Gov 50: 13. Midterm Review + Prediction

Matthew Blackwell

Harvard University

Roadmap

1. Midterm Review: Estimating effects
2. Prediction
3. Evaluating the predictions

1/ Midterm Review: Estimating effects

Minimum wage study

- Does increasing the minimum wage affect employment?

Minimum wage study

- Does increasing the minimum wage affect employment?
    - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.

Minimum wage study

- Does increasing the minimum wage affect employment?
    - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
    - But that's a theoretical concern, can we give empirical evidence?

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that’s a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ

Minimum wage study

- Does increasing the minimum wage affect employment?
    - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
    - But that's a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
    - In 1992, NJ raised minimum wage from \$4.25/hr to \$5.05/hr.

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that's a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from \$4.25/hr to \$5.05/hr.
  - What is the effect of this change?

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that’s a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from $4.25/hr to $5.05/hr.
  - What is the effect of this change?
- Three research designs:

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that's a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from \$4.25/hr to \$5.05/hr.
  - What is the effect of this change?
- Three research designs:
  - **Cross-sectional**: Compare NJ employment to neighbor PA employment in 1993 (after).

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that's a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from \$4.25/hr to \$5.05/hr.
  - What is the effect of this change?
- Three research designs:
  - **Cross-sectional**: Compare NJ employment to neighbor PA employment in 1993 (after).
  - **Before-and-after**: Compare changes in NJ employment bewteen 1991 (before) and 1993 (after).

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that’s a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from $4.25/hr to $5.05/hr.
  - What is the effect of this change?
- Three research designs:
  - **Cross-sectional**: Compare NJ employment to neighbor PA employment in 1993 (after).
  - **Before-and-after**: Compare changes in NJ employment bewteen 1991 (before) and 1993 (after).
  - **Difference-in-differences**: Compare changes in NJ employment bewteen 1991 (before) and 1993 (after) to changes in PA in the same period.

Minimum wage study

- Does increasing the minimum wage affect employment?
  - Economists worry that requiring higher wages will lead employers shifting workers from full time to part time.
  - But that's a theoretical concern, can we give empirical evidence?
- Canonical study: Card and Krueger (1994) on minimum wage laws in NJ
  - In 1992, NJ raised minimum wage from \$4.25/hr to \$5.05/hr.
  - What is the effect of this change?
- Three research designs:
  - **Cross-sectional**: Compare NJ employment to neighbor PA employment in 1993 (after).
  - **Before-and-after**: Compare changes in NJ employment bewteen 1991 (before) and 1993 (after).
  - **Difference-in-differences**: Compare changes in NJ employment bewteen 1991 (before) and 1993 (after) to changes in PA in the same period.
- RCT or observational study?

Data

Name	Description
`chain`	Name of the fast-food restaurant chain
`location`	Location of the restaurant
`wageBefore`	Average wage at the restaurant before NJ minimum wage law
`wageAfter`	Average wage at the restaurant after NJ minimum wage law
`fullBefore`	Number of full-time employees before NJ minimum wage law
`fullAfter`	Number of full-time employees after NJ minimum wage law
`partBefore`	Number of full-time employees before NJ minimum wage law
`partAfter`	Number of full-time employees after NJ minimum wage law

Loading the data

```
library(tidyverse)
library(qss)
data(minwage)
minwage <- as_tibble(minwage)
minwage
```

```
## # A tibble: 358 x 8
##    chain  location wageBefore wageAfter fullBefore fullAfter
##    <chr>  <chr>         <dbl>     <dbl>      <dbl>     <dbl>
##  1 wendys PA             5         5.25       20           0
##  2 wendys PA             5.5       4.75        6          28
##  3 burge~ PA             5         4.75       50          15
##  4 burge~ PA             5         5          10          26
##  5 kfc    PA             5.25      5           2           3
##  6 kfc    PA             5         5           2           2
##  7 roys   PA             5         4.75        2.5         1
##  8 burge~ PA             5         5          40           9
##  9 burge~ PA             5         4.5         8           7
## 10 burge~ PA             5.5       4.75       10.5        18
## # i 348 more rows
## # i 2 more variables: partBefore <dbl>, partAfter <dbl>
```

Creating a treatment vector

```r
minwage |>
  count(location)
```

```
## # A tibble: 5 x 2
##   location     n
##   <chr>    <int>
## 1 PA          67
## 2 centralNJ   45
## 3 northNJ    146
## 4 shoreNJ     33
## 5 southNJ     67
```

```r
minwage <- minwage |>
  mutate(
    state = if_else(location == "PA", "PA", "NJ"), ## PA is control
    full_prop_after = fullAfter / (fullAfter + partAfter) ## proportion ful
  )
```

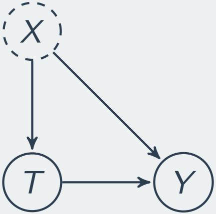
Cross-sectional estimate

```r
ate_cs <- minwage |>
  group_by(state) |>
  summarize(full_mean = mean(full_prop_after)) |>
  pivot_wider(
    names_from = state,
    values_from = full_mean
  ) |>
  mutate(ATE = NJ - PA)
ate_cs
```

```
## # A tibble: 1 x 3
##      NJ    PA    ATE
##   <dbl> <dbl>  <dbl>
## 1 0.320 0.272 0.0481
```

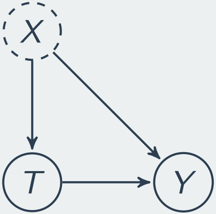
Interpretation: The minimum wage law increased the percent of full-time employment by 4.81 percentage points if the cross sectional assumptions hold.

Confounders



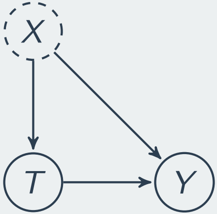
- Could there be **confounders** between having a minimum wage law at \$5.05 and employment?

Confounders



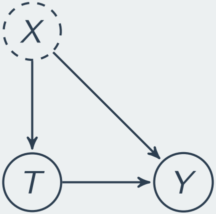
- Could there be **confounders** between having a minimum wage law at $5.05 and employment?
    - A confounder is a pre-treatment variable that affects both treatment and the outcome.

Confounders



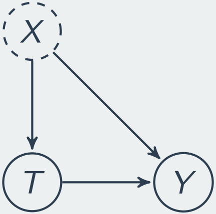
- Could there be **confounders** between having a minimum wage law at \$5.05 and employment?
  - A confounder is a pre-treatment variable that affects both treatment and the outcome.
- One possibility: different chain types.

Confounders



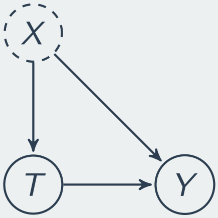
- Could there be **confounders** between having a minimum wage law at \$5.05 and employment?
  - A confounder is a pre-treatment variable that affects both treatment and the outcome.
- One possibility: different chain types.
  - Imagine if Burger King requires fewer workers to operate than other chains and if for historical reasons there are more BKs in PA than in NJ.

Confounders



- Could there be **confounders** between having a minimum wage law at $5.05 and employment?
    - A confounder is a pre-treatment variable that affects both treatment and the outcome.
- One possibility: different chain types.
    - Imagine if Burger King requires fewer workers to operate than other chains and if for historical reasons there are more BKs in PA than in NJ.
    - Then the difference we see in employment might be due to the differece in BKs rather than the MW law.

Confounders



- Could there be **confounders** between having a minimum wage law at \$5.05 and employment?
    - A confounder is a pre-treatment variable that affects both treatment and the outcome.
- One possibility: different chain types.
    - Imagine if Burger King requires fewer workers to operate than other chains and if for historical reasons there are more BKs in PA than in NJ.
    - Then the difference we see in employment might be due to the differece in BKs rather than the MW law.
    - We can check this by comparing chain distribution across states.

Balance of chains across states

```
minwage |>
  group_by(state, chain) |>
  summarize(n = n(), .groups = "drop_last") |>
  mutate(prop = n / sum(n)) |>
  pivot_wider(
    id_cols = chain,
    names_from = state,
    values_from = prop
  )
```

```
## # A tibble: 4 x 3
##   chain        NJ    PA
##   <chr>     <dbl> <dbl>
## 1 burgerking 0.405 0.463
## 2 kfc       0.223 0.149
## 3 roys      0.251 0.224
## 4 wendys    0.120 0.164
```

Some differences here: more BK in PA and more KFC in NJ. What to do? We can perform **statistical control** by estimating ATEs within groups.

ATE by chain

```r
minwage |>
  group_by(state, chain) |>
  summarize(full_mean = mean(full_prop_after)) |>
  pivot_wider(
    names_from = state,
    values_from = full_mean
  ) |>
  mutate(ATE = NJ - PA)
```

```
## # A tibble: 4 x 4
##   chain          NJ    PA    ATE
##   <chr>       <dbl> <dbl>  <dbl>
## 1 burgerking 0.358 0.321 0.0364
## 2 kfc        0.328 0.236 0.0918
## 3 roys       0.283 0.213 0.0697
## 4 wendys     0.260 0.248 0.0117
```

Before-and-after design

- Maybe there are difference between NJ and PA that we can't observe.

Before-and-after design

- Maybe there are difference between NJ and PA that we can't observe.
    - Called **unmeasured confounding**

Before-and-after design

- Maybe there are difference between NJ and PA that we can't observe.
  - Called **unmeasured confounding**
- Before and after design compares NJ before the law to after the law.

Before-and-after design

- Maybe there are difference between NJ and PA that we can't observe.
    - Called **unmeasured confounding**
- Before and after design compares NJ before the law to after the law.
    - Anything fixed about NJ cannot be causing the the differences.

Estimating ATE with before-and-after

```r
minwage <- minwage |>
  mutate(full_prop_before = fullBefore / (fullBefore + partBefore))

minwage |>
  filter(state == "NJ") |>
  summarize(ATE = mean(full_prop_after) - mean(full_prop_before))
```

```
## # A tibble: 1 x 1
##      ATE
##    <dbl>
## 1 0.0239
```

Interpretation: we estimate the MW law increase the full-time employment percentage by 2.39% if there are no **time-varying confounders**.

Difference-in-differences

- Before and after designs could be affected by time-varying confounders.

Difference-in-differences

- Before and after designs could be affected by time-varying confounders.
- If the whole US economy is shifting to full time employment due to a good economy, then it’s not the MW law that is driving things.

Difference-in-differences

- Before and after designs could be affected by time-varying confounders.
- If the whole US economy is shifting to full time employment due to a good economy, then it's not the MW law that is driving things.
- We can account for trends that are affecting all units by comparing the trends in the treated group to the trends in the control group.

Difference-in-differences estimate

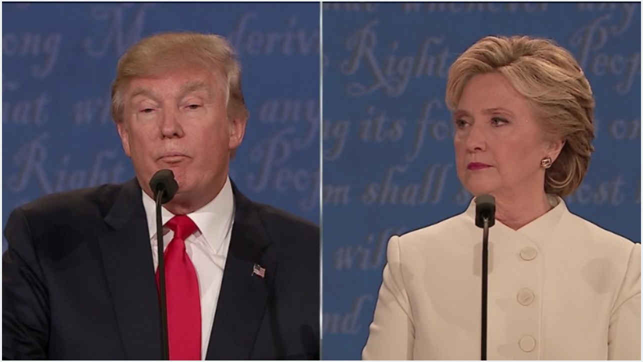
```
minwage |>
  group_by(state) |>
  summarize(trend = mean(full_prop_after) - mean(full_prop_before)) |>
  pivot_wider(
    names_from = state,
    values_from = trend
  ) |>
  mutate(DID = NJ - PA)
```

```
## # A tibble: 1 x 3
##       NJ      PA    DID
##    <dbl>   <dbl>  <dbl>
## 1 0.0239 -0.0377 0.0616
```

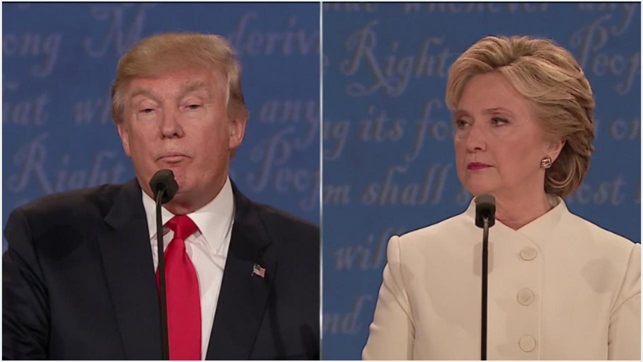
Interpretation: minimum wage laws increased percent full-time in NJ by 6.16 percentage points if trends in PA are a good proxy for trends in NJ if it didn't enact a MW law.

2/ Prediction

2016 US Presidential Election

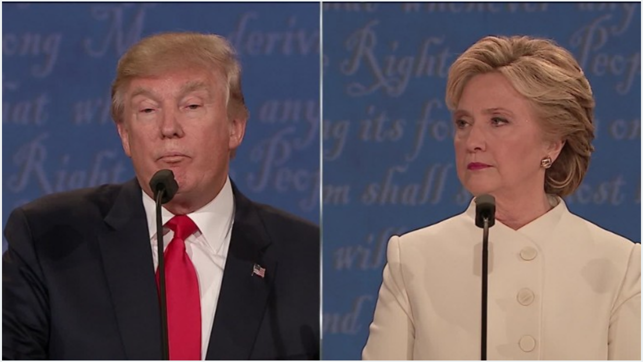


2016 US Presidential Election



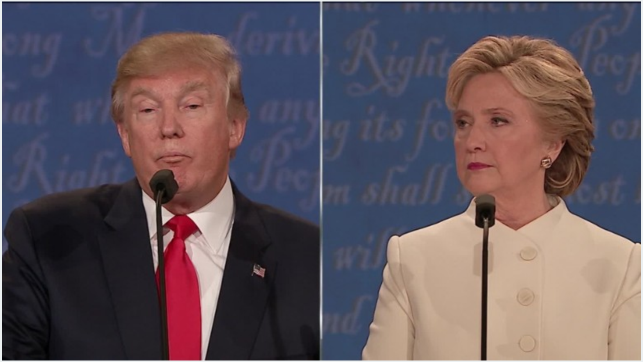
- 2016 election popular vote:

2016 US Presidential Election



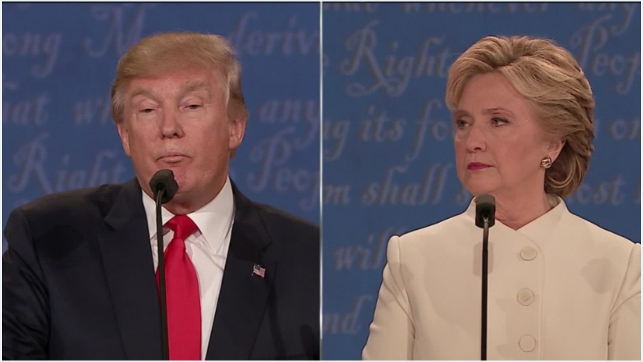
- 2016 election popular vote:
    - Clinton: 65,853,516 (48.2%)

2016 US Presidential Election



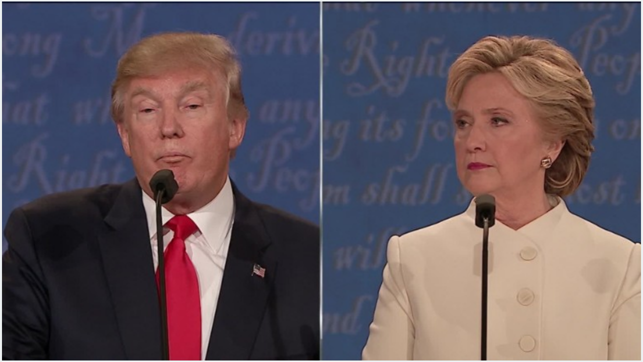
- 2016 election popular vote:
    - Clinton: 65,853,516 (48.2%)
    - Trump: 62,984,825 (46.1%)

2016 US Presidential Election



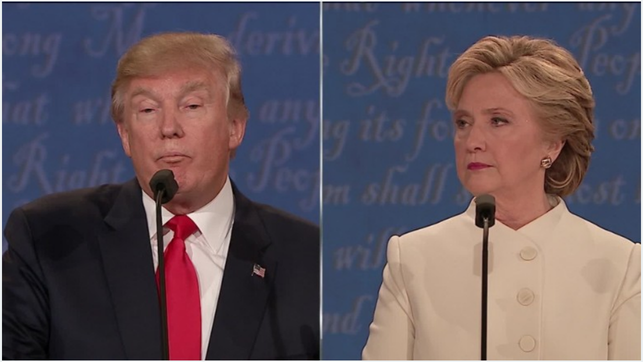
- 2016 election popular vote:
    - Clinton: 65,853,516 (48.2%)
    - Trump: 62,984,825 (46.1%)
- Why did Trump win? **Electoral college**

2016 US Presidential Election



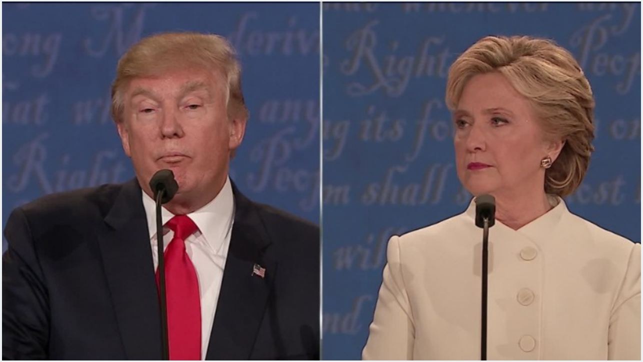
- 2016 election popular vote:
  - Clinton: 65,853,516 (48.2%)
  - Trump: 62,984,825 (46.1%)
- Why did Trump win? **Electoral college**
  - Trump: 304, Clinton: 227

2016 US Presidential Election



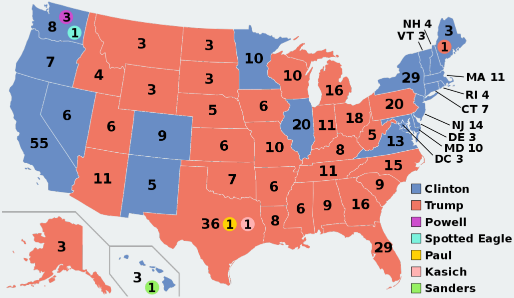
- 2016 election popular vote:
    - Clinton: 65,853,516 (48.2%)
    - Trump: 62,984,825 (46.1%)
- Why did Trump win? **Electoral college**
    - Trump: 304, Clinton: 227
- Election determined by 77,744 votes (margins in WI, MI, and PA)

2016 US Presidential Election



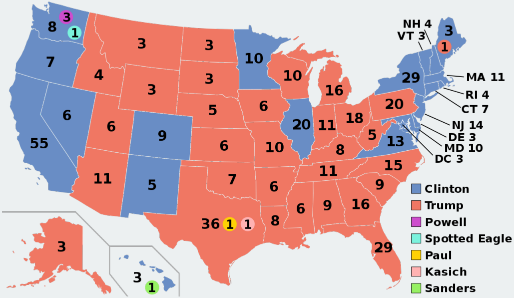
- 2016 election popular vote:
    - Clinton: 65,853,516 (48.2%)
    - Trump: 62,984,825 (46.1%)
- Why did Trump win? **Electoral college**
    - Trump: 304, Clinton: 227
- Election determined by 77,744 votes (margins in WI, MI, and PA)
    - 0.056% of the electorate (~136 million)

Predicting US Presidential Elections



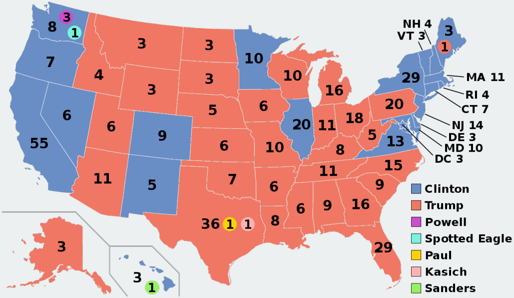
- **Electoral college system**

Predicting US Presidential Elections



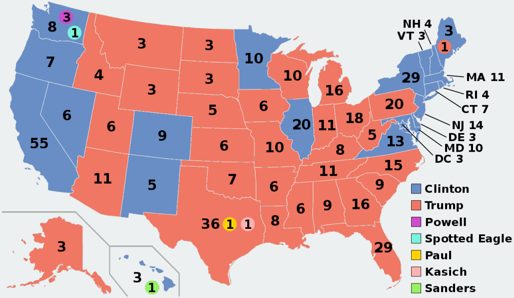
- **Electoral college system**
  - Must win an absolute majority of 538 electoral votes

Predicting US Presidential Elections



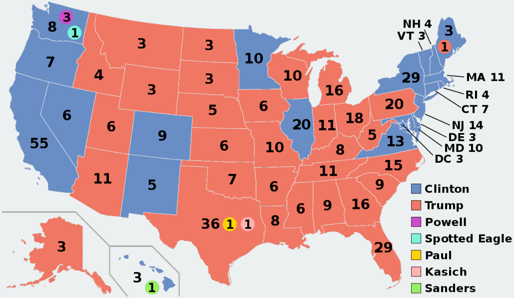
- **Electoral college system**
  - Must win an absolute majority of 538 electoral votes
  - 538 = 435 (House of Representatives) + 100 (Senators) + 3 (DC)

Predicting US Presidential Elections



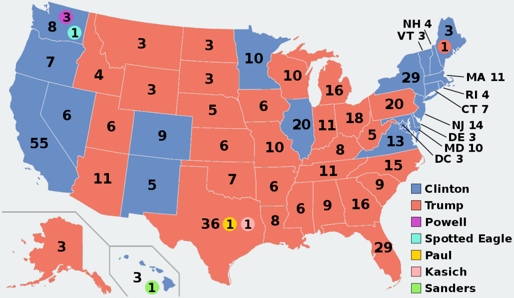
- **Electoral college system**
  - Must win an absolute majority of 538 electoral votes
  - 538 = 435 (House of Representatives) + 100 (Senators) + 3 (DC)
  - Must win at least 270 votes

Predicting US Presidential Elections



- **Electoral college system**
  - Must win an absolute majority of 538 electoral votes
  - 538 = 435 (House of Representatives) + 100 (Senators) + 3 (DC)
  - Must win at least 270 votes
  - nobody wins an absolute majority $\rightsquigarrow$ House vote

Predicting US Presidential Elections



- **Electoral college system**
  - Must win an absolute majority of 538 electoral votes
  - 538 = 435 (House of Representatives) + 100 (Senators) + 3 (DC)
  - Must win at least 270 votes
  - nobody wins an absolute majority $\rightsquigarrow$ House vote
- Must predict winner of each state

Prediction strategy

- Predict state-level support for each candidate using polls

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:
    1. For each state, subset to polls within that state.

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:
    1. For each state, subset to polls within that state.
    2. Further subset the latest polls

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:
    1. For each state, subset to polls within that state.
    2. Further subset the latest polls
    3. Average the latest polls to estimate support for each candidate

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:
  1. For each state, subset to polls within that state.
  2. Further subset the latest polls
  3. Average the latest polls to estimate support for each candidate
  4. Allocate the electoral votes to the candidate who has greatest support

Prediction strategy

- Predict state-level support for each candidate using polls
- Allocate electoral college votes of that state to its predicted winner
- Aggregate EC votes across states to determine the predicted winner
- Coding strategy:

    1. For each state, subset to polls within that state.
    2. Further subset the latest polls
    3. Average the latest polls to estimate support for each candidate
    4. Allocate the electoral votes to the candidate who has greatest support
    5. Repeat this for all states and aggregate the electoral votes

2020 polling prediction

Election data: `pres20`

Name	Description
`state`	abbreviated name of state
`biden`	Biden's vote share (percentage)
`trump`	Trump's vote share (percentage)
`ev`	number of electoral college votes for the state

Polling data `polls20`:

Name	Description
`state`	state in which poll was conducted
`end_date`	end date the period when poll was conducted
`daysleft`	number of days between end date and election day
`pollster`	name of organization conducting poll
`sample_size`	name of organization conducting poll
`biden`	predicted support for Biden (percentage)
`trump`	predicted support for Trump (percentage)

Some preprocessing

```r
library(gov50data)
glimpse(polls20)
```

```
## Rows: 2,445
## Columns: 7
## $ end_date    <date> 2020-11-02, 2020-11-02, 2020-11-02, 2~
## $ state       <chr> "FL", "PA", "FL", "FL", "NV", "GA", "S~
## $ days_left   <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ pollster    <chr> "The Political Matrix/The Listener Gro~
## $ sample_size <dbl> 966, 499, 400, 1054, 1024, 1041, 817, ~
## $ biden       <dbl> 44.2, 48.4, 47.0, 47.3, 48.4, 45.4, 39~
## $ trump       <dbl> 48.0, 49.2, 48.2, 49.4, 49.1, 49.7, 51~
```

Easy to iterate with tidyverse

```
poll_pred <- polls20 |>
  group_by(state) |>
  filter(days_left == min(days_left)) |>
  summarize(margin_pred = mean(biden - trump))
poll_pred
```

```
## # A tibble: 51 x 2
##    state margin_pred
##    <chr>       <dbl>
##  1 AK         -9
##  2 AL        -26
##  3 AR        -23
##  4 AZ          4.25
##  5 CA         26
##  6 CO         11
##  7 CT         22
##  8 DC         89
##  9 DE         22
## 10 FL          0.0800
## # i 41 more rows
```

3/ Evaluating the predictions

Polling errors

Prediction error = actual outcome − predicted outcome

```
poll_pred <- poll_pred |>
  left_join(pres20) |>
  mutate(margin = biden - trump) |>
  mutate(errors = margin - margin_pred)
poll_pred
```

```
## # A tibble: 51 x 8
##    state margin_pred    ev biden trump  other  margin errors
##    <chr>       <dbl> <dbl> <dbl> <dbl>  <dbl>   <dbl>  <dbl>
##  1 AK         -9         3  42.8 52.8  0.732  -10.1   -1.06
##  2 AL        -26         9  36.6 62.0  0.699  -25.5    0.538
##  3 AR        -23         6  34.8 62.4  0.257  -27.6   -4.62
##  4 AZ          4.25     11  49.4 49.1  0.263    0.309 -3.94
##  5 CA         26        55  63.5 34.3  0.244   29.2    3.16
##  6 CO         11         9  55.0 41.6  0.161   13.4    2.41
##  7 CT         22         7  59.3 39.2  0.129   20.1   -1.93
##  8 DC         89         3  92.1  5.40 0.491   86.8   -2.25
##  9 DE         22         3  58.7 39.8  0.0780  19.0   -3.03
## 10 FL          0.0800   29  47.9 51.2  0.0835  -3.36  -3.44
## # i 41 more rows
```

Assessing the prediction error

Bias: average prediction error

```
mean(poll_pred$errors)
```

```
## [1] -3.98
```

Assessing the prediction error

Bias: average prediction error

```
mean(poll_pred$errors)
```

```
## [1] -3.98
```

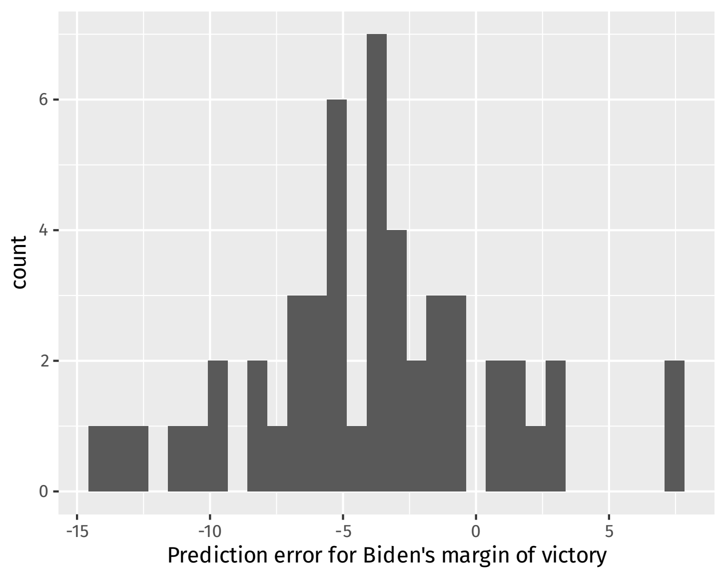
Root mean-square error: average magnitude of the prediction error

```
sqrt(mean(poll_pred$errors^2))
```

```
## [1] 6.07
```

Histogram of the errors

```r
ggplot(poll_pred, aes(x = errors)) +
  geom_histogram() +
  labs(
    x = "Prediction error for Biden's margin of victory"
  )
```



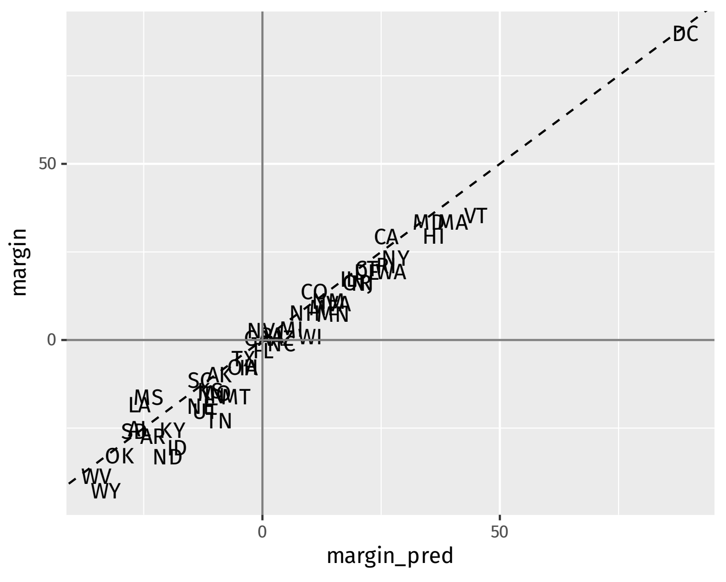
count
6
4
2
0
-15
-10
-5
0
5
Prediction error for Biden's margin of victory

Comparing polls to outcome

Sometimes we want plot text labels instead of point and we use `geom_text` and the `label` aesthetic:

```
## merge the actual results
ggplot(poll_pred, aes(x = margin_pred, y = margin)) +
  geom_text(aes(label = state)) +
  geom_abline(xintercept = 0, slope = 1, linetype = 2) +
  geom_hline(yintercept = 0, color = "grey50") +
  geom_vline(xintercept = 0, color = "grey50")
```

Comparing polls to outcome



Classification

Election prediction: need to predict winner in each state:

```r
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```r
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification
  1. **true positive**: predict Trump wins when he actually wins.

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification
    1. **true positive**: predict Trump wins when he actually wins.
    2. **false positive**: predict Trump wins when he actually loses.

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification
    1. **true positive**: predict Trump wins when he actually wins.
    2. **false positive**: predict Trump wins when he actually loses.
    3. **true negative**: predict Trump loses when he actually loses.

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification
    1. **true positive**: predict Trump wins when he actually wins.
    2. **false positive**: predict Trump wins when he actually loses.
    3. **true negative**: predict Trump loses when he actually loses.
    4. **false negative**: predict Trump loses when he actually wins.

Classification

Election prediction: need to predict winner in each state:

```
poll_pred |>
  filter(margin > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 306
```

```
poll_pred |>
  filter(margin_pred > 0) |>
  summarize(sum(ev)) |> pull()
```

```
## [1] 328
```

- Prediction of binary outcome variable = **classification problem**
- Wrong prediction $\rightsquigarrow$ misclassification
  1. **true positive**: predict Trump wins when he actually wins.
  2. **false positive**: predict Trump wins when he actually loses.
  3. **true negative**: predict Trump loses when he actually loses.
  4. **false negative**: predict Trump loses when he actually wins.
- Sometimes false negatives are more/less important: e.g., civil war.

Classification based on polls

Accuracy: `sign()` returns 1 for a positive number, `-1` for a negative number, and `0` for 0.

```
poll_pred |>
  summarize(prop_correct = mean(sign(margin_pred) == sign(margin))) |>
  pull()
```

```
## [1] 0.922
```

Classification based on polls

Accuracy: `sign()` returns 1 for a positive number, `-1` for a negative number, and `0` for 0.

```
poll_pred |>
  summarize(prop_correct = mean(sign(margin_pred) == sign(margin))) |>
  pull()
```

```
## [1] 0.922
```

Which states did polls call wrong?

```
poll_pred |>
  filter(sign(margin_pred) != sign(margin))
```

```
## # A tibble: 4 x 8
##   state margin_pred    ev biden trump  other margin errors
##   <chr>       <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1 FL         0.0800    29  47.9  51.2 0.0835 -3.36   -3.44
## 2 GA        -1.15      16  49.5  49.2 0.0759  0.236   1.39
## 3 NC         3.95      15  48.6  49.9 0.296  -1.35   -5.30
## 4 NV        -0.350      6  50.1  47.7 0.759   2.39    2.74
```