3. Consistency

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

30 Aug 2022

Responding to feedback

Things that are good

- Group work
- Moving around with new people
- Clarifying what we learned at the end

Things that could be better

- Zoom audio quality
- Some things are a little fast

How was Problem Set 1?

Learning goals for today

At the end of class, you will be able to:

1. Reason about when one unit's treatment affects another unit's outcome (interference)
2. Reason about treatments that hide distinct versions
3. Formalize the assumption of treatment variation irrelevance

Hernán (2016): Does Water Kill?

London cholera epidemic, 1854.
John Snow deduced that the water was the cause of death.

Source: Wikimedia Commons

Hernán (2016): Does Water Kill?

Does drinking water kill?

Hernán (2016): Does Water Kill?

Does drinking **fresh** water kill?

Hernán (2016): Does Water Kill?

Does drinking **a swig of** fresh water kill?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water **from the Broad Street pump** kill?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump **between August 31 and September 10** kill?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump between August 31 and September 10 kill

compared with drinking all your water from other pumps?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump between August 31 and September 10 **and not initiating a rehydration treatment if diarrhea starts** kill

compared with drinking all your water from other pumps?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump between August 31 and September 10 and not initiating a rehydration treatment if diarrhea starts kill

compared with drinking all your water from other pumps?

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump between August 31 and September 10 and not initiating a rehydration treatment if diarrhea starts kill

compared with drinking all your water from other pumps?

The **definition** of the causal effect is unclear without details.

Hernán (2016): Does Water Kill?

Does drinking a swig of fresh water from the Broad Street pump between August 31 and September 10 and not initiating a rehydration treatment if diarrhea starts kill

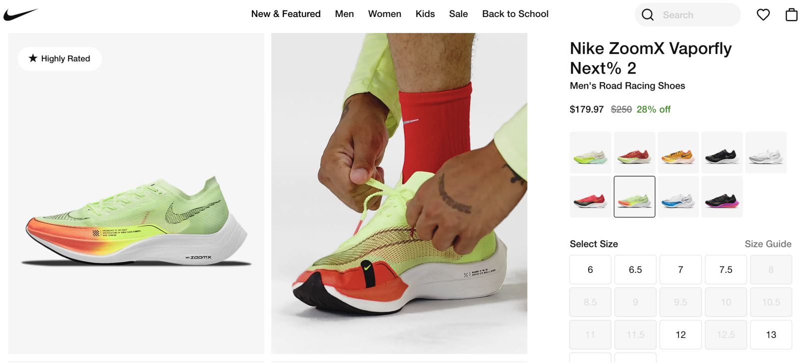
compared with drinking all your water from other pumps?

The **definition** of the causal effect is unclear without details.

Recommendation: Specify versions
"until no meaningful vagueness remains,"
(Hernan 2016)

Interference

Interference Example 1: Springy Shoes



[1]

[1]Image source: Nike

Interference Example 1: Springy Shoes

- The Nike Vaporfly is springy. These shoes make you faster.

Interference Example 1: Springy Shoes

- The Nike Vaporfly is springy. These shoes make you faster.
- Suppose two closely-matched people run a race.

Interference Example 1: Springy Shoes

- The Nike Vaporfly is springy. These shoes make you faster.
- Suppose two closely-matched people run a race.
- Would the Nike Vaporfly affect the outcome?

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2.
We will focus on the outcome for person 1.

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2.
We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: ${Y_1}^{TT} = \text{Win}$

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2. We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: $Y_1^{TT} = \text{Win}$
- If neither wears them, the faster one wins: $Y_1^{FF} = \text{Win}$

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2. We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: $Y_1^{TT} = \text{Win}$
- If neither wears them, the faster one wins: $Y_1^{FF} = \text{Win}$
- If only the faster person wears them, the faster one wins $Y_1^{TF} = \text{Win}$

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2. We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: $Y_1^{TT} = \text{Win}$
- If neither wears them, the faster one wins: $Y_1^{FF} = \text{Win}$
- If only the faster person wears them, the faster one wins $Y_1^{TF} = \text{Win}$
- If only the slower person wears them, the faster one loses: $Y_1^{FT} = \text{Lose}$

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2. We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: $Y_1^{TT} = \text{Win}$
- If neither wears them, the faster one wins: $Y_1^{FF} = \text{Win}$
- If only the faster person wears them, the faster one wins $Y_1^{TF} = \text{Win}$
- If only the slower person wears them, the faster one loses: $Y_1^{FT} = \text{Lose}$

Interference Example 1: Springy Shoes

Let the faster person be Person 1. Let the slower be Person 2. We will focus on the outcome for person 1.

- If they both wear them, the faster one wins: $Y_1^{TT} = \text{Win}$
- If neither wears them, the faster one wins: $Y_1^{FF} = \text{Win}$
- If only the faster person wears them, the faster one wins $Y_1^{TF} = \text{Win}$
- If only the slower person wears them, the faster one loses: $Y_1^{FT} = \text{Lose}$

Potential outcomes depend on the treatments of both units (sometimes termed **interference**)

Interference

With springy shoes, each potential outcome depended on two people's treatments?

Interference

With springy shoes, each potential outcome depended on two people's treatments?

What if each potential outcome depends on the whole population's treatments?

Interference Example 2: College

Why does college lead to higher pay?

Interference Example 2: College

Why does college lead to higher pay?

Human Capital Story: It makes you more productive.

Interference Example 2: College

Why does college lead to higher pay?

Human Capital Story: It makes you more productive.

Sorting Story: It helps you jump the line to a better job

Interference Example 2: College: Human Capital Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Interference Example 2: College: Human Capital Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

Interference Example 2: College: Human Capital Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

In words: Unit i's outcome is the same under any population-level treatment assignment vector $\vec{a}$ where unit i is assigned to a.

Interference Example 2: College: Human Capital Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

In words: Unit i's outcome is the same under any population-level treatment assignment vector $\vec{a}$ where unit i is assigned to a.

$$Y_i^{\vec{a}} = Y_i^{a} \qquad \text{for all} \quad \vec{a} \quad \text{such that} \quad a_i = a$$

Interference Example 2: College: Human Capital Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

In words: Unit i's outcome is the same under any population-level treatment assignment vector $\vec{a}$ where unit i is assigned to a.

$$Y_i^{\vec{a}} = Y_i^a \qquad \text{for all} \quad \vec{a} \quad \text{such that} \quad a_i = a$$

In words again: Unit i's outcome depends only on unit i's treatment

Interference Example 2: College: Sorting Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

Interference Example 2: College: Sorting Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

Maybe unit i's outcome is only a function of:

- Whether unit i has a degree
- The proportion of the population with degrees

Interference Example 2: College: Sorting Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

Maybe unit i's outcome is only a function of:

- Whether unit i has a degree
- The proportion of the population with degrees

$$Y_i^{\vec{a}} = Y_i^{a,\pi}$$

Interference Example 2: College: Sorting Story

Let $Y_i^{\vec{a}}$ be unit i's potential outcome
if the population of treatments are set to $\vec{a}$

Can we simplify that potential outcome?

Maybe unit i's outcome is only a function of:

- Whether unit i has a degree
- The proportion of the population with degrees

$$Y_i^{\vec{a}} = Y_i^{a,\pi} \qquad \text{for all} \quad \vec{a} \quad \text{such that} \quad a_i = a \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^{n} a_j = \pi$$

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.
- In this case, we should really talk about the outcome $Y_i^{\vec{a}}$ when the population of treatments is set to a vector $\vec{a}$

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.
- In this case, we should really talk about the outcome $Y_i^{\vec{a}}$ when the population of treatments is set to a vector $\vec{a}$
- But that is unmanageably huge. Often we make simplifying assumptions.

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.
- In this case, we should really talk about the outcome $Y_i^{\vec{a}}$ when the population of treatments is set to a vector $\vec{a}$
- But that is unmanageably huge. Often we make simplifying assumptions.
- Most common assumption: Unit i's outcome depends only on unit i's treatment
$$Y_i^{\vec{a}} = Y_i^a$$

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.
- In this case, we should really talk about the outcome $Y_i^{\vec{a}}$ when the population of treatments is set to a vector $\vec{a}$
- But that is unmanageably huge. Often we make simplifying assumptions.
- Most common assumption: Unit i's outcome depends only on unit i's treatment
$$Y_i^{\vec{a}} = Y_i^{a}$$

Interference: Conclusion

- Often, each unit's response to treatment depends on the treatments of others.
- In this case, we should really talk about the outcome $Y_i^{\vec{a}}$ when the population of treatments is set to a vector $\vec{a}$
- But that is unmanageably huge. Often we make simplifying assumptions.
- Most common assumption: Unit i's outcome depends only on unit i's treatment
$$Y_i^{\vec{a}} = Y_i^{a}$$

Part of the consistency assumption.
Sometimes called the Stable Unit Treatment Value Assumption (SUTVA)

Multiple versions of treatment

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:
  - Pay us lots of money

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Question: What is misleading about the ad?

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Question: What is misleading about the ad?

Earnings don't just go up from any degree.
Earnings go up from a **Cornell** degree.

Multiple versions: Example

Consider the following scenario:

- A researcher at Cornell provides evidence that a Cornell degree causes higher earnings, compared to no college degree
- A business person at a for-profit college see the study.
  They make an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Question: What is misleading about the ad?

Earnings don't just go up from any degree.
Earnings go up from a **Cornell** degree.

Then how should we define the treatment?

Multiple versions: Example

Let A_i be a detailed treatment. Values of A_i include

- No BA
- Cornell BA
- UPenn BA
- ITT Tech BA
- (many more)

Multiple versions: Example

Let A_i be a detailed treatment. Values of A_i include

- No BA
- Cornell BA
- UPenn BA
- ITT Tech BA
- (many more)

Let D_i be a dichotomized treatment. Values of D_i include

- No BA
- Any BA

Multiple versions: Example

Let A_i be a detailed treatment. Values of A_i include

- No BA
- Cornell BA
- UPenn BA
- ITT Tech BA
- (many more)

Let D_i be a dichotomized treatment. Values of D_i include

- No BA
- Any BA

A researcher might write $Y_i^{\text{Any BA}}$

Multiple versions: Example

Let A_i be a detailed treatment. Values of A_i include

- No BA
- Cornell BA
- UPenn BA
- ITT Tech BA
- (many more)

Let D_i be a dichotomized treatment. Values of D_i include

- No BA
- Any BA

A researcher might write $Y_i^{\text{Any BA}}$
OK if all BAs have equal effects

Multiple versions: Example

Let A_i be a detailed treatment. Values of A_i include

- No BA
- Cornell BA
- UPenn BA
- ITT Tech BA
- (many more)

Let D_i be a dichotomized treatment. Values of D_i include

- No BA
- Any BA

A researcher might write $Y_i^{\text{Any BA}}$
OK if all BAs have equal effects
Otherwise, this definition is problematic

Treatment variation irrelevance (Vanderweele 2009)[2]

Within the analyzed treatments,
all variations of the treatment
have the same effect

[2] VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. Epidemiology, 20(6), 880-883.

Treatment variation irrelevance (Vanderweele 2009)[2]

Within the analyzed treatments,
all variations of the treatment
have the same effect

In Vanderweele's notation,

- x is the treatment
- k_x and k'_x are versions of the treatment x
- Potential outcomes are in parentheses

$$Y_i(x, k_x) = Y_i(x, k'_x) \quad \text{for all} \quad k_x, k'_x$$

[2]VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. Epidemiology, 20(6), 880-883.

Multiple versions: What can go wrong

Consider two degrees: **Cornell** and **ITT Tech**

Multiple versions: What can go wrong

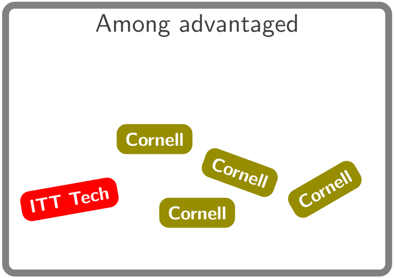
Consider two degrees: **Cornell** and **ITT Tech**

Suppose these are distributed unequally across subgroups.

Multiple versions: What can go wrong

Consider two degrees: **Cornell** and **ITT Tech**

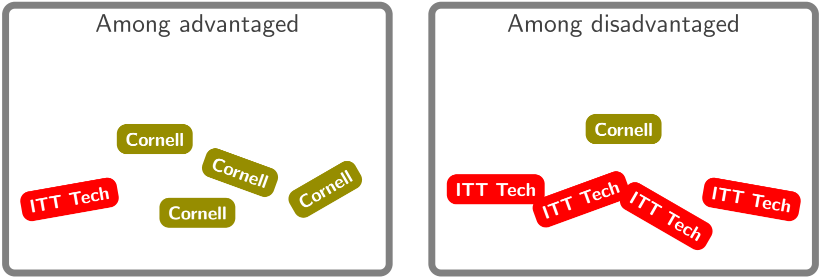
Suppose these are distributed unequally across subgroups.



Multiple versions: What can go wrong

Consider two degrees: **Cornell** and **ITT Tech**

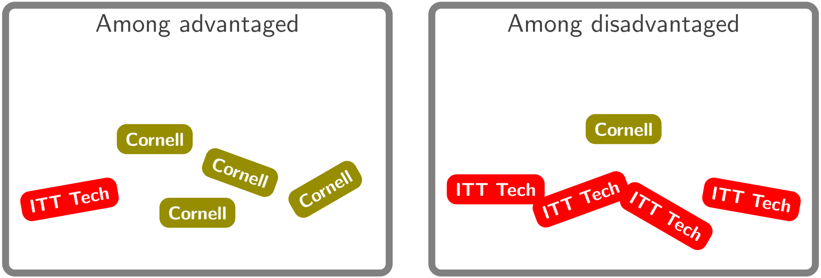
Suppose these are distributed unequally across subgroups.



Multiple versions: What can go wrong

Consider two degrees: **Cornell** and **ITT Tech**

Suppose these are distributed unequally across subgroups.

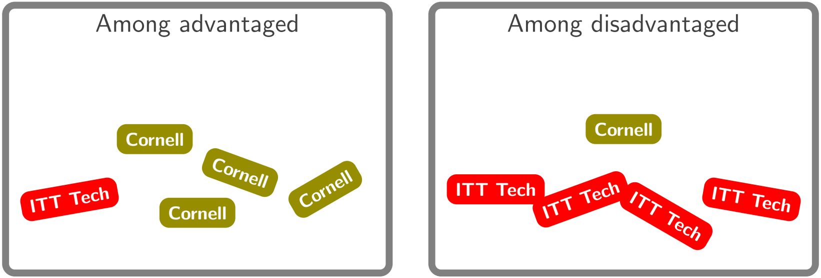


Mistaken claim: College is more helpful to the advantaged.

Multiple versions: What can go wrong

Consider two degrees: **Cornell** and **ITT Tech**

Suppose these are distributed unequally across subgroups.



Mistaken claim: College is more helpful to the advantaged.
Better claim: The advantaged disproportionately take the effective treatment.

Multiple versions: A few resolutions

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)
2. Measure very detailed treatments

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)
2. Measure very detailed treatments
   - Case study of Cornell degrees rather than degrees writ large

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)
2. Measure very detailed treatments
   - Case study of Cornell degrees rather than degrees writ large
3. Study an aggregation of potential outcomes

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)
2. Measure very detailed treatments
   - Case study of Cornell degrees rather than degrees writ large
3. Study an aggregation of potential outcomes
   - Effect of a college degree, averaged over the marginal distribution of colleges

Multiple versions: A few resolutions

1. Assume treatment variation irrelevance
   - If two distinct treatments (e.g., Cornell BA and Yale BA) map to the same treatment value in the data (BA), then assume they have the same effect.
   - More plausible in some settings (Cornell and Yale) than others (Cornell and ITT Tech)
2. Measure very detailed treatments
   - Case study of Cornell degrees rather than degrees writ large
3. Study an aggregation of potential outcomes
   - Effect of a college degree, averaged over the marginal distribution of colleges
   - Requires great care to specify the estimand here

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

- Interference: Maybe the treatment is a population vector
  Example: $Y_i^{100001110}$, perhaps denoted $Y_i^{\vec{a}}$

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

- Interference: Maybe the treatment is a population vector
  Example: $Y_i^{100001110}$, perhaps denoted $Y_i^{\vec{a}}$
- Multiple versions: Maybe the treatment is a non-binary scalar
  Example: Y_i^{Cornell} Y_i^{Yale} $Y_i^{\text{Princeton}}$ $Y_i^{\text{ITT Tech}}$

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

- Interference: Maybe the treatment is a population vector
  Example: $Y_i^{100001110}$, perhaps denoted $Y_i^{\vec{a}}$
- Multiple versions: Maybe the treatment is a non-binary scalar
  Example: Y_i^{Cornell} Y_i^{Yale} $Y_i^{\text{Princeton}}$ $Y_i^{\text{ITT Tech}}$

In each case, there are many more potential outcomes.

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

- Interference: Maybe the treatment is a population vector
  Example: $Y_i^{100001110}$, perhaps denoted $Y_i^{\vec{a}}$
- Multiple versions: Maybe the treatment is a non-binary scalar
  Example: Y_i^{Cornell} Y_i^{Yale} $Y_i^{\text{Princeton}}$ $Y_i^{\text{ITT Tech}}$

In each case, there are many more potential outcomes.

To collapse the potential outcomes requires theoretical argument.

Closing comparison

It is nice to have two potential outcomes: Y_i^1 and Y_i^0

But often the world is not so simple:

- Interference: Maybe the treatment is a population vector
  Example: $Y_i^{100001110}$, perhaps denoted $Y_i^{\vec{a}}$
- Multiple versions: Maybe the treatment is a non-binary scalar
  Example: Y_i^{Cornell} Y_i^{Yale} $Y_i^{\text{Princeton}}$ $Y_i^{\text{ITT Tech}}$

In each case, there are many more potential outcomes.

To collapse the potential outcomes requires theoretical argument.

More research on these questions!

Learning goals for today

At the end of class, you will be able to:

1. Reason about when one unit's treatment affects another unit's outcome (interference)
2. Reason about treatments that hide distinct versions
3. Formalize the assumption of treatment variation irrelevance

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!