UNIVERSITAT DE BARCELONA

CAUSAL INFERENCE AND MACHINE LEARNING

About the course

01 **Introduction**
Observational and Interventional Distributions

02 **Potential Outcomes**
Fundamental Problem of Causal Inference

03 **Causal Graphs**
Do Calculus

04 **Estimand-based Estimation**
Metalearners

05 **Estimand-agnostic Estimation**
Counterfactuals

06 **Causal Machine Learning**
Supervised and Reinforcement Learning

07 **Practical Causal Inference**
Exercises

The relationship between causality and artificial intelligence can be seen from two points of view: how causality can help solve some of the current problems of AI and how causal inference can leverage machine learning techniques. In this course we will review the two points of view with special emphasis on examples and practical cases.

02 - Potential Outcomes

The Fundamental Problem of Causal Inference

UNIVERSITAT DE BARCELONA

In this section we'll see

- → The Potential Outcomes (PO) framework intuition and definition
- → The fundamental problem of Causal Inference
- → How PO can be used to get around the fundamental problem:
    - → Naive case
    - → Randomised Controlled Trials
    - → Unconfoundedness
- → A real example from the PO perspective

Initial remarks

→ The goal in causal inference is to assess the causal effect of some **potential cause** (e.g. an institution, intervention, policy, or event) on some **outcome.**

Initial remarks

→ The goal in causal inference is to assess the causal effect of some **potential cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

→ The **Potential Outcomes** framework is a way of thinking about causation and how to measure it

Initial remarks

→ The goal in causal inference is to assess the causal effect of some **potential cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

→ The **Potential Outcomes** framework is a way of thinking about causation and how to measure it

→ Even though it is a different perspective, it is **equivalent with Pearl's graphical models**, in the sense that they reach the same conclusions

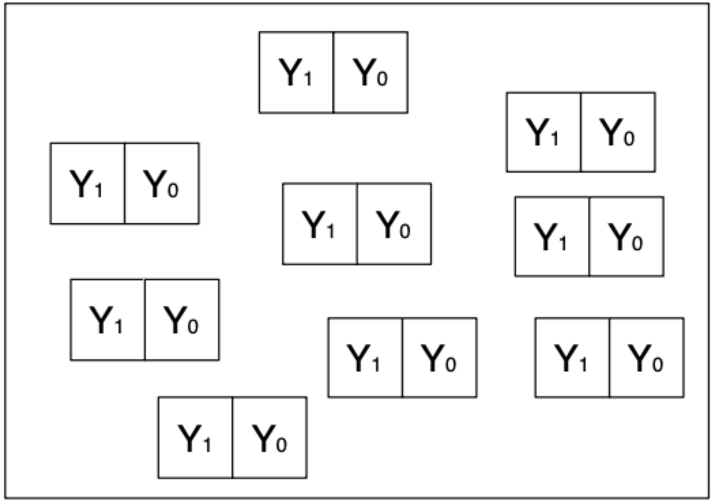
→ It is common in the literature to only use one of the perspectives, so it can be a bit confusing at first.

Potential Outcomes intuition and notation

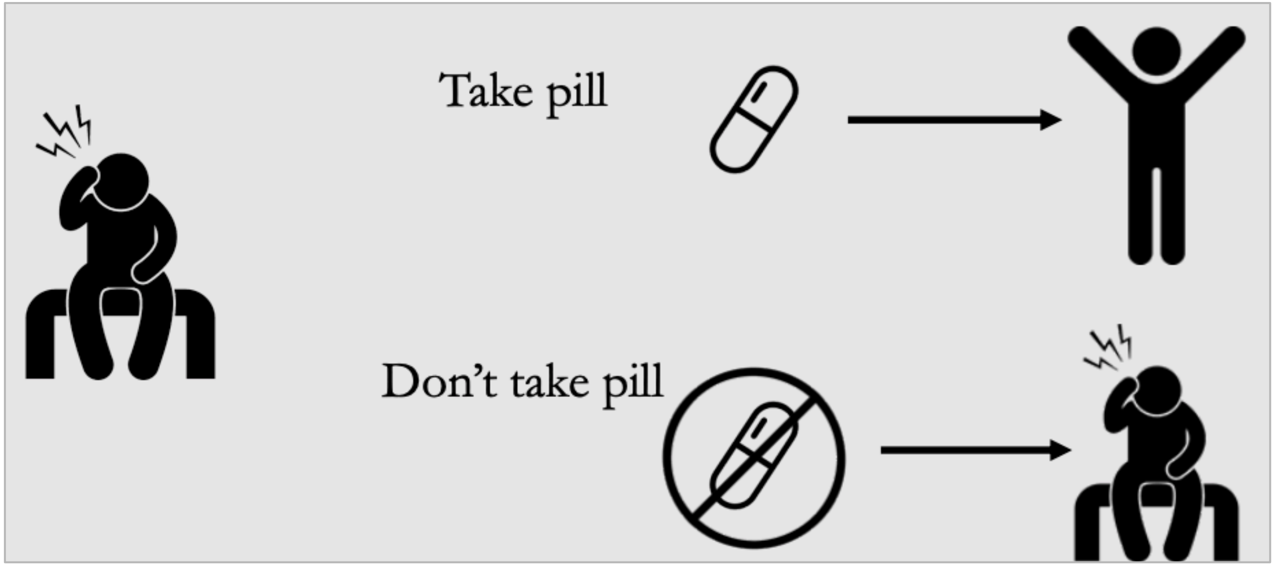
The Potential Outcomes Framework

→ It is based on the idea of potential outcomes, which are **the outcomes that would have happened** under different scenarios or interventions.

→ The causal effect is the difference between the two potential outcomes

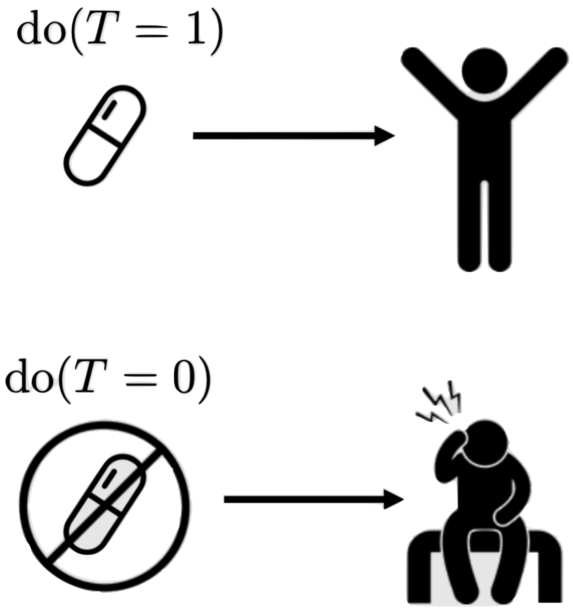


Framework intuition and notation



https://www.bradyneal.com/causal-inference-course

Framework intuition and notation



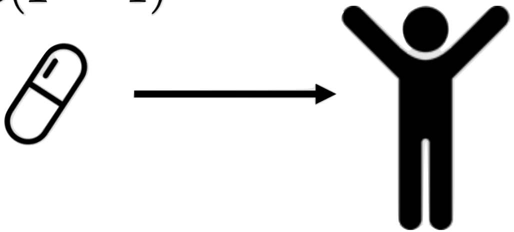
T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

https://www.bradyneal.com/causal-inference-course

Framework intuition and notation

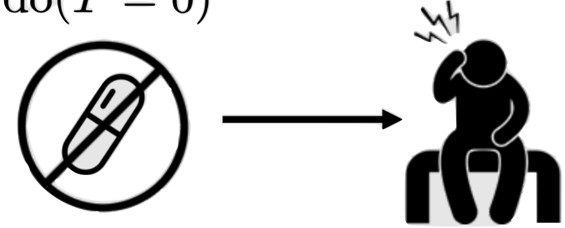
$$Y_i|_{\mathrm{do}(T=1)} \triangleq Y_i(1)$$

$\mathrm{do}(T=1)$



$$Y_i|_{\mathrm{do}(T=0)} \triangleq Y_i(0)$$

$\mathrm{do}(T=0)$



T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Causal effect

$$Y_i(1) - Y_i(0)$$

https://www.bradyneal.com/causal-inference-course

The fundamental problem of Causal Inference

The fundamental problem of Causal Inference

Dog example

Imagine you want to know if your happiness will increase by getting a dog.

The fundamental problem of Causal Inference

Dog example

Imagine you want to know if your happiness will increase by getting a dog.

→ You could observe $Y(1)$ by getting a dog and observing your happiness after getting a dog.

→ Alternatively, you could observe $Y(0)$ by not getting a dog and observing your happiness.

The fundamental problem of Causal Inference

Dog example

Imagine you want to know if your happiness will increase by getting a dog.

→ You could observe Y(1) by getting a dog and observing your happiness after getting a dog.

→ Alternatively, you could observe Y(0) by not getting a dog and observing your happiness.

However, you cannot observe both Y(1) and Y(0), unless you have a time machine

Missing data interpretation

i	T	Y	Y(1)	Y(0)	Y(1) - Y(0)
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?
7	0	1	?	1	?
8	1	1	1	?	?

Causal Inference is difficult because it involves missing data

Association is not causation

Back to clickthrough example

	Ad 0	Ad 1
Male	108/120 (90%)	340/400 (85%)
Female	266/380 (70%)	65/100 (65%)
Total	374/500 (75%)	405/500 (81%)

Table 1. Clickthroughs in the AdBot setting striated by the ad shown to participants in a focus group, and the sex of the viewer.

In this example, the gender is a confounder that makes it impossible to measure the causal effect if we don't take it into account

Association is not causation

In general, the answer to

$$\mathbb{E}[Y|do(X=1)] - E[Y|do(X \neq 1)]$$

is not

$$\mathbb{E}[Y|X=1] - E[Y|X \neq 1]$$

Association is not causation

In general, the answer to

$$\mathbb{E}[Y|do(X=1)] - E[Y|do(X \neq 1)]$$

is not

$$\mathbb{E}[Y|X=1] - E[Y|X \neq 1]$$

We need to add business / expert knowledge to interpret associative relationships in the data as causal relationships

Getting around the fundamental problem

Getting around the fundamental problem - assumptions

→ The Potential Outcomes framework uses **expert knowledge encoded as assumptions** to get around the fundamental problem

Getting around the fundamental problem - assumptions

→ The Potential Outcomes framework uses **expert knowledge encoded as assumptions** to get around the fundamental problem

→ These assumptions are **properties about the data** that might or might not be directly observable and allow us to transform **associational relations to causal relations**.

Before we start - Identifiability

- → The goal of causal inference is to estimate $P(Y \mid do(T = t))$
- → $P(Y \mid do(T = t))$ requires an intervention, so it's not directly observable from the data
- → The Potential Outcomes framework gives us tools to estimate $P(Y \mid do(T = t))$ in terms of the observed data by deriving an estimand using assumption

Before we start - Identifiability

- The goal of causal inference is to estimate $P(Y \mid do(T = t))$
- $P(Y \mid do(T = t))$ requires an intervention, so it's not directly observable from the data
- The Potential Outcomes framework gives us tools to estimate $P(Y \mid do(T = t))$ in terms of the observed data by deriving an estimand using assumption

Identification is the process of reducing a causal question to a purely statistical expression

Before we start - Basic assumptions for CI

The following three assumptions are required any time we are doing Causal Inference under the Potential Outcomes framework:

→ No interference

→ Consistency

→ Positivity

These assumptions are known by several names, a common name is **Stable Unit Treatment Value Assumption (SUTVA)**, which is a combination of no interference and consistency assumptions

Basic assumptions for CI - No interference

No interference means that my outcome is unaffected by anyone else's treatment

No interference: $Y_i(t_1, \ldots, t_{i-1}, t_i, t_{i+1}, \ldots, t_n) = Y_i(t_i)$

Intuition: In the dog example (where the treatment is getting or not a dog and the outcome is happiness), no interference could be violated if, for example, the happiness of getting a dog depends on wether or not your friends get dogs too so they can play together

Basic assumptions for CI - Consitency

The outcome we observe Y is actually the potential outcome under the observed treatment T

$$\textbf{Consistency: } T = t \implies Y = Y(t)$$

Intuition: In the dog example (where the treatment is getting or not a dog and the outcome is happiness), consistency could be violated if, for example, getting a dog is not specific enough and the outcome could vary depending on whether the dog is a puppy or an old dog.

Basic assumptions for CI - Positivity

Positivity: $0 < P(T = 1) | X = x) < 1$

Intuition: Positivity is required when there are subgroups of the data with different covariates X. Positivity is the condition that all such subgroups must have some probability of receiving any value of the treatment

Positivity - unconfoundedness tradeoff: It is common to be interested in the causal effects of small, specific subgroups of the population. As the subgroup gets smaller, there is a higher chance that the whole subgroup is assigned with the same treatment and thus violate the positivity assumption (Ex. In a subgroup of a single sample positivity is guaranteed to not hold)

Getting around the fundamental problem - Naive Case

Assumptions - Ignorability: Naive Case

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

- Assuming ignorability is ignoring how subjects ended up selecting a treatment
- Formally this means: $\mathbb{E}[Y(1) \mid T = 0] = \mathbb{E}[Y(1) \mid T = 1]$ and $\mathbb{E}[Y(0) \mid T = 0] = \mathbb{E}[Y(0) \mid T = 1]$
- Ignorability assumption allows us to identify $P(Y \mid do(T = t))$ as $P(Y \mid T = t)$ but ...

Assumptions - Ignorability: Naive Case

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

How realistic of an assumption is it?

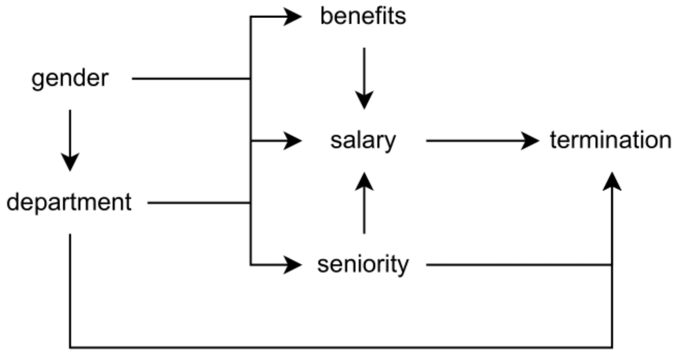
In general, it is completely unrealistic because there is likely to be confounding in most data we observe

Getting around the fundamental problem - Randomised Controlled Trials (RTC)

Graphical perspective intuition (More on this later)

A causal graph is a directed acyclic graph (DAG) that **represents the causal relationships** between variables in a system. We refer to ”causal graph” as a DAG that satisfies the causal edges assumptions, i.e. that **all parents are causes of their children**.

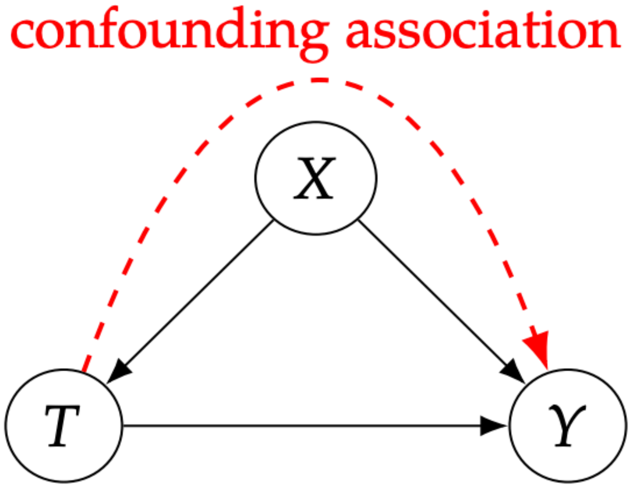
We use the following notation:
- T denotes the treatment variable
- Y denotes the outcome variable
- X denotes the vector of covariates
- Pa(i) denotes parents of the node i
- Ch(i) denotes children of the node i



Causal Graph example

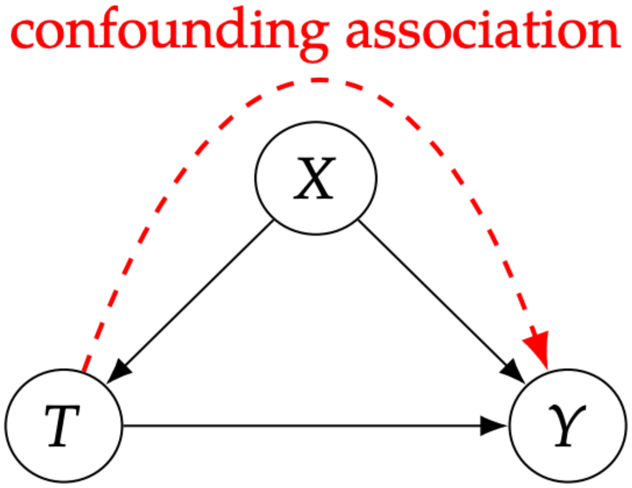
Getting around the fundamental problem - Case 2 RTC

- The graph on the right is a common setting where there are confounders affecting both treatment and outcome
- In this case the ignorability assumption does not hold, so we can't compute directly causal effects



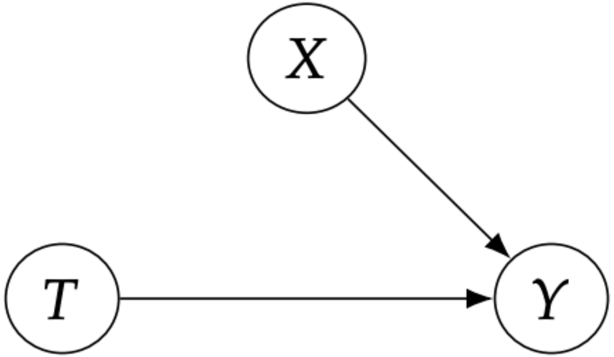
Getting around the fundamental problem - Case 2 RTC

- →The graph on the right is a common setting where there are confounders affecting both treatment and outcome
- →In this case the ignorability assumption does not hold, so we can't compute directly causal effects
- →**Running a RCT is a solution to obtain ignorability in this case**



Getting around the fundamental problem - Case 2 RTC

- By randomising the treatment we remove the arrow from X to T, making the treatment independent from the rest of confounders
- In a RCT, **association is causation**

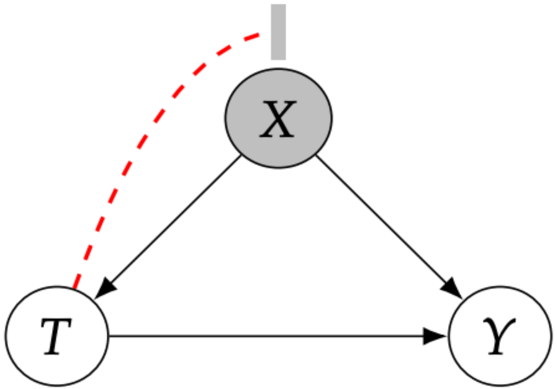


Getting around the fundamental problem - Unconfoundedness

Assumptions - Conditional ignorability / Unconfoundedness

Unconfoundedness: $(Y(1), Y(0)) \perp T \mid X$

Intuition: In observational data it is unrealistic to assume that treatment groups are the same in **all relevant variables (X)** other than the treatment. But if we control by those variables, then the resulting subrgroups may be exchangeable



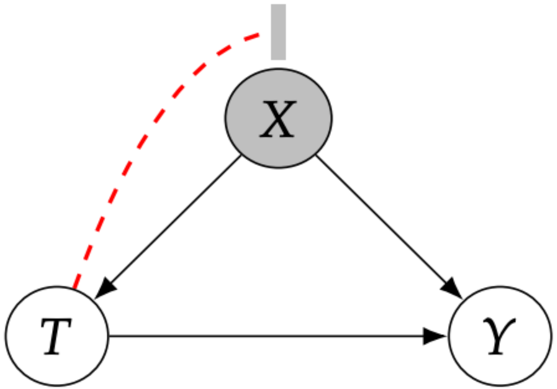
Conditioning on X blocks information that flows indirectly from T to Y

Assumptions - Conditional ignorability / Unconfoundedness

Unconfoundedness: $(Y(1), Y(0)) \perp T \mid X$

Intuition: In observational data it is unrealistic to assume that treatment groups are the same in **all relevant variables (X)** other than the treatment. But if we control by those variables, then the resulting subrgroups may be exchangeable

But how can we know which are all relevant variables?



Conditioning on X blocks information that flows indirectly from T to Y

The adjustment formula

In the Potential Outcomes framework,
selecting X is a matter of expert knowledge

$$\boxed{\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}_X[\mathbb{E}[Y \mid Y = 1, X] - \mathbb{E}_X[\mathbb{E}[Y \mid Y = 0, X]]}$$

But …

Beyond PO - Backdoor adjustment

→ In the next section we'll see the causal graphs perspective on Causal Inference.

→ This perspective allows for a more defined set of properties of the variables X to be used in the adjustment formula

Backdoor Criterion: A set of variables X satisfies the backdoor criterion relative to T and Y if the following are True:

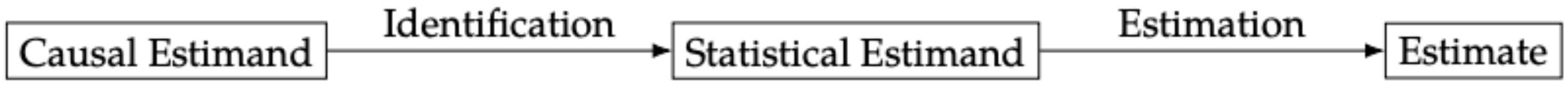
1. X blocks all backdoor paths from T to Y
2. X does not contain any descendants of T

Potential Outcomes closing remarks

Potential Outcomes summary

→ The Potential Outcomes framework uses assumptions to interpret associational relations in the data as causal relations

→ The goal the framework is to go from a causal question to a **statistical estimand** that uses only observational data



Causal Inference Workflow under the Potential Outcomes framework

Untestable assumptions - Expert knowledge

There are <u>two ways</u> to asses the validity of assumptions:

1. Run statistical tests on the data
2. Have external expert knowledge on the mechanisms underlying the data generation

But..

Untestable assumptions - Expert knowledge

There are <u>two ways</u> to asses the validity of assumptions:

1. Run statistical tests on the data
2. Have external expert knowledge on the mechanisms underlying the data generation

But..

Some assumptions, such as the causal graph or Unconfoundedness are **untestable.** In Causal Inference, expert knowledge about the data is extremely important

Causality example: Google Pay Study Case

https://www.npr.org/2019/03/05/700288695/google-pay-study-finds-its-underpaying-men-for-some-jobs

> When Google conducted its annual pay equity analysis for 2018, the tech company found something nobody expected: It was underpaying men for doing similar work as women.

→ This case was controversial because Google was facing a class action lawsuit filed by women who allege systemic underpayment

→ The controversy can be tracked up to ambiguous conclusions that depend on data not available or ethical reasons, but ..

→ **Causality can help us obtain a clearer picture by giving us the appropriate tools to reason about the situation**

Google Case - Summary

Since 2012, Google has conducted a yearly companywide analysis to ensure pay is "equitable across gender and racial lines," Barbato said. She offered an

How we run our pay equity analysis at Google

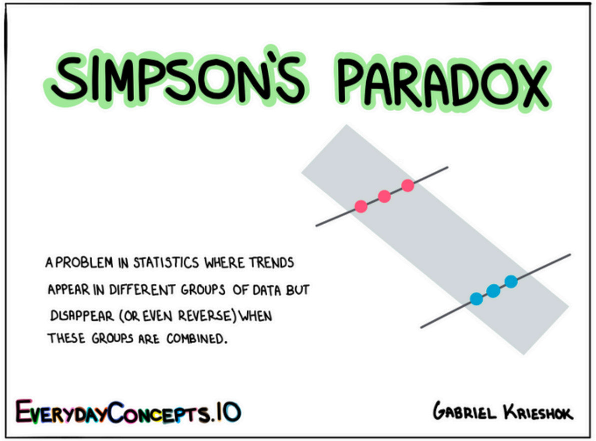
To ensure we can produce results that translate to meaningful action, we run our analyses at the job code level, adjusting for job function and level. Here's how it works:

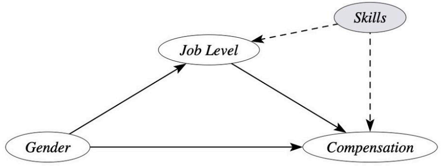
- At the end of our annual compensation planning process (for salary, bonus, and equity) we ran rigorous statistical analyses to check the outcomes before any amounts were final. We conducted separate ordinary least squares (OLS) regressions to check for pay equity in each job group—a job group is made up of job family (like Software Engineer) and level (like Level 4).
- The OLS method allows us to account for factors that should influence pay (e.g., tenure, location, performance ratings) and look for unexplained differences in total compensation (salary, bonus, and equity) across demographic groups. Specifically, we looked for pay differences based on gender (for which we have information worldwide) and, in the U.S., by race/ethnicity.
- Our analyses covered every job group with at least 30 Googlers total and at least five Googlers per demographic group for which we have data (e.g., at least five men and at least five women). These n-count minimums ensure statistical rigor (e.g., higher statistical power, narrower confidence intervals)

Google Case - A causality based critique by Paul Hünermund

https://twitter.com/PHuenermund/status/1540262891890278402?s=20

- The key point is that they are controlling by Job Level
- It is known from the literature that variables such as job level can be impacted by Gender
- This makes Job level a descendant of the treatment (Gender) and thus controlling for it can introduce bias to the computation
- In the PO perspective, they are assuming ignorability but there are reasons to think it does not hold
- Depending on the data, it could even mean they are increasing discrimination by performing these corrections





Google Case - How can opposite conclusions arise from the same analysis?

→ In the original approach there is an implicit assumption that ignorability holds in the local level of the job-level groups, and that the goal of the corrections is to fix unexplained discrimination at this local level

→ Paul Hünermund’s critique affects at the global level of the company

→ **Causality is a powerful tool to make reasoning in this kind of scenarios and understand its implications in both perspectives**

Pay equity analyses. The Compensation and People Analytics teams conduct pay equity analyses to identify any unexplained differences between groups of Googlers who are doing the same job. We do these analyses before pay changes for the following year are finalized, and where differences are observed, action is taken.