Causal Analysis in SAS/STAT®

Propensity Score Analysis and Graphical Causal Models

Overview

If you work with nonrandomized trials or observational data in which treatment conditions are not randomly assigned to individuals, confounding covariates can bias your causal effect estimation by inducing extraneous associations between the treatment and outcome. A regression analysis with a simple covariate adjustment might not always lead to valid causal inferences, especially when there are systematic differences between the distributions of the covariates in the treatment and control groups. SAS/STAT offers you several procedures that help you make valid causal inferences under the assumptions of the Neyman-Rubin potential outcome framework (Rubin 1990) or a counterfactual outcome framework (Pearl 2009). These procedures provide a variety of advanced numerical and graphical techniques for causal inference, including propensity score matching and weighting, analysis of graphical causal models, doubly robust estimation, and a regression approach to causal mediation analysis.

Propensity Scores and Matching

The PSMATCH procedure provides a variety of tools for propensity score analysis. You use PROC PSMATCH to compute propensity scores, which estimate the probability that a subject is assigned to treatment given a set of pretreatment covariates. You can use the propensity scores to adjust for the confounding variables and to create a new data set that has balanced treatment conditions. PROC PSMATCH provides the following methods:

- matching treated and control units to adjust for the confounding and create a new data set that has balanced treatment conditions
- computing inverse probability of treatment weights for subsequent estimation of treatment effect with a weighted analysis
- stratification of observations that have similar propensity scores for use in a subsequent outcome analysis

PROC PSMATCH provides statistical and graphical methods of assessing the balance of the propensity scores and the covariates between the treatment and control groups, as shown by the two graphs on this page.

Graphical Causal Models

The CAUSALGRAPH procedure examines the structure of graphical causal models and suggests statistical strategies that enable you to compute unbiased estimates of causal effects. With PROC CAUSALGRAPH, you can obtain a valid adjustment set that can remove noncausal influences.
You use PROC CAUSALGRAPH to define graphical causal models in the form of directed acyclic graphs (DAGs). PROC CAUSALGRAPH lets you explore formal properties of a causal graph, enabling you to take the following actions:

- list or test adjustment sets that can be used to remove or block noncausal associations between the treatment and outcome variables
- list or test sets of variables that can be used as instruments to estimate a causal effect
- clarify causal and noncausal (associative) paths between the treatment and outcome variables
- enumerate conditional independence assumptions encoded in a causal model

PROC CAUSALGRAPH analyzes multiple causal models simultaneously to obtain common identification criteria. The identification results from PROC CAUSALGRAPH enable you to devise sound statistical strategies to estimate causal effects in confounding situations.

Causal Estimation

The CAUSALTRT procedure estimates the average causal effect of a binary treatment variable $T$ on a continuous or discrete outcome $Y$. Depending on the application, the variable $T$ can represent an intervention, an exposure to a condition, or an existing characteristic of subjects. PROC CAUSALTRT estimates two types of causal effects: the average treatment effect and the average treatment effect for the treated.

You can adjust for confounding by using PROC CAUSALTRT to model the treatment assignment or the outcome or both. Modeling the treatment assignment leads to inverse probability weighting methods, and modeling the outcome leads to regression adjustment methods. Modeling both leads to doubly robust methods that can provide unbiased estimates for the treatment effect even if one of the models is misspecified.

Causal Mediation Analysis

Causal mediation analysis considers the decomposition of a treatment’s total causal effect on an outcome. You use the CAUSALMED procedure to estimate causal mediation effects from observational data. In a causal mediation analysis, there are four main variables of interest:

- an outcome variable $Y$
- a treatment variable $T$ that is hypothesized to have direct and indirect causal effects on the outcome variable $Y$
- a mediator variable $M$ that is hypothesized to be causally affected by the treatment variable $T$ and itself has an effect on outcome variable $Y$
- a set of background covariates $C$ that confound the observed relationships among $Y$, $T$, and $M$

The main goal of the analysis is to obtain unbiased estimates of the direct causal effect of the treatment $T$ on $Y$ and the indirect causal effect of $T$ on $Y$ through the mediator $M$.

PROC CAUSALMED uses a regression-based method of estimation (Valeri and VanderWeele 2013; VanderWeele 2015) within the counterfactual framework of Robins and Greenland (1992) and Pearl (2001). This framework provides a unified foundation for defining direct and indirect effects in a very general setting, which allows for treatment and mediator interactions, binary and continuous outcomes, and binary and continuous mediators. The framework also provides conditions for a valid causal mediation effect estimation.

For More Information

For complete information about all SAS/STAT releases and procedures, see the documentation available at support.sas.com/statistics.