

완벽한 모델을 향해서



$$y = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

빅콘테스트 예선 탈락

하지만 세상 어딘가 분명 영화관객수를 정확히 예측할 수 있는 모델이 있지 않을까



일단... 내 모델이 필요한 독립변수
들을 다 가지고 있는건가?

- 필요한 변수가 빠진 경우
- 필요없는 변수가 있는 경우

Unbiasedness

$$E(X_1) = \mu$$

필요한 독립변수가 없는 경우

Biased estimator

필요없는 독립변수가 있는 경우

Unbiased but with much more variance

Consistency

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$$

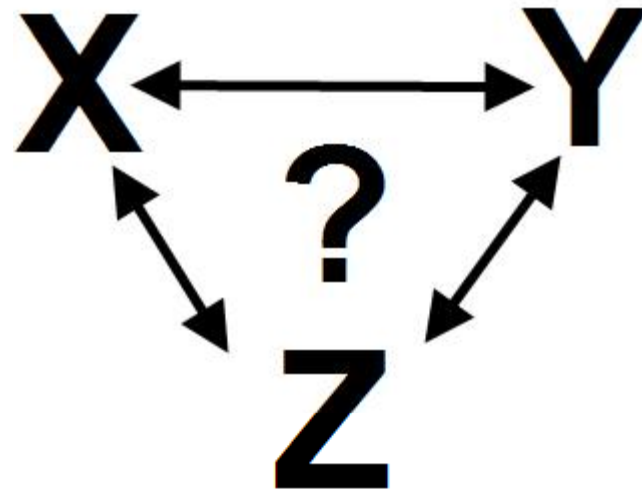
애초부터 내 모델이 필요한 독립변수를
모두 가지고 있지 않다면 샘플을 더 늘려봤자
다른 변수를 넣어봤자 큰 의미가 없다.

$$y = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

그렇다면 어떤 변수가 유의미하고
무의미 한지 알 수 있을까?

이 독립변수가 진짜로 내 종속변수
에 영향을 줄까..?

Causality



변수들간에 아무리 복잡한 관계가 있어도 회귀 모델 하나로 인과관계를 알 수 있다.

$$y = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$