Hypothesis testing

Section 6

Sima Biondi

Spring 2025

Gov 51: Data Analysis and Politics

Overview

Housekeeping

- Reach out to me if you don't have a group for the final project
  - OH if you have questions about the project or about finding a group
- My midterm review session → March 11st 4:00 PM - 6:00 PM in CGIS K105
- Midterm → March 13th
  - 50% Conceptual + 50% Coding
  - Exam is open note (personal materials, course materials) but not open-internet

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

→ Use penalized regressions like Lasso which adds a penalty term to our loss function to only keep relevant covariates in our model

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

→ Use penalized regressions like Lasso which adds a penalty term to our loss function to only keep relevant covariates in our model

→ Tackles two problems: high-dimensionality and overfitting

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

→ Use penalized regressions like Lasso which adds a penalty term to our loss function to only keep relevant covariates in our model

→ Tackles two problems: high-dimensionality and overfitting

→ Lasso is better than ridge regression but isn't perfect: can induce bias if OLS assumptions are met

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

→ Use penalized regressions like Lasso which adds a penalty term to our loss function to only keep relevant covariates in our model

→ Tackles two problems: high-dimensionality and overfitting

→ Lasso is better than ridge regression but isn't perfect: can induce bias if OLS assumptions are met

Roadmap

Last week: how to estimate *ground-truth* relationships with lots of predictors

- → Use penalized regressions like Lasso which adds a penalty term to our loss function to only keep relevant covariates in our model
- → Tackles two problems: high-dimensionality and overfitting
- → Lasso is better than ridge regression but isn't perfect: can induce bias if OLS assumptions are met

In this section, we try to quantify the degree to which our results are due to randomness

Regression again, with confidence

Linear regression is a tool to help us estimate the relationship between two variables

Regression again, with confidence

Linear regression is a tool to help us estimate the relationship between two variables

$\hookrightarrow$ regression provides an numerical estimate ($\widehat{\beta}_0$ and $\widehat{\beta}_1$)

Regression again, with confidence

Linear regression is a tool to help us estimate the relationship between two variables

$\hookrightarrow$ regression provides an numerical estimate ($\widehat{\beta}_0$ and $\widehat{\beta}_1$)

Regression again, with confidence

Linear regression is a tool to help us estimate the relationship between two variables

$\hookrightarrow$ regression provides an numerical estimate ($\widehat{\beta}_0$ and $\widehat{\beta}_1$)

Problem: how do we know our results ($\widehat{\beta}_0$ and $\widehat{\beta}_1$) are not just due to randomness?

Regression again, with confidence

Linear regression is a tool to help us estimate the relationship between two variables

↪ regression provides an numerical estimate ($\widehat{\beta}_0$ and $\widehat{\beta}_1$)

Problem: how do we know our results ($\widehat{\beta}_0$ and $\widehat{\beta}_1$) are not just due to randomness?

↪ More precisely, how sure can we be of our $\widehat{\beta}_0$ and $\widehat{\beta}_1$ estimates match up to β_0 and β_1?

How do we read the table?

	(1)
(Intercept)	0.247***
	(0.007)
inc	0.391***
	(0.011)
Num.Obs.	1154
R2	0.513
R2 Adj.	0.513
Log.Lik.	329.063
F	1215.638
RMSE	0.18

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Variables: inc (incumbency), outcome (vote share in congressional election)

Statistical hypothesis testing

- **Solution**: hypothesis testing!
  - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
  - **Null hypothesis** (H_0): “a hypothesis we would like to refute” (Imai 2017)

Statistical hypothesis testing

- **Solution**: hypothesis testing!
  - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
  - **Null hypothesis** (H_0): "a hypothesis we would like to refute" (Imai 2017)
    - "some statement about the population parameters" (Blackwell 2022)

Statistical hypothesis testing

- **Solution**: hypothesis testing!
  - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
  - **Null hypothesis** (H_0): "a hypothesis we would like to refute" (Imai 2017)
    - "some statement about the population parameters" (Blackwell 2022)
    - Ex: incumbency has no effect on election results

Statistical hypothesis testing

- **Solution**: hypothesis testing!
    - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
    - **Null hypothesis** (H_0): “a hypothesis we would like to refute” (Imai 2017)
        - “some statement about the population parameters” (Blackwell 2022)
        - Ex: incumbency has no effect on election results
    - **Alternative hypothesis** (H_1 or H_2 or H_3...): a hypothesis we would like to prove

Statistical hypothesis testing

- **Solution**: hypothesis testing!
  - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
  - **Null hypothesis** (H_0): "a hypothesis we would like to refute" (Imai 2017)
    - "some statement about the population parameters" (Blackwell 2022)
    - Ex: incumbency has no effect on election results
  - **Alternative hypothesis** (H_1 or H_2 or H_3...): a hypothesis we would like to prove
    - "the statement we hope or suspect is true" (Blackwell 2022)

Statistical hypothesis testing

- **Solution**: hypothesis testing!
  - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
  - **Null hypothesis** (H_0): "a hypothesis we would like to refute" (Imai 2017)
    - "some statement about the population parameters" (Blackwell 2022)
    - Ex: incumbency has no effect on election results
  - **Alternative hypothesis** (H_1 or H_2 or H_3...): a hypothesis we would like to prove
    - "the statement we hope or suspect is true" (Blackwell 2022)
    - Ex: incumbency has an effect on election results

Statistical hypothesis testing

- **Solution**: hypothesis testing!
    - A statistical thought experiment that compares what would have happened to what did happen
- Key concepts
    - **Null hypothesis** (H_0): "a hypothesis we would like to refute" (Imai 2017)
        - "some statement about the population parameters" (Blackwell 2022)
        - Ex: incumbency has no effect on election results
    - **Alternative hypothesis** (H_1 or H_2 or H_3...): a hypothesis we would like to prove
        - "the statement we hope or suspect is true" (Blackwell 2022)
        - Ex: incumbency has an effect on election results
        - But there are many other potential examples!

Example of null and alternative hypotheses

Question: What is the effect of incumbency on vote share?

- H_0: A candidate's vote share does not change depending on incumbency status

$$\widehat{\beta}_1 = \overline{Y}(1) - \overline{Y}(0) = 0$$

Example of null and alternative hypotheses

Question: What is the effect of incumbency on vote share?

- H_0: A candidate's vote share does not change depending on incumbency status

$$\widehat{\beta}_1 = \overline{Y}(1) - \overline{Y}(0) = 0$$

- H_1: A candidate's vote share does change depending on incumbency status

$$\widehat{\beta}_1 = \overline{Y}(1) - \overline{Y}(0) \neq 0$$

How do we test?

- We need to try to disprove the null hypothesis to argue there is a treatment effect.

How do we test?

- We need to try to disprove the null hypothesis to argue there is a treatment effect.
- Assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?

How do we test?

- We need to try to disprove the null hypothesis to argue there is a treatment effect.
- Assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?
    - We can sort of guess - probably not scientific

How do we test?

- We need to try to disprove the null hypothesis to argue there is a treatment effect.
- Assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?
    - We can sort of guess - probably not scientific
    - We can set a cutoff - also probably not very systematic

How do we test?

- We need to try to disprove the null hypothesis to argue there is a treatment effect.
- Assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?
    - We can sort of guess - probably not scientific
    - We can set a cutoff - also probably not very systematic
    - We can simulate!

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it?

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it? Or how unlikely is the null hypothesis given that $\widehat{\beta}_1$ exists?

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it? Or how unlikely is the null hypothesis given that $\widehat{\beta}_1$ exists?

Remember: Random variables, like our $\widehat{\beta}_1$, have nice properties!

- The distribution of averages (or expectations) approximate a **normal distribution**

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it? Or how unlikely is the null hypothesis given that $\widehat{\beta}_1$ exists?

Remember: Random variables, like our $\widehat{\beta}_1$, have nice properties!

- The distribution of averages (or expectations) approximate a **normal distribution**
  - $\Rightarrow$ Central Limit Theorem,

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it? Or how unlikely is the null hypothesis given that $\widehat{\beta}_1$ exists?

Remember: Random variables, like our $\widehat{\beta}_1$, have nice properties!

- The distribution of averages (or expectations) approximate a **normal distribution**
  - $\Rightarrow$ Central Limit Theorem,

What is the null distribution?

We use simulation to show, assuming the null hypothesis is true, what values would $\widehat{\beta}_1$ have to be in order to disprove it? Or how unlikely is the null hypothesis given that $\widehat{\beta}_1$ exists?

Remember: Random variables, like our $\widehat{\beta}_1$, have nice properties!

- The distribution of averages (or expectations) approximate a **normal distribution**
  - $\Rightarrow$ Central Limit Theorem, decomposition is outside the scope of this course
- Simulating can give a rough idea of expected values under the null hypothesis

Leveraging the null distribution

Steps:

1. Specify a null hypothesis

Leveraging the null distribution

Steps:

1. Specify a null hypothesis
2. Use properties of existing data, namely its variation, to estimate likely values for $\widehat{\beta}_1$

Leveraging the null distribution

Steps:

1. Specify a null hypothesis
2. Use properties of existing data, namely its variation, to estimate likely values for $\widehat{\beta}_1$
    - Leverage the null distribution (the distribution of $\widehat{\beta}_1$ if the null hypothesis is true) to test the likelihood of $\widehat{\beta}_1$ occuring if the null hypothesis is true

Leveraging the null distribution

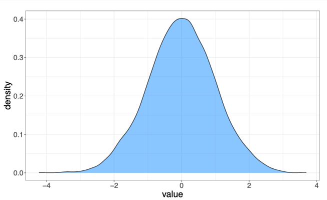
Steps:

1. Specify a null hypothesis
2. Use properties of existing data, namely its variation, to estimate likely values for $\widehat{\beta}_1$
    - Leverage the null distribution (the distribution of $\widehat{\beta}_1$ if the null hypothesis is true) to test the likelihood of $\widehat{\beta}_1$ occuring if the null hypothesis is true

Leveraging the null distribution

Steps:

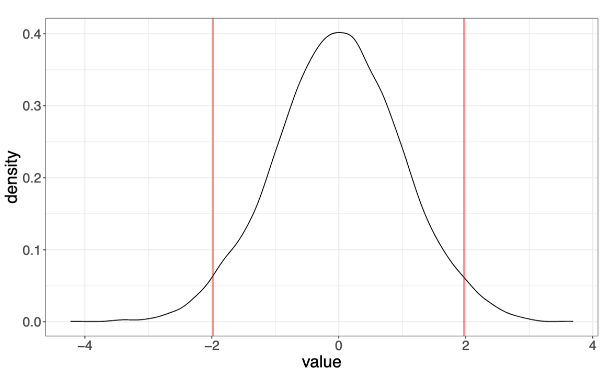
1. Specify a null hypothesis
2. Use properties of existing data, namely its variation, to estimate likely values for $\widehat{\beta}_1$
   - Leverage the null distribution (the distribution of $\widehat{\beta}_1$ if the null hypothesis is true) to test the likelihood of $\widehat{\beta}_1$ occuring if the null hypothesis is true



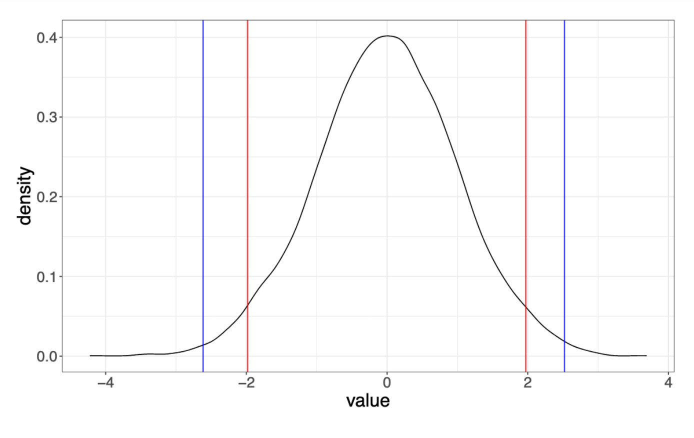
p-value

- We have a distribution of $\widehat{\beta}_1$ values under the null - now what?
- Generally reject the null when our estimates are extreme in the null distribution
- **p-value**: the probability under the null hypothesis that we observe data as extreme as ours
- Scientific community cutoffs to reject the null
  - $p < .05 \approx$ *
  - $p < .01 \approx$ **
  - $p < .001 \approx$ ***
- Norms change - in the past, $p < .1$ was significant, now it's $p < .05$
- Quick check: relationship between p-value and α (not β_0?

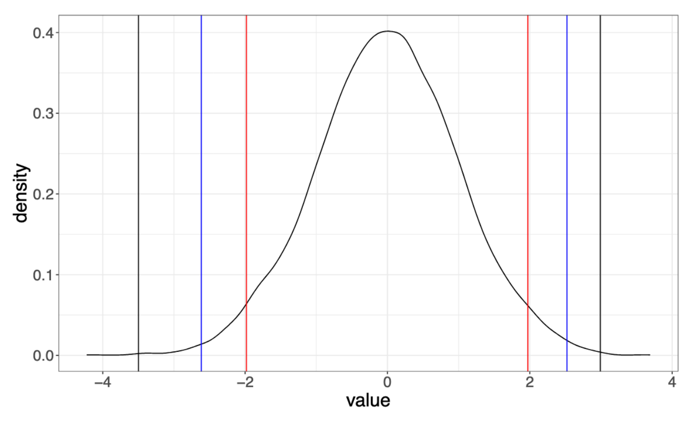
# p < .05



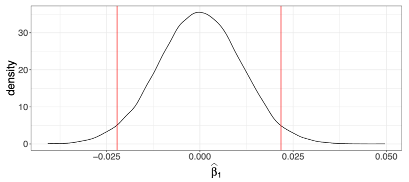
# p < .01



# p < .001



Estimating the null for congressional elections



- Back to our incumbency question, what estimate of a treatment effect is sufficient to reject the null
  - If our estimate is positive, 0.0217468 is a sufficient $\widehat{\beta}_1$
- Instead of a normal, we use a Student T-Distribution, which is a bit more conservative

In table form

```r
model <- lm(voteshare ~ inc, data = cand20)
summary(model
```

	(1)
(Intercept)	0.247***
	(0.007)
inc	0.391***
	(0.011)
Num.Obs.	1154
R2	0.513
R2 Adj.	0.513
Log.Lik.	329.063
F	1215.638
RMSE	0.18

$+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Error Types

- Type I Error: Reject null hypothesis when it is true

Error Types

- Type I Error: Reject null hypothesis when it is true

Error Types

- Type I Error: Reject null hypothesis when it is true (False Positive)

Error Types

- Type I Error: Reject null hypothesis when it is true (False Positive)
- Type II Error: Accept the null when it is false

Error Types

- Type I Error: Reject null hypothesis when it is true (False Positive)
- Type II Error: Accept the null when it is false

Error Types

- Type I Error: Reject null hypothesis when it is true (False Positive)
- Type II Error: Accept the null when it is false (False Negative)

Error Types

- Type I Error: Reject null hypothesis when it is true (False Positive)
- Type II Error: Accept the null when it is false (False Negative)
- What situations might Type I or Type II errors be worse?

Quick recap

- Hypothesis testing provides framework to evaluate $\widehat{\beta}$ point estimates

Quick recap

- Hypothesis testing provides framework to evaluate $\widehat{\beta}$ point estimates
  - Provides evidence against null hypothesis of no effect

Quick recap

- Hypothesis testing provides framework to evaluate $\widehat{\beta}$ point estimates
  - Provides evidence against null hypothesis of no effect
- Relies on properties like Central Limit Theorem to estimate a null distribution

Quick recap

- Hypothesis testing provides framework to evaluate $\widehat{\beta}$ point estimates
  - Provides evidence against null hypothesis of no effect
- Relies on properties like Central Limit Theorem to estimate a null distribution
- Use cut-offs to determine when to reject a null hypothesis

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share
    - Strategic behavior by politicians

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share
  - Strategic behavior by politicians
  - Strategic behavior by challengers

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share
  - Strategic behavior by politicians
  - Strategic behavior by challengers
- Is it identified?

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share
  - Strategic behavior by politicians
  - Strategic behavior by challengers
- Is it identified?

Drawbacks to our congressional elections example

- Confounders in relationship between incumbency and vote share
  - Strategic behavior by politicians
  - Strategic behavior by challengers
- Is it identified? No
  - Further reading: Gelman and King (1990), Levitt and Wolfram (1997), Ansolabehere, Snyder, and Stewart (2000)

Quick note on bias-variance tradeoff

- Decomposition from lecture
    1. Increase model complexity: better in-sample, overfitting concerns
    2. Decrease model complexity: worse in-sample, better out-of-sample
- Similar to Lasso unit problems

Implementation in R for regression

- Mean squared error is a combination of our variance and bias
    - Decomposition gets us a bias quantity that includes an impossible to know parameter β
- We can use an estimator to get our best estimate

```
mean(model$ˆresiduals2)
## [1] 0.03310148
```

Notation

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \rightarrow \text{True } \beta\text{'s}$$

$$Y_i = \boxed{\widehat{\beta}_0 + \widehat{\beta}_1 X_i} + \epsilon_i \rightarrow \text{Estimated } \beta\text{'s}$$

$$\widehat{Y}_i = \boxed{\widehat{\beta}_0 + \widehat{\beta}_1 X_i} \rightarrow \text{Relationship between estimated } \beta\text{'s and predicted } \widehat{Y}_i$$

Recap

- Hypothesis testing is incredibly important
  - Contextualize effect estimates in OLS
- Types of errors in hypothesis testing
- Bias and variance make up Mean Squared Error (MSE) - their tradeoff
- Midterm on March 13th
- Review on March 11st
  - CGIS K105