Causal Inference from Observational Data using R

Training dates: 06-13 May 2023

Anteneh Tessema, Diribsa Tsegaye and Yebelay Berelie

Ethiopian Statistical Association (ESA)

2023-05-06

- Mainly based on Prof R Machekano's Lecture notes for Analysis of Observational Data: Causal Inference Module in the Division of Epidemiology and Biostatistics at the University of Stellenbosch

Outlines

1. Introduction to causal inference
2. Randomised experiments
3. Observational studies
4. Method for estimating causal effects in observational data
5. Marginal structural models (MSM)
6. Doubly Robust (DR) estimators - combination methods
7. Marginal Mean Weighting through Stratification (MMWS)
8. Causal inference from survival and longitudinal data

1 Introduction to causal inference

Outline

- Motivation for causal inference
- Causal inference framework
- Potential outcomes framework / counterfactuals model
- Fundamental problem of causal inference
- Defining and measures of causal/treatment effects (individual and average)
- Causation versus association
- Assumptions for causal inference
- Random variability
- Roadmap to causal effect estimation

Readings:

- Chapter 1. MH&JR
- Rubin DR. 1974
- Holland, 1986
- Chapter 1. Rosenbaum PR - Design of Observational Studies

Objectives

- Motivate reasons for causal inference
- Present a framework for Causal inference
- Define causal effect
- Define measures of causal effect
- Distinguish between causation and association
- Assumptions for causal inference
- Understand random variability in causal inference
- Roadmap to causal effect estimation

Introduction

- Aim of most epidemiological studies is to search for causes of diseases.
- Clinical studies aim to establish effects of treatments or interventions.
- The existence of other factors related to outcomes and exposures distort relationships of interest causal inference.
- There are various data analysis approaches to estimate the causal effect of interest under a particular set of assumptions when data are collected on each individual in a population.
- Numerical quantities that measure changes in the distribution of an outcome under different interventions.

Motivation for causal inference

We ask questions so that we can take action.

Example

- What is the effect of smoking on lung cancer?
- What is the effect of exercise on hypertension?
- What is the effect of Prevention of Mother to Child Transmission (PMTCT) intervention on transmission of HIV from infected mother to baby?
- What is the effect of nurse task shifting on quality of HIV care of infected patients?
- What is the effect of transport reimbursement on antenatal care (ANC) attendance among pregnant women?

HIV causes AIDS

- How would you establish this fact?
- We cannot do such an experiment in humans.
- We can do it with mice or monkeys.
- In humans, we observe those infected versus not infected.
- There are potential differences between people with and those without HIV - not comparable.
- However, we can easily determine association.

Observational studies

- Treatment groups not directly comparable.
- Biased treatment or exposure effects.
- Statistical adjustment for confounders.
    - Does coffee drinking cause lung cancer?
- If a coffee drinker is more likely to be a smoker but the study only measured coffee drinking, results may show coffee drinking increases the risk of lung cancer.
- With smoking recognized as a `confounder`, adjustments can be made in study design or data analysis.
- Without adjustment, risk of false positive (Type I) error.
- What is a `confounder`?

Simpson's Paradox

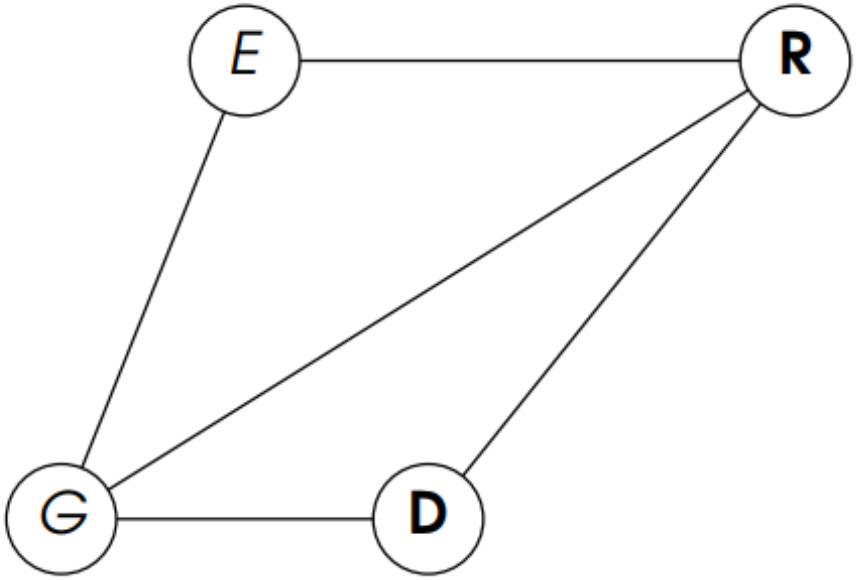
- Results from a study into a new drug on recovery among sick patients

	Drug	No Drug
Men	81/87(93%)	234/270 (87%)
Women	192/263(73%)	55/80 (69%)
Combined	273/350(78%)	289/350(83%)

- Drug on recovery better in men and women but not in a combined sex.
- What is policy recommendation - to treat or not to treat?

- Always understand the story behind the data - `causal mechanism`.

- We need additional information/fact



- Women taking drug = 77% compared to 24% in men.
- Randomly selecting a drug user likely to be a woman less likely to recover
- Randomly selecting a non-drug user likely to be a man more likely to recover
- Effectiveness require that you compare like to like - remove effect of estrogen

- Abbreviations: D: Drug, R: Recovery, E: Estrogen, G: Gender

Example: Relationship between Cholesterol, Exercise and Age

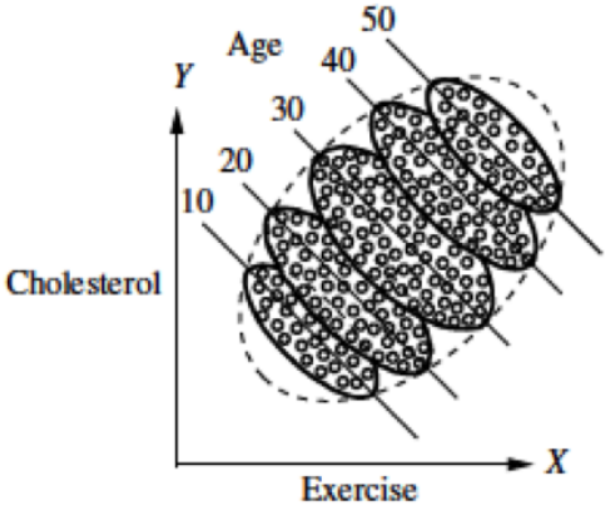


Figure 1.1: Results of the exercise-cholesterol study, segregated by age

- Relationship between cholesterol and exercise is negative within age group.

Relationship between Cholestrol and Exercise (without considering age)

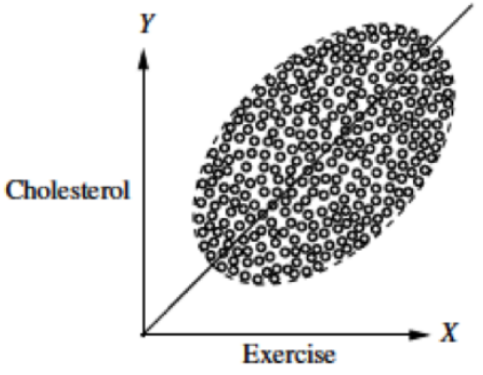


Figure 1.2: Results of the exercise-cholesterol study, unsegregated. The data points are identical to those of Figure 1.1, except the boundaries between the various age groups are not shown

- The relationship distorts/in opposite due to confounder age.

Defining causal inference

Definition

- Variable `X is a cause of a variable Y` if Y in anyway relies on X for its value.
- X is a cause of Y if Y listens to X and decides its value on what it hears.
- Causal inference using the `Neyman-Rubin` potential outcome framework allows adjustment for `confounders` under structural `causal assumptions`.

The story of Zeus and Hera

- Zeus is a heart transplant patient.
- Jan 1, receives a heart transplant `(A = 1)`, dies 5 days later `(Y = 1)`.
- Suppose somehow we know (by divine revelation), had Zeus not received heart transplant `(A = 0)`, he would be alive 5 days later `(Y = 0)`.
- Based on this info, we can conclude that the heart transplant caused Zeus's demise.
- The heart transplant intervention had a causal effect on Zeus's five-day survival.

- Another patient,Hera, is alive `(Y = 1)` 5 days after receiving heart plant. We also know (mysteriously), had she not received a heart transplant, she would still be alive 5 days later `(Y = 1)`. Then the heart transplant did not have a causal effect on Hera.

Notation

- Let A represent an intervention, treatment or exposure or policy.

Assume A takes values 1: treated, 0: untreated.

- Let Y represent the outcome with values 1:death, 0: survival
- A and Y are random variables.
- Let $Y^{a=1}$ (read as Y under treatment $a = 1$) be the outcome variable that would be observed under treatment value `a = 1`.
- Let $Y^{a=0}$ be the outcome variable that would be observed under treatment value `a = 0`.
- $Y^{a=1}$ and $Y^{a=0}$ are also random variables.
- $Y^{a=1}$ and $Y^{a=0}$ are known as `potential outcomes` or `counterfactual` outcomes
- What are Zeus's and Hera's potential outcomes?
- Z: $Y^1 = 1 \neq Y^0 = 0$ ... Treat cause of death vs H: $Y^1 = Y^0 = 0$ ... not cause of death

Individual causal effect

Definition

- Treatment A has a causal effect on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$ for the individual.
- Causal effect for individual i: $Y_i^{a=1} \neq Y_i^{a=0}$
- The action A has a causal effect, causative or preventive on the outcome.

Fundamental problem of causal inference

- We estimate individual causal effects by comparing the `counterfactuals`. But for each individual, we observe only one of the `counterfactuals` corresponding to the treatment the individual received.
- We have a missing data problem - hence individual causal effects cannot be estimated from the observed data

Average Treatment Effect (ATE)

- $E[Y^{a=1} - Y^{a=0}] = E[Y^{a=1}] - E[Y^{a=0}] = P_1 - P_0$ - Risk difference
  - $P_1 - P_0$ if Y is binary: Proportion difference.
  - $\mu_1 - \mu_0$ if Y is continuous: mean difference.
- $\frac{E[Y^{a=1} \times (1 - E[Y^{a=0}])]}{(1 - E[Y^{a=1}) \times E[Y^{a=0}]}$

Measures of causal effect

Consider a binomial outcome Y taking values {0, 1}. Causal Effect measures could be:

1. $Pr[Y^1 = 1] - Pr[Y^0 = 1] \Longrightarrow$ risk difference
2. $\frac{Pr[Y^1=1]}{Pr[Y^0=1]} \Longrightarrow$ risk ratio
3. $\frac{Pr[Y^1=1]=Pr[Y^1=0]}{Pr[Y^0=1]=Pr[Y^0=0]} \Longrightarrow$ odds ratio
4. $\frac{-1}{Pr[Y^1=1]-Pr[Y^0=1]} \Longrightarrow$ number needed to treat (NNT).

- NNT is equal to the reciprocal of the absolute value of the causal risk difference.

Average Causal Effect (ACE)

Hypothetical complete data

Hypothetical complete data

Unit i	Pretreat inputs W_i		Potential ouctomes y_i^0	y_i^1	Treat effect $y_i^1 - y_i^0$
1	1	50	69	75	6
2	1	98	111	108	-3
3	2	80	92	102	10
4	1	98	112	111	-1
⋮	⋮	⋮	⋮	⋮	⋮
100	1	104	111	114	3

- If everyone had received treatment, the average outcome would be $E(Y^1) = \bar{y}_1$
- If everyone had not received treatment, the average outcome would be $E(Y^0) = \bar{y}_0$
- Then an average causal effect of treatment A on outcome Y exists if $E(Y^1) \neq E(Y^0)$
- Absence of average causal effect does not mean absence of individual causal effect

Hypothetical complete data

Unit i	Pretreat inputs W_i		Potential outcomes y_i^0	y_i^1	Treat effect $y_i^1 - y_i^0$
1	1	50	69	75	6
2	1	98	111	108	-3
3	2	80	92	102	10
4	1	98	112	111	-1
⋮	⋮	⋮	⋮	⋮	⋮
100	1	104	111	114	3

$\mathcal{X} = (W, Y^1, Y^0) = (W, Y^a : a \in \mathcal{A})$

Observed data

Unit i	Pretreat inputs W_i		Treat indic T_i	Potential outcomes y_i^0	y_i^1	Obs Outc Y	Trec effe $y_i^1 -$
1	1	50	0	69	?	69	?
2	1	98	0	111	?	111	?
3	2	80	1	?	102	102	?
4	1	98	1	?	111	111	?
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1	104	1	?	114	114	?

$\mathcal{O} = (W, A, Y)$

Notation

Complete data:

$$X = (W; Y^1; Y^0) = (W; Y^a : a \in A)$$

Observed data: missing data problem

$$O = (W; A; Y)$$

- We only get to observe one of the `counterfactuals` (either receive treatment or not)

Association

- From the observed data $O = (W, A, Y)$:
- Estimate conditional means: $E(Y|A = a)$
- When $E(Y|A = 1) = E(Y|A = 0) \Rightarrow$ independence i.e. treatment A and outcome Y are not `associated`

Independence - Associational measures on different scales

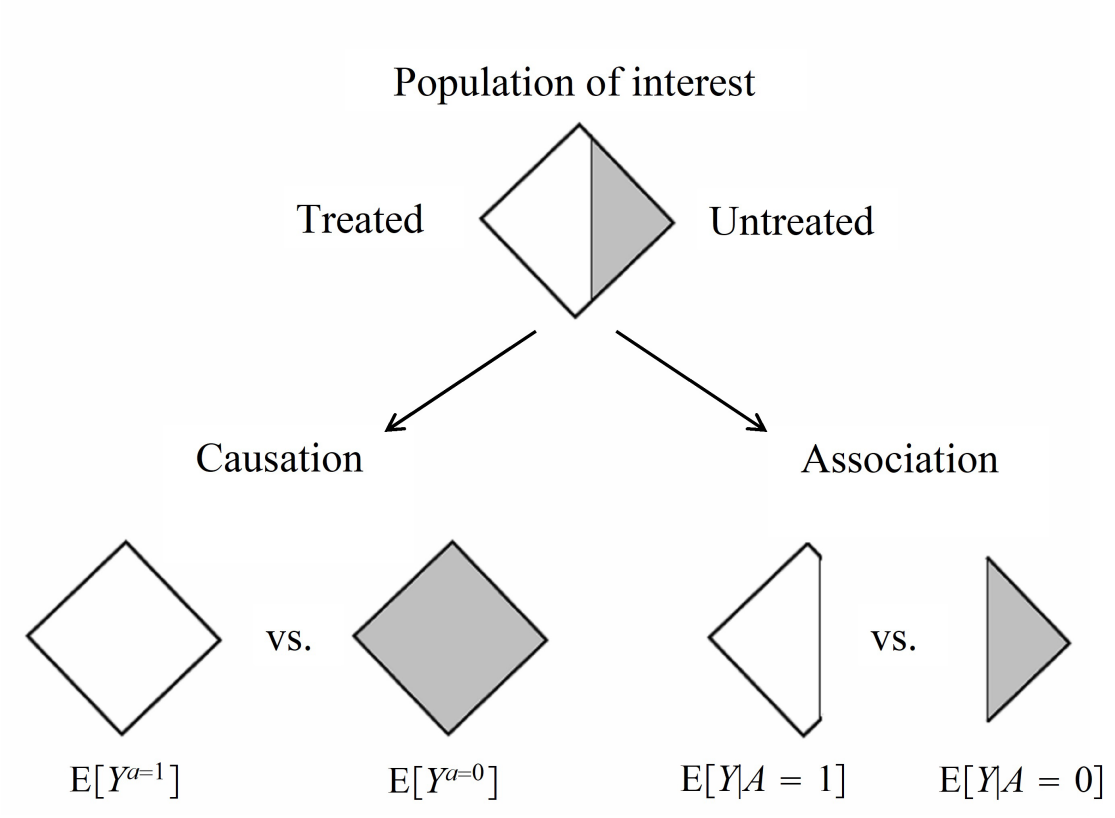
i. Associational risk difference (RD)

ii. Associational risk ratio (RR)

iii. Associational odds ratio (OR)

- Association (categorical) or correlation (continuous)
- No association when: RD = 0, RR = 1, OR = 1.

Illustration of causation and association



Causation versus association

- In association, we deal with conditional expectations - in a subset of the population
- In causation, we deal with unconditional or marginal expectations
    - in entire population
- Association implies different risks in two disjoint subsets of the population determined by subjects' actual treatment
- Causation is defined by a different risk in entire population under two different treatments

Assumptions

To identify ATE, some assumptions are required:

- Conditional exchangeability
- Positivity
  - $Pr(A = a) > 0$
- Consistency
  - For each subject, one of the `counterfactual` outcomes is actually factual.
  - A subject with observed treatment A = a has observed outcome Y equal to the `counterfactual` outcome Y^a.
  - $Y = Y^A$
- Noninterference
  - No interaction between units - independence.

- STUVA (stable-unit-treatment-value assumption)- treatment variation irrelevance
  - Implicit assumption in defining `counterfactual` outcomes under treatment value a is that there is only one version of treatment value A = a
  - If there are different versions of same treatment (surgery performed by a different surgeon), then the `counterfactuals` on same person could be different
  - The assumption of no multiple versions of treatment is included in the STUVA

The presence of random variability

- We only have information on a sample of the population - sampling variability.
- This prevents us from obtaining the exact proportion (or mean) in the population who had the outcome under treatment a
- Random error due to sampling variability prevents us from getting the exact population proportion
- We therefore use $\hat{Pr}[Y^a = 1]$ to estimate $Pr[Y^a = 1]$
- $\hat{Pr}[Y^a = 1]$ is a consistent estimator of $Pr[Y^a = 1]$ because as $n \to \infty$

$$\hat{Pr}[Y^a = 1] - Pr[Y^a = 1]) < \epsilon$$

- We therefore need a statistical test to quantify the chance that any difference between the `counterfactuals` is wholly due to sampling variability

Roadmap to causal effect estimation

- Study design - randomisation
- Regression
- G-formula (standardisation)
- G-computation
- Propensity scores (inverse weighting, stratification, matching)

- Double Robust methods (AIPTW and TMLE)

  Abbreviation:

  - AIPTW: Augmented Inverse Proportional/Probability of Weights
  - TMLE: Targeted Maximum Likelihood Estimator

2 Randomised experiments and observational studies

Outline

- Introduction
  - Randomised experiments and exchangeability
- Conditional Randomisation
  - Example study
  - Stratification
  - Estimating average causal effects
  - Standardisation
  - Inverse probability weighting

Objectives

- To demonstrate why randomised studies are convincing for causal inference.
- Understand the concept of exchangeability.
- Estimating causal effects under conditional randomisation.

Readings

- Chpt 2. MH JR
- Holland, 1986
- Chpt 1. PRR - DOS

2.1 Introduction

- Fundamental problem of causal inference - missing data
- Since we can not observe all potential outcomes, causal inference becomes a prediction problem
- How?
  - Find close substitutes for the potential outcomes (e.g. rats experiments, pre-post etc)
  - Randomisation
  - Statistical adjustment

2.2 Randomised experiments

- All studies generate data with missing counterfactuals.
- Randomisation ensures that those missing values occurred by chance.
- Effects can be consistently estimated in randomised studies despite the missing data.

2.3 Exchangeability

Study

- Suppose treatment A is randomised to our study sample using a coin toss: heads - treatment (A = 1) and tails - no treatment (A = 0).
- If the coin is not fair, we end up with more people in one group than the other.
- Research assistants administer treatment to the groups. After 5 days, we compute risk of mortality in the groups.
- $P[Y = 1|A = 0] = 0.3$ and
- $P[Y = 1|A = 1] = 0.6$
- Associational effect
  - risk ratio, $RR = p_1/p_0 = 0.6/0.3 = 2$ and
  - risk difference, $RD = p_1 - p_0 = 0.6 - 0.3 = 0.3$
- `What would have happened if the research assistants had misheard the instructions and reversed the treatments?` The risk under potential treatment value a among the treated equals the risk under potential treatment value a among the untreated. $P[Y^a = 1|A = 1] = P[Y^a = 1|A = 0] = P[Y^a = 1] \Rightarrow Y^a \perp\!\!\!\perp A$

2.4 Counterfactuals under exchangeability

In an ideal randomized experiment:

- No loss to follow-up.
- Full adherence to the assigned treatment over the duration of the study.
- Single variation of treatment (STUVA).
- Double blinded assignment

- We can compute counterfactual risk under treatment in the population $P[Y^{a=1} = 1] = P[Y = 1/A = 1]$
- It is equal to the risk in the treated $P[Y = 1|A = 1] = 0.6$

- Similarly, $P[Y^{a=0} = 1] = P[Y = 1|A = 0] = 0.3$

- In an ideal randomised experiment, association is causation
    - Association $\Rightarrow$ causation

2.5 Full Exchangeability and Independence

- Randomisation makes Y^a jointly independent of A

$$Y^a \perp\!\!\!\perp A$$

- Also written as
  - Proportion exchangeability: $Pr[Y^a = 1|A = 1] = Pr[Y^a = 1|A = 0]$ OR
  - Mean exchangeability: $E[Y^a = 1|A = 1] = E[Y^a = 1|A = 0]$

However,

$$Y^a \perp\!\!\!\perp A \nRightarrow Y \perp\!\!\!\perp A : Counterfactuals \nRightarrow Outcomes$$

Randomised study on the effectiveness of a new drug to prevent HIV infections in infants

- We wish to study the effectiveness of a new HIV drug for preventing transmission of HIV from infected mothers to their infants. We have two possible randomized designs we can use:

1. We can randomly select 60% of the study population and give them the new drug. The rest we keep them on the standard of care.
2. We can consider classifying pregnant women into two groups: `high viremia (L=1)` `and low viremia (L=0)`.
    - Randomly select 70% of women with high viremia and assign them the new drug and the rest of the women with high viremia to standard
    - Randomly select 40% of women with low viremia and assign them the new drug and the rest of the women to standard

- Design 2 is a conditionally randomised experiment (why?) - stratification/blocking - randomisation is restricted within each stratum.
- Design 1 is a marginally randomised experiment

2.6 Extending Exchangeability

- Randomisation in design 1 is expected to induce exchangeability (counterfactuals independent of treatment)
- Conditional randomisation (design 2) will not generally result in exchangeability of the treated and untreated (different prognosis for each treatment group).

- Design 2 is simply a combination of two marginally randomised experiments.

- In each subset $(L = l)$, treated and untreated are exchangeable
  $Pr[Y^a = 1|A = 1; L = l] = Pr[Y^a|A = 0; L = l]$

  $$\Rightarrow Y^a \perp\!\!\!\perp A/L$$

  Conditional exchangeability

Session 2 - Practical exercise (RCT)

A new anti retroviral drug has just been developed. We are interested in the effectiveness of the new drug in reducing HIV transmission to infant at birth compared to the current standard of treatment. A randomised controlled trial (RCT) is initiated where HIV positive pregnant women are randomised to new drug with probability 0.4. Baseline viral load is also assessed at baseline. The infant's HIV infection status is assessed at birth using DNA PCR. Using R, simulate the study data and estimate the effect of the new drug using the following assumptions:

- Let W be the baseline log viral load from a normal distribution with mean=3 and standard deviation=1.
- A is the treatment indicator variable with 0=standard of care and 1=new drug
- Y is the infant's HIV infection status with 0=not infected and 1=infected.
- Assume

$$logit(E[Ya = 1|W]) = \beta_0 + \beta_1 a + \beta_2 w$$

where

$$\beta_0 = -2, \beta_1 = -1, \beta_2 = 0.1$$

Assume a sample size n = 1000.

Questions

1. Generate the pair of potential outcomes given the information above
2. Generate the observed data from the study
3. Check that randomisation of treatment was achieved
4. Demonstrate that the data satisfies exchangeability condition
5. Estimate the causal relative risk and risk difference
6. Repeat the experiment with n = 1,000,000 – what is the effect of increasing sample size?

2.7 Computing causal effects under conditional randomisation

Under marginal exchangeablility, we know how to compute causal risk ratio, causal risk difference etc.

How do we compute causal risk ratio or causal risk difference in a conditionally randomized experiment?

1. Compute average causal risk ratio for each L = l using the associational risk ratio
    - Stratification
    - If CRR in strata L = 1 differs from CRR in strata L = 0, effect modification
2. Compute average causal effect in the entire population

2.8 Standardisation or G-formula

- The marginal counterfactual risk $Pr[Y^a = 1]$ is the weighted average of the stratum specific risks $Pr[Y^a = 1|L = l]$:

$$Pr[Y^a = 1] = \sum_{l=1}^{L} Pr[Y^a = 1|L = l]Pr[L = l]$$

- The weights are the proportion of each stratum in the population $Pr[L = l] = N_l/N$
- By conditional exchangeability,

$$Pr[Y^a = 1|L = l] = Pr[Y^a = 1|A = a, L = l]$$

2.9 Inverse probability weighting

Heart Transplant Study

i	*L*	*A*	*Y*
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- $Pr[Y^{a=0} = 1]$ is the counterfactual risk of death had everybody in the population remain untreated.

1. In strata $L = 0$, $\frac{1}{4}$ untreated died
2. If all 8 in strata $L = 0$ had been untreated $\frac{2}{8}$ would have died (conditional exchangeability)
3. In strata $L = 1$, $\frac{2}{3}$ untreated died
4. If all 12 in strata $L = 1$ had been untreated $\frac{8}{12}$ would have died

   5 Therefore $\frac{10}{20}$ would have died had all 20 not been treated $\Rightarrow Pr[Y^{a=0} = 1] = 0.5$

- Similarly for $Pr[Y^{a=1} = 1]$

Pseudo-population

How does this work?

- We have created a hypothetical population (pseudo-population) in which every individual appears treated and untreated
- The 4 untreated in $L = 0$ are used to create a pseudo-population of untreated by weighting each individual by 2 which comes from $1/0.5$.
- $0.5 = Pr[A = 0|L = 0]$
- The pseudo-population is created by weighting each individual by the inverse of the conditional probability of receiving treatment level $A = a$ that she indeed received.
- Under $Y^a \perp\!\!\!\perp A|L$ in the original population, treated and untreated are unconditionally exchangeable in the pseudo-population because L is independent of A.

Standardisation and inverse probability weighting (IPW)

- Standardisation and IPW are equivalent
- IPW uses conditional probability of treatment A given covariate L
- Standardisation use probability of covariate L and conditional probability of outcome Y given A and L
- Adjusting for L

3. Observational studies

Outline

- Randomised assignment (Recap)
- Instrumental variables
- Regression discontinuity design
- Difference-in-differences

Introduction

- Randomisation forms gold standard for causal inference, because it balances observed and unobserved confounders.

- A randomised experiment is an experiment with the following properties:

1. Positivity: assignment is probabilistic: $0 < p_i < 1$
   - No deterministic assignment.
   - Making treatment or intervention groups as similar as possible within subgroups.

1. Unconfoundedness: $P[A_i = 1|Y(1), Y(0)] = P[A_i = 1]$
   - Treatment assignment does not depend on any potential outcomes (baseline characteristics and outcomes).
   - Sometimes written as $A_i \perp\!\!\!\perp (Y(1), Y(0))$

Why do Experiments Help?

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

$$= \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{ATT} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{Selection\ bias}$$

- In an experiment we know that treatment is randomly assigned. Thus we can do the following:

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

- When all goes well, an experiment eliminates selection bias.

Why we need observational studies to evaluate interventions

- Randomized controlled studies are considered the `gold standard` for causal effects estimation but are at times:

    1. Unnecessary
    2. Inappropriate - HIV as a causal agent for AIDS; smoking as a cause of lung cancer
    3. Impossible - infrequent outcomes and rare events
    4. Inadequate

What are the main challenges of causal inferencein observational study?

1. `Lack of control`
    - lack of balance - confounding bias
    - lack of comparability - selection bias
2. `Unmeasured confounders`
3. `Time-varying confounders`

Estimating causal effects from observational data

- Unlike randomised experiments, in observational studies researchers cannot assign study subjects into treatment or control groups using a random mechanism,
    - makes it very difficult to draw a causal relationship between the treatment and the observed outcomes.
- Therefore, to control confounding and arrive at a causal estimate, broadly there are two approaches
    - **Use statistical adjustment:** (rely on the assumption, no remaining unmeasured confounding)
    - **Use design-based methods:** (to address unmeasured confounding)

- In observational studies treatment is not independent of potential outcomes
- Individuals receiving treatment A may not be comparable to those receiving treatment B
  - sicker, older, poorer, less adherent
  - differences in outcomes may be a reflection of these differences
- Difference of observed average response may be a biased estimate of the causal effect

Solution:

1. Identify all confounders W
2. Potential responses $(Y_0; Y_1)$ are independent of treatment exposure among subject with the same W values
3. Estimate differences within strata

What is confounding?

- Confounding is the bias caused by common causes of the treatment and outcome.
  - arise when exposure and outcome share an uncontrolled common cause.
- Observed confounders refer to confounders for which measures are available in the study data.
- Residual confounding is any confounding bias that remains after conditioning on observed confounders, either due to
  - variables not observed in the data (unmeasured or unobserved confounding),
  - inadequate measurement or
  - modelling of observed confounders.
- In observational studies, the goal is to avoid confounding inherent in the data.
- No unmeasured confounding assumes that we've measured all sources of confounding.

The most common methods to reduce the impact of confounding are:

- **Restricting** the study sample to one level of the confounding variable,
- **Stratifying** (analysing each gender separately) or
- **Matching** (selecting the sample so that the exposed and unexposed groups have the same gender balance).

Other methods for confounder adjustment include:

- **multivariable regression** (including confounders as covariates) and
- **inverse probability** (or propensity score) weighting.
- **G-methods:**
  - `G-computation` uses a statistical model (eg, a regression model)
    - relies on the statistical model being correctly specified.
  - `Marginal structural models` (commonly using IPWT).
  - `G-estimation` - predicts the counterfactual outcome at each time point.

Causal effects estimation in observational studies

- Analyze the data as if treatment was randomized, conditional on measured covariates.

Condition 1: **exchangeability**

- In marginally randomized experiments, $Y^a | A$
- In conditionally randomized experiments, $Y^a \perp\!\!\!\perp A | L$
- In observational studies
  - Reasons for receiving treatment are likely associated with some outcome predictors
  - Distribution of outcome predictors vary between treated and untreated groups.
  - Conditional exchangeability will not hold if there exits unmeasured independent predictors U of outcome such that probability of receiving treatment A depends on U with strata L
  - Exchangeability is not verifiable in observational studies

Condition 2: positivity

- There is probability greater than zero of being assigned to each of the treatment levels $Pr[A = a] > 0$ or $Pr[A = a|L = l] > 0, \forall l : Pr[L = l] \neq 0$
- In observational studies, positivity is not guaranteed, however, it can be sometimes empirically verified

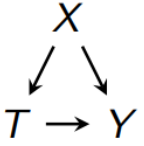
Condition 3: Consistency

- A defined standardized treatment exists with no variation (no multiple versions of the same treatment)
- In observational data, we have no control over the versions of treatments - use restriction

Causality with Unmeasured Confounding

Unmeasured Confounding

Consider cases of measured confounding



- In this case we block the backdoor path $T \leftarrow X \rightarrow Y$ by conditioning on X.
- What happens in the general case where X is unobserved?
- The above methods rely on an assumption of no unmeasured confounding
  - (ie, `conditional exchangeability`), **which is often not plausible in observational study designs**.

- If there exist **unmeasured confounders** that may be a common cause of both the outcome and the treatment,
    - impossible to accurately estimate the causal effect using above methods.
    - We will use Design methods.
    - These designs are often called quasi-experimental designs or natural experiments:
        - `difference-in-difference (DiD)`,
        - `regression discontinuity (RD)`, and
        - `instrumental variables (IV)`

Problem

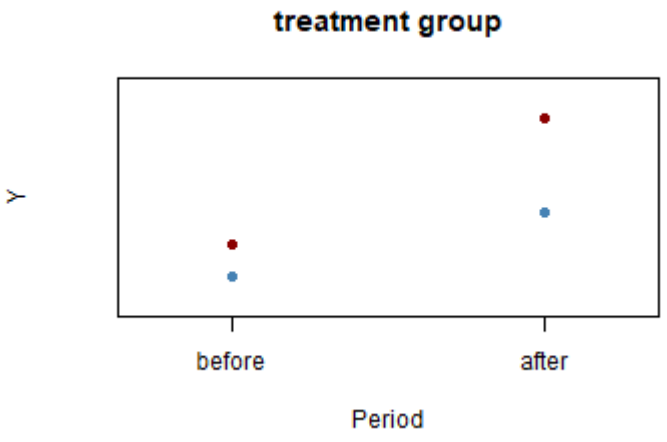
- Subject to certain unprovable assumptions,
- By exploiting some assignment mechanism

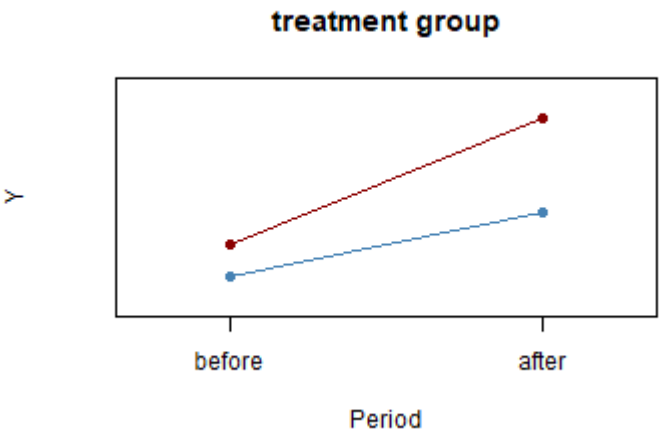
Difference-in-Differences (DiD)

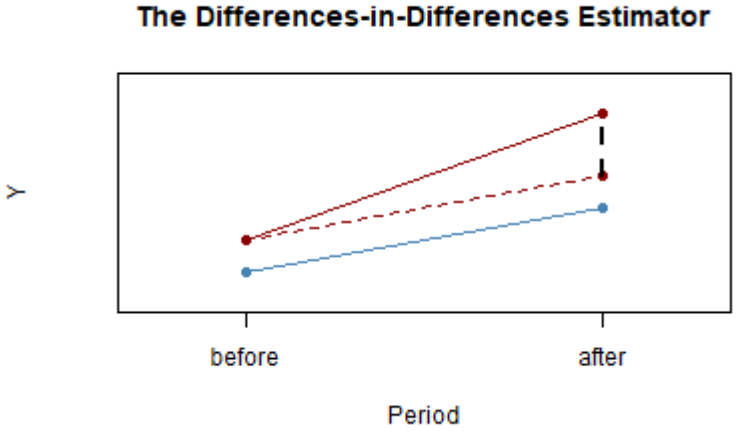
- The DiD design is a quasi-experimental alternative to the well understood and straight forward RCT design.
    - use data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.
- DiD methods exploit variation in time `(before vs. after)` and across groups `(treated vs. untreated)` to recover causal effects of interest.
- Pre vs. Post comparisons
    - `Compares:` same individuals before and after program.
    - **Limitation:** Does not account for potential trends in outcomes.
- Treated vs. Untreated comparisons
    - `Compares:` participants to those who have not experienced treatment (at least not yet).
    - **Limitation:** Selection - is participation driven by other factors?

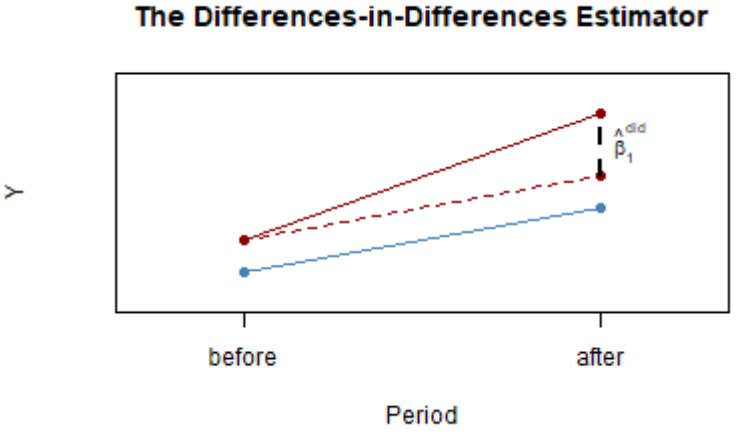
- DiD combines these two approaches to **avoid their pitfalls**.
- The DiD approach includes a before-after comparison for a treatment and control group.
- This is a combination of:
  - a cross-sectional comparison (= compare a sample that was treated to an non-treated control group)
  - a before-after comparison (= compare treatment group with itself, before and after the treatment)
- The before-after difference in the treatment group gets a correction, by accounting for the before-after difference in the control group, eliminating the trend problem.

DiD- Graphically









Derivation

- We assume that the outcome is determined by $Y_{it} = c_i + d_t + \delta D_{it} + \eta_{it}$, where i indexes the unit of observation and t indexes time.
- c_i is a variable(s) do not change by time but do change by unit.
- d_t is a variable(s) do not change by unit but can change by time.
- η_{it} is an unexplained, random error.

Differencing

1) Treated group after and before:

$$E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] = c_i + d_1 + \delta_1 - (c_i + d_0 + \delta_0)$$

2) Control group after and before:

$$E[Y_{i1}|D_i = 0] - E[Y_{i0}|D_i = 0] = c_i + d_1 - (c_i + d_0) = d_1 - d_0$$

- The difference of the differences (1)-(2) is $ATE = d_1 - d_0 + \delta - (d_1 - d_0) = \delta$.
- we can extend our notation to condition for a vector of covariates X_{it},
- So now we have:

$$Y_{it} = c_i + d_t + \delta D_{it} + X_{0it}\beta + \eta_{it}$$

.

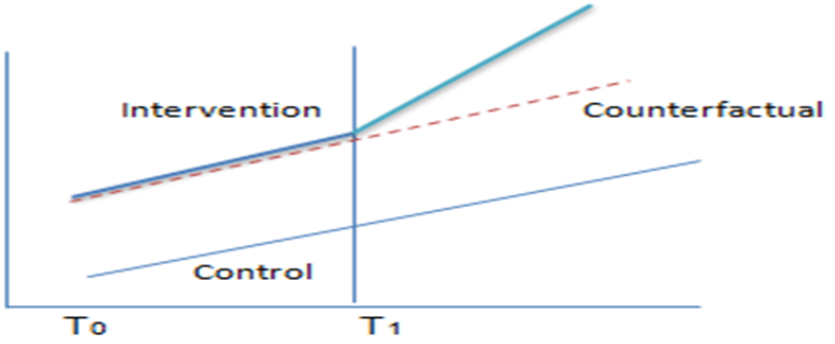
Estimation

- We often do a before and after comparison, even when we have more time points.
  - So we only need four means to estimate a DiD design.
- A before and after comparison of outcome Y for the treated is: $E[Y_{tpost}] - E[Y_{tpre}]$.
- We want to compare that difference with the difference in the control: $E[Y_{cpost}] - E[Y_{cpre}]$.
- The estimate of interest is: $DiD = E[Y_{tpost}] - E[Y_{tpre}] - E[Y_{cpost}] - E[Y_{cpre}]$.
- No regression model here yet, but we could estimate those four means
  - parametrically or nonparametrically or semiparametrically.
- The difference above is the same as: $DiD = E[Y_{tpost}] - E[Y_{cpost}] - E[Y_{tpre}] - E[Y_{cpre}]$

Assumptions of DiD

1. Parallel Trend Assumption

- Without treatment, the average change/trend in the outcome variable would be the same in the two groups (Mora 2015).



- To obtain an unbiased estimate of the treatment effect one needs to make **a parallel trend assumption**.
  - i.e. Treatment group would have had the same changes as the control group in absence of the treatment (counterfactual outcome).

2. Stable unit treatment value assumption (SUTVA)

- Subject's potential outcome depends only on its own treatment status, not by treatment status by the other unit.

3. No Spill-Over Effects assumption

- The members of the comparison group should not be affected by the intervention.

4. Covariate balance test

On average, both `observable` and `unobservable` characteristics should not vary between both groups.

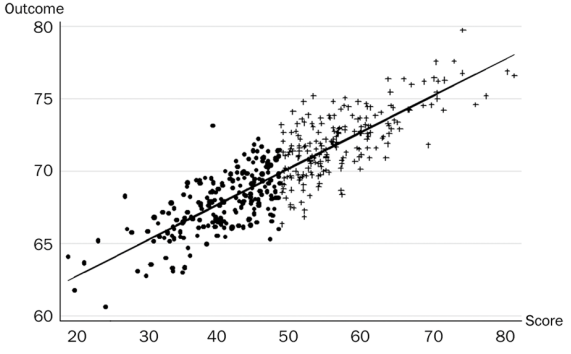
DiD Practicalities in : 2 Period, 2 Group Design

Regression Discontinuity Design

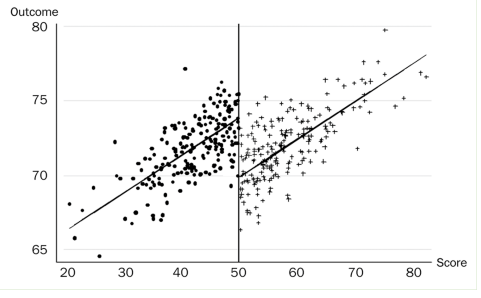
- Regression discontinuity design (RDD) is a quasi-experimental method used to estimate the causal effect of an intervention by examining the impact of a threshold on an outcome variable.
- Many programs determine eligibility through the use of continuous indices or scores:
- Anti-poverty programs $\Rightarrow$ Targets households under a specific income level or poverty line
- Education $\Rightarrow$ Scholarship for the best students based on a standardized test
- Agriculture $\Rightarrow$ Fertilizer program targeted to small farms less than given number of hectares)
  - Farmers with ≤ 50 hectares are eligible
  - Farmers with > 50 hectares are ineligible

Regression Discontinuity Design

- **At Baseline**



- **At Post Intervention**



- We need a continuous eligibility index with a defined eligibility cutoff point
- The basic idea is to compare the outcomes of individuals who are just above and below a threshold that determines whether they receive an intervention or not.
- RDD is useful when it is impossible or unethical to randomly assign individuals to treatment and control groups.

- RDD can be used to establish causal inference by controlling for confounding variables that might otherwise distort the causal relationship.
- In particular, RDD can control for selection bias that can occur when individuals self-select or are selected for an intervention based on their characteristics.
- The key steps in RDD are as follows:

1. **Identify the threshold:** Determine the threshold value of a continuous variable that separates individuals into treatment and control groups.
2. **Determine the outcome variable:** Specify the outcome variable that reflects the impact of the intervention.

- Then estimate the causal effect using a regression model by comparing the outcomes of individuals just above and below the threshold.
- Test the robustness of the results `by varying the bandwidth of the threshold`, controlling for additional covariates, and checking for violations of assumptions.

Practice in R

Here is an example of how to implement these steps using the RDD package in R:

```
# Load the data into R
data <- read.csv("data.csv")# Explore the data
plot(data$running_var, data$outcome)# Estimate the treatment effect using a linear re
model_linear <- rdd::rdd_data(x = data$running_var, y = data$outcome, cutpoint = 0, s
summary(model_linear)# Estimate the treatment effect using a nonparametric model
model_np <- rdd::rdd_data(x = data$running_var, y = data$outcome, cutpoint = 0, slope
summary(model_np)# Plot the results
plot(model_linear)# Conduct sensitivity   analyses
rdd::rdd_bw(model_linear)
rdd::rdd_rdplot(model_linear)
```

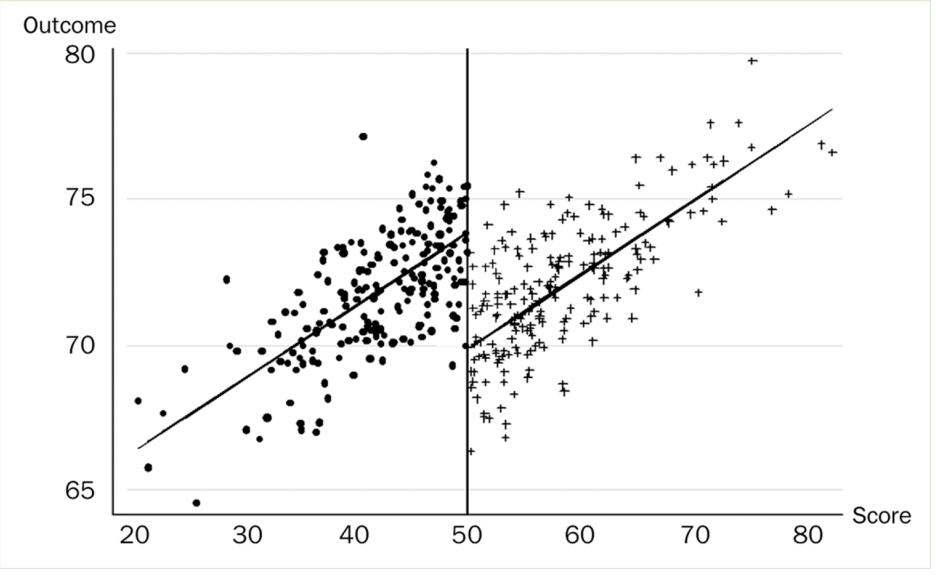
Instrumental Variables

- An IV is a variable that causes some variation in the exposure and is unrelated to the outcome except through the exposure.
- IVs are often used to identify the causal effect of a particular exposure or treatment on an outcome of interest.
- If we have an instrument, we can deal with unmeasured confounding in the treatment-outcome relationship.
- It is going to turn out that the same construction will let us deal with non-compliance in experiments.

Assumptions for using IV

1. The **IV is correlated with the exposure** or treatment of interest:
    - The IV should be able to predict the level of the exposure or treatment.
2. The IV is **not directly related to the outcome**:
    - The IV should only be associated with the outcome through its effect on the exposure or treatment.
3. The IV is **independent of the error term**:
    - There should be *no correlation between the IV and the error term* in the outcome equation.
    - If these assumptions are met, then using IVs can lead to more accurate estimates of the causal effect of interest.

Assumption



- In Figure, Z is an IV to the effect of X on outcome.
- IV relies on three main conditions.
- A valid IV, Z, must (i) predict treatment status (**relevance**), (ii) only affect outcome through X (**exclusion**), and (iii) be as good as randomly assigned (**independence**).
- These conditions are met as there's a causal path $Z \rightarrow X$ and no open paths between Z and outcome except through X.

Example

4. Method for estimating causal effects in observational data

Outline

- Introduction
- Inverse Probability Weighting
- Stabilized Weights
- Standardisation/G-Computation

Objectives

- Establish conditions under which observational studies can be used to estimate causal effects
- Highlight some commonly used methods in causal inference with observational data

Introduction

- Observational data are often the basis for epidemiological and other investigations seeking to make inference on the effect of treatment exposure on a response.
- Randomized studies aim to balance distributions of subject characteristics across groups, so that groups are similar except for the treatments.

With observational data:

- Treatment exposure may be associated with covariates that are also associated with potential response
- Groups may be seriously imbalanced Unbiased treatment comparisons from observational data require methods that adjust for such confounding
- Inferences with a causal interpretation cannot be made without appropriate adjustment.

Estimating Causal Effects from observational data

- In randomized controlled studies, average causal effect = difference in outcomes between the two groups
- The mean outcome among treated equals mean treated counterfactual because treatment is independent of potential outcome In observational studies treatment is not independent of potential outcomes
- In observational studies, individuals receiving treatment A may not be comparable to those receiving treatment B
    - sicker, older, poorer, less adherent
    - differences in outcomes may be a reflection of these differences

Estimating causal effects from observational data

- Treatment may not be independent of counterfactuals - same characteristics that led to treatment exposure may also be associated with potential response.
- Difference of observed average response may be a biased estimate of the causal effect.
  Solution:
  - Identify all confounders W
  - Potential responses (Y^0, Y^1) are independent of treatment exposure among subject with the same W values
  - Estimate differences within strata

Designing observational studies

`How would the study be conducted if it were possible to do it by controlled experimentation?`

- Define study population
- Eligibility and exclusion criteria - restriction
- Define Exposure - what are the treatments/intervention groups
- Pre-exposure covariates
- Define outcome

Challenge

- Achieving balance between intervention groups

Example: observational study

- In an observational study of heart transplant and mortality, those who receive a transplant have a severe heart condition
- Those who receive the transplant would be expected to have a greater risk of mortality had they not received the transplant compared to patients who did not received a transplant.
- This is a violation of $Pr[Y^a = 1 = 1 | A = 1] = Pr[Y^a = 1 = 1 | A = 0]$ (exchangeability).
- An associational effect therefore does not estimate a causal effect.

Addressing selection and confounding challenges

`Use novel methods to mimic randomization`

- Stratification
- Matching Propensity scoring
  - Matching or stratifying (Rosenbaum and Rubin, JASA 1984)
  - Inverse Probability Weighting (IPW)
- Instrumental Variables (IV)
- Sensitivity Analysis (SA)

Causal effects estimation in observational studies

Analyze the data as if treatment was randomized, conditional on measured covariates.

Conditions necessary:

- Values of treatment under comparison correspond to well define interventions corresponding to versions of treatment in the data
- Conditional probability of receiving every value of the treatment depends only on the measured covariates
- Conditional probability of receiving every value of treatment is greater than zero
- With these conditions, an observational study can emulate a conditionally randomized experiment

Causal effect of smoking cessation on weight gain

- To estimate the effect of smoking cessation on weight gain we will use real data from the NHEFS
- NHEFS stands for National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study
- Can be found at wwwn.cdc.gov/nchs/nhanes/nhefs/
- A subset of the NHEFS data will be used with 9 measured covariates (L) to estimate the ACE (`1566 individuals`)

Causal effect of smoking cessation on weight gain

	seqn	qsmk	age	sex	race	university	wt71	smokeintensity	smokeyrs	exercise	active
1	233	0	42	0	1	0	79.04	30	29	2	0
2	235	0	36	0	0	0	58.63	20	24	0	0
3	244	0	56	1	1	0	56.81	20	26	2	0
4	245	0	68	0	1	0	59.42	3	53	2	1
5	252	0	40	0	0	0	87.09	20	19	1	1
6	257	0	43	1	1	0	99.00	10	21	1	1

- A = smoking cessation (qsmk: 1: quit smoking, 0: still smoking)
- L = (*age*, *sex*, *race*, *university*, *wt71*, *smokeintensity*, *smokeyrs*, *exercise*, *active*)
- Y = weight gain (wt8287)

Mean of Potential Outcomes/Counterfactuals

- The mean weight gain that would have been observed if all individuals in the population had quit smoking before the follow-up visit, $E[Y^{a=1}]$
- The mean weight gain that would have been observed if all individuals in the population had not quit smoking, E$E[Y^{a=0}]$.
- The average causal effect on the additive scale: $E[Y^{a=1}] - E[Y^{a=0}]$.
- The difference in mean weight that would have been observed if everybody had been treated compared with untreated.

Associational Effect

- Mean weight gain among quitters, $E[Y|A=1] = 4.5$.
- Mean weight gain among non-quitters, $E[Y|A=0] = 2.0$.
- Associational Effect $E[Y|A=1] - E[Y|A=0] = 2.5$ with 95% CI 1.7 to 3.4

Use R to estimate the associational effect and the associated 95% confidence interval

Average causal effect

- $E[Y^{a=1}] - E[Y^{a=0}]$
- If quitters and non-quitters are different wrt characteristics that affect weight gain, then the associational effect will not be equal with the average causal effect.
- Check the distribution of covariates L between levels of A Check covariates association with Y.
- Age independently associated with both quitting and weight gain (regardless of quitting status) $\Rightarrow$ confounder of effect of A on Y.

Estimation

There are two formulations for $Pr[Y^a = 1]$

$$Pr[Y^a = 1] = \sum_w Pr[Y^a = 1|W = w]Pr[W = w]$$

$$Pr[Y^a = 1] = \sum_w Pr[Y^a = 1|W = w]\frac{Pr[A = a]}{Pr[A|W = w]}$$

Estimating IP weights using models ..... (1)

- IP weighting creates a pseudo-population in which the association between A and L is removed.
- The pseudo-population is created by weighting each individual by the inverse (reciprocal) of the conditional probability of receiving the treatment level that he/she indeed received.
- The associational effect in the pseudo-population would estimate the causal effect of A on Y.
- $W^A = 1/f(A|L), f(A|L) = Pr[A = 1|L]$ ..... then dividual-specific IP weights for treatment A.
- With a multidimensional L, we can no longer estimate $Pr[A = 1|L]$ non-parametrically

Estimating IP weights using models ..... (2)

- The denominator $f(A|L)$ of the IP weights is the probability of quitting conditional on the measured confounders, $Pr[A = 1|L]$, for the quitters, and $Pr[A = 0|L]$, for the non-quitters.
- For a dichotomous treatment A, we only need to estimate Pr [A = 1 | L] since $Pr[A = 0|L] = 1 - Pr[A = 1|L]$ Use logistic regression model for the probability of quitting smoking with all 9 confounders included as covariates.
- Use linear and quadratic terms for the (quasi-)continuous covariates No product terms
- The model restricts the possible values of $Pr [A = 1 | L] such that, on the logit scale, the conditional relation between the continuous covariates and the risk of quitting can be represented by a parabolic curve
- Each covariate's contribution to the risk is independent of that of the other covariates.

Estimating IP weights using models ... (3)

- Under these parametric restrictions, an estimate for $Pr[A = 1|L]$ will be obtained for each combination of L values, and therefore for each individuals (1566) in the study population
- All members of the study population are replaced by two copies of themselves.
- One copy receives treatment value $A = 1$ and the other copy receives treatment value $A = 0$
- The estimated IP weights WA ranges from 1.05 to 16.7, and the mean is 2

Average causal effect using IP weights

- To estimate $E[Y|A=1] - E[Y|A=0]$ in the pseudo population, the saturated linear mean model can be fitted $E[Y|A] = \beta_0 + \beta_1 A$.
- Weighted least squares, with individuals weighted by their estimated IP weights.
- The estimated β_1 is 3.4 that is, we estimated that quitting smoking increases weight by 3.4 on average.
- To obtain a 95% confidence interval around the point estimate, we need a method that takes IP weighting into account

95% confidence interval for ACE

- Use statistical theory to derive the corresponding variance estimator
- Approximate the variance by non-parametric bootstrapping
- Use robust variance estimator (standard option in most statistical software packages)
- The 95% confidence intervals based on the robust variance estimator are valid but conservative–they cover the super-population parameter more than 95% of the time
- If the model for $Pr[A = 1|L]$ is misspecified, the estimates of β_0 and β_1 will be biased and, the confidence intervals may cover the true values less than 95% of the time

Estimating stabilized IP weights using models ... (1)

- The IP weights $W^A = 1/f(A|L)$ adjust for confounding by L
- All individuals have the same probability of receiving A = 1 (prob. = 1) and A = 0 (prob. = 1)
- A and L are independent in the pseudo population
- There are other ways of creating pseudo population in which A and L are independent
- A pseudo population in which all individuals have a probability of receiving $A = 1$ equal to 0.5 rather than 1 and a probability of receiving $A = 0$ also equal to 0.5, regardless of their values of L Stabilized IP weights $(SW^A) = 0.5/f(A|L)$

Estimating stabilized IP weights using models ... (2)

- Pseudo-population would be of the same size as the study population.
- The expected mean of the weights $0.5/f(A|L)$ is 1.
- Deviations from 1 indicate model misspecification or possible violations, or near violations, of positivity.
- The ACE obtained in the pseudo-population is the same for both weights (W^A and SW^A).
- The same goes for any other IP weights p/f (A|L) with $0 \leq p \leq 1$.
- The IP weights $f(A)/f(A|L)$ range from 0.33 to 4.30, whereas the IP Weights $1/f(A|L)$ range from 1.05 to 16.70
- Narrower range of the $f(A)/f(A|L)$ weights because of the stabilizing factor f(A) in the numerator
- `Stabilized weights`, SW^A.
- `Non-stabilized weights`, WA.

Average causal effect using stabilized IP weights

- An estimate of the conditional probability $Pr[A = 1|L]$ to construct the denominator of the weights
- The logistic regression will be used
- Estimate $Pr[A = 1]$ for the numerator of the weights: Non-parametric estimate or fitting a saturated logistic regression model for $Pr[A = 1]$ with an intercept and no covariates
- Estimate the causal difference $E[Ya = 1] - E[Ya = 0]$ by fitting the mean model $E[Y|A] = \beta_0 + \beta_1 A$ with individuals weighted by their estimated stabilized IP weights
- The estimated ACE = 3.4 kg (95% CI: 2.4, 4.5) on average
- The same estimate obtained using non-stabilized weights W^A

Why we use stabilized IP weights?

- Stabilized weights result in narrower 95% confidence intervals than non-stabilized weights.
- In many settings (time-varying or continuous treatments), the weighted model cannot possibly be saturated and therefore stabilized weights are used.

Using logistic regression to estimate f (A|L)

```
#Pr[A=1|L]
# quadratic terms
age2=age*age
si2=smokeintensity*smokeintensity
sy2=smokeyrs*smokeyrs
wt2=wt71*wt71

qsmk.pr = glm(qsmk~sex+race+university+exercise+active+age+age
qsmk.hat = predict.glm(qsmk.pr,type=c("response"))
pr.qsmk.hat = ifelse(qsmk==1,qsmk.hat,1-qsmk.hat)

# Weights = inverse probability of treatment
w=1/pr.qsmk.hat

# Estimate E[Y|A=1]-E[Y|A=0] is pseudo-population
qsmk.lm.wt = lm(wt82_71~qsmk, weights=w)
```

Estimated Causal Effect of quitting smoking on weight gain

```
> # Estimate E[Y|A=1]-E[Y|A=0] is pseudo-population
> qsmk.lm.wt = lm(wt82_71~qsmk, weights=w)
> display(qsmk.lm.wt)
lm(formula = wt82_71 ~ qsmk, weights = w)
            coef.est coef.se
(Intercept) 1.78     0.29
qsmk        3.43     0.41
---
n = 1566, k = 2
residual sd = 11.41, R-Squared = 0.04
>
> # Associational effect
> display(lm1)
lm(formula = wt82_71 ~ qsmk)
            coef.est coef.se
(Intercept) 1.98     0.23
qsmk        2.54     0.45
---
n = 1566, k = 2
residual sd = 7.80, R-Squared = 0.02
```

Estimating standardized mean outcome

$$f(A)/f(A|L)$$

Allows for pseudo-population to be the same size as the original population

```
# Stabilized weights
> fa=ifelse(qsmk==1,sum(qsmk)/length(qsmk),1-(sum(qsmk)/length(qsmk)))
> sw=fa/pr.qsmk.hat
>
> qsmk.lm.swt = lm(wt82_71~qsmk, weights=sw)
> display(qsmk.lm.swt)
lm(formula = wt82_71 ~ qsmk, weights = sw)
            coef.est coef.se
(Intercept) 1.78     0.23
qsmk        3.43     0.45
---
n = 1566, k = 2
residual sd = 7.80, R-Squared = 0.04

> mean(sw)
[1] 0.9994
```

5 Marginal structural models (MSM)

- Introduce marginal structural models
- Motivation for G-estimation
- Out
- Propensity Score (PS) Methods
  - Introduction and estimation of propensity scores
  - PS Stratification
  - PS Standardisation
  - PS Matching (PSM)
  - Weighting by the inverse propensity
  - Practical: propensity score estimation (stratification, matching)

Readings:

- Chpt 14. MH & JR
- Chpt 15. MH & JR
- Joffe M, 2004
- Lunceford J, 2004
- D'Agostino RB, 1998
- Rosenbaum PR, 1998
- Rosenbaum PR, 1987

Objectives

- Introduce marginal structural models
- Motivate for G-estimation
- Review outcome regression
- Introduce propensity scores
- Estimation of propensity scores
- Applications of propensity score to control for counfounding

Introduction

Recap

- Interested in estimating the effect of a treatment or exposure
- Standard regression models $E[Y|A, W]$.
- We include confounders in the structural model
- But some of the W maybe mere `nuisance` variables of no interest

Potential outcome densities

- Let $f(Y|L)$ be the joint density of the potential outcomes $Y = Y^a$.
- Let $f(Y|L)$ be the density of potential outcome Y^a - marginal density.
- Joint density is not estimable but marginal density is estimable from observed data.
- Comparison of marginal densities $f(Y^a|L)$ and $f(Y^{a'}|L)$ for $(a \neq a')$ represent causal effects

Marginal structural models

- Consider the model: $E[Y^a] = \beta_0 + \beta_1 a$
- Keep variables of little interest out of the structural part of the model while still controlling for counfounding
- The model is known as a marginal structural mean model
- The parameter β_1 correspond to the average causal effect

Conditions for MSM estimation

1. A is discrete
2. $Y^a \perp\!\!\!\perp A \Longrightarrow$ conditional exchangeability
3. $Pr(A = a|W, Y^a = Pr(A = a|W) > 0 \Longrightarrow$ positivety

Estimation of Marginal Structural Models

i. Model association between W and A

ii. Derive weights from model

iii. Use weights to create pseudo-population in which W are not associated with A

iv. Estimate parameters in the MSM using weighted methods

Example: Estimating effect of change in smoking intensity on weight gain

- In this case, treatment takes many values
- Interest: difference in average change in weight under different changes in smoking intensity
- $E[Y^a] = \beta_0 + \beta_1 a + \beta_2 a^2$
- Use IP weights to estimate the parameters β
- Stabilized weights $f(A)/f(A/L)$
- But $f(A/L)$ is difficult to estimate for continuous A - you have to rely on strong assumptions on the pdf of A

Exercise 4: G-computation simulation: Snowden paper simulation

Propensity scores methods

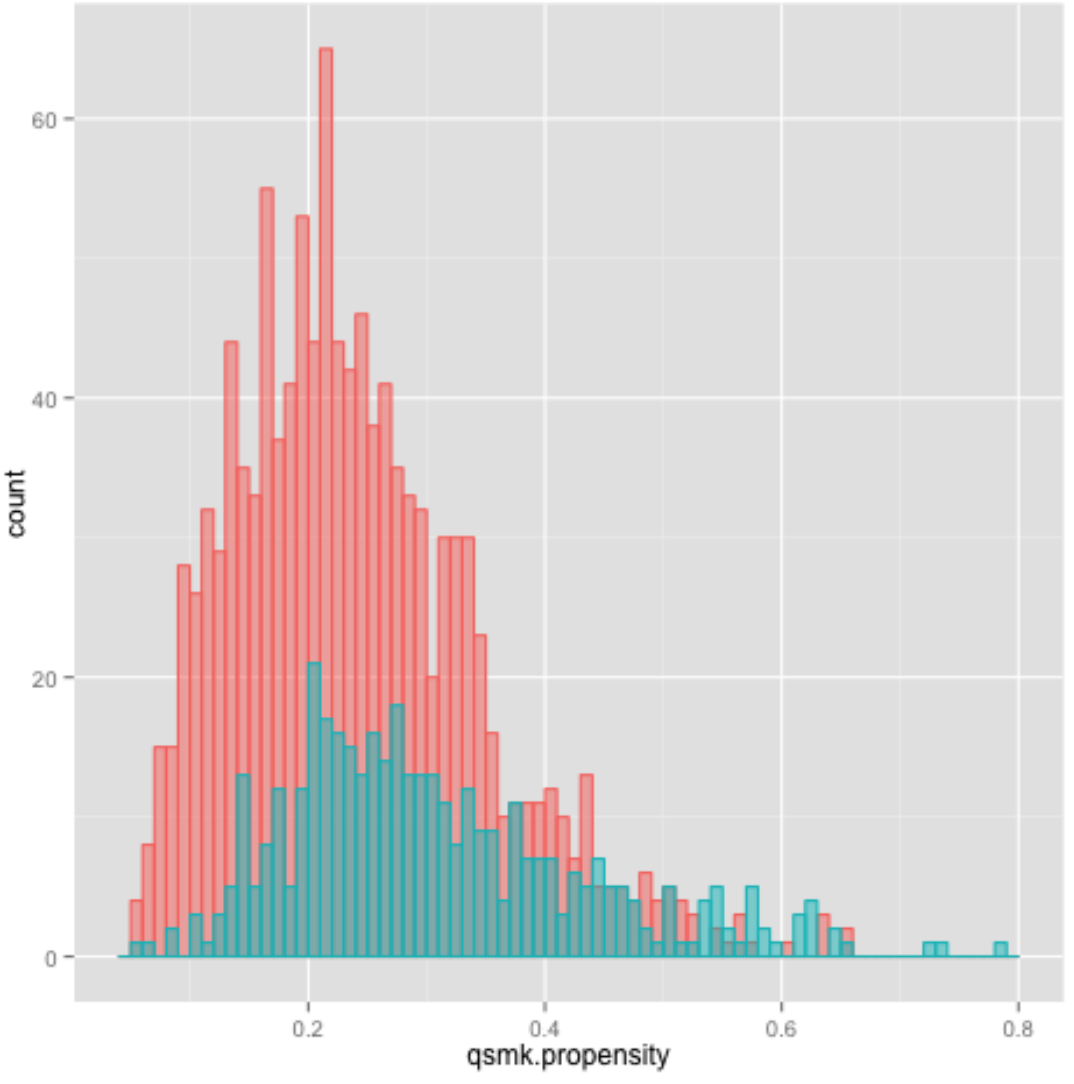
Outcome Regression

- Consider the structural model: $E[Y^a|L] = \beta_0 + \beta_1 a + \beta_2 aL + \beta_3 L$
- The causal parameter are β_1 and β_2.
- Our estimates of β_1 and β_2 are consistent if only β_0 and $\beta_3 L$ correctly models the dependence of $E[Y^a|L]$ on L.
- β_0 and β_3 are known as nuisance parameters
- If exchangeability, positivity and well define treatments holds, then the causal parameters can be estimated using the usual regression model on the observed.

Propensity scores: an introduction

- Let $p(L) = Pr[A = 1|L]$
- $0 \leq p(L) \leq 1$
- In randomized studies $p(L)$ is the same within strata of L
- In observational studies, $p(L)$ is unknown and has to be estimated Common estimate is the logistic model for binary treatments

Distribution of propensity scores between quitters and non-quitters



Example: propensity scores for quitters

- Quitters have a higher propensity to quit smoking
- Difference in the distribution of the propensity $\Rightarrow$ confounding by L
- Similar propensity scores between individuals does not imply equal L
- Within levels of $p(L)$, covariates (measured) are balanced between treated and untreated $A \perp\!\!\!\perp L|p(L)$

Key results:

1. $(Y^a) \perp\!\!\!\perp A|L) \Rightarrow (Y^a) \perp\!\!\!\perp A|p(L)$
2. Positivity with levels of p(L) holds $\iff$ positivity in levels of L
3. $p(L)$ can be used to estimate causal effects using:
    - stratification
    - standardization
    - matching

Estimating average causal effects with $p(L)$ stratification

- Under exchangeability and positivity,

$$E[Y^a = 1|p(L) = s] - E[Y^a = 0|p(L) = s] = E[Y|A = 1, p(L) = s] - E[Y|A = 0, p(L) =$$

- $p(L)$ is a continuous variable, therefore unlikely to have two individuals with same score
- Create strata of individuals with similar propensity scores
- Use deciles or quintiles and estimate strata specific effects
- Average across strata using standardization

Potential complications

- If distribution of $p(L)$ differs between treated and untreated in some strata, then exchangeability is violated

Regression with propensity scores

- Fit a regression with A, the 9 indicators and 9 product terms between A and decile indicators of the deciles:

$$E[Y|A, l(p(L)]$$

- Can also use $p(L)$ as a continuous variable in the model
- To guard against misspecification of the model use flexible models (e.g. cubic splines rather than linear models in $p(L)$)

Propensity score matching

- Create a matched population in which treated and untreated are exchangeable because they have same distribution of $p(L)$.
- Under exchangeability and positivity give $p(L)$, association measures in the matched population are consistent estimates of causal effect measures.
- Close values in $p(L)$ - caliper matching $(s \pm 0.005)$.

Software

- R: Matchit
- Stata: teffects psmatch

Potential problems with propensity matching

- Bias-variance tradeoff based on closeness
- Matching does not distinguish between random (assumed in regression modeling) and structural nonpositivity in the non-overlap region
- Matched population might be different from target population
- Hard to transport results to other populations when restriction is based on propensity score

Weighting using PS: Recap

- $p(L) = Pr[A = a|L]$
- Weights $W^A = 1/Pr[A|L]$ or stabilized weights
- Estimation in pseudo-population using weighted regression

Application to quit smoking data

- Estimate propensity score
- Check for balance of covariates between intervention arms
- Create strata (5 or 10)
- Use standardization to get marginal effect
- Use the regression approach
- Using matching
- Use inverse weighting

6. Doubly Robust (DR) estimators - combination methods

- The concept of DR estimators
- Augmented Inverse Probability of Treatment Weighted (AIPTW) Estimator
- Targeted Maximum Likelihood Estimation (TMLE)
- Practical: Introduce implementation of estimators in R

7. Marginal Mean Weighting through Stratification (MMWS)

Outline

- Objectives
- Introduction
- Marginal Mean Weighting through Stratification
- Example: Implementation in R

Objectives

1. Introduce marginal mean weighting and stratification estimators
2. Demonstrate with examples
3. Highlight strengths and weaknesses of MMWS

Readings:

1. Linden A, 2014
2. Huang, I, 2005
3. Hong G, 2010 & 2012

Recap propensity score methods

$$p(W) = Pr[A|W]$$

Comparability between treatment groups achieved through:

1. Matching (one-to-one, $1:k, nn$, mahalanobis distance, kernel density matching)
2. Stratification (deciles or quintiles (90% of bias removed)
3. Weighting (IPTW - standardize treatment groups to population for which treatment is intended, AIPTW)

MMWS

Combine propensity stratification and IPTW

i. Stratify sample into quantiles of propensity score

ii. Generate weight for each individual based on corresponding stratum and treatment assignment

iii. Estimate the strata-specific weighted means

Effect of maternal smoking intensity during pregnancy on birth weight

mkbwt.dta

variables: description & codes

- bweight: infant birthweight (grams)
- mmarried: 1 if mother married
- mhisp: 1 if mother hispanic
- fhisp: 1 if father hispanic
- foreign: 1 if mother born abroad
- alcohol: 1 if alcohol consumed during pregnancy
- deadkids: previous births where newborn died
- mage: mother's age
- medu: mother's education attainment
- fage: father's age
- fedu: father's education attainment
- nprenatal: number of prenatal care visits
- monthslb: months since last birth

- order: order of birth of the infant
- msmoke: cigarettes smoked during pregnancy
- mbsmoke: 1 if mother smoked
- mrace: 1 if mother is white
- frace: 1 if father is white
- prenatal: trimester of first prenatal care visit
- birthmonth: month of birth
- lbweight: 1 if low birthweight baby
- fbab: 1 if first baby
- prenatal1: 1 if first prenatal visit in 1 trimester

Estimate propensity score

- E[mbsmoke = 1|mmarried, mage, fbaby, medu]
- Use logistic regression
- Save predicted probabilities for each individual
- Generate weights of treatment received (WA)

Stratification

- Identify regions of common support
- Out of support individuals get a weight of zero
- Stratify propensity score into quintiles (5) or deciles (10) (equal size)

Marginal mean weights (MMW)

- Calculate MMW for each treatment group by stratum

$$W^s = \frac{n_s * Pr(A = a)}{n_{A=a,s}}$$

  Assign weights for each individual corresponding to their stratum and treatment assignment

$$W = W^A * W^s$$

8. Causal inference from survival and longitudinal data