Day 09: Sampling units and generalizability; internal and external validity

Erin Rossiter

March, 2022

Announcements

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Announcements

- PolMeth
  - conference
  - listserv
- Summers
- Draft of final paper due next week! Say, Wednesday March 23 at 3pm
  - Github and/or email

Today

Today

Two interconnected themes:

1. Sampling
2. Generalizability (internal vs. external validity)

Note these slides draw heavily from Erin Hartman's "Generalizing Experimental Results" chapter of *Advances in Experimental Political Science*, Eds. James Druckman and Donald Green that we read for today.

Today

Two interconnected themes:

1. Sampling
2. Generalizability (internal vs. external validity)

Note these slides draw heavily from Erin Hartman's "Generalizing Experimental Results" chapter of *Advances in Experimental Political Science*, Eds. James Druckman and Donald Green that we read for today.

Today

Two interconnected themes:

1. Sampling
2. Generalizability (internal vs. external validity)

Note these slides draw heavily from Erin Hartman's "Generalizing Experimental Results" chapter of *Advances in Experimental Political Science*, Eds. James Druckman and Donald Green that we read for today.

Today

Two interconnected themes:

1. Sampling
2. Generalizability (internal vs. external validity)

Note these slides draw heavily from Erin Hartman's "Generalizing Experimental Results" chapter of *Advances in Experimental Political Science*, Eds. James Druckman and Donald Green that we read for today.

Overview

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - Board
  - Sampling important for descriptive tasks, too!

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - Board
  - Sampling important for descriptive tasks, too!

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - Board
  - Sampling important for descriptive tasks, too!

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - » so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - » Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - » Board
  - » Sampling important for descriptive tasks, too!

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - » so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - » Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - » Board
  - » Sampling important for descriptive tasks, too!

Sampling

- Our goal is to make inferences about some population **based on data from a sample**
  - » so far, we've glossed over this point
- Specifically, we're interested in causal inferences using a randomized experiment with our sample
  - » Note: we'll be talking about randomly choosing **units for the study**, which comes *before* randomly assigning units to treatment and control
  - » Board
  - » Sampling important for descriptive tasks, too!

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

   - construct validity
   - measurement
   - not the focus today

2. How results change if we change units, treatment, outcome, or setting.

   - experiments are designed to be internally valid
   - generalizability is a question of external validity
   - sampling is one big experimental ingredient for generalizability
     - » focus today is extrapolating results to *units* outside the experiment's context
     - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

   - construct validity
   - measurement
   - not the focus today

2. How results change if we change units, treatment, outcome, or setting.

   - experiments are designed to be internally valid
   - generalizability is a question of external validity
   - sampling is one big experimental ingredient for generalizability
     - » focus today is extrapolating results to *units* outside the experiment's context
     - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

   - construct validity
   - measurement
   - not the focus today

2. How results change if we change units, treatment, outcome, or setting.

   - experiments are designed to be internally valid
   - generalizability is a question of external validity
   - sampling is one big experimental ingredient for generalizability
     - » focus today is extrapolating results to *units* outside the experiment's context
     - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

   - construct validity
   - measurement
   - not the focus today

2. How results change if we change units, treatment, outcome, or setting.

   - experiments are designed to be internally valid
   - generalizability is a question of external validity
   - sampling is one big experimental ingredient for generalizability
     - » focus today is extrapolating results to *units* outside the experiment's context
     - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Generalizability

What people mean when they talk about "generalizability":

1. How results connect to broad theoretical understanding of the concepts represented in the study.

- construct validity
- measurement
- not the focus today

2. How results change if we change units, treatment, outcome, or setting.

- experiments are designed to be internally valid
- generalizability is a question of external validity
- sampling is one big experimental ingredient for generalizability
  - » focus today is extrapolating results to *units* outside the experiment's context
  - » ex: GOTV campaign in IA have similar results if applied to people in IN?

Big picture

Randomized experiment results help us understand the impact of a **specific treatment** on a **certain outcome** in a given context (including **time, place, sample**) $\rightarrow$ **Internal validity**

But for our knowledge-building goals, we ought to be interested in how our results can speak to (i.e., generalize to) other individuals and settings beyond our experiment. $\rightarrow$ **External validity**

Does our experiment answer a causal question **outside** the context of our specific experiment? Maybe. These conclusions depend on strong assumptions. (akin to observational causal inference)

This is the internal validity, external validity tension/tradeoff!

Big picture

Randomized experiment results help us understand the impact of a **specific treatment** on a **certain outcome** in a given context (including **time, place, sample**) $\rightarrow$ **Internal validity**

But for our knowledge-building goals, we ought to be interested in how our results can speak to (i.e., generalize to) other individuals and settings beyond our experiment. $\rightarrow$ **External validity**

Does our experiment answer a causal question **outside** the context of our specific experiment? Maybe. These conclusions depend on strong assumptions. (akin to observational causal inference)

This is the internal validity, external validity tension/tradeoff!

Big picture

Randomized experiment results help us understand the impact of a **specific treatment** on a **certain outcome** in a given context (including **time, place, sample**) $\rightarrow$ **Internal validity**

But for our knowledge-building goals, we ought to be interested in how our results can speak to (i.e., generalize to) other individuals and settings beyond our experiment. $\rightarrow$ **External validity**

Does our experiment answer a causal question **outside** the context of our specific experiment? Maybe. These conclusions depend on strong assumptions. (akin to observational causal inference)

This is the internal validity, external validity tension/tradeoff!

Big picture

Randomized experiment results help us understand the impact of a **specific treatment** on a **certain outcome** in a given context (including **time, place, sample**) $\rightarrow$ **Internal validity**

But for our knowledge-building goals, we ought to be interested in how our results can speak to (i.e., generalize to) other individuals and settings beyond our experiment. $\rightarrow$ **External validity**

Does our experiment answer a causal question **outside** the context of our specific experiment? Maybe. These conclusions depend on strong assumptions. (akin to observational causal inference)

This is the internal validity, external validity tension/tradeoff!

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:/pause

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**

More on the interval, external validity tradeoff

Example:

- In a highly controlled setting, like a lab, we can be very confident that our manipulation is the causal agent.
  - » Intentionally contrived for this reason
  - » Controlled, artificial environment allows for internal validity
  - » Not built for external validity, but we ought to think about it nonetheless
- Because a lab setting is *very different* than the real world, our statements about how the causal effect applies *outside* the lab aren't as rigorous
  - » field experiments often better at making externally valid claims than lab experiments, for example
  - » **Why?**
    - **sampling bias** Participants of the study differ substantially from the population in ways that explain treatment effects, so our estimate of the ATE using our sample will be a biased estimate of the ATE in the population

Methods of sampling

1. Simple random sampling

- Each sample of size n from a population of size N has an equal chance of being selected
- Implies that each subject has an equal chance of being included in the sample
- Ex:
  - Select a random sample of size 4 from this class.

1. Simple random sampling

- Each sample of size n from a population of size N has an equal chance of being selected
- Implies that each subject has an equal chance of being included in the sample
- Ex:
  - Select a random sample of size 4 from this class.

1. Simple random sampling

- Each sample of size n from a population of size N has an equal chance of being selected
- Implies that each subject has an equal chance of being included in the sample
- Ex:
  - » Select a random sample of size 4 from this class.

2. Computer-generated STS

- Obtain list of the units in the population (**sampling frame**):
- Generate a random number for each unit and sort by random number
  - » i.e., randomly shuffle list
- Sample is the first n units

2. Computer-generated STS in R

```
zips <- usa::zipcodes
N <- nrow(zips)
print(N)
```

```
## [1] 44336
```

```
n <- 10
sample(x = 1:N, size = n, replace = F)
```

```
##  [1] 33546 33180 10637 35554 24897  5578   670 36193 42276  8
```

2. Computer-generated STS in R

```
zips <- usa::zipcodes
N <- nrow(zips)
print(N)
```

```
## [1] 44336
```

```
n <- 10
sample(x = 1:N, size = n, replace = F)
```

```
##  [1] 33546 33180 10637 35554 24897  5578   670 36193 42276  8
```

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

3. Stratified random sampling

- The population is divided into subgroups, or strata
- A SRS is selected from each strata
- Ensures representation from each strata; no guarantee with SRS
- (Similar motivation as blocking. How is this different?)
- Ex:
  - Stratify ND freshman by gender, randomly select 100 women and 100 men
  - Voter poll – stratify the nation by state, randomly select 100 voters in each state

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

4. Cluster sampling

- The subjects in the population may be grouped in a natural way
- Ex:
  - households
  - students in a class
  - media markets
- Cluster sampling involves taking a SRS of clusters, then using *all units* in the cluster

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

5. Convenience sample

- A sample that is simple or easy to obtain without randomization
- Ex:
  - online markets/panels
  - posting to your social media network
  - sitting in a park and grabbing passers-by
  - surveying your intro to polisci class
- These samples are often biased. Why? Examples?

Population effects

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Population effects

We need to think and plan carefully when trying to justify inferences about a *population effect* using experimental data.

Why? Because we estimate a *sample effect*

Estimands so far in this class:

- ATE = average treatment effect
- CATE = conditional average treatment effect
- **SATE = sample average treatment effect**
- **PATE = population average treatment effect**

We'll formally define these later. First, how to think and plan for population effect inferences:

1. Define the target population to which you'll generalize your results
2. Make assumptions! These largely pertain to sampling
3. Estimating the PATE

Steps

Step 1: Target population

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 1: Target population

Clearly define a **target population** to which you'll generalize your results

- PATE is meaningless without this
- external validity is meaningless without this
- population isn't automatically defined, you must define it!

Example:

- GOTV experiment with voters in northwest IA county
- What is the population? no right answer! depends on your RQ:
  - all people in northwest IA?
  - all people in IA?
  - all people in the midwest?
  - all people in the US?

What might a target population be in your research?

Step 2: Assumptions for generalization

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

- potential outcomes do not depend on whether the unit is in experimental sample or not
- violated if:
  - » treatment operationalized differently in the sample vs. target population
  - » Hawthorne effects

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

- potential outcomes do not depend on whether the unit is in experimental sample or not
- violated if:
  - » treatment operationalized differently in the sample vs. target population
  - » Hawthorne effects

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

- potential outcomes do not depend on whether the unit is in experimental sample or not
- violated if:
  - » treatment operationalized differently in the sample vs. target population
  - » Hawthorne effects

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

- potential outcomes do not depend on whether the unit is in experimental sample or not
- violated if:
  - » treatment operationalized differently in the sample vs. target population
  - » Hawthorne effects

Step 2: Assumptions for generalization

New notation!

- $S_i \in 0, 1$ sampling indicator
- $Y_i(s, d)$ potential outcome for unit i under context s and treatment d
  - » context either experimental sample or target population

Assumption 1:

$$Y_i(s, d) = Y_i(s', d) = Y_i(d) \forall s \neq s'$$

What does this mean?

- potential outcomes do not depend on whether the unit is in experimental sample or not
- violated if:
  - » treatment operationalized differently in the sample vs. target population
  - » Hawthorne effects

Step 2: Assumptions for generalization

Assumption 2:

$$Y_i(1), Y_i(0), \mathbf{X}_i \perp D_i | S_i = 1$$

- potential outcomes, and pre-treatment covariates, are statistically independent from treatment assignment *for the experimental sample*
- implies that we have a known randomization scheme

Step 2: Assumptions for generalization

Assumption 2:

$$Y_i(1), Y_i(0), \mathbf{X}_i \perp D_i | S_i = 1$$

- potential outcomes, and pre-treatment covariates, are statistically independent from treatment assignment *for the experimental sample*
- implies that we have a known randomization scheme

Step 2: Assumptions for generalization

Assumption 2:

$$Y_i(1), Y_i(0), \mathbf{X}_i \perp D_i | S_i = 1$$

- potential outcomes, and pre-treatment covariates, are statistically independent from treatment assignment *for the experimental sample*
- implies that we have a known randomization scheme

Step 2: Assumptions for generalization

Assumption 2:

$$Y_i(1), Y_i(0), \mathbf{X}_i \perp D_i | S_i = 1$$

- potential outcomes, and pre-treatment covariates, are statistically independent from treatment assignment *for the experimental sample*
- implies that we have a known randomization scheme

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Adjustment set: assume $\mathbf{W}_i$ is a set of observable variables that exist and can be measured for *both* the population and experimental samples.

Why?

- needed to recover the PATE
- make the joint distribution of $\mathbf{W}_i$ in the experimental sample match that of the target population
- intuition: *make the treatment effect moderator independent of the sampling process*
- (more on the math of this later!)

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Step 2: Assumptions for generalization

Assumption 3:

(a) Conditional ignorability

$$Y_i(1) - Y_i(0) \perp S_i | \mathbf{W}_i$$

- after conditioning on covariates, treatment effect heterogeneity and sampling indicator are independent
  - » put simply, if we condition on covariates, we can't tell if you're in the sample or population from looking at your τ_i

(b) Positivity

$$0 < Pr(S_i = 1 | \mathbf{W}_i) < 1$$

- everyone (i.e., all $\mathbf{W}_i$ values) in the target population needs to represented in the experimental sample

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Quantity of interest, or estimand

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

- this has been our ATE definition all along
- see how SATE just articulates that it is the ATE for sample?
  - (recall, randomization of treatment allows for causal identification of SATE)
- making this clear allows us to think about how sample and population might be different!

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

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

- if experimental sample is a random sample from the population
- if there is no treatment effect heterogeneity

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

- if experimental sample is a random sample from the population
- if there is no treatment effect heterogeneity

Putting it together

Generalizability is a statement of how the SATE relates to the PATE

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

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

When might SATE = PATE?

- if experimental sample is a random sample from the population
- if there is no treatment effect heterogeneity

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

Step 3: Estimation

Under these assumptions, the PATE is:

$$\hat{PATE} = \sum_{w} \Big\{ E[Y_i | T_i = 1, S_i = 1, \mathbf{W}_i = w] - E[Y_i | T_i = 0, S_i = 1, \mathbf{W}_i = w] \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Estimator is:

$$\hat{PATE}_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

Standard error:

$$\hat{SE}(\hat{PATE})_{ps} = \sum_{w} \Big\{ \hat{SATE}_w \Big\} Pr(\mathbf{W}_i = w | S_i = 0)$$

- "weighted mean of subgroup effects within values of $\mathbf{W} = w$, with weights determined by the probability that $\mathbf{W} = w$" (Hartman pg 12)
  - what does this mean?

A note on $\mathbf{W}$

Note:

- need this info about target population!
- need coarsely defined variables in W
- need all levels of $\mathbf{w}$ represented in experimental sample

A note on **W**

Note:

- need this info about target population!
- need coarsely defined variables in W
- need all levels of $\mathbf{w}$ represented in experimental sample

A note on $\mathbf{W}$

Note:

- need this info about target population!
- need coarsely defined variables in W
- need all levels of $\mathbf{w}$ represented in experimental sample

A note on **W**

Note:

- need this info about target population!
- need coarsely defined variables in W
- need all levels of **w** represented in experimental sample

Step 3: Estimation

See how this is akin to an observational causal inference problem?

- selection on observables
  - » aims to make treatment independent of potential outcomes, conditional on an observable set of variables **X**
- estimating the PATE with experimental sample
  - » aims to make the difference in potential outcomes independent of the sampling mechanism, conditional on observable covariates **W** (measured in sample and pop)

Step 3: Estimation

See how this is akin to an observational causal inference problem?

- selection on observables
  - » aims to make treatment independent of potential outcomes, conditional on an observable set of variables **X**
- estimating the PATE with experimental sample
  - » aims to make the difference in potential outcomes independent of the sampling mechanism, conditional on observable covariates **W** (measured in sample and pop)

Step 3: Estimation

See how this is akin to an observational causal inference problem?

- selection on observables
  - » aims to make treatment independent of potential outcomes, conditional on an observable set of variables **X**
- estimating the PATE with experimental sample
  - » aims to make the difference in potential outcomes independent of the sampling mechanism, conditional on observable covariates **W** (measured in sample and pop)

Step 3: Estimation

See how this is akin to an observational causal inference problem?

- selection on observables
  - » aims to make treatment independent of potential outcomes, conditional on an observable set of variables **X**
- estimating the PATE with experimental sample
  - » aims to make the difference in potential outcomes independent of the sampling mechanism, conditional on observable covariates **W** (measured in sample and pop)

Step 3: Estimation

See how this is akin to an observational causal inference problem?

- selection on observables
  - » aims to make treatment independent of potential outcomes, conditional on an observable set of variables **X**
- estimating the PATE with experimental sample
  - » aims to make the difference in potential outcomes independent of the sampling mechanism, conditional on observable covariates **W** (measured in sample and pop)

Summing up

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - » but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

Experiments have a known treatment assignment mechanism

- thus, internal validity!
- unbiased SATE

Experiments often don't know the sampling mechanism by which individuals are selected to participate

- probably not a random sample from target population...
  - but yay if it is!!!
- so, external validity/generalizability not a given!
- estimating PATE can be done, but should be thoughtful and careful, even at design phase (esp when determining sampling procedure)

Summing up

- We talked about one straightforward PATE estimator
  - uses post-stratification
- There are alternatives, like Inverse Probability of Sampling Weights

Summing up

- We talked about one straightforward PATE estimator
  - uses post-stratification
- There are alternatives, like Inverse Probability of Sampling Weights

Summing up

- We talked about one straightforward PATE estimator
  - uses post-stratification
- There are alternatives, like Inverse Probability of Sampling Weights

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it

Summing up

- Convenience samples
  - » Recall, "Non-representative samples only impact generalizability inasmuch as the sample inclusion propensity is related, or aligned, to treatment effect moderation" (pg 17)
    - We don't know why MTurkers end up in our sample compared to other people in the US
  - » But, uf no heterogeneity, easier to make generalizability claims.
    - Evidence MTurk experiments replicate with TESS (Coppock, Leeper and Mullinix, 2018).
  - » How does this affect design decisions?
    - ensure sufficient variation on likely moderators in sample!
    - if you want a good estimate of PATE, make sure your design allows it