interpretation: *marginal structural models and structural nested models*. The statistical theory of estimation under these models will be discussed. Time permitting, the discussion will include the derivation of

i) the class of inverse weighted probability estimators for parameters of marginal structural models,
ii) the class of g-estimators for parameters of structural nested mean models and, i
iii) locally semiparametric efficient, doubly-robust estimators for the parameters in each of the models.