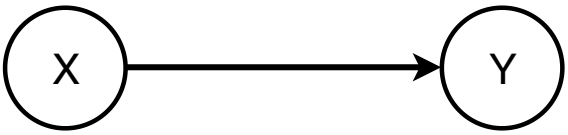
Statistical Learning Group – Causality

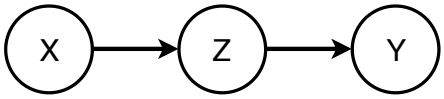
What is causality?

Following Lewis (1973), "X causes Y" can be defined by the notion of multiple words and **counter factuals.**

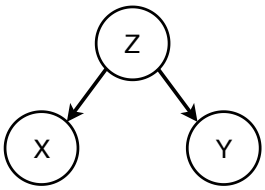


"If not for X, Y would not have occurred"

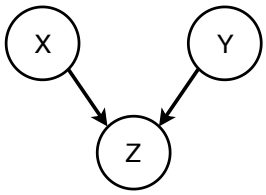
Three Fundamental Forms



Mediator



Confounder

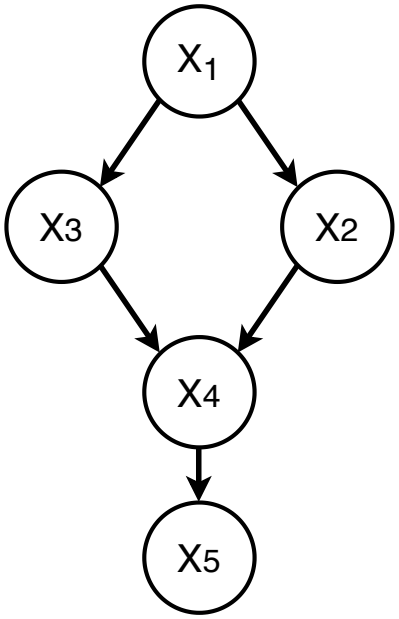


Collider

You must control for confounders and you cannot control for colliders (either will introduce bias).

Structural Causal Model (SCM)

A SCM is a triple $M =< U, V, F >$



U – Set of exogenous variables
V – Set of endogenous variables
F – Set of functions relating each V_i to its parents and u_i

$$V_i = f(PA_i, u_i)$$

e.g. $X_4 = \alpha X_3 + \beta X_2 + u_4 \quad u_4 \sim \mathcal{N}$

We can ask questions from these models
e.g. What is $\mathbb{E}[X_5] =$?

We can also climb the Ladder of Causation

1. Association – *What if I see?*
2. Intervention – *What if do? How?*
3. Counter factual – *What if I had done? Why?*