

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, \bar{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, \bar{X}_T
 - Sample average in the control group, \bar{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, \bar{X}_T
 - Sample average in the control group, \bar{X}_C
- Estimated **average treatment effect**

$$\widehat{ATE} = \bar{X}_T - \bar{X}_C$$

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: $\widehat{ATE} = \bar{X}_T - \bar{X}_C$

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: $\widehat{ATE} = \bar{X}_T - \bar{X}_C$
- \bar{X}_T has a distribution centered on μ_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: $\widehat{ATE} = \bar{X}_T - \bar{X}_C$
- \bar{X}_T has a distribution centered on μ_T
- \bar{X}_C has a distribution centered on μ_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: $\widehat{ATE} = \bar{X}_T - \bar{X}_C$
- \bar{X}_T has a distribution centered on μ_T
- \bar{X}_C has a distribution centered on μ_C
- $\rightsquigarrow \bar{X}_T - \bar{X}_C$ has a distribution centered on $\mu_T - \mu_C$

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: $\widehat{ATE} = \bar{X}_T - \bar{X}_C$
- \bar{X}_T has a distribution centered on μ_T
- \bar{X}_C has a distribution centered on μ_C
- $\rightsquigarrow \bar{X}_T - \bar{X}_C$ has a distribution centered on $\mu_T - \mu_C$
 - 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

```
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
```

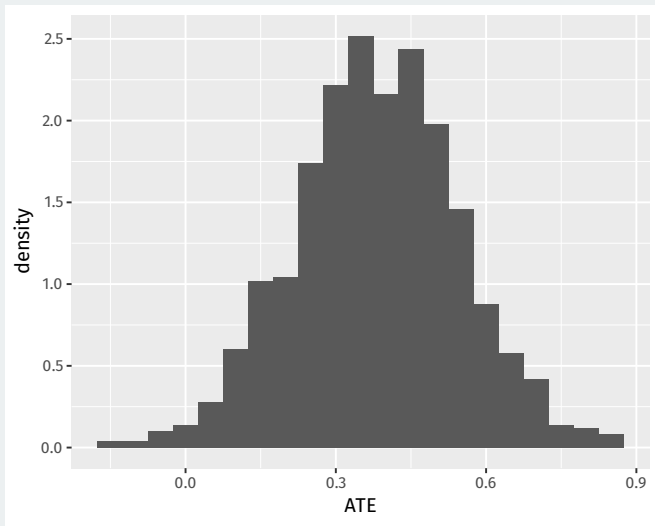
Bootstrap for the difference in means

```
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  2.79  2.91 0.121
## 2         2  2.63  3.09 0.455
## 3         3  2.67  2.92 0.245
## 4         4  2.75  3.17 0.427
## 5         5  2.70  3.40 0.694
## 6         6  2.70  3.21 0.507
## 7         7  2.89  3.10 0.207
## 8         8  2.83  2.91 0.0758
## 9         9  2.77  3.09 0.322
```

Visualizing the bootstraps

```
dim_boots |>  
  ggplot(aes(x = ATE)) +  
  geom_histogram(aes(y = after_stat(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  0.0697    0.707
```

What about change in views as the outcome?

```
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 -0.00167  0.610
```

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 0.637
```

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

```
## upper_ci  
## 1 0.612
```

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  0.0735     0.715
```

2/ Bootstrap CIs for a difference in ATEs

Interactions

We have also estimated conditional ATEs:

$$\begin{aligned}ATE_{\text{college}} &= \bar{X}_{T,\text{college}} - \bar{X}_{C,\text{college}} \\ATE_{\text{noncollege}} &= \bar{X}_{T,\text{noncollege}} - \bar{X}_{C,\text{noncollege}}\end{aligned}$$

Interactions

We have also estimated conditional ATEs:

$$\begin{aligned}ATE_{\text{college}} &= \bar{X}_{T,\text{college}} - \bar{X}_{C,\text{college}} \\ATE_{\text{noncollege}} &= \bar{X}_{T,\text{noncollege}} - \bar{X}_{C,\text{noncollege}}\end{aligned}$$

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:

$$\begin{aligned}ATE_{\text{college}} &= \bar{X}_{T,\text{college}} - \bar{X}_{C,\text{college}} \\ATE_{\text{noncollege}} &= \bar{X}_{T,\text{noncollege}} - \bar{X}_{C,\text{noncollege}}\end{aligned}$$

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

```
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     1 0.259 -0.667     0.926
## 2     2     2 0.432 -0.775     1.21
## 3     3     3 0.239 -0.337     0.576
## 4     4     4 0.559 -0.214     0.773
## 5     5     5 0.641 -0.120     0.761
## 6     6     6 0.758 -0.556     1.31
## 7     7     7 0.807 -0.286     1.09
## 8     8     8 0.472 -0.171     0.643
## 9     9     9 0.470 -1         1.47
## 10    10    10 0.182 -0.525     0.707
## # i 990 more rows
```

Getting the confidence interval

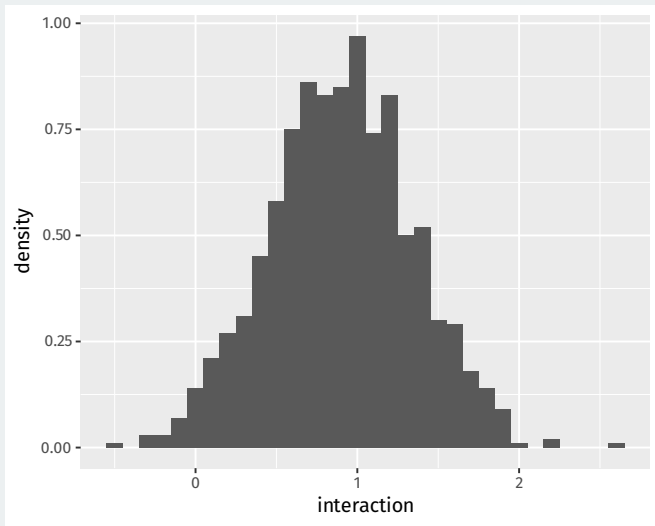
We have to drop NA values because sometimes the bootstrap gets a draw of all college or all noncollege and we can't calculate the interaction:

```
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  0.0269    1.77
```

Visualizing the bootstrap

```
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 Biden 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 Biden 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 Biden of $p = 0.4$.
 - Calculate 95% confidence intervals in each sample.
 - See how many overlap with the true population approval.

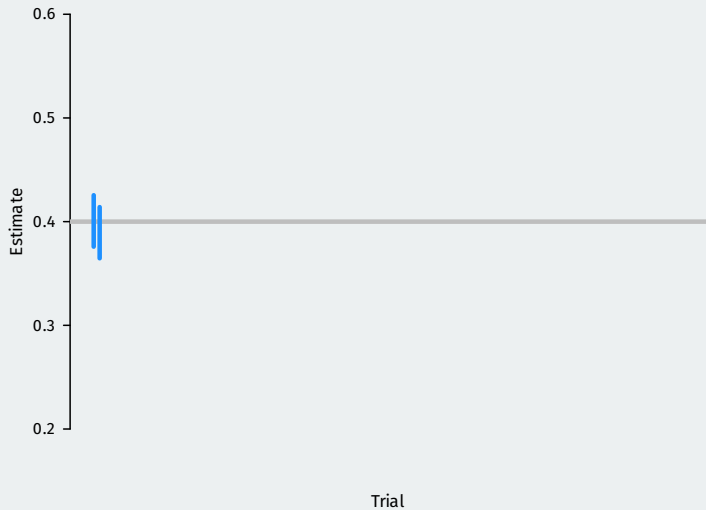
Plotting the CIs



Plotting the CIs



Plotting the CIs



Plotting the CIs



Plotting the CIs

