DUKE
COMPUTER SCIENCE

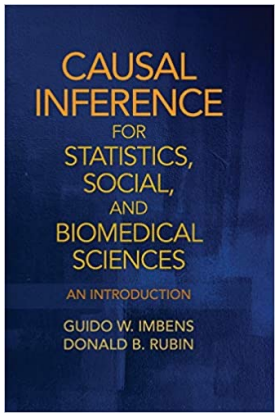
# CompSci 590.01

# Causal Inference in Data Analysis with Applications to Fairness and Explanations

## Lecture 5:
## Potential Outcome Framework,
## Statistical Causal Inference,
## And
## Matching Methods

Sudeepa Roy

# Reading




First three chapters are relevant

- Survey by Stuart (2010) – Matching methods for causal inference: A review and a look forward
- Survey by Sekhon (2007): The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods
- Rubin (2005)
- Rosenbaum-Rubin (1983)
- Several online articles

Acknowledgement (big thanks!):

Many slides from a joint talk with Profs. Cynthia Rudin and Alexander Volfivsky that have been modified here.

# Announcements

- Please add your first and second choice of topics for presentation – we need to have a balance of causal inference-fairness-explanations, and cover important topics

- See your fellow classmates' interests in the google doc and start discussing with them about project and paper presentation
  - Added "presentation topic" options to help you choose a topic
  - Send a private message on Ed or an email to Sudeepa if you would like to discuss

- Please send Sudeepa your slides for presentation 24-48 hours before you present to get feedback – we will try to meet over zoom for 15 mins to discuss before you present (say 5 pm the day before your presentation or another time)

# Announcements

Timeline:

- Start talking to your fellow students – in person or on Ed
- Presentation topic (1st and 2nd choice) & date due: Thursday 2/2
  - On spreadsheet
  - Then Sudeepa checks all and makes a balance of topics covering important papers
- Initial project ideas & teammates' names due: Tuesday 2/7
  - Please share on Google doc or Overleaf (latex)
- Project proposal due: Tuesday 2/14

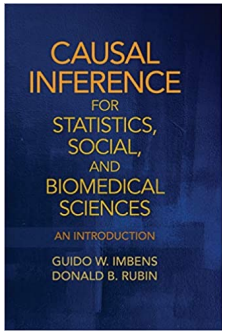
# Homework 1 (2/3rd of HW grade5): Paper Reviews

- Submit a ½ page to 1 page short review of papers that are being presented by others
- For each presentation day, you can review the main paper and skip the other papers, if both are main papers, you can choose either of them – Sudeepa will announce these main paper(s)
- Include
  - **Summary (30 points)** – what the paper is studying and main contributions
  - **One result (35 points):** Talk about one of the main results from the paper that you like and the main techniques used
  - **Scope for improvement (35 points)**: Talk about some limitation / scope for improvement / future work that you have noticed for this paper or topic
  - **Collaboration:** If you have discussed with any other student mention their names – copying or same submission is not allowed
- You should try to do it while you read the paper "before the presentation" taking notes – reading the entire paper might not be needed, and we will try to identify relevant sections
- Reports are due (on sakai) within 2 days after the presentation
- We will discuss the scope for improvements you have noted in the class after you submit
- Individual submissions needed with your own writing, but feel free to discuss with others and add their names in your submission
- You can skip reviews for 2 presentation days in the entire semester of your choice (so will review about 10 papers)

# Potential Outcome Framework

- Referred to as Neyman-Rubin's model or Rubin's model
  - First proposed in Neyman's Ph.D. thesis (1923)
  - A model for "Randomized Experiments" by Fisher (1920s-30s)
  - Further developed by Rubin (1978) and others

- Establish a causal relationship between a potential cause (treatment) and its effect (outcome)



# Potential Outcome Framework

Widely used in

- Medicine
  - Christakis and Iwashyna 2003; Rubin 1997
- Economics
  - Abadie and Imbens 2006; Galiani, Gertler, and Schargrodsky 2005; Dehejia and Wahba 2002, 1999
- Political science
  - Bowers and Hansen 2005; Imai 2005; Sekhon 2004b
- Sociology
  - Morgan and Harding 2006; Diprete and Engelhardt 2004; Winship and Morgan 1999; Smith 1997
- Law
  - Rubin 2001

References in [Sekhon 2007]

# Units

- N “units”
  - physical objects at particular points in time
  - e.g., individual people, one person at different points of time, plots of lands

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Treatment and Control

- Each unit i can be exposed or not to a treatment $T_i$
  - e.g. individuals taking an Aspirin vs. placebo ,
- "Active Treatment" or "Treatment" ($T_i = 1$)
  - if exposed
- "Control Treatment" or "Control" ($T_i = 0$)
  - if not exposed

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Covariates

- Variables that take their values before the treatment assignment
- Cannot be affected by the treatment
  - e.g., pre-aspirin headache pain, gender, blood-pressure

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Potential Outcome

- $Y_1$ (for treatment, $T_i = 1$)
- $Y_0$ (for control, $T_i = 0$)
- for i-th unit : $Y_{1i}$ and $Y_{0i}$
- Observed outcome $Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Unit-level causal effect

- The comparisons of $Y_{1i}$ and $Y_{0i}$
  - difference or ratio
  - Typically $Y_{1i} - Y_{0i}$
- For any unit i, only one of them can be observed
  - we cannot go back in time and expose it to the other treatment
- Fundamental problem of causal inference

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Summary of causal effect

- Defined for a collection of units
- e.g.
  - the mean (or expected) unit-level causal effect -- standard
  - the median unit-level causal effect for all males
  - the difference between the median $Y_{1i}$ and $Y_{0i}$ for all females

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Remark..

- To be a causal effect, the comparisons of $Y_1$ and $Y_0$ should be for a common set of units
  - e.g., females
  - we cannot apply control to males and treatment to females

| Units | Covariates | Treatment assignment | Potential Outcome: Treatment | Potential Outcome: Control | Unit-level causal effects | Summary of causal effects |
|---|---|---|---|---|---|---|
| 1 | $X_1$ | $T_1$ | $Y_{11}$ | $Y_{01}$ | $Y_{11} - Y_{01}$ | $E[Y_1 - Y_0]$ |
| 2 | $X_2$ | $T_2$ | $Y_{12}$ | $Y_{02}$ | $Y_{12} - Y_{02}$ | |
| … | | | | | | |
| N | $X_n$ | $T_N$ | $Y_{1N}$ | $Y_{0N}$ | $Y_{1N} - Y_{0N}$ | |

# Randomized Controlled Experiments

[Rubin, ‘74 ‘05]

Treatment
(vaccine)

Population

Control
(placebo)

At random

$T \perp Y(1), Y(0)$

(Only one of Y(1), Y(0) is observed)

Average Treatment Effect (ATE) = $\mathbf{E}[Y(1) - Y(0)]$

$= E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0]$

Can be estimated from experimental observed data!

# Average Treatment Effect (ATE)

- ATE = $E[Y_1 - Y_0]$
- Recall observed outcome $Y = T\,Y_1 + (1\text{-}T)\,Y_0$
- Suppose Treatment Assignment (T) is independent of $Y_1$ , $Y_0$
- Then

$E[Y_1 - Y_0]$

$= E[Y_1] - E[Y_0]$

$= E[Y_1 \mid T = 1] - E[Y_0 \mid T = 0]$

$= E[Y \mid T = 1] - E[Y \mid T = 0]$

- e.g., in a Randomized Experiment (Fisher 1920-30), when each unit is randomly assigned to a Treatment or Control Group
- Still need additional assumptions

# SUTVA

Stable Unit Treatment Value Assumptions

- Cox 1958, Rubin 1978

1. No "interference" or "spill-over effect" among units
   - For unit i, $Y_{1i}$ and $Y_{10}$ are NOT affected by what action any other unit j received
2. Unique Treatment Level or "Dose"
   - There are no hidden versions of treatments
   - No matter how (mechanism) unit i received treatment 1, the outcome that would be observed would be $Y_{1i}$ -- similarly for treatment 0

# Violations of SUTVA

1. No interference
   - (wiki) Two units Joe and Mary for effect of a drug for high blood pressure
   - They share the same household
   - Mary cooks
   - Mary got drug (treatment) – her pressure reduces – cooks salty food
     - In practice, Mary may not know if she got the drug or placebo
   - Joe's pressure increases
2. Unique Treatment Level or "Dose"
   - Different doses of the medicine for blood pressure

# More assumptions

- Compliance issue
  - People assigned to treatment may refuse it
  - People assigned to control may try to get treatment
    - Barnard, Frangakis, Hill, and Rubin 2003
  - People started taking a medicine, then stopped in the middle because it made them too sick to work

# Notes on Neyman-Rubin Model

- At least half of the potential outcomes are missing
  - Still it is important to explicitly represent both potential outcomes
  - Considered to be a significant contribution by Neyman (Rubin 2005)

- Assumptions are critical
  - without them the causal inferences are meaningless

# The Power of Randomized Experiments

Recall

- Covariates (X) represent the set of variables that take their values before the assignment of the units into treatment or control groups
  - e.g., the gender of a human subject
  - cannot be affected by treatments

- What do we get by randomly assigning units to treatment/control groups?

# The Power of Randomized Experiments

- The assigned treatment is statistically independent of any (measured or unmeasured) covariate in the population before the experiment has been started
  - The distribution of any covariate is the same in the treatment and control groups
- Any difference in outcomes is due to the treatment and not any other pre-existing differences
- The average of control/treatment group outcomes is an unbiased estimate of average outcome under control/treatment for whole population
  - $ATE = E[Y_1 - Y_0] = E[Y \mid T = 1] - E[Y \mid T = 0]$

# But, Randomized Experiments are not always feasible

1. Infeasibility or high cost
   - e.g., how allocation of government funding in different research areas will affect the number of academic jobs in these areas
2. Ethical reasons
   - e.g., effect of availability to better resources during childhood on higher education in the future
3. Prohibitive delay
   - e.g., effect of childhood cholesterol on teen obesity)
4. In some scenarios randomization may not estimate effects for the groups we are interested in
5. Experiments can be on a small population, may have a large variance

# Observational Study

- Alternative to true randomized experiments
  - Tries to simulate the ideal situation

- Create treatment and control groups that appear to be random
  - at least on observed/measured variables by choosing individuals with similar covariate values
  - do not use the outcome while selecting the groups

# Observational Study

Covariates (X)

| T | Y1 | Y0 | Age ($X_1$) | Race ($X_2$) | Gender ($X_3$) | State ($X_4$) | Edu ($X_5$) |
|---|---|---|---|---|---|---|---|
| 1 | 130 | ? | 20s | W | M | NC | College |
| 0 | ? | 125 | 20s | W | M | NC | College |
| 1 | 127 | ? | 30s | B | F | MA | PhD |
| 0 | ? | 130 | 30s | L | F | CA | PhD |

Average Treatment Effect (ATE) = $\mathbf{E}[Y(1) - Y(0)]$

~~$T \perp Y1, Y0 \mid X$~~

(strong ignorability)

[Rosenbaum-Rubin, '83]

~~$= E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0]$~~

$= E_X[E[Y1 \mid T = 1, X] - E[Y0 \mid T = 0, X]]$

Can be (again) estimated from observed data

# “Matching” – Exact Matching

Covariates (X)

| T | Y1 | Y0 | Age ($X_1$) | Race ($X_2$) | Gender ($X_3$) | State ($X_4$) | Edu ($X_5$) |
|---|---|---|---|---|---|---|---|
| 1 | 130 | ? | 20s | W | M | NC | College |
| 0 | ? | 125 | 20s | W | M | NC | College |
| 1 | 127 | ? | 30s | B | F | MA | PhD |
| 0 | ? | 130 | 30s | L | F | CA | PhD |

Average Treatment Effect (ATE) = **E**[Y(1) - Y(0)]

Valid group

$= E_X[E[Y1 \mid T = 1, X] - E[Y0 \mid T = 0, X]]$

Each valid matched group must have

- at least one treated unit
- at least one control unit

# Exact Matching = Interpretability

Why exact matching?

- Highlights overlap between treatment and control populations
- Helps us to find uncertainty and determine what type of additional data must be collected
- Can estimate "individualized/conditional average treatment effects (CATE)"

*"As a method of multivariate adjustment, subclassification has the advantage that it involves direct comparisons of ostensibly comparable groups of units within each subclass and therefore can be both understandable and persuasive to an audience with limited statistical training... "*

- Subclassification = exact matching
- Direct comparisons = individualized effects
- Persuasive = intuitive, uncomplicated, reproducible

*"A major problem with subclassification is that as the number of confounding variables increases, the number of sublcasses grows dramatically, so that even with only two categories per variable, yielding $2^P$ classes for P variables, most subclasses will not contain both treated and control units... "*

- Confounders = variables of potential interest
- Number of subclasses = types of individualized effects
- Empty subclasses = impossible to draw causal conclusions

[Rosenbaum-Rubin'83]

https://www.jstor.org › stable

The Central Role of the Propensity Score in ... - jstor

by PR Rosenbaum · 1983 · Cited by 33425 — Biometrika (**1983**), 70, 1, pp. ... has been called the stable unit-treatment value assumption (**Rubin**, 1980a), and will ... 42 P. R. **ROSENBAUM**...

# Several Matching Techniques

- Rosenbaum and Rubin (1983) — propensity score matching
- Rosenbaum (2002) — Full matching
- Iacus and King (2008) — Coarsened Exact Matching
- Hansen (2008) — Prognostic scores
- Schneeweiss et al (2009) — high dimensional PS
- Diamond and Sekhon (2013) — Genetic matching
- Rosenbaum, Zubizaretta, and others — Optimal matching, linear balance constraint matching (mixed integer programming approaches)
- Other approaches to observational causal inference: (re)weighting, direct modeling of outcomes—parametric, nonparametric.
- Our work from the AME lab: “Almost Exact Matching” - FLAME (2021) and DAME (2019)

# General steps in implementing matching methods

1. Define “closeness” – a distance measure to determine whether an individual is a good match for another,
2. Implement a matching method, given that measure of closeness
3. Assess the quality of the resulting matched samples, and perhaps iterating with Steps (1) and (2) until well-matched samples result, and
4. Analysis of the outcome and estimation of the treatment effect, given the matching done in Step (3).

[Reference – Survey by Stuart, 2010]

# Exact to “close-enough” match

- Dealing with multiple covariates was a challenge due to both computational and data problems. With more than just a few covariates it becomes very difficult to find matches with close or exact values of all covariates.
- Chapin (1947) finds that with initial pools of 671 treated and 523 controls there are only 23 pairs that match exactly on six categorical covariates.
- An important advance was made in 1983 with the introduction of the “propensity score”, defined as the probability of receiving the treatment given the observed covariates
  - Advantage: does not require close or exact matches on individual covariates
  - Limitation: Not much interpretable & requires a model, which may not be correct

[Reference – Survey by Stuart, 2010]

Rosenbaum-Rubin 1983

# Balancing Scores

- A balancing score b(X) is a function of the observed covariates X such that
  - the conditional distributions of X given b(X) are the same on the treatment (T = 1) and the control groups (T = 0), i.e.,
  - $X \perp T \mid b(X)$

- b(X) = X
  - The finest balancing score
- Propensity score b(X) = e(X)
  - The coarsest balancing score
  - Make coarse (bigger) groups
  - May not match on all measured covariates
  - But the distributions of covariates are the same for treatment and control
  - Cannot say anything about unmeasured/unobserved covariates

# Propensity Score

- The conditional probability of assignment to treatment given the covariates
  - $e(X) = Pr(T = 1 \mid X)$

- Known for Randomized Experiments
- Not known for Observational Study

# Strongly Ignorable Treatment Assignment

- Treatment assignment is

"strongly ignorable given a vector of covariates V"

if for all V

1. $(Y1, Y0) \perp T \mid V$
2. $0 < Pr[T = 1 \mid V] < 1$

- Simply "strongly ignorable" when V = X

[Rosenbaum-Rubin 1983]

1. If treatment assignment is strongly ignorable given X, then it is strongly ignorable given any balancing score b(X)
2. For any function b(X) of X, b(X) is a balancing score if and only if e(X) = f(b(X)) for some function f
   - In particular, $X \perp T \mid e(X)$

# Three methods for using balancing score on observational data

1. Pair matching on balancing scores
   - Sample b(X) at random
   - Then sample one treated and one control units with this value of b(X)
   - The expected difference in response equals the ATE at this b(X)
   - the mean of matches pair differences in this two-step process is an unbiased estimator of the ATE

2. Sub-classification on balancing scores
   - Sample a group of units using b(X) such that b(X) is constant for all units in this group and at least one unit in the group received each treatment (T = 1, 0).
   - The expected difference in treatment means equals the ATE at this b(X)
   - the weighted average of such differences (weight = fraction of population at b(X)) is an unbiased estimator of the ATE.

3. Covariance adjustment on balancing scores
   - Assumes that the conditional expectation of $Y_t$ given b(X) is linear
   - $E[Y_t \mid b(X), S = t] = \alpha t + \beta t b(X)$ for t = 0, 1
   - Gives an unbiased estimator of the treatment effect at b(X) = $E[Y_1 - Y_0 | b(X)]$ in terms of unbiased estimators of α1, β1, α0, β0

# Prognostic Scores [Hansen 2008]

- If $Y0 \perp X \mid \Psi(X)$, then $\Psi(X)$ is a prognostic score
- Intuition: in many settings, information about response Y in the absence of treatment is more available than information on treated subjects (otherwise need to flip)
- Need to fit a model to $\Pr(y_0 \mid x)$

# Comparing Rubin's and Pearl's Models

# Neyman-Rubin vs. Pearl's Model

- Potential Outcome (Neyman-Rubin) = Do Operator/counterfactual (Pearl)
- Treatment (Neyman-Rubin) ≋ intervention (Pearl)
- Structural causal graph on variables assumed by Pearl
  - Causal inference is on (variable-value) pairs
- No causal structure assumed in Neyman-Rubin's model
  - Infers causal relationships by experiments or from evidence
- Pearl's method gives a systematic way to find the covariates to adjust for --- but you may not have a reliable causal DAG available.. In practice the directions might not be known
- Mathematically the two frameworks are connected, but each has different established goals, tools and applicable areas (Richardson and Robins, 2013)

# Neyman-Rubin vs. Pearl's Model

Disclaimer: only some excerpts, not exhaustive views and not the most recent ones..

Some authors (e.g., Greenland, Pearl, and Robins 1999; Dawid 2000) call the potential outcomes "counterfactuals," borrowing the term from philosophy (e.g., Lewis 1973). I much prefer Neyman's implied term "potential outcomes," because these values are not counterfactual until after treatments are assigned, and calling all potential outcomes "counterfactuals" certainly confuses quantities that can never be observed (e.g., your height at age 3 if you were born yesterday in the Arctic) and so are truly a priori counterfactual, with unobserved potential outcomes that are not a priori counterfactual (see Frangakis and Rubin 2002; Rubin 2004; and the discussion and reply for more on this point).

Despite other approaches advocated by people whom I greatly respect (e.g., Dawid 2000; Lauritzen 2004; Pearl 2000), the potential outcomes formulation of causal effects, whether in randomized experiments or in observational studies, has achieved widespread acceptance. The potential outcomes, together with

(Rubin, JASA, 2005, p325 & p329 )

*"Formally, the two frameworks are logically equivalent; a theorem in one is a theorem in the other, and every assumption in one can be translated into an equivalent assumption in the other. Therefore, the two frameworks can be used interchangeably and symbiotically, as it is done in the advanced literature in the health and social sciences….In summary, the PO framework offers a useful analytical tool (i.e.. an algebra of counterfactuals) when used in the context of a symbiotic SCM analysis. It may be harmful however when used as an exclusive and restrictive subculture that discourages the use of process-based tools and insights."*

*(Pearl 2012)*

*Read [http://causality.cs.ucla.edu/blog/index.php/2012/12/03/judea-pearl-on-potential-outcomes/](http://causality.cs.ucla.edu/blog/index.php/2012/12/03/judea-pearl-on-potential-outcomes/) for a detailed version*