GOV 51 Section

Week 6: Uncertainty

Pranav Moudgalya

Harvard College

Agenda

- Housekeeping
- Hypothesis Testing Review
- Conclusion

Housekeeping

- Problem set 1 returned tomorrow
- Material: will cover material until Lecture 10
- Acceptable materials: Both parts are semi-closed book
- Place: you can take the exam either (1) online, OR (2) in-person in CGIS South 020
- Submission: on Gradescope and if you're taking it in-person submit the paper copy
- Midterm review session: 3/11 from 4-6 pm in CGIS Knafel 105

Statistical Inference

- A **sample statistic**, $T(X)$, is a scalar value we derive from the sample (i.e., a function of the data $X = \{x_1, x_2, \ldots, x_n\}$) to summarize the data in some way. For example, $\hat{p}$, $\bar{x}$, s, etc.

Statistical Inference

- A **sample statistic**, $T(X)$, is a scalar value we derive from the sample (i.e., a function of the data $X = \{x_1, x_2, \ldots, x_n\}$) to summarize the data in some way. For example, $\hat{p}$, $\bar{x}$, s, etc.
- The **expected value** of a sample statistic, $\mathbb{E}[T(X))]$, is the average value of the statistic across repeated samples.

Statistical Inference

- A **sample statistic**, $T(X)$, is a scalar value we derive from the sample (i.e., a function of the data $X = \{x_1, x_2, \ldots, x_n\}$) to summarize the data in some way. For example, $\hat{p}$, $\bar{x}$, s, etc.
- The **expected value** of a sample statistic, $\mathbb{E}[T(X))]$, is the average value of the statistic across repeated samples.
  - **Note**: *Expected value* is distinct from the *sample mean*. The sample mean, $\bar{x}$ is a function of the sample data, $X = \{x_1, x_2, \ldots, x_n\}$, meaning it is computed from a sample of size n. However, the *expected value* is the average **across many samples!** Formally, that means we take the average across $X_1, X_2, \ldots, X_N$, where $X_n = \{x_1, x_2, \ldots, x_n\}$. We can get a single *sample mean* from a single sample and then get the *expected value* across samples.

Statistical Inference

- A **sample statistic**, $T(X)$, is a scalar value we derive from the sample (i.e., a function of the data $X = \{x_1, x_2, \ldots, x_n\}$) to summarize the data in some way. For example, $\hat{p}$, $\bar{x}$, s, etc.
- The **expected value** of a sample statistic, $\mathbb{E}[T(X))]$, is the average value of the statistic across repeated samples.
  - **Note**: *Expected value* is distinct from the *sample mean*. The sample mean, $\bar{x}$ is a function of the sample data, $X = \{x_1, x_2, \ldots, x_n\}$, meaning it is computed from a sample of size n. However, the *expected value* is the average **across many samples!** Formally, that means we take the average across $X_1, X_2, \ldots, X_N$, where $X_n = \{x_1, x_2, \ldots, x_n\}$. We can get a single *sample mean* from a single sample and then get the *expected value* across samples.
- The **standard error** is the standard deviation of the sample statistic across repeated samples. (In other words, $\sqrt{\mathbb{V}[T(X)]}$)

Statistical Inference

- A **sample statistic**, $T(X)$, is a scalar value we derive from the sample (i.e., a function of the data $X = \{x_1, x_2, \ldots, x_n\}$) to summarize the data in some way. For example, $\hat{p}$, $\bar{x}$, s, etc.
- The **expected value** of a sample statistic, $\mathbb{E}[T(X))]$, is the average value of the statistic across repeated samples.
  - **Note**: *Expected value* is distinct from the *sample mean*. The sample mean, $\bar{x}$ is a function of the sample data, $X = \{x_1, x_2, \ldots, x_n\}$, meaning it is computed from a sample of size n. However, the *expected value* is the average **across many samples!** Formally, that means we take the average across $X_1, X_2, \ldots, X_N$, where $X_n = \{x_1, x_2, \ldots, x_n\}$. We can get a single *sample mean* from a single sample and then get the *expected value* across samples.
- The **standard error** is the standard deviation of the sample statistic across repeated samples. (In other words, $\sqrt{\mathbb{V}[T(X)]}$)
  - **Note**: *Standard error* is distinct from the *sample standard deviation*. The sample standard deviation, s is a function of the sample data, $X = \{x_1, x_2, \ldots, x_n\}$, meaning it is computed from a sample of size n.

Sampling Distributions

- Suppose $X_1, X_2, \ldots, X_n$ is an SRS from a population distribution with mean μ and variance σ^2.

Sampling Distributions

- Suppose $X_1, X_2, \ldots, X_n$ is an SRS from a population distribution with mean μ and variance σ^2.

$$\text{Sample Mean} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sampling Distributions

- Suppose $X_1, X_2, \ldots, X_n$ is an SRS from a population distribution with mean μ and variance σ^2.

$$\text{Sample Mean} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- This is a **random variable with a distribution**!
- The distribution of $\bar{X}_n$ has:
  - **Expected value** of μ
  - **Standard error** (remember that's the s.d. of $\bar{X}_n$!) of $\sigma/\sqrt{n}$

Understanding these Distributions

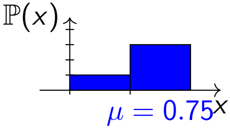


Figure 1: Population

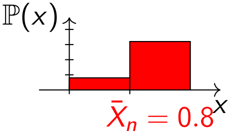


Figure 2: One Sample

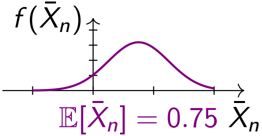


Figure 3: Sampling

Understanding these Distributions

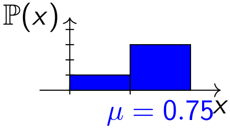


Figure 1: Population

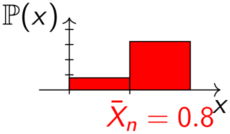


Figure 2: One Sample

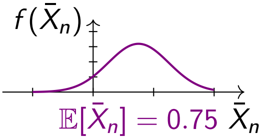


Figure 3: Sampling

- The **population distribution** is in Figure 1. It shows the proportion of 0s and 1s in the population.

Understanding these Distributions

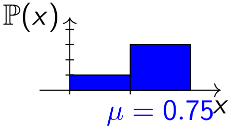


Figure 1: Population

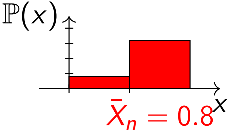


Figure 2: One Sample

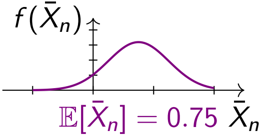


Figure 3: Sampling

- The **population distribution** is in Figure 1. It shows the proportion of 0s and 1s in the population.
- We take **samples** from the distribution to get a sample, which has its own distribution as in Figure 2. Note this changes from sample to sample!

Understanding these Distributions

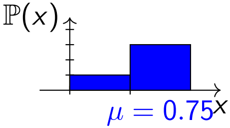


Figure 1: Population

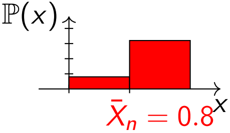


Figure 2: One Sample

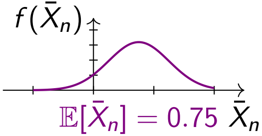


Figure 3: Sampling

- The **population distribution** is in Figure 1. It shows the proportion of 0s and 1s in the population.
- We take **samples** from the distribution to get a sample, which has its own distribution as in Figure 2. Note this changes from sample to sample!
- We summarize the distribution of the **sample means** from all samples in Figure 3, the **sampling distribution**! Note that the sampling distribution is continuous even though the population and sample distributions are discrete because means can take on any continuous value!

Understanding these Distributions

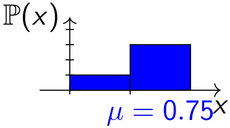


Figure 1: Population

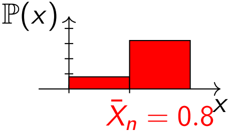


Figure 2: One Sample

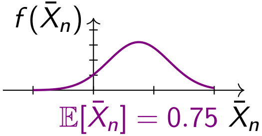


Figure 3: Sampling

- The **population distribution** is in Figure 1. It shows the proportion of 0s and 1s in the population.
- We take **samples** from the distribution to get a sample, which has its own distribution as in Figure 2. Note this changes from sample to sample!
- We summarize the distribution of the **sample means** from all samples in Figure 3, the **sampling distribution**! Note that the sampling distribution is continuous even though the population and sample distributions are discrete because means can take on any continuous value!
- **Intuition**: The sampling distribution will be centered about the population distribution mean because many of our samples will have means close to the true population proportion, but some may be extremely off (very few, though!)

Understanding these Distributions

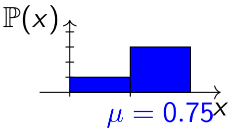


Figure 1: Population

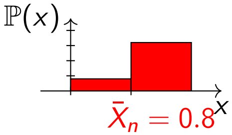


Figure 2: One Sample

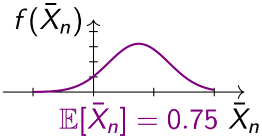


Figure 3: Sampling

- The **population distribution** is in Figure 1. It shows the proportion of 0s and 1s in the population.
- We take **samples** from the distribution to get a sample, which has its own distribution as in Figure 2. Note this changes from sample to sample!
- We summarize the distribution of the **sample means** from all samples in Figure 3, the **sampling distribution**! Note that the sampling distribution is continuous even though the population and sample distributions are discrete because means can take on any continuous value!
- **Intuition**: The sampling distribution will be centered about the population distribution mean because many of our samples will have means close to the true population proportion, but some may be extremely off (very few, though!)

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.
  (3) Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.
  (3) Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  (4) Compute the test statistic's probability under the null from reference distribution.

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.
  (3) Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  (4) Compute the test statistic's probability under the null from reference distribution.
  (5) Conclude whether you have sufficient evidence to reject the null, based on a confidence level. **IS THE OBSERVED STATISTIC SO INCONSISTENT WITH WHAT WE'D EXPECT UNDER THE NULL THAT WE CAN REJECT**

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  1. Specify a test statistic.
  2. Specify a null hypothesis.
  3. Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  4. Compute the test statistic's probability under the null from reference distribution.
  5. Conclude whether you have sufficient evidence to reject the null, based on a confidence level. **IS THE OBSERVED STATISTIC SO INCONSISTENT WITH WHAT WE'D EXPECT UNDER THE NULL THAT WE CAN REJECT**
- Statistical hypothesis testing is a **thought experiment**. We quantify the probability that we observed a result as extreme as we actually retrieved in our sample under the null hypothesis due to *random chance alone*.

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.
  (3) Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  (4) Compute the test statistic's probability under the null from reference distribution.
  (5) Conclude whether you have sufficient evidence to reject the null, based on a confidence level. **IS THE OBSERVED STATISTIC SO INCONSISTENT WITH WHAT WE'D EXPECT UNDER THE NULL THAT WE CAN REJECT**
- Statistical hypothesis testing is a **thought experiment**. We quantify the probability that we observed a result as extreme as we actually retrieved in our sample under the null hypothesis due to *random chance alone*.
- **Null hypothesis**: Some statement about the population parameters (often H_0 : Statistic $= 0$).

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  (1) Specify a test statistic.
  (2) Specify a null hypothesis.
  (3) Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  (4) Compute the test statistic's probability under the null from reference distribution.
  (5) Conclude whether you have sufficient evidence to reject the null, based on a confidence level. **IS THE OBSERVED STATISTIC SO INCONSISTENT WITH WHAT WE'D EXPECT UNDER THE NULL THAT WE CAN REJECT**
- Statistical hypothesis testing is a **thought experiment**. We quantify the probability that we observed a result as extreme as we actually retrieved in our sample under the null hypothesis due to *random chance alone*.
- **Null hypothesis**: Some statement about the population parameters (often H_0 : Statistic $= 0$).
- **Alternative hypothesis**: The statement we hope or suspect is true instead of H_0.

Hypothesis Testing

- What is a hypothesis test in a nutshell?
  1. Specify a test statistic.
  2. Specify a null hypothesis.
  3. Approximate the reference distribution – distribution of test statistic under the null (i.e., permutation)
  4. Compute the test statistic's probability under the null from reference distribution.
  5. Conclude whether you have sufficient evidence to reject the null, based on a confidence level. **IS THE OBSERVED STATISTIC SO INCONSISTENT WITH WHAT WE'D EXPECT UNDER THE NULL THAT WE CAN REJECT**
- Statistical hypothesis testing is a **thought experiment**. We quantify the probability that we observed a result as extreme as we actually retrieved in our sample under the null hypothesis due to *random chance alone*.
- **Null hypothesis**: Some statement about the population parameters (often H_0 : Statistic $= 0$).
- **Alternative hypothesis**: The statement we hope or suspect is true instead of H_0.

p-value

- The *p*-**value** is the probability of observing data as or more extreme as our data if the null hypothesis is true.

p-value

- The p-**value** is the probability of observing data as or more extreme as our data if the null hypothesis is true.
- Low p-value $\rightsquigarrow$ reject the null (dependent on significance level)

p-value

- The *p*-**value** is the probability of observing data as or more extreme as our data if the null hypothesis is true.
- Low *p*-value $\rightsquigarrow$ reject the null (dependent on significance level)
- **NOT** the probability that the null is true!

p-value

- The *p*-**value** is the probability of observing data as or more extreme as our data if the null hypothesis is true.
- Low *p*-value $\rightsquigarrow$ reject the null (dependent on significance level)
- **NOT** the probability that the null is true!
- *p*-values are typically **two-sided**

p-value

- The *p*-**value** is the probability of observing data as or more extreme as our data if the null hypothesis is true.
- Low *p*-value $\rightsquigarrow$ reject the null (dependent on significance level)
- **NOT** the probability that the null is true!
- *p*-values are typically **two-sided**
- Common thresholds for rejecting the null:
  - $p \geq$ "not statistically significant"
  - $p < 0.05$ "statistically significant"
  - $p < 0.01$ "highly significant"

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.
- Test errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.
- Test errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.
- Test errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

- Type I error because it's the worst
  - "Convicting" an innocent null hypothesis

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.
- Test errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

- Type I error because it's the worst
  - "Convicting" an innocent null hypothesis
- Type II error less serious
  - Missed out on an awesome finding

Testing errors

- A p-value of 0.05 says that data this extreme would only happen in 5% of repeated samples if the null were true.
  - $\rightsquigarrow$ 5% of the time we'll reject the null when it is actually true.
- Test errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

- Type I error because it's the worst
  - "Convicting" an innocent null hypothesis
- Type II error less serious
  - Missed out on an awesome finding

What kind of significance?

There are different types of significance that don't all have to be true together:

What kind of significance?

There are different types of significance that don't all have to be true together:

1. **Statistical significance:** we can reject the null of no effect.

What kind of significance?

There are different types of significance that don't all have to be true together:

1. **Statistical significance:** we can reject the null of no effect.
2. **Causal significance**: we can interpret our estimated difference in means as a causal effect.

What kind of significance?

There are different types of significance that don't all have to be true together:

1. **Statistical significance:** we can reject the null of no effect.
2. **Causal significance**: we can interpret our estimated difference in means as a causal effect.
3. **Practical significance**: the estimated effect is meaningfully large.

Regression Again

- Linear regression helps us estimate the relationship between two variables

Regression Again

- Linear regression helps us estimate the relationship between two variables
- **Problem:** how do we know our results are not just due to randomness?

Regression Again

- Linear regression helps us estimate the relationship between two variables
- **Problem:** how do we know our results are not just due to randomness?
- In our regression notation, how sure can we be of our $\widehat{\beta}_0$ and $\widehat{\beta}_1$ estimates match up to β_0 and β_1?

Regression Again

- Linear regression helps us estimate the relationship between two variables
- **Problem:** how do we know our results are not just due to randomness?
- In our regression notation, how sure can we be of our $\widehat{\beta}_0$ and $\widehat{\beta}_1$ estimates match up to β_0 and β_1?
- We know that our $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have a distribution, and we can use this fact to compute a range of plausible values for these estimates.

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

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

Example of Null and Alternative Hypotheses

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

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

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

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

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

Null Distribution

- Now, we need to try to disprove the null hypothesis to argue there is a treatment effect.
- Let's assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?

Null Distribution

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

Null Distribution

- Now, we need to try to disprove the null hypothesis to argue there is a treatment effect.
- Let's 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

Null Distribution

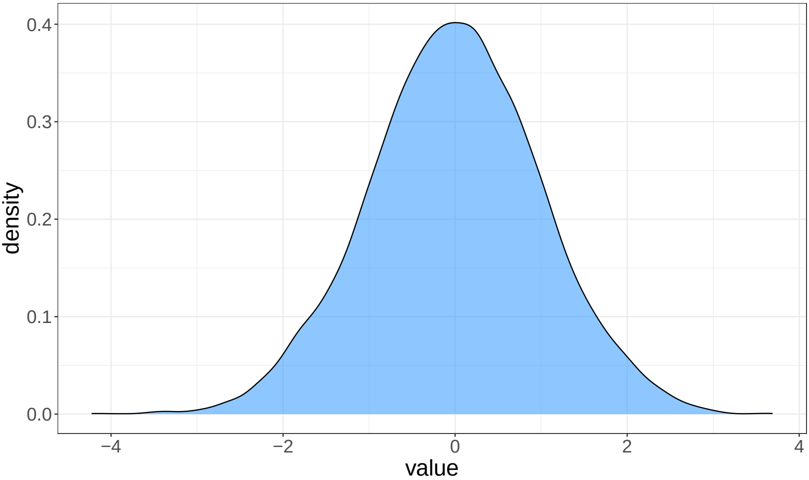
- Now, we need to try to disprove the null hypothesis to argue there is a treatment effect.
- Let's assume the null hypothesis is true - what values of $\widehat{\beta}_1$ disprove it?
  - We can sort of guess - probably not scientific
  - We can set a cutoff - also probably not very systematic
- We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

1. Specify a null hypothesis
2. Use our existing data, namely variation, to estimate likely values we might get

Null Distribution

- Random variables, like our $\widehat{\beta}_1$, have nice properties!
  - The distribution of averages (or expectations) approximate to a **normal distribution**
  - So, simulating can give us a rough idea of what values to expect under the null hypothesis
- This is also known as the Central Limit Theorem
  - Decomposition is outside the scope of this course

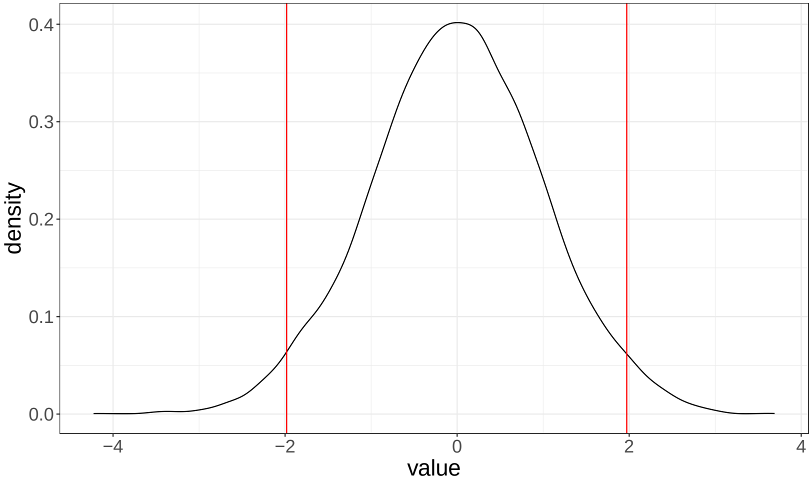
Example Normal Distribution



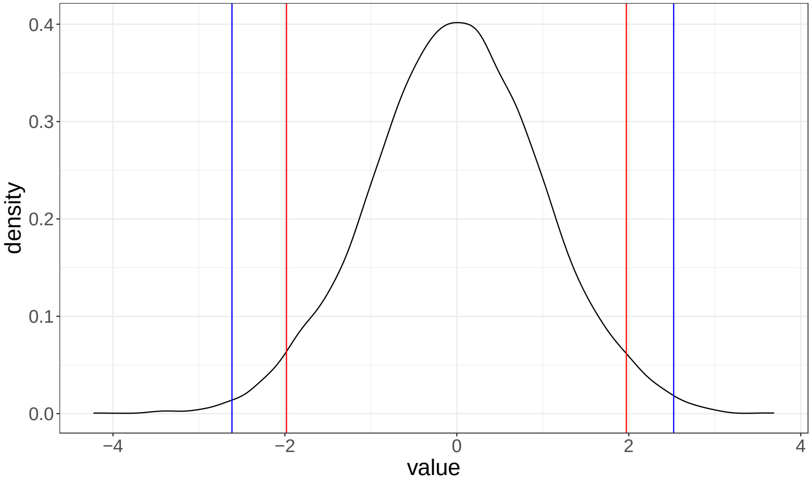
p-value

- We have a distribution of $\widehat{\beta}_1$ values under the null - now what?
- We generally reject the null, when our estimates are at some extremity in the null distribution
- **p-value**: the probability under the null hypothesis, we observe data as extreme as ours
- As a scientific community, we have some arbitrary cutoffs to reject the null
  - p $< .05 \approx$ *
  - p $< .01 \approx$ **
  - p $< .001 \approx$ ***
- Norms change - in the past, p $< .1$ was considered significant, but lower bound is now p $< .05$
- Relationship between p-value and α

# p < .05



# $p < .01$



# $p < .001$