

Fundamental Problem of Causal Inference

We cannot observe two values of a variable in the same unit at the same time.

Since only one value can be observed, the other value only exists in a **Counterfactual**.



Assumptions:

Unit Homogeneity

The expected values of the dependent variable are the same in two units which have the same values of the independent variable.

This helps us estimate the **Counterfactual**.

(Conditional) Independence

The values of each independent variable are assigned independently from the values of the dependent variable (and independently from the values of the other independent variables).

This assumes there are no problems of **Endogeneity** (or **Multicollinearity**).

Definitions:

Endogeneity: The dependent variable causes the independent variable.

Multicollinearity: Two independent variables are highly correlated with each other.

Measurement Error

Uncertainty

All measurements and conclusions are uncertain.

Providing estimates of our uncertainty allows others to interpret our results.

Types of Bias:

Omitted Variable Bias

Our conclusions will be biased if we do not include a relevant independent variable.

This can be a major problem, and can only be avoided by careful theory-building

Selection Bias

Our conclusions will be biased if our sample of cases is different from the population as a whole in important ways.

Bias makes our estimated relationship between variables different from their true relationship.

Some sources of **Selection Bias**:

Causes of Effects

A question asking “What causes Y?” instead of “What is the relationship between X and Y?”

Selecting on the Dependent Variable

No valid inferences can be made from a sample of cases that all have the same value of the dependent variable.