Problem Set 4-Write Up

Olivia Bogiages

2025-12-03

1. What is the difference between a confounder and a collider? How should you address each in your models? A confounder is a variable that has a causal relationship with both X and Y. A collider is a variable that has an association with both X and Y where the relationship direction is towards the collider. In your model, you should always include a confounder, whether or not you are looking at total or direct affect. You never want to include a collider in your model because this would open up pathways the illustrate false relationships.

2. How can conditioning on a collider create bias? Conditioning on a collider creates bias by opening up false relationships between variables within your model. Conditioning in this way could create bias because x and y have a directional association towards z and thus conditioning on it would create a non-existant association.

3. Why can't statistical summaries or correlations alone tell us whether to control for a variable? Statistical summaries and correlations cannot tell us whether to control for a vairable because they do not tell us what the relationship that variable has with the varaibles of interest in the model. You do not know if it is a confounder, in which case you should control for it, or a mediator, or a collider, in which case you should evaluate if you want the direct effect (for the mediator) or you should leave it out altogether (with the collider.)

4. What is meant by a "kitchen sink" regression, and what is wrong with this approach to modeling? Kitchen sink regression refers to including all available variables in a multivariable regression model, regardless of if those variables are associated with the outcome of interest. This issue with doing this is that, if you do not evaluate the nature of varaible (confounder, mediator, collider etc.) then you might create bias within your model by unintentionally opening pathways that are not aspects of the true relationship.

5. What is a "backdoor path" and how does multiple regression help block these paths? A backdoor path is when a pathway runs from the confounder to your independent variable. Multiple regression includes confounders and thus close the paths in order to isolate x's and y's.