Gov 50: 19. More Confidence Intervals

Matthew Blackwell

Harvard University

Roadmap

1. Bootstrap CIs for a difference in means

2. Bootstrap CIs for a difference in ATEs

3. Interpreting confidence intervals

1/ Bootstrap CIs for a difference in means

Comparison between groups

- Last time: confidence intervals for means.

Comparison between groups

- Last time: confidence intervals for means.
- More interesting to compare across groups.

Comparison between groups

- Last time: confidence intervals for means.
- More interesting to compare across groups.
    - Differences in public opinion across groups

Comparison between groups

- Last time: confidence intervals for means.
- More interesting to compare across groups.
    - Differences in public opinion across groups
    - Difference between treatment and control groups.

Comparison between groups

- Last time: confidence intervals for means.
- More interesting to compare across groups.
    - Differences in public opinion across groups
    - Difference between treatment and control groups.
- Bedrock of causal inference!

Trains experiment

- Back to the Boston trains example.

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.
- Control group: no confederates.

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.
- Control group: no confederates.
- Outcome: X_i change in views on immigration.

Trains experiment

- Back to the Boston trains example.
  - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.
- Control group: no confederates.
- Outcome: X_i change in views on immigration.
  - Sample average in the treated group, $\overline{X}_T$

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.
- Control group: no confederates.
- Outcome: X_i change in views on immigration.
    - Sample average in the treated group, $\overline{X}_T$
    - Sample average in the control group, $\overline{X}_C$

Trains experiment

- Back to the Boston trains example.
    - Boston commuter rail platform setting.
- Treatment group: presence of native Spanish-speaking confederates.
- Control group: no confederates.
- Outcome: X_i change in views on immigration.
    - Sample average in the treated group, $\overline{X}_T$
    - Sample average in the control group, $\overline{X}_C$
- Estimated **average treatment effect**

$$\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
    - μ_T: Average outcome in the population if everyone received treatment.

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Average outcome in the population if everyone received treatment.
  - μ_C: Average outcome in the population if everyone received control.

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Average outcome in the population if everyone received treatment.
  - μ_C: Average outcome in the population if everyone received control.
- Difference-in-means estimator: $\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
    - μ_T: Average outcome in the population if everyone received treatment.
    - μ_C: Average outcome in the population if everyone received control.
- Difference-in-means estimator: $\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$
- $\overline{X}_T$ is a r.v. with mean $\mathbb{E}[\overline{X}_T] = \mu_T$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Average outcome in the population if everyone received treatment.
  - μ_C: Average outcome in the population if everyone received control.
- Difference-in-means estimator: $\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$
- $\overline{X}_T$ is a r.v. with mean $\mathbb{E}[\overline{X}_T] = \mu_T$
- $\overline{X}_C$ is a r.v. with mean $\mathbb{E}[\overline{X}_C] = \mu_C$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Average outcome in the population if everyone received treatment.
  - μ_C: Average outcome in the population if everyone received control.
- Difference-in-means estimator: $\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$
- $\overline{X}_T$ is a r.v. with mean $\mathbb{E}[\overline{X}_T] = \mu_T$
- $\overline{X}_C$ is a r.v. with mean $\mathbb{E}[\overline{X}_C] = \mu_C$
- $\rightsquigarrow \overline{X}_T - \overline{X}_C$ is a r.v. with mean $\mu_T - \mu_C$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Average outcome in the population if everyone received treatment.
  - μ_C: Average outcome in the population if everyone received control.
- Difference-in-means estimator: $\widehat{\text{ATE}} = \overline{X}_T - \overline{X}_C$
- $\overline{X}_T$ is a r.v. with mean $\mathbb{E}[\overline{X}_T] = \mu_T$
- $\overline{X}_C$ is a r.v. with mean $\mathbb{E}[\overline{X}_C] = \mu_C$
- $\rightsquigarrow \overline{X}_T - \overline{X}_C$ is a r.v. with mean $\mu_T - \mu_C$
  - Sample difference in means is on average equal to the population difference in means.

Trains data

```
library(gov50data)
trains
```

```
## # A tibble: 115 x 14
##      age  male income white college usborn treatment
##    <dbl> <dbl>  <dbl> <dbl>   <dbl>  <dbl>     <dbl>
##  1    31     0 135000     1       1      1         1
##  2    34     0 105000     1       1      0         1
##  3    63     1 135000     1       1      1         1
##  4    45     1 300000     1       1      1         1
##  5    55     1 135000     1       1      1         0
##  6    37     0  87500     1       1      1         1
##  7    53     0  87500     1       0      1         0
##  8    36     1 135000     1       1      1         1
##  9    54     0 105000     1       0      1         0
## 10    42     1 135000     1       1      1         1
## # i 105 more rows
## # i 7 more variables: ideology <dbl>, numberim.pre <dbl>,
## #   numberim.post <dbl>, remain.pre <dbl>,
## #   remain.post <dbl>, english.pre <dbl>,
## #   english.post <dbl>
```

Estimating the difference in means

```r
diff_in_means <- trains |>
  group_by(treatment) |>
  summarize(post_mean = mean(numberim.post)) |>
  pivot_wider(names_from = treatment, values_from = post_mean) |>
  mutate(ATE = `1` - `0`)
diff_in_means
```

```
## # A tibble: 1 x 3
##     `0`   `1`   ATE
##   <dbl> <dbl> <dbl>
## 1  2.73  3.12 0.383
```

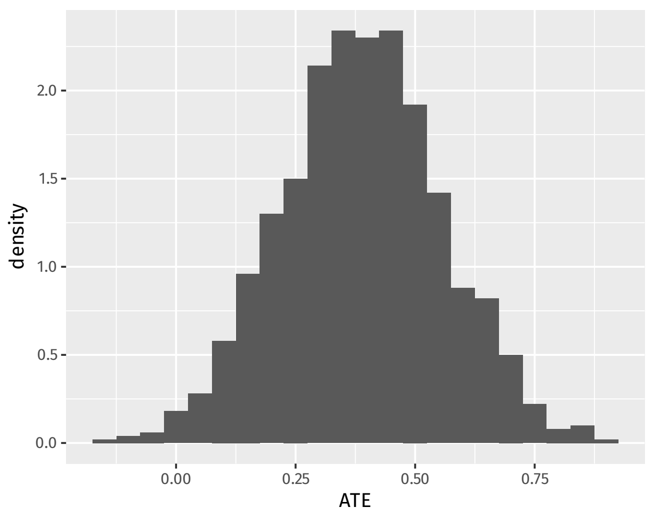
Bootstrap for the difference in means

```r
library(infer)
dim_boots <- trains |>
  rep_slice_sample(prop = 1, replace = TRUE, reps = 1000) |>
  group_by(replicate, treatment) |>
  summarize(post_mean = mean(numberim.post)) |>
  pivot_wider(names_from = treatment, values_from = post_mean) |>
  mutate(ATE = `1` - `0`)
dim_boots
```

```
## # A tibble: 1,000 x 4
## # Groups:   replicate [1,000]
##    replicate   `0`   `1`   ATE
##        <int> <dbl> <dbl> <dbl>
##  1         1  2.65  3.35 0.695
##  2         2  2.75  3.08 0.328
##  3         3  2.58  3.41 0.835
##  4         4  2.90  3.10 0.209
##  5         5  2.83  2.94 0.117
##  6         6  2.74  3.12 0.380
##  7         7  2.73  3.27 0.538
##  8         8  2.69  3.19 0.497
##  9         9  2.66  3.15 0.489
```

Visualizing the bootstraps

```r
dim_boots |>
  ggplot(aes(x = ATE)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.05)
```



Calculating the percentile CI

You can use `get_confidence_interval()` with your “hand-rolled” bootstraps, but you have to make sure you only pass it the variable of interest using `select`:

```
dim_ci_95 <- dim_boots |>
  select(replicate, ATE) |>
  get_confidence_interval(level = 0.95, type = "percentile")

dim_ci_95
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1   0.0668    0.715
```

What about change in views as the outcome?

```r
change_ci_95 <- trains |>
  rep_slice_sample(prop = 1, replace = TRUE, reps = 1000) |>
  group_by(replicate, treatment) |>
  summarize(change_mean = mean(numberim.post - numberim.pre)) |>
  pivot_wider(names_from = treatment, values_from = change_mean) |>
  mutate(ATE = `1` - `0`) |>
  select(replicate, ATE) |>
  get_confidence_interval(level = 0.95, type = "percentile")
change_ci_95
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1   0.0217    0.606
```

What's different?

Let's look at the width of the two confidence intervals:

```
## Post outcome width
dim_ci_95[2]-dim_ci_95[1]
```

```
##   upper_ci
## 1    0.649
```

```
## Change outcome width
change_ci_95[2] - change_ci_95[1]
```

```
##   upper_ci
## 1    0.585
```

Width of CI depends on outcome variability

Change CI is narrower! Why? Because the change is less variable than the post outcome:

```
trains |> summarize(sd_post = sd(numberim.post),
                    sd_change = sd(numberim.post - numberim.pre))
```

```
## # A tibble: 1 x 2
##   sd_post sd_change
##     <dbl>     <dbl>
## 1   0.917     0.826
```

`infer` workflow

For `infer`, we have to do a bit of massaging. It wants the treatment variable to be a vector and we have to tell it what order we take the difference:

```
dim_boots_infer <- trains |>
  mutate(treatment = if_else(treatment == 1, "Treated", "Control")) |>
  specify(numberim.post ~ treatment) |>
  generate(reps = 1000, type = "bootstrap") |>
  calculate(stat = "diff in means", order = c("Treated", "Control"))
dim_boots_infer |>
  get_confidence_interval(level = 0.95, type = "percentile")
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1   0.0610    0.708
```

2/ Bootstrap CIs for a difference in ATEs

Interactions

We have also estimated conditional ATEs:

$$ATE_{\text{college}} = \overline{X}_{T,\text{college}} - \overline{X}_{C,\text{college}}$$
$$ATE_{\text{noncollege}} = \overline{X}_{T,\text{noncollege}} - \overline{X}_{C,\text{noncollege}}$$

Interactions

We have also estimated conditional ATEs:

$$ATE_{\text{college}} = \overline{X}_{T,\text{college}} - \overline{X}_{C,\text{college}}$$
$$ATE_{\text{noncollege}} = \overline{X}_{T,\text{noncollege}} - \overline{X}_{C,\text{noncollege}}$$

An **interaction** between treatment and college is the difference between these two effects:

$$ATE_{\text{college}} - ATE_{\text{noncollege}}$$

Interactions

We have also estimated conditional ATEs:

$$ATE_{\text{college}} = \overline{X}_{T,\text{college}} - \overline{X}_{C,\text{college}}$$
$$ATE_{\text{noncollege}} = \overline{X}_{T,\text{noncollege}} - \overline{X}_{C,\text{noncollege}}$$

An **interaction** between treatment and college is the difference between these two effects:

$$ATE_{\text{college}} - ATE_{\text{noncollege}}$$

This is a random variable and has a **sampling distribution**.

Estimating the interaction

To estimate the interaction, we need to pivot both treatment and college to the columns.

```
trains |>
  mutate(
    treatment = if_else(treatment == 1, "Treated", "Control"),
    college = if_else(college == 1, "College", "Noncollege")
  ) |>
  group_by(treatment, college) |>
  summarize(post_mean = mean(numberim.post)) |>
  pivot_wider(
    names_from = c(treatment, college),
    values_from = post_mean
  )
```

```
## # A tibble: 1 x 4
##   Control_College Control_Noncollege Treated_College
##             <dbl>              <dbl>           <dbl>
## 1            2.63               3.57            3.11
## # i 1 more variable: Treated_Noncollege <dbl>
```

Estimating the interaction

```
trains |>
  mutate(
    treatment = if_else(treatment == 1, "Treated", "Control"),
    college = if_else(college == 1, "College", "Noncollege")
  ) |>
  group_by(treatment, college) |>
  summarize(post_mean = mean(numberim.post)) |>
  pivot_wider(
    names_from = c(treatment, college),
    values_from = post_mean
  ) |>
  mutate(
    ATE_c = Treated_College - Control_College,
    ATE_nc = Treated_Noncollege - Control_Noncollege,
    interaction = ATE_c - ATE_nc
  ) |>
  select(ATE_c, ATE_nc, interaction)
```

```
## # A tibble: 1 x 3
##   ATE_c ATE_nc interaction
##   <dbl>  <dbl>       <dbl>
## 1 0.482 -0.429       0.911
```

Bootstrapping the interaction

```r
int_boots <- trains |>
  mutate(
    treatment = if_else(treatment == 1, "Treated", "Control"),
    college = if_else(college == 1, "College", "Noncollege")
  ) |>
  rep_slice_sample(prop = 1, replace = TRUE, reps = 1000) |>
  group_by(replicate, treatment, college) |>
  summarize(post_mean = mean(numberim.post)) |>
  pivot_wider(
    names_from = c(treatment, college),
    values_from = post_mean
  ) |>
  mutate(
    ATE_c = Treated_College - Control_College,
    ATE_nc = Treated_Noncollege - Control_Noncollege,
    interaction = ATE_c - ATE_nc
  ) |>
  select(replicate, ATE_c, ATE_nc, interaction)
```

```
int_boots
```

```
## # A tibble: 1,000 x 4
## # Groups:   replicate [1,000]
##    replicate ATE_c  ATE_nc interaction
##        <int> <dbl>   <dbl>       <dbl>
##  1         1 0.570 -0.286        0.856
##  2         2 0.695 -0.528        1.22
##  3         3 0.48  -0.333        0.813
##  4         4 0.365 -0.0417       0.407
##  5         5 0.568 -0.429        0.997
##  6         6 0.299 -0.200        0.499
##  7         7 0.428 -0.167        0.594
##  8         8 0.612 -0.5          1.11
##  9         9 0.649 -1.14         1.79
## 10        10 0.718 -0.232        0.950
## # i 990 more rows
```

Getting the confidence interval

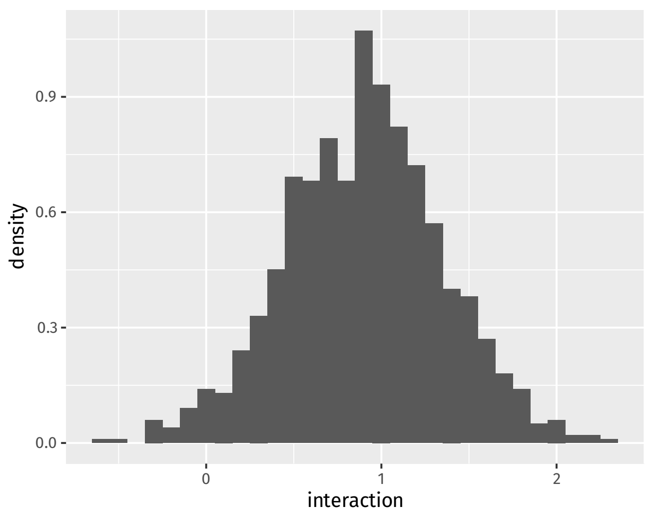
We have to drop NA values because sometimes the bootstrap gets a draw of all college or all noncollege and we can't calculate the interaction:

```
int_boots |>
  select(replicate, interaction) |>
  drop_na() |>
  get_confidence_interval(level = 0.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1  -0.0125     1.78
```

Visualizing the bootstrap

```r
int_boots |>
  ggplot(aes(x = interaction)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.1)
```



3/ Interpreting confidence intervals

Interpretation and simulation

- Be careful about interpretation:

Interpretation and simulation

- Be careful about interpretation:
    - A 95% confidence interval will contain the true value in 95% of repeated samples.

Interpretation and simulation

- Be careful about interpretation:
    - A 95% confidence interval will contain the true value in 95% of repeated samples.
    - For a particular calculated confidence interval, truth is either in it or not.

Interpretation and simulation

- Be careful about interpretation:
  - A 95% confidence interval will contain the true value in 95% of repeated samples.
  - For a particular calculated confidence interval, truth is either in it or not.
- A simulation can help our understanding:

Interpretation and simulation

- Be careful about interpretation:
  - A 95% confidence interval will contain the true value in 95% of repeated samples.
  - For a particular calculated confidence interval, truth is either in it or not.
- A simulation can help our understanding:
  - Draw samples of size 1500 assuming population approval for Trump of $p = 0.4$.

Interpretation and simulation

- Be careful about interpretation:
  - A 95% confidence interval will contain the true value in 95% of repeated samples.
  - For a particular calculated confidence interval, truth is either in it or not.
- A simulation can help our understanding:
  - Draw samples of size 1500 assuming population approval for Trump of $p = 0.4$.
  - Calculate 95% confidence intervals in each sample.

Interpretation and simulation

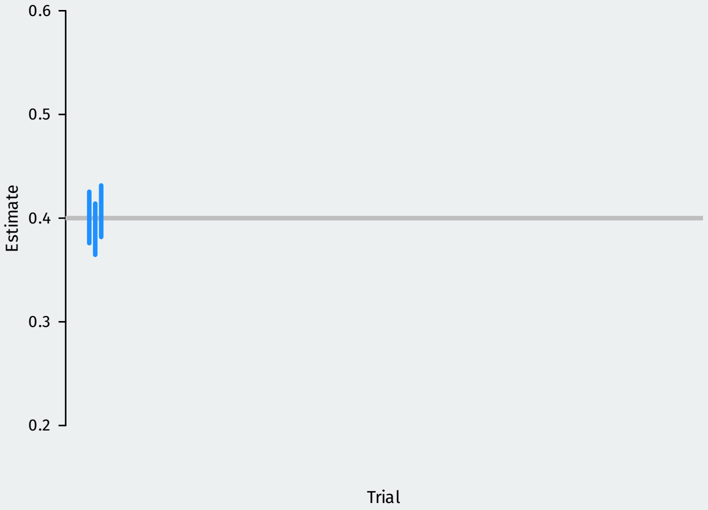
- Be careful about interpretation:
  - A 95% confidence interval will contain the true value in 95% of repeated samples.
  - For a particular calculated confidence interval, truth is either in it or not.
- A simulation can help our understanding:
  - Draw samples of size 1500 assuming population approval for Trump of $p = 0.4$.
  - Calculate 95% confidence intervals in each sample.
  - See how many overlap with the true population approval.

Plotting the CIs

Plotting the CIs

Plotting the CIs

Plotting the CIs



Plotting the CIs