Lecture 3 - Potential Outcome Framework

Treatment Assignments and Counterfactuals

Dr. Xi Chen

Rotterdam School of Management

2025-06-21

Outline

Think about this problem...

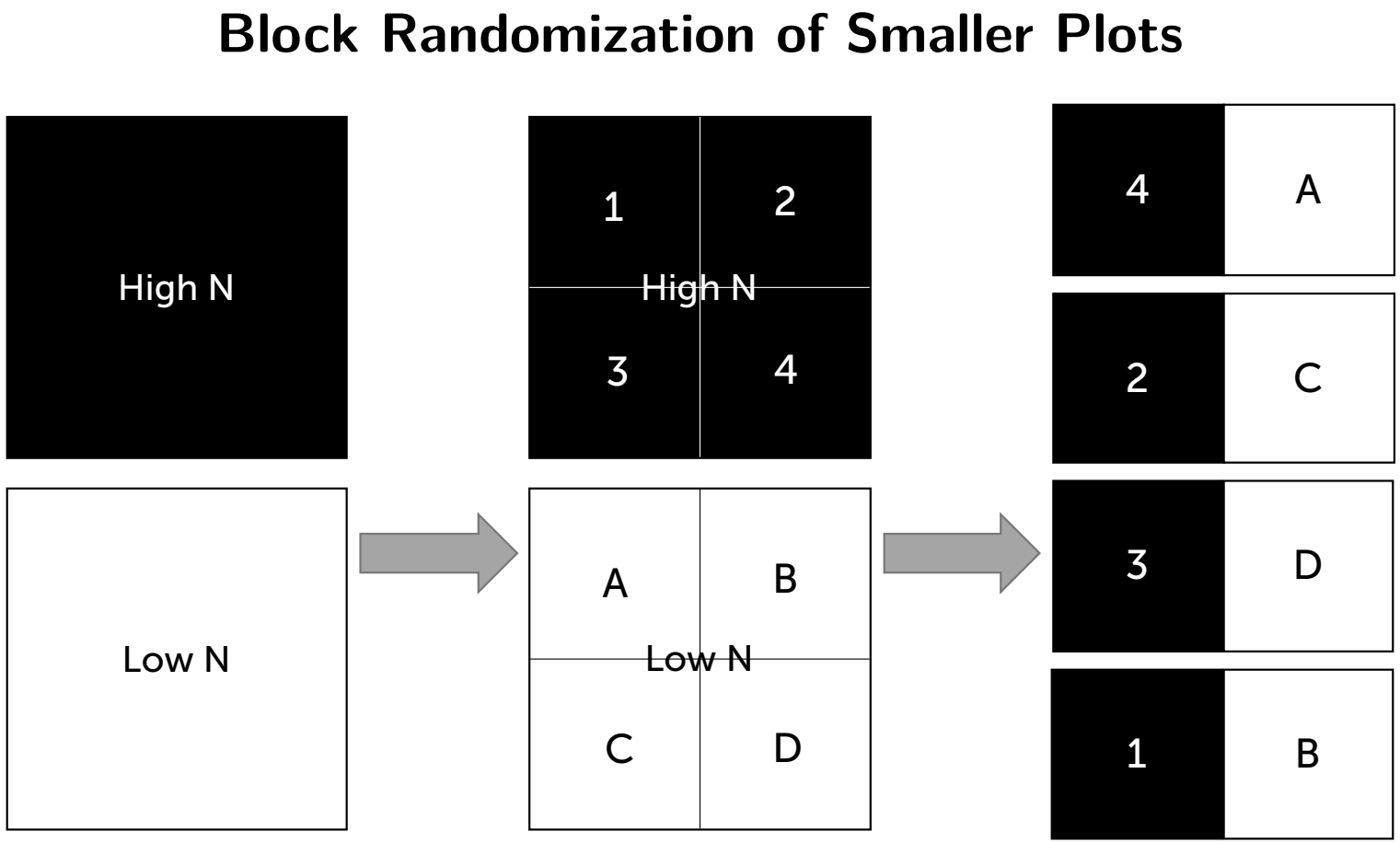
Ronald Fisher at Rothamsted[1] in 1919

Problems with some of the agricultural experiments

- Same field, same treatment, but plant performance is uneven...

[1] Rothamsted Agricultural Experiment Station is and was a household word among agricultural scientists, especially during Fisher's day.

Fisher's proposal: the physical randomization



Fisher's contribution: the central role of randomization

A quote from Fisher (1926):

- "*One way of making sure that* ***a valid estimate of error will be obtained is to arrange the plots deliberately at random****, so that no distinction can creep in between pairs of plots treated alike and pairs treated differently; in such a case an estimate of error, derived in the usual way from the variation of sets of plots treated alike, may be applied to test the significance of the observed difference between the averages of plots treated differently.*"

Basic concepts I

Definition (**Unit**)

The person, place or thin upon which a treatment will operate, at a particular time.

Definition (**Treatment/Intervention/Policy**)

An intervention (X), the effects of which (on some particular measurement of the units, Y) the investigator wishes to assess relative to no intervention (i.e., the "control").

- Effectively, the treatment (or the lack of it) set $X = x$ or $\mathtt{do}(X) = x$.
- A treatment set is the possible treatments the investigator intend to compare, e.g., {Ads, No Ads}.

Basic concepts II

Definition (**Causal State**)

A causal state with respect to a unit and a treatment (in the treatment set) in a system is the ***hypothetical scenario*** where the treatment is applied to the unit.

Definition (**Potential Outcome**)

A potential outcome of a causal state is the value of the measure of interest Y in that causal state.

Basic concept III

Units	**Treatments**		**Causal Effect**		
	Y(**Ads**)	Y(**No Ads**)	Y(**Ads**) $-$ Y(**No Ads**)	$Y^{\text{(Ads)}}/Y^{\text{(No Ads)}}$	...
Timmy	€75	€25	€50	300%	...

Definition (**Causal Effect**)

For a unit, the comparison of the potential outcomes under two causal states is a causal effect.

The fundamental problem of causal inference

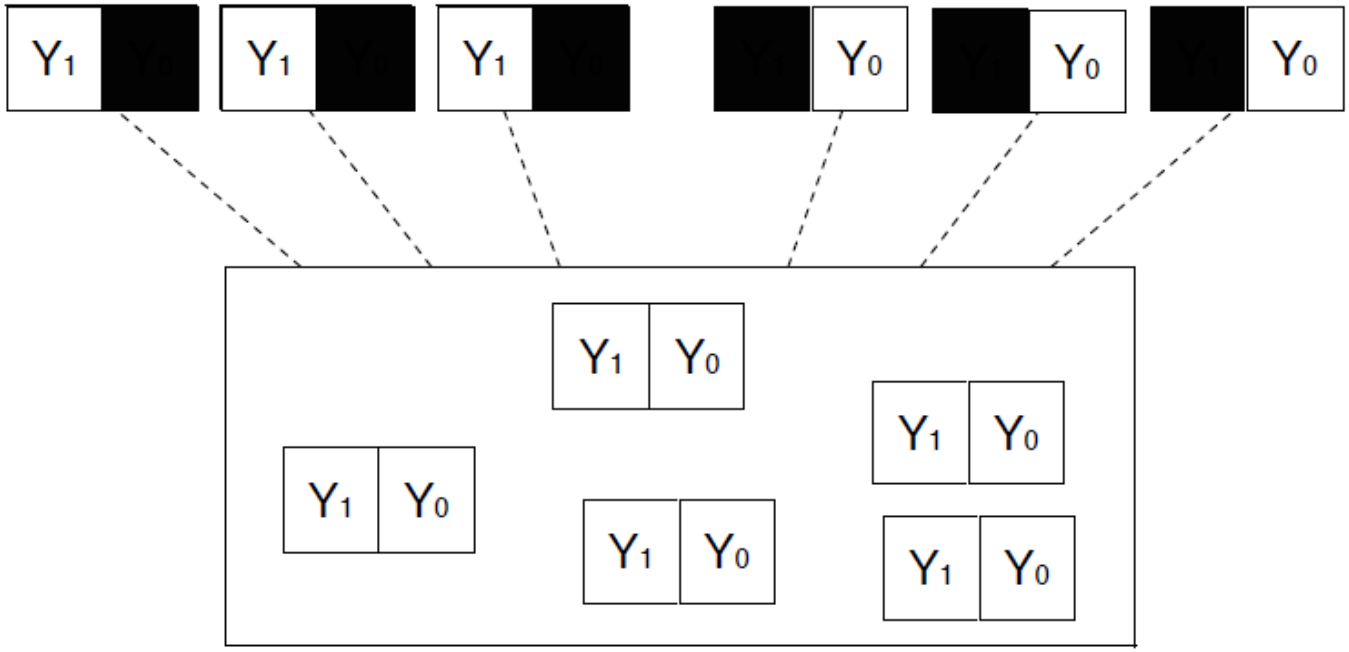
Units	**Treatments**		**Causal Effect**		
	Y(**Ads**)	Y(**No Ads**)	Y(**Ads**) $-$ Y(**No Ads**)	$Y^{\text{(Ads)}}/Y^{\text{(No Ads)}}$	...
Timmy	€75	?	?	?	?

Fact (The Fundamental Problem of Causal Inference)

For a unit, only one causal state can be realized, and the investigator can only observe the potential outcome from the realized causal state.

"Potential outcome" is credited to Neyman

- The idea of "potential outcome" is credited to Jerzy Neyman (1894-1981)[2].
- Neyman noticed a plot has two states - fertilized or not, but only one outcome was observed.



[2]Neyman earned his Doctor of Philosophy degree at University of Warsaw in 1924 for a dissertation titled "On the Applications of the Theory of Probability to Agricultural Experiments".

Potential outcomes: a bit of math

D_i: indicator of treatment assignment to unit i.

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ receives the treatment} \\ 0 & \text{otherwise} \end{cases}$$

$Y_i^{D_i}$: The potential outcome for unit i, under D_i.

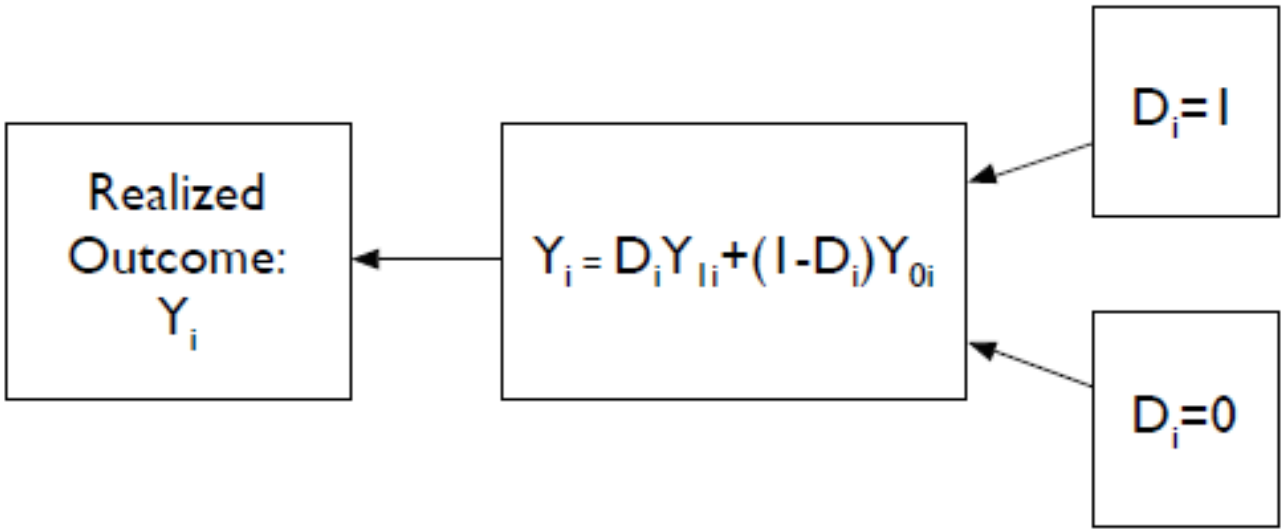
$$\begin{cases} Y_i^1 & \text{the potential outcome if } D_i = 1 \\ Y_i^0 & \text{the potential outcome if } D_i = 0 \end{cases}$$

τ_i: individual treatment effect (ITE) of unit i, with $\tau_i = Y_i^1 - Y_i^0$.

Potential outcomes: a bit of math

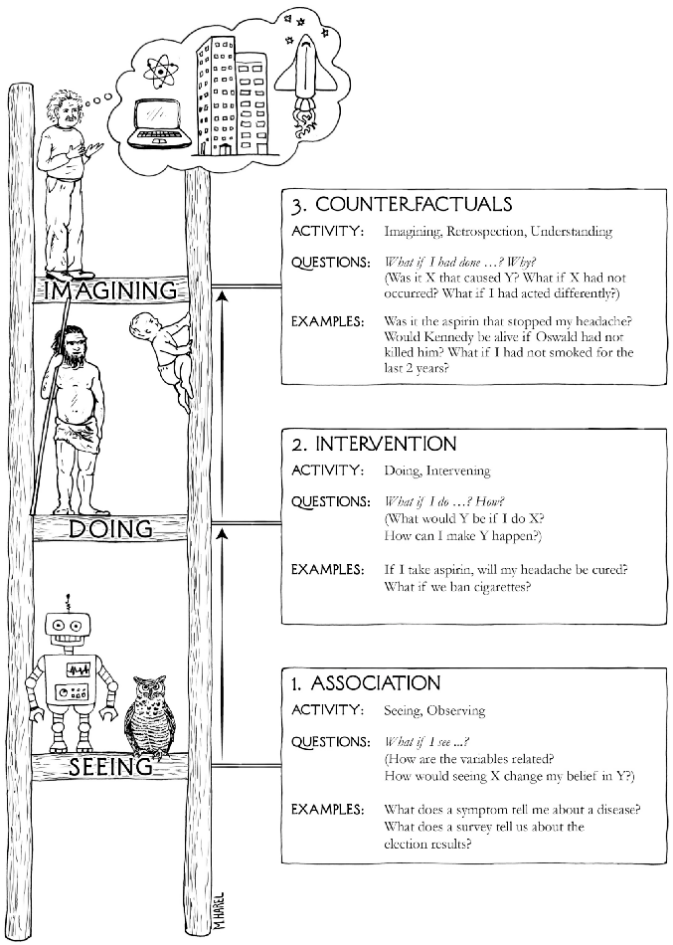
Formalizing the fundamental problem

$$Y_i = D_i \cdot Y_i^1 + (1 - D_i) \cdot Y_i^0, \text{ with } Y_i = \begin{cases} Y_i^1 & \text{if } D_i = 1 \\ Y_i^0 & \text{if } D_i = 0 \end{cases}$$



The ladder of causation

Counterfactual / potential outcomes take the key role in both causal graph and potential outcome framework.



The ladder of causation, from Pearl and Mackenzie (2018)

ITE ($\tau_i = Y_i^1 - Y_i^0$) and SUTVA

Stale Unit Treatment Value Assumption (SUTVA)

The SUTVA assumption consists of two parts: (a) there is only one form of the treatment and one form of the control, and (b) there is no interference among units.

How to understand these assumptions?

ITE ($\tau_i = Y_i^1 - Y_i^0$) and SUTVA

Part (a): An Exclusion restriction to rule out different treatments.

- Do not mix up treatments: one consistently treatment for all treated units.
- Or else, we do not know which treatment is working.

A mix of two treatments for an individual i

The outcome depends on the mix of the two treatments:

Y_i	$D_i^2 = 0$	$D_i^2 = 1$
$D_i^1 = 0$	$Y_i^0\left(D_i^2 = 0\right)$	$Y_i^0\left(D_i^2 = 1\right)$
$D_i^1 = 1$	$Y_i^1\left(D_i^2 = 0\right)$	$Y_i^1\left(D_i^2 = 1\right)$

ITE ($\tau_i = Y_i^1 - Y_i^0$) and SUTVA

Part (b): An Exclusion restriction to rule out interference

- Can be unrealistic in some settings: *New product adoptions, activities on social media, fertilizers on plot yields etc*.
- But, it greatly simplifies our analysis...
- Traditional models **almost always implicitly** assume SUTVA.

Two individuals i and j under interference

My outcomes depend on your causal states:

Y_i	$D_j = 0$	$D_j = 1$
$D_i = 0$	$Y_i^0(D_j = 0)$	$Y_i^0(D_j = 1)$
$D_i = 1$	$Y_i^1(D_j = 0)$	$Y_i^1(D_j = 1)$

ITE ($\tau_i = Y_i^1 - Y_i^0$) and SUTVA

Two individuals without SUTVA

$$Y_i = \begin{cases} Y_i^{00} \\ Y_i^{01} \\ Y_i^{10} \\ Y_i^{11} \end{cases} \Rightarrow \tau_i(D) = \begin{cases} Y_i^{11} - Y_i^{00} \\ Y_i^{11} - Y_i^{01} \\ Y_i^{10} - Y_i^{00} \\ Y_i^{10} - Y_i^{01} \\ Y_i^{11} - Y_i^{10} \\ Y_i^{01} - Y_i^{00} \end{cases}$$

However, we only observe one potential outcome...
The no. of ITEs $\uparrow^{\text{Exponentially}}$ with more individuals!

How to solve the fundamental problem?

Assuming homogeneity to get ITE $\tau_i = Y_i^1 - Y_i^0$:

- $\{Y_i^0, Y_i^1\}$ **constant across individuals**: cross-sectional comparisons.

Subject	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Joe	130	?	?
Mary	?	125	?

How to solve the fundamental problem?

Assuming homogeneity to get ITE $\tau_i = Y_i^1 - Y_i^0$:

- $\{Y_i^0, Y_i^1\}$ **constant across individuals**: cross-sectional comparisons.

Subject	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Joe	130	?	?
Mary	?	125	?

Subject	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Joe	130	**125**	**5**
Mary	**130**	125	**5**

How to solve the fundamental problem?

Assuming homogeneity to get ITE $\tau_i = Y_i^1 - Y_i^0$:

- $\{Y_i^0, Y_i^1\}$ **constant across time**: before-after comparisons.

Time	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Time 1	130	?	?
Time 2	?	125	?

How to solve the fundamental problem?

Assuming homogeneity to get ITE $\tau_i = Y_i^1 - Y_i^0$:

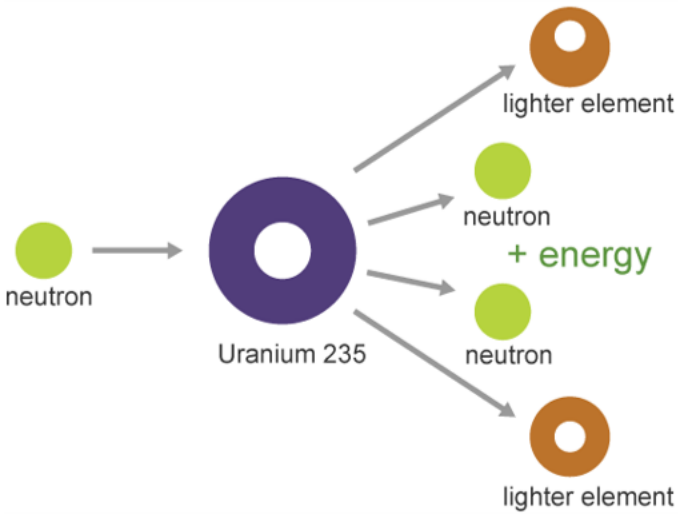
- $\{Y_i^0, Y_i^1\}$ **constant across time**: before-after comparisons.

Time	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Time 1	130	?	?
Time 2	?	125	?

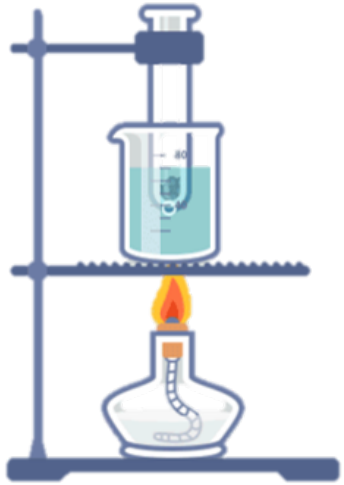
Time	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Time 1	130	**125**	**5**
Time 2	**130**	125	**5**

Homogeneity assumptions

Homogeneous Units



Homogeneous Time



However, in many cases, the assumption of homogeneity is unreasonable.

How to solve the fundamental problem?

Subject	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Joe	130	?	?
Mary	?	125	?
Sally	100	?	?
Bill	?	130	?
James	?	120	?
Mean	**115**	**125**	**-10**

Fact (**The Expectations**)

ITEs are unknowable, but the expected potential outcome of a causal state is observable with multiple units.

How to solve the fundamental problem?

$\tau_{\text{ATE}} = E\left[Y_i^1 - Y_i^0\right]$ is unobserved, and we need (EO = Expected Outcome):

$$\begin{cases} \mathbf{E}\left[\mathbf{Y_i^1}\right] & \text{EO with treatment} \\ \mathbf{E}\left[\mathbf{Y_i^0}\right] & \text{EO without treatment} \end{cases}$$

From treatments and realized outcomes, we have:

$$\begin{cases} \mathbf{E}\left[\mathbf{Y_i^1} \mid \mathbf{D_i = 1}\right] & \text{EO of the treatment group} \\ \mathbf{E}\left[\mathbf{Y_i^0} \mid \mathbf{D_i = 0}\right] & \text{EO of the control group} \end{cases}$$

To identify τ_{ATE}, need a **sufficient condition**:

$$\begin{cases} \mathbf{E}\left[\mathbf{Y_i^1}\right] & \mathbf{= E}\left[\mathbf{Y_i^1} \mid \mathbf{D_i = 1}\right] \\ \mathbf{E}\left[\mathbf{Y_i^0}\right] & \mathbf{= E}\left[\mathbf{Y_i^0} \mid \mathbf{D_i = 0}\right] \end{cases}$$

A closer look at the condition

To identify τ_{ATE}, need a **sufficient condition**:

$$\begin{cases} \mathbf{E}\left[\mathbf{Y_i^1}\right] & = \mathbf{E}\left[\mathbf{Y_i^1} \mid \mathbf{D_i = 1}\right] \\ \mathbf{E}\left[\mathbf{Y_i^0}\right] & = \mathbf{E}\left[\mathbf{Y_i^0} \mid \mathbf{D_i = 0}\right] \end{cases}$$

This condition requires **an assumption** called

Unconfoundedness

The **causal states** for all units are **independent of all potential outcomes**, observed or unobserved.

The role of randomization

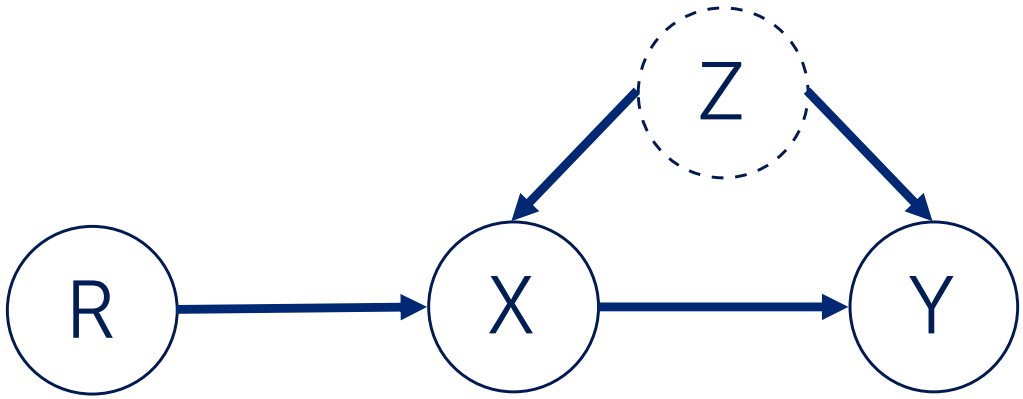
"One way of making sure that a valid **estimate of error** will be obtained is to **arrange the plots deliberately at random**, so that **no distinction can creep in** between pairs of plots treated alike and pairs treated differently; ..."

– R. Fisher (1926), p. 503.

The real reason of doing randomization is beyond Fisher

Fisher sees randomization as a tool to guarantee "no distinction" between treatment groups, or else a valid estimate of error cannot be obtained.

Randomization as an instrumental variable design



- Draw a random variable $R \sim N(0, 1)$.
- Get the treatment if $R > 0$.
- R needs to be **exogenous, excluded and relevant**.

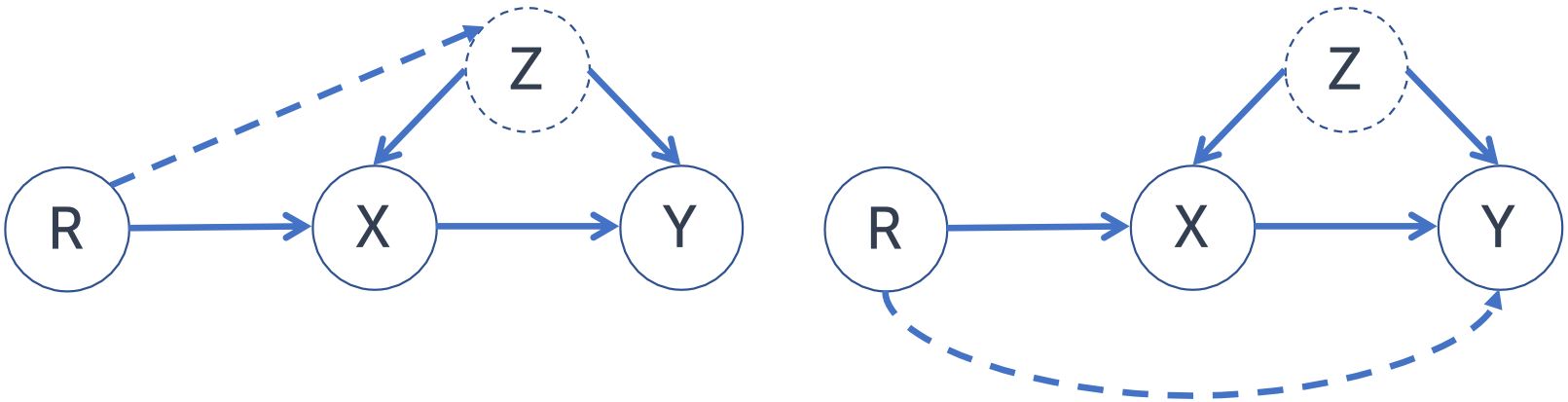
An IV perspective of experiments

Agricultural or medical experiments, $do(X) = x$ or directly change X.

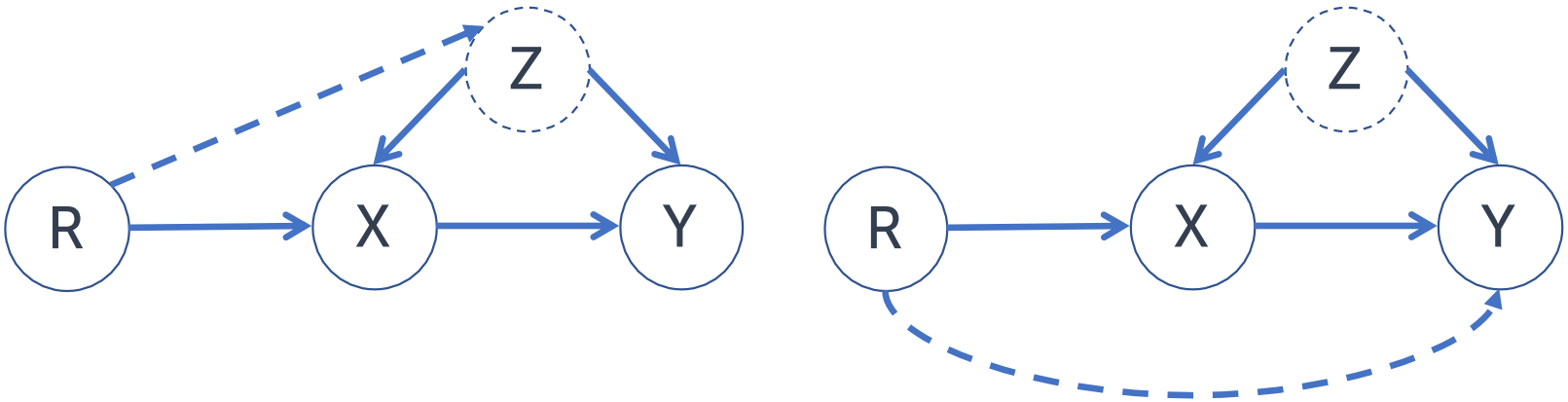
- E.g., Fertilizers or treatments or medication (dosage) etc.

Psych experiments: using "manipulations" in hope to change X.

- E.g., to use happy face (☺) or sad face (☹) to manipulate emotions.



An IV perspective of experiments

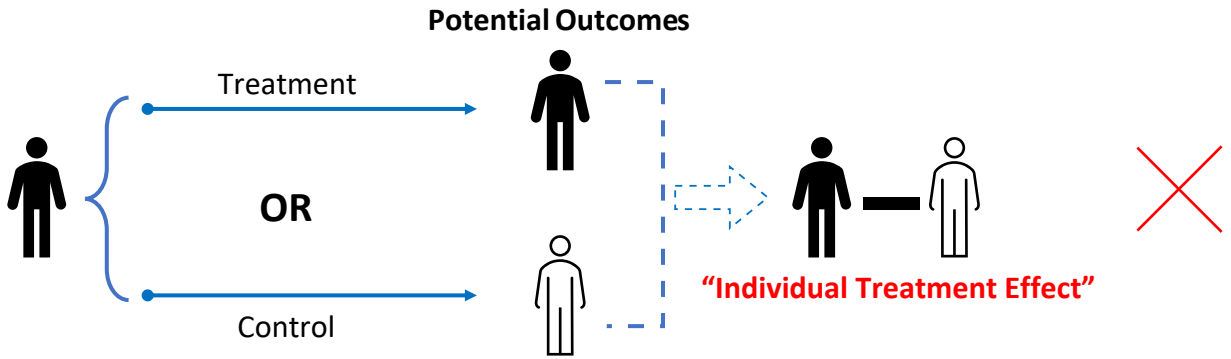


Violations of the exclusion assumption

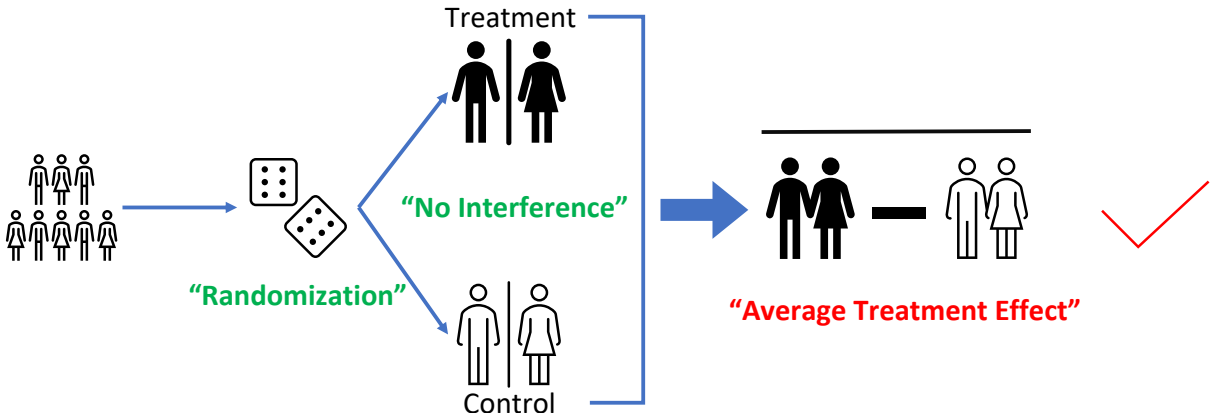
The exogeneity still holds (randomization).

Exclusion might be violated.

Put it all together...



(a) ITE is not identified



(b) ATE is identified

Understanding assignment

Definition (Assignment mechanism)

The process that determines which units receive which treatments. With N units, the assignment mechanism is a probability $P\left(D \mid X, Y^0, Y^1\right)$, with X as the pre-assignment characteristics of units.

$$\sum_{D \in \{0,1\}^N} P\left(D \mid X, Y^0, Y^1\right) = 1, \forall X, Y^0, Y^1.$$

Definition (Unit Assignment Probability)

The assignment probability for a unit i is

$$p_i\left(X, Y^0, Y^1\right) = \sum_{\{D \mid D_i = 1\}} P\left(D \mid X, Y^0, Y^1\right)$$

An example of three units

With 3 units, all possible values for D are:

$$D \in \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Suppose we observe characteristics X with $X_i = i \in \{1, 2, 3\}$. The assignment process is like this:

1. The unit with highest X is assigned to treatment, so $p_3(1 \mid X) = 1$.
2. Unit 2 is assigned with a coin toss with $p_2(1) = 0.5$.
3. Unit 1 is assigned to treatment, if the sum of observed outcomes of Unit 2 and 3 are positive.

An example of three units

The final assignment function is,

$$P\left(D \mid X, Y^0, Y^1\right) = \begin{cases} \frac{1}{2} & \text{if } \left(Y_2^0 + Y_3^1\right) \leq 0, \text{ and } D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ \frac{1}{2} & \text{if } \left(Y_2^1 + Y_3^1\right) \leq 0, \text{ and } D = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ \frac{1}{2} & \text{if } \left(Y_2^0 + Y_3^1\right) > 0, \text{ and } D = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\ \frac{1}{2} & \text{if } \left(Y_2^1 + Y_3^1\right) > 0, \text{ and } D = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{cases}$$

Featuring some assignment mechanisms

Definition (Individualistic Assignment)

An assignment mechanism $P\left(D \mid X, Y^0, Y^1\right)$ is individualistic if

$$p_i\left(X, Y^0, Y^1\right) = p_i\left(X_i, Y_i^0, Y_i^1\right)$$

Definition (Probabilistic Assignment)

An assignment mechanism $P\left(D \mid X, Y^0, Y^1\right)$ is probabilistic if

$$0 < p_i\left(X, Y^0, Y^1\right) < 1, \forall X, Y^0, Y^1$$

Featuring some assignment mechanisms

Definition (Unconfounded Assignment)

An assignment mechanism is unconfounded if it does not depend on the potential outcomes:

$$P\left(D \mid X, Y^0, Y^1\right) = P\left(D \mid X\right)$$

Definition (Regular Assignment)

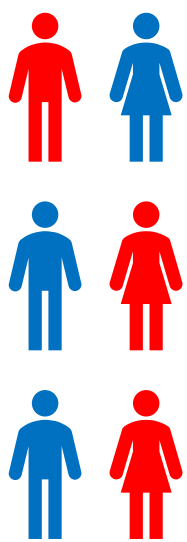
An assignment mechanism is regular, if the assignment is individualistic, probabilistic and unconfounded.

Randomized experiments: some designs

- A randomized experiment is a regular assignment mechanism.
- The functional form of $P\left(D \mid X, Y^0, Y^1\right)$ may not be known to experimenters.

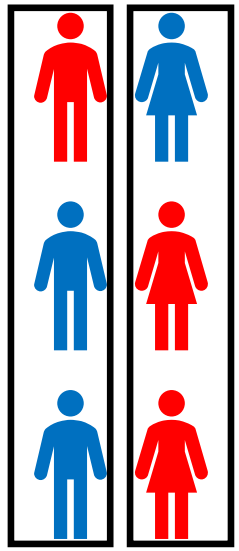
Bernoulli Trial

$p_i = p$



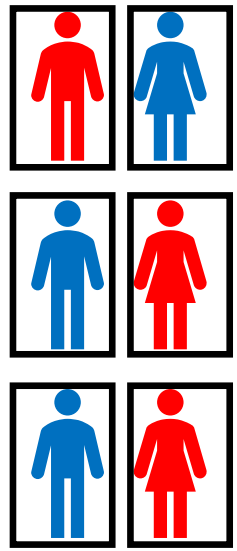
Complete Random Trial

$P = \binom{N}{N_1}^{-1}$



Block Random Trial

$P = \prod_{b=1}^{B} \binom{N_b}{N_{1b}}^{-1}$

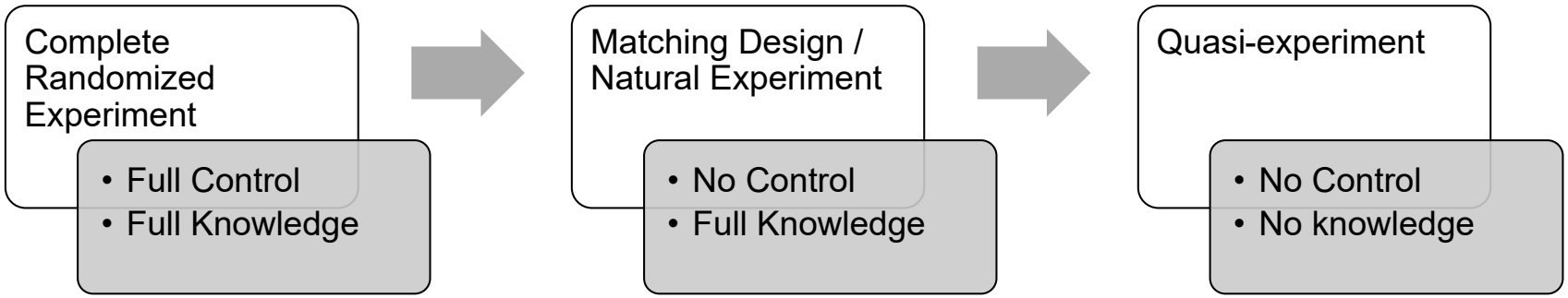


Paired Random Trial

$P = 2^{-N}$

Why do we study assignment mechanism?

The **control** over and the **knowledge** about the **assignment mechanism** determine the credibility of your causal inference efforts.



Fisher's contribution: the inferential method

Fisher's test: logic

Definition (Proof by Contradiction)

1. Start by assuming the opposite of what you want to prove.
2. Work from the assumption, arrive at a contradiction.
3. Conclude that your initial assumption is wrong, and conclude the proof.

Fisher's test for completely random experiments: 6 steps

1. Specify a **sharp null hypothesis** (e.g., no treatment effect or $H_0 : Y_i^1 - Y_i^0 = 0$).
2. Specify a **test statistic** (e.g., $T : \overline{Y^1} - \overline{Y^0}$).
3. Calculate the value of the test statistic with the observed outcomes.
4. Fill in potential outcomes using the sharp H_0 and the observed outcomes.
5. **Permute to get other possible assignments**, calculate the value of T under **the permuted assignments**.
6. Determine how extreme the value in the observations (**Step 3**) by calculating the exact p-value.

Fisher's test: An example

A completely randomized experiment is done to examine the exposure to an ad on the purchase intentions of consumers.

Unit	D_i	Y_i	Y_i^0	Y_i^1
1	0	55.0	55.0	?
2	0	72.0	72.0	?
3	0	72.7	72.7	?
4	1	70.0	?	70.0
5	1	66.0	?	66.0
6	1	78.9	?	78.9

1. Specify a sharp null hypothesis H_0: the ad leads to an increase of 5.0 on purchase intentions: $Y_i^1 - Y_i^0 = 5.0$.
2. Specify a test statistic: the difference in means $T = \overline{Y^1} - \overline{Y^0}$.
3. Calculate the observed value of the statistic: $T^{\text{Obs}} = 5.1$; bigger means more extreme.

Fisher's test: An example

4. Fill in the missing values given H_0:

Unit	D_i	Y_i	Y_i^0	Y_i^1
1	0	55.0	55.0	(60.0)
2	0	72.0	72.0	(77.0)
3	0	72.7	72.7	(77.7)
4	1	70.0	(65.0)	70.0
5	1	66.0	(61.0)	66.0
6	1	78.9	(73.9)	78.9

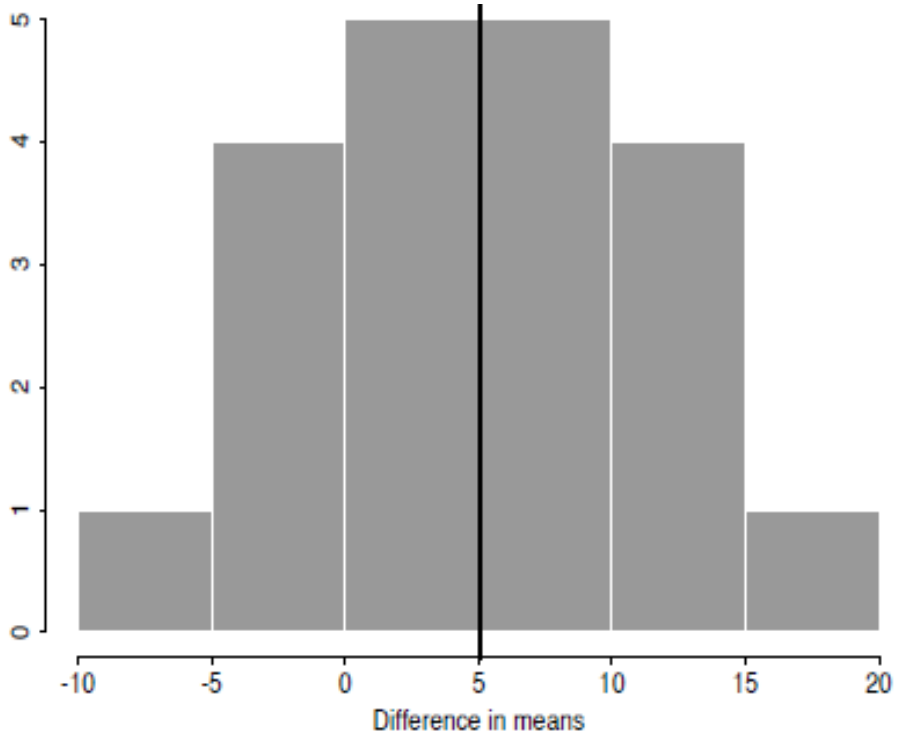
Fisher's test: An example

5. For each possible assignment (with 3 treated units and 3 control units, and $C_6^3 = 20$), calculate the test statistic.

D	Prob of D	T	Calculation
1 1 1 0 0 0	$1/20$	4.9	$\frac{60+77+77.7}{3} - \frac{65+61+73.9}{3}$
1 1 0 1 0 0	$1/20$	-0.2	
1 1 0 0 1 0	$1/20$	-2.9	
1 1 0 0 0 1	$1/20$	5.7	
1 0 1 0 1 0	$1/20$	0.3	
...	$1/20$		...
0 0 0 1 1 1	$1/20$	5.1	...

Fisher's test: An example

6. Calculate the exact p-value under H_0 or the probability of the average treatment effect is larger than the observed 5.1 (p-value $10/20 = 0.5$).



Fisher's test: summary

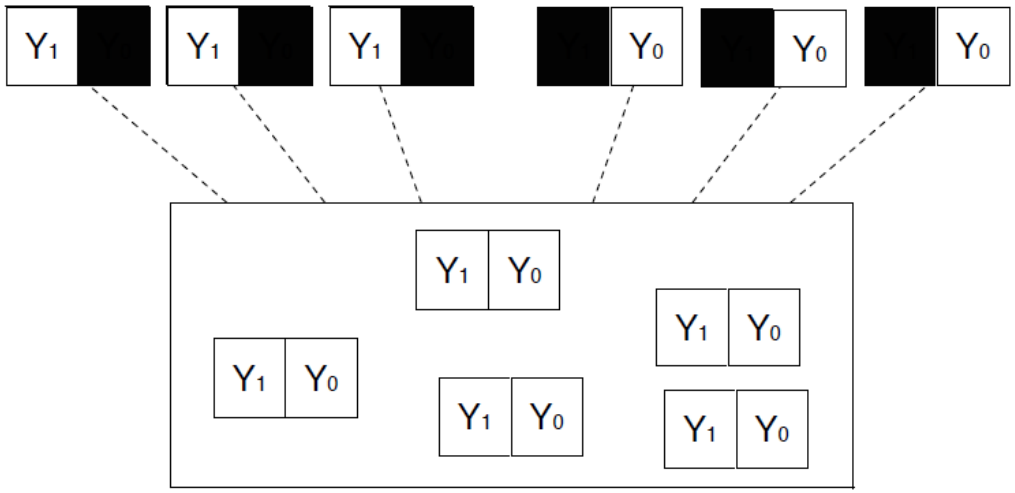
- Works from sharp null hypotheses (no alternative hypotheses or H_1).
- Requires full knowledge of the assignment mechanism.
  - SUTVA must be assumed to simplify analysis.
- The variation is from the randomness of the assignment mechanism.
- With many units, hard to calculate the exact value.

Neyman's repeated sampling approach

Neyman cares about the **"generalizable effects"** of treatments.

- Neyman defines the effect as **"the diff in average yields of plots."**
- The **difference in individual plots** is uninteresting to him.

$$\overline{Y^1} - \overline{Y^0} = \frac{1}{N}\sum_i Y_i^1 - \frac{1}{N}\sum_i Y_i^0, \text{ with potential outcomes}$$



Neyman's approach: the sample analog

- However, $\overline{\tau} = \overline{Y^1} - \overline{Y^0} = \frac{1}{N}\sum_i Y_i^1 - \frac{1}{N}\sum_i Y_i^0$ is not directly observed (**the fundamental problem**).
- Neyman constructs a "sample analog",

$$\hat{\tau} = \hat{Y}^1 - \hat{Y}^0 = \frac{1}{N_1}\sum_i Y_i^1 - \frac{1}{N_0}\sum_j Y_j^0.$$

  - $\hat{\tau}$ is unbiased for the ATE under complete randomization.
- Neyman wants to check how the sample ATE **varies with repeated sampling** for generalization.

Neyman's approach: to derive variance of $\hat{\tau}$

Neyman derives the "sampling variance" of $\hat{\tau}$, assuming all potential outcomes are observed[3]:

$$V_D(\hat{\tau}) = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{01}^2}{N}, \text{ with}$$

$$\begin{cases} \underbrace{S_0^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i^0 - \overline{Y^0}\right)}_{\text{Variance of } Y_i^c} \text{and } \underbrace{S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i^1 - \overline{Y^1}\right)}_{\text{Variance of } Y_i^t} \\ \underbrace{S_{01}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\left(Y_i^1 - Y_i^0\right) - \overline{\tau}\right)}_{\text{Variance of ITEs}} \end{cases}$$

[3] For the derivation, please see Imbens and Rubin (2015), p. 89.

Neyman's approach: to derive variance of $\hat{\tau}$

We can consistently estimate S_0^2 and S_1^2 from observed experiments, but S_{01}^2 is not identified (the fundamental problem):

$$S_{01}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\left(Y_i^1 - Y_i^0 \right) - \overline{\tau} \right).$$

Neyman thus proposes **a conservative estimator**:

$$\begin{aligned} V_D^{\text{Neyman}}(\hat{\tau}) &= \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} \\ &> V_D(\hat{\tau}) \\ &= \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{01}^2}{N} \end{aligned}$$

Neyman's approach: to derive variance of $\hat{\tau}$

Some notes $V_D^{\text{Neyman}}(\hat{\tau}) = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1}$

1. $V_D^{\text{Neyman}}(\hat{\tau})$ is only unbiased if the ITEs are constant (homogeneity).
2. Fisher's sharp null hypothesis of no treatment effects ($Y_i^t - Y_i^c = 0$) also makes $V_D^{\text{Neyman}}(\hat{\tau})$ unbiased.
3. Other approximations are also proposed, omitted here.
4. Neyman applies the **central limit theorem** and devises a t-test based on $\hat{\tau}$ and $V_D^{\text{Neyman}}(\hat{\tau})$.

The Fisher-Neyman exchange

Neyman: *So long as the average yields of any treatments are identical, the question as to whether these treatments affect separate yields on single plots seems to be uninteresting.*

Fisher: *It may be foolish, but that is what the Z test was designed for, and the only purpose for which it has been used.*

Neyman: *I am considering problems which are important from the point of view of agriculture.*

Fisher: *It may be that the question which Dr. Neyman thinks should be answered is more important than the one I have proposed and attempted to answer. I suggest that before criticizing previous work it is always wise to give enough study to the subject to understand its purpose.*

Fisher vs. Neyman: hypotheses

Neyman's approach:

$$\begin{cases} H_0^{\text{Neyman}} : & \frac{1}{N}\sum_i Y_i^1 - \frac{1}{N}\sum_i Y_i^0 = 0 \\ H_1^{\text{Neyman}} : & \frac{1}{N}\sum_i Y_i^1 - \frac{1}{N}\sum_i Y_i^0 \neq 0 \end{cases}$$

Test statistic:

$$t\text{-stat} = \frac{\hat{Y}^1 - \hat{Y}^0}{\sqrt{V_D^{\text{Neyman}}(\hat{\tau})}}$$

Fisher's approach:

$$\begin{cases} H_0^{\text{Fisher}} : & Y_i^1 - Y_i^0 = 0, \forall i \\ H_1^{\text{Fisher}} : & \text{N.A.} \end{cases}$$

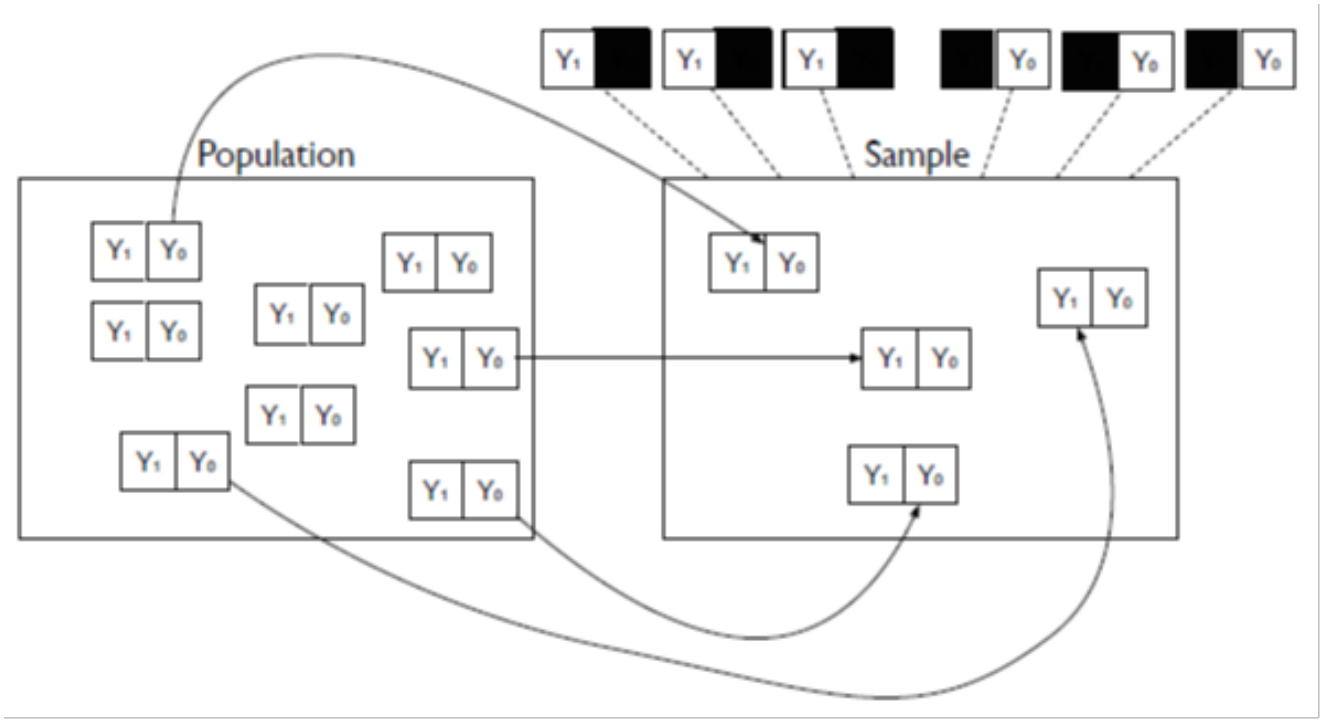
Test statistic:

Any statistic with the exact p-value.

Fisher vs. Newman: sample vs. population ATE

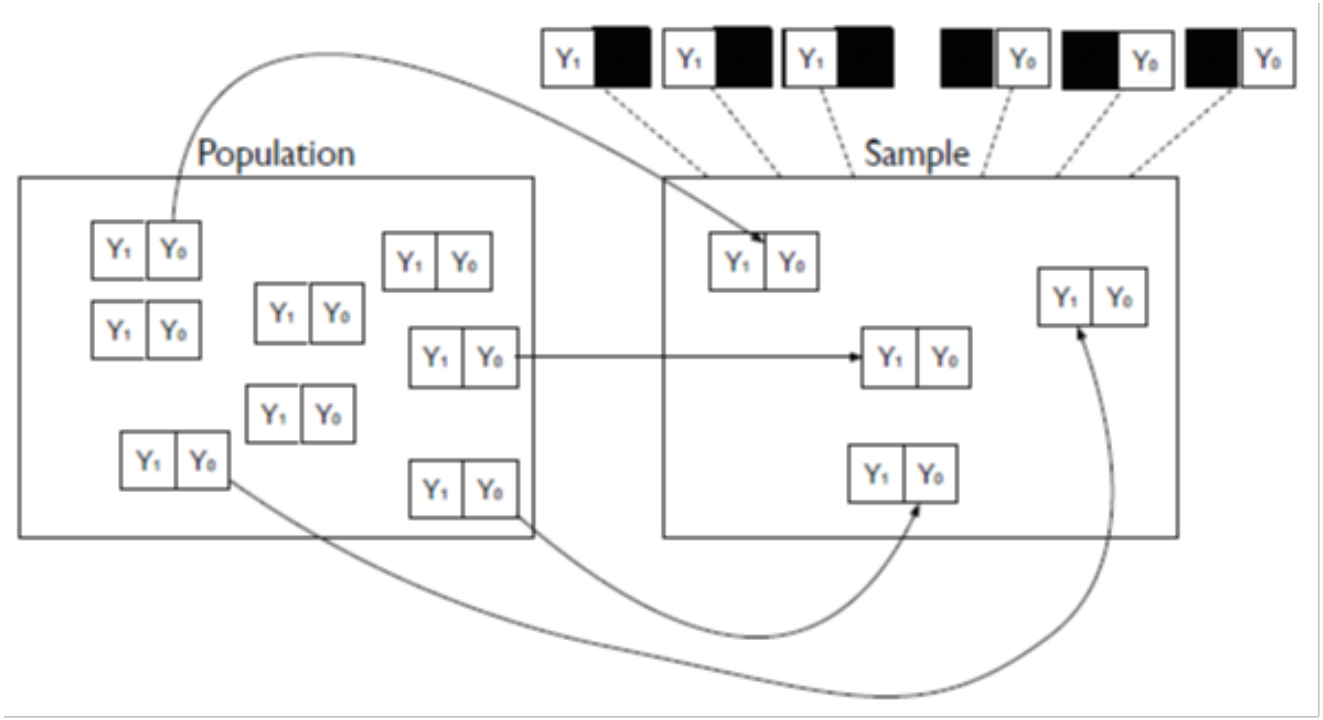
Two sources of uncertainties:

1. Variation from the assignment mechanism.
2. Variation from selecting units into your study (from the super-population).



Fisher vs. Newman: sample vs. population ATE

Sample ATE: Variation 1 and inference **limited to samples in the study**.
Population ATE: Variation 1 plus 2, to **generalize SATE** to the population.



Unconfoundedness vs. ignorability

Here I adopt the conceptualization by Rubin (2005). Unconfoundedness refers to the independence of treatment assignments with all potential outcomes, and ignorability to the independence of treatment assignments with only the unobserved potential outcomes.

Ignorable but confounded assignment

Suppose Bol.com is doing a field experiment to compare two versions of their recommender system: version A and version B. The experiment is implemented in two phases.

- Phase I: Version A or B are randomly assigned to consumers ($p = 0.5$), until the data scientists observe a 10% difference in the purchases between A and B.
- Phase II: Phase I supports Version A and Version A is assigned to all consumers until the end of the experiment.

Because of random assignment in Phase I and the fact all consumers in Phase II are given Version A (no assignment), the assignment of A is not related to the actual purchases. For Group B, because of the random assignment in Phase I, the assignment of B is also unrelated to the actual purchases. However, the assignment of A are not independent from the potential purchases of Group A people, as A is assigned to all people in Phase II exactly because $E[Y_i \mid A] > E[Y_i \mid B]$.

The inference: Fisher's take

Lady Tasting Tea (Fisher 1935)

At a tea party, a lady claims she can discriminate by tasting a cup of tea whether the milk (M) or the tea (T) was poured first. To verify her claim, a statistician presents her in random order with 8 cups of tea, and tells her there are 4 of each type M and T. The lady correctly identifies the four M types (and thus also correctly identifies the four T types). How can the statistician verify if her claim is true?

The inference: Fisher's take

- 8 cups of tea, with 4 cups tea-first (T) and 4 milk-first (M).
- The lady is given 4 cups randomly drawn from the 8 cups.
- The lady tastes the 4 cups.
- Then, we have the lady's answer: how many cups are T's.

The inference: Fisher's take

Null Hypothesis (H_0): The lady does not know and guesses at random[4].
We can then calculate the probabilities of the outcome under H_0!

- Suppose all 2 cups are T's and 2 cups are M's.
- The lady's answer: 1 T.

$$P(\text{Obs} \mid H_0) = \frac{\binom{4}{2} \times \binom{4}{2}}{\binom{8}{4} \times \binom{5}{1}} = 0.10$$

The value $P(\text{Obs} \mid H_0)$ is called **Fisher's exact p-value**.

[4] An implication form H_0 is that given any (random) selection of 4 cups (out of 8 cups), the "random" lady always takes one of 5 guesses (i.e., 0 T's, 1 T's, ..., 4 T's) with equal probabilities.

Counterfactual in Rubin Causal Model (RCM)

- **The fundamental problem in causal inference.**
- **Solution**: to avoid the individual treatment effect as it is unknowable.

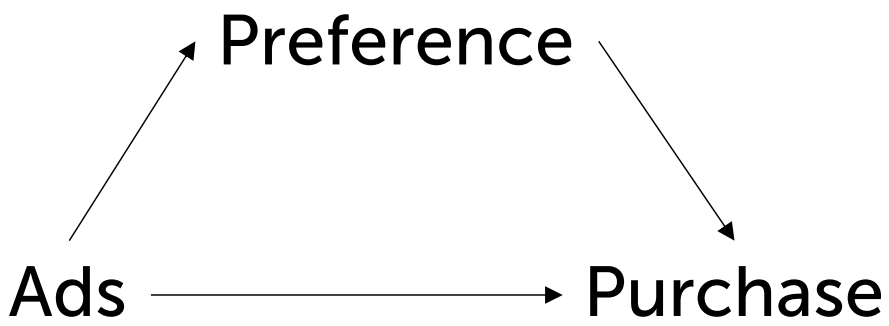
Subject	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Joe	130	?	?
Mary	?	125	?
Sally	100	?	?
Bill	?	130	?
James	?	120	?
Mean	**115**	**125**	**-10**

Counterfactual in Pearlian Causal Model (PCM)

Pre-conditions of counterfactual prediction

1. We have full knowledge of the DAG.
2. We have observations of (some) variables of the DAG.
3. We know the causal effect is identified given data (i.e., do-calculus).

Counterfactual in PCM: An example



The structural equations of the DAG are (E stochastic terms or unmeasured ancestors):

$$\begin{cases} \text{Purchase} &= f\left(\text{Ads}, \text{Preference}, E_{\text{Purchase}}\right) \\ \text{Preference} &= g\left(\text{Ads}, E_{\text{Preference}}\right) \end{cases}$$

Counterfactual in PCM: An example

The additional assumption of "**separability**":

$$\begin{cases} \text{Purchase} &= f\left(\text{Ads}, \text{Preference}\right) + E_{\text{Purchase}} \\ \text{Preference} &= g\left(\text{Ads}\right) + E_{\text{Preference}} \end{cases}$$

Prediction of counterfactual

Suppose $\text{Ads} = \{0, 1\}$ and a consumer i is exposed $\text{Ads}_i = 1$, and we want the counterfactual Purchase_i^0

- $\hat{E}^i_{\text{Preference}} = \text{Preference}_i^1 - \hat{g}\left(\text{Ads}_i = 1\right)$.
- $\hat{E}^i_{\text{Purchase}} = \text{Purchase}_i^1 - \hat{f}\left(\text{Ads}_i = 1, \text{Preference}_i^1\right)$.
- $\text{Purchase}_i^0 = \hat{f}\left(\text{Ads}_i = 1, \hat{g}\left(\text{Ads}_i = 0\right) + \hat{E}^i_{\text{Preference}}\right) + \hat{E}^i_{\text{Purchase}}$.

Counterfactual in PCM: general steps[5]

General steps for predicting counterfactuals in PCM

1. *Abduction*
   - Use an observation to determine the value of error terms.
2. *Action*
   - Modify the structural equations, by replacing the cause variable X with $X := x$.
3. *Prediction*:
   - Use the value of errors in Step 1 and the modified equations from Step 2 to compute the value of $Y(X = x)$.

[5] There is also a probabilistic version of the general steps. For more details in the prediction procedure, please see Chapter 4 of Pearl,Glymour and Jewell (2016).

Reflections on the PCM view of counterfactual

Two assumptions are essential to get individual-level counterfactual:

- Separability of errors from the observed variables.
- Errors (or their distributions) are invariant as to the cause variables (treatments).

Similar assumptions are also made in other areas:

- E.g., structural models in economics.
- Knowing the DAG $\approx$ Applying theories in structural models.

Counterfactuals: In summary

- The idea of counterfactuals is pivotal for both models.
- The main difference is more in methodologies.
  - PCM: designed to solve problems in AI reasoning.
  - RCM: designed to solve problems in medical research.

References I

- Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Hachette UK.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- Fisher, R. (1926). The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain, 33, 503-513.
- Rubin, D. (2005), Basic Concepts in Causal Inference, unpublished manuscript, Link: `www.stat.columbia.edu/~cook/qr33.pdf`.
- Roy, A. D. (1951). Some Thoughts On The Distribution Of Earnings. Oxford Economic Papers, 3(2), 135-146.f

References II

 Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology, 66(5), 688-701.