Correlation vs Causation

Correlation

Definition

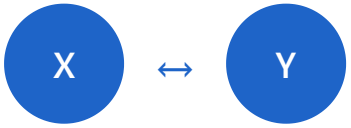
Statistical association between variables

Pearson Correlation

$$\rho = \mathrm{Cov}(X,Y) \;/\; (\sigma_x \sigma_Y)$$

$\rho = -1$	$\rho = 0$	$\rho = 1$
Perfect negative	No linear	Perfect positive

Association only



Causation

Definition

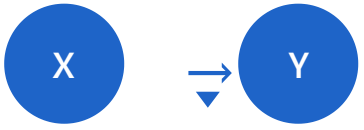
X directly influences Y

Causal Relationship

Requires:

- Temporal precedence
- Mechanism
- Control of confounders

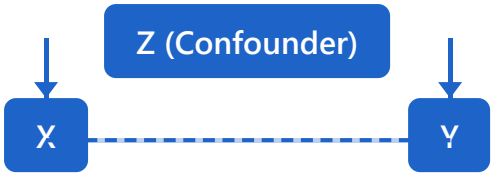
Direct influence



⚠ Key Warning

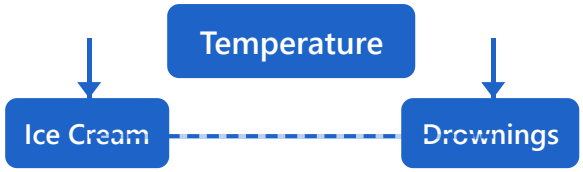
Correlation does NOT imply causation

Confounding Variable



Z affects both X and Y,
creating spurious correlation

Classic Example



Correlated, but ice cream
doesn't cause drownings

Regression Limitation

Regression shows association, not necessarily causation