# Random variability

Hwichang Jeong

2021. 07.29

Seoul National University

# Introduction

- There are two qualitatively different reasons why causal inference may be wrong.
  - Systemic bias
  - Random variability
- Three types of systemic bias :
  - Confounding bias (Chap7)
  - Selection bias (Chap8)
  - Measurement bias (Chap9)

# Introduction

- We have ignored random variability in previous nine chapters.
- We compute causal effects in study populations of near infinite size (unrealistic).

# Causal effect

- Association

$$Pr(Y = 1|A = 0) - Pr(Y = 1|A = 1)$$

- Exchangeability

$$Pr(Y^a = 1|A = 0) = Pr(Y^a = 1|A = 1)$$

- When we assume exchageability, association is equal to causal effect.
- Here $Pr$ is not probability but proportion.

# Causal effect - non-infinite population

Table 2.1

| | $A$ | $Y$ | $Y^0$ | $Y^1$ |
|---|---|---|---|---|
| Rheia | 0 | 0 | 0 | ? |
| Kronos | 0 | 1 | 1 | ? |
| Demeter | 0 | 0 | 0 | ? |
| Hades | 0 | 0 | 0 | ? |
| Hestia | 1 | 0 | ? | 0 |
| Poseidon | 1 | 0 | ? | 0 |
| Hera | 1 | 0 | ? | 0 |
| Zeus | 1 | 1 | ? | 1 |
| Artemis | 0 | 1 | 1 | ? |
| Apollo | 0 | 1 | 1 | ? |
| Leto | 0 | 0 | 0 | ? |
| Ares | 1 | 1 | ? | 1 |
| Athena | 1 | 1 | ? | 1 |
| Hephaestus | 1 | 1 | ? | 1 |
| Aphrodite | 1 | 1 | ? | 1 |
| Cyclope | 1 | 1 | ? | 1 |
| Persephone | 1 | 1 | ? | 1 |
| Hermes | 1 | 0 | ? | 0 |
| Hebe | 1 | 0 | ? | 0 |
| Dionysus | 1 | 0 | ? | 0 |

- Treatment $A$ : heart transplant
- Outcome $Y$ : survive or die.
- Assign $A_i = 0$ or 1 for all individual $i$ randomly.
- Not exchangeable although we do randomization.

$$Pr(Y^{a=0} = 1 | A = 0) = \frac{3}{7}$$

$$Pr(Y^{a=0} = 1 | A = 1) = \frac{7}{13}$$

- If population size becomes infinitely large, groups become exchangeable.

# Assumption & Estimation

- Suppose that there is a super-population so large that we can regard it as infinite.
- Our goal is to make inferences about the super-population.
- Estimand : The parameter of interest in the super-population ($e.g$ $Pr(Y = 1|A = a)$).
- Estimator : A rule that produces a numerical value for the estimand from samples.($e.g$ $\hat{Pr}(Y = 1|A = a)$).
- As sample size increases, the estimates get closer to the estimand (consistency).

# Example

Table 2.1

| | $A$ | $Y$ | $Y^0$ | $Y^1$ |
|---|---|---|---|---|
| Rheia | 0 | 0 | 0 | ? |
| Kronos | 0 | 1 | 1 | ? |
| Demeter | 0 | 0 | 0 | ? |
| Hades | 0 | 0 | 0 | ? |
| Hestia | 1 | 0 | ? | 0 |
| Poseidon | 1 | 0 | ? | 0 |
| Hera | 1 | 0 | ? | 0 |
| Zeus | 1 | 1 | ? | 1 |
| Artemis | 0 | 1 | 1 | ? |
| Apollo | 0 | 1 | 1 | ? |
| Leto | 0 | 0 | 0 | ? |
| Ares | 1 | 1 | ? | 1 |
| Athena | 1 | 1 | ? | 1 |
| Hephaestus | 1 | 1 | ? | 1 |
| Aphrodite | 1 | 1 | ? | 1 |
| Cyclope | 1 | 1 | ? | 1 |
| Persephone | 1 | 1 | ? | 1 |
| Hermes | 1 | 0 | ? | 0 |
| Hebe | 1 | 0 | ? | 0 |
| Dionysus | 1 | 0 | ? | 0 |

- We randomly choose $n$ samples drawn from super-population. ($n = 20$)
- Each individuals in super-population are assigned $A_i$ randomly.
- $\hat{Pr}(Y = 1|A = a)$ is consistent and unbiased estimator of $Pr(Y = 1|A = a)$
- Since super-population is large enough and $A_i$ is assigned randomly, exchangeability holds.
- $\hat{Pr}(Y = 1|A = 1)$ is consistent and unbiased estimator of $Pr(Y^{a=1} = 1)$ and $\hat{Pr}(Y = 1|A = 0)$ is consistent and unbiased estimator of $Pr(Y^{a=0} = 1)$
- We can estimate $Pr(Y^{a=1} = 1) - Pr(Y^{a=0} = 1)$ as $\hat{Pr}(Y = 1|A = 1) - \hat{Pr}(Y = 1|A = 0)$

# Conditionality principle

- Suppose there is a variable $L$ which is associated to treatment $A$.
- We should adjust for $L$ because of the confounding.
- In statistics, the $(L, A)$ is said to be an ancillary statistic for causal risk difference.
- Conditionality principle : Inference on a parameter should be performed conditional on ancillary statistics.

# Conditionality principle

- Assume $sRD = \Pr(Y = 1 \mid L = l, A = 1) - \Pr(Y = 1 \mid L = l, A = 0)$ known to be constant across strata $L$.
- The parameter of interest is the stratum-specific causal risk difference.
- The likelihood of the data $\{Y_i, A_i, L_i\}_{i=1}^{n}$ is

$$\prod_{i=1}^{n} f\left(Y_i \mid L_i, A_i; sRD, p_0\right) \times f\left(A_i \mid L_i; \alpha\right) \times f\left(L_i; \rho\right)$$

  Where $p_0 = (p_{01}, p_{02})$ with $p_{0l} = \Pr(Y = 1 \mid L = l, A = 0)$, $\alpha$, and $\rho$ are nuisance parameters associated with the conditional density of $Y$ given $A$ and $L$.
- $A$ and $L$ are ancillary statistics for the parameter of interest when the $f\left(Y_i \mid L_i, A_i; p_0\right)$ depends on the parameter of interest, but the joint density of $L$ and A does not share parameters with $f\left(Y_i \mid L_i, A_i; sRD, p_0\right)$.

# Curse of dimensionality

- Suppose the investigators had measured 100 pre-treatment binary variables.
- Pre-treatment variable $L$ is formed by combining the 100 variables.
- We say that the data is of high dimensionality when there are many possible combinations of values of the pre-treatment variables.

# Curse of dimensionality

- There is $2^{100}$ strata, a few strata contain both a treated and an untreated individual.
- Curse of dimensionality : The conditional estimator is uninformative when there are many pre-treatment variables.