Political Methodology I

Justin Grimmer

Assistant Professor
Department of Political Science
Stanford University

November 26th, 2012
Hypothesis Tests

Estimators to infer population parameter $\theta$ from data $X$

- **Point estimation:**
  - Given data, single most likely value
  - What happens in large samples? (consistency)
  - On average over repeated sampling? (bias)

- **Interval estimation:**
  - Construct intervals that cover true parameter with some fixed probability
  - **Coverage probability:** how likely is it to cover true parameter value under worst case scenarios?
  - **Length:** how long is the interval
  - **Bootstrap:** nonparametric method for constructing confidence intervals

- **Hypothesis tests**
  - Divide parameter space into two components: null and alternative
  - **Ask:** how weird is it, under null, of observing these data?
  - **Do we have sufficient evidence to reject the null?**

Plan:

Definition $\rightarrow$ Example 1 (psychic) $\rightarrow$ Example 2 (treatment effects)
Definitions

Definition

**Hypothesis**: A *statement about a population parameter*
Definitions

Definition

*Hypothesis:* A *statement about a population parameter*

- Causal inference:
Definitions

Definition

_Hypothesis:_ A statement about a population parameter

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)

- Parameters from samples
  - Is support for Obama above 50%?
  - Is the committee composition weird?
  - Are characteristics of treatment/control groups different?
  - Is there evidence of electoral manipulation?
Hypothesis: A statement about a population parameter

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
Definitions

Definition

Hypothesis: A statement about a population parameter

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)
Definitions

Definition

**Hypothesis**: A *statement about a population parameter*

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)
- Parameters from samples
Definitions

Definition

**Hypothesis**: A *statement about a population parameter*

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)

- Parameters from samples
  - Is support for Obama above 50%?
Definitions

Definition

Hypothesis: A statement about a population parameter

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)

- Parameters from samples
  - Is support for Obama above 50%?
  - Is the committee composition weird
Definitions

Definition

**Hypothesis:** A statement about a population parameter

- **Causal inference:**
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)

- **Parameters from samples**
  - Is support for Obama above 50%?
  - Is the committee composition **weird**
  - Are characteristics of treatment/control groups different?
Definitions

Definition

Hypothesis: A statement about a population parameter

- Causal inference:
  - Does voter mobilization increase turnout? (turnout higher among treatment group?)
  - Do treaties constrain countries? (is behavior distinct among treaty signers?)
  - Do electoral incentives affect redistribution? (are there differences in pork barrel spending?)

- Parameters from samples
  - Is support for Obama above 50%?
  - Is the committee composition weird
  - Are characteristics of treatment/control groups different?
  - Is there evidence of electoral manipulation?
Definitions

Definition

Types of Hypotheses We will divide parameter space into two mutually exclusive components:

- Null hypothesis: $H_0$
- Alternative hypothesis: $H_1$
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Causal inference
Definitions

Definition

**Types of Hypotheses** We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: \( H_0 \)
- **Alternative hypothesis**: \( H_1 \)

- Causal inference
  - Does voter mobilization increase turnout?

- Does treaties constrain countries?
  - \( H_0 \): Treaties do not affect human rights behavior (no difference between treaty/non treaty groups)
  - \( H_1 \): Treaties do affect human rights behavior (difference between treaty/non treaty groups)
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Causal inference
  - Does voter mobilization increase turnout?
    - $H_0$: Mobilization does not increase turnout (no difference between mobilized/non mobilized groups)
Definitions

Definition

**Types of Hypotheses** We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Causal inference
  - Does voter mobilization increase turnout?
    - $H_0$: Mobilization does not increase turnout (no difference between mobilized/non mobilized groups)
    - $H_1$: Mobilization increases turnout (difference between mobilized/non mobilized groups)
Definitions

Definition

Types of Hypotheses We will divide parameter space into two mutually exclusive components:

- **Null hypothesis:** $H_0$
- **Alternative hypothesis:** $H_1$

- Causal inference
  - Does voter mobilization increase turnout?
    - $H_0$: Mobilization does not increase turnout (no difference between mobilized/non mobilized groups)
    - $H_1$: Mobilization increases turnout (difference between mobilized/non mobilized groups)
  - Do treaties constrain countries?
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Causal inference
  - Does voter mobilization increase turnout?
    - $H_0$: Mobilization does not increase turnout (no difference between mobilized/non mobilized groups)
    - $H_1$: Mobilization increases turnout (difference between mobilized/non mobilized groups)
  - Do treaties constrain countries?
    - $H_0$: Treaties do not affect human rights behavior (no difference between treaty/non treaty groups)
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- *Null hypothesis*: $H_0$
- *Alternative hypothesis*: $H_1$

- Causal inference
  - Does voter mobilization increase turnout?
    - $H_0$: Mobilization does not increase turnout (no difference between mobilized/non mobilized groups)
    - $H_1$: Mobilization increases turnout (difference between mobilized/non mobilized groups)
  - Do treaties constrain countries?
    - $H_0$: Treaties do not affect human rights behavior (no difference between treaty/non treaty groups)
    - $H_1$: Treaties do affect human rights behavior (difference between treaty/non treaty groups)
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- *Null hypothesis*: $H_0$
- *Alternative hypothesis*: $H_1$

- Parameters from samples
Definitions

**Definition**

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Parameters from samples
  - Committee composition
Definitions

Definition

Types of Hypotheses We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Parameters from samples
  - Committee composition
    - $H_0$: committee is representative of legislature (no difference between committee characteristics and legislature)
Definitions

Definition

*Types of Hypotheses* We will divide parameter space into two mutually exclusive components:

- **Null hypothesis**: $H_0$
- **Alternative hypothesis**: $H_1$

- Parameters from samples
  - Committee composition
    - $H_0$: committee is representative of legislature (no difference between committee characteristics and legislature)
    - $H_1$: committee is not representative of legislature (difference between committee characteristics and legislature)
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data.
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data
Use a test statistic: $t(X)$
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data

Use a test statistic: \( t(X) \)

Definition

*Hypothesis testing procedure* is a rule that specifies:

1) *Value of test statistic where decision is made to not reject* \( H_0 \)
2) *Value of test statistic where* \( H_0 \) *is rejected*
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data
Use a **test statistic:** \( t(X) \)

**Definition**

*Hypothesis testing procedure is a rule that specifies:*

1) *Value of test statistic where decision is made to not reject \( H_0 \)*
2) *Value of test statistic where \( H_0 \) is rejected*

- Values of \( t(X) \) where \( H_0 \) not rejected: **acceptance region**
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data

Use a test statistic: $t(X)$

Definition

*Hypothesis testing procedure* is a rule that specifies:

1) Value of test statistic where decision is made to not reject $H_0$
2) Value of test statistic where $H_0$ is rejected

- Values of $t(X)$ where $H_0$ not rejected: *acceptance* region
- Values of $t(X)$ where $H_0$ rejected: *rejection* region
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data
Use a test statistic: $t(X)$

Definition

*Hypothesis testing procedure* is a rule that specifies:

1) Value of test statistic where decision is made to not reject $H_0$
2) Value of test statistic where $H_0$ is rejected

- Values of $t(X)$ where $H_0$ not rejected: acceptance region
- Values of $t(X)$ where $H_0$ rejected: rejection region
- Value separating regions: critical value of test
Definitions

Specify a set of rules to adjudicate between null and alternative hypotheses, using observed data.

Use a test statistic: $t(X)$

Definition

*Hypothesis testing procedure is a rule that specifies:*

1) *Value of test statistic where decision is made to not reject $H_0$*
2) *Value of test statistic where $H_0$ is rejected*

- Values of $t(X)$ where $H_0$ not rejected: acceptance region
- Values of $t(X)$ where $H_0$ rejected: rejection region
- Value separating regions: critical value of test

Define the remaining terms in context of example
Extended Example: Prediction

**Forecasting:** can we use information in order to predict the future?

- Election forecasts
- Roll call vote predictions
- Incidence of war

We'll focus on a classic example, in the context of forecasts:
- We flip a large number of coins (use normal distribution)
- Psychic: guesses (heads/tails)
- Inference problem: better than at random?
Extended Example: Prediction

**Forecasting**: can we use information in order to predict the future?
- Election forecasts
Extended Example: Prediction

Forecasting: can we use information in order to predict the future?
- Election forecasts
- Roll call vote predictions
Extended Example: Prediction

Forecasting: can we use information in order to predict the future?
- Election forecasts
- Roll call vote predictions
- Incidence of war
Extended Example: Prediction

**Forecasting**: can we use information in order to predict the future?

- Election forecasts
- Roll call vote predictions
- Incidence of war

We’ll focus on a classic example, in the context of forecasts:
Forecasting: can we use information in order to predict the future?
- Election forecasts
- Roll call vote predictions
- Incidence of war

We’ll focus on a classic example, in the context of forecasts:
- We flip a large number of coins (use normal distribution)
Extended Example: Prediction

**Forecasting**: can we use information in order to predict the future?
- Election forecasts
- Roll call vote predictions
- Incidence of war

We’ll focus on a classic example, in the context of forecasts:
- We flip a large number of coins (use normal distribution)
- Psychic: guesses (heads/tails)
Forecasting: can we use information in order to predict the future?
- Election forecasts
- Roll call vote predictions
- Incidence of war

We’ll focus on a classic example, in the context of forecasts:
- We flip a large number of coins (use normal distribution)
- Psychic: guesses (heads/tails)
- Inference problem: better than at random?
Psychic Hotline

- Flip coin and psychic guess

- \( X_i = 1 \) if agree

- \( X_i = 0 \) if disagree

- \( X = (X_1, X_2, \ldots, X_N) \), parameter of interest

- \( \pi = \text{probability of true guess} \)

- \( \pi = 0.5 \) guessing at random

- \( \pi > 0.5 \) psychic

- \( \pi < 0.5 \) bad at job
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
  - $X = (X_1, X_2, \ldots, X_N)$
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
- $X = (X_1, X_2, \ldots, X_N)$
- Parameter of interest $\pi = \text{probability of true guess}$
Psychic Hotline

- Flip coin and psychic guess
  - If agree then \( X_i = 1 \)
  - If disagree then \( X_i = 0 \)
- \( X = (X_1, X_2, \ldots, X_N) \)
- Parameter of interest \( \pi = \) probability of true guess
  - \( \pi = 0.5 \) guessing at random

- Null and Alternative hypothesis:
  - \( H_0: \pi \leq 0.5 \)
  - \( H_1: \pi > 0.5 \)
Psychic Hotline

- Flip coin and psychic guess
  - If agree then \( X_i = 1 \)
  - If disagree then \( X_i = 0 \)
- \( X = (X_1, X_2, \ldots, X_N) \)
- Parameter of interest \( \pi = \) probability of true guess
  - \( \pi = 0.5 \) guessing at random
  - \( \pi > 0.5 \) psychic
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
  - $X = (X_1, X_2, \ldots, X_N)$
  - Parameter of interest $\pi = \text{probability of true guess}$
    - $\pi = 0.5$ guessing at random
    - $\pi > 0.5$ psychic
    - $\pi < 0.5$ bad at job
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
- $X = (X_1, X_2, \ldots, X_N)$
- Parameter of interest $\pi =$ probability of true guess
  - $\pi = 0.5$ guessing at random
  - $\pi > 0.5$ psychic
  - $\pi < 0.5$ bad at job

- Null and Alternative hypothesis:
Psychic Hotline

- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
  - $X = (X_1, X_2, \ldots, X_N)$
  - Parameter of interest $\pi = \text{probability of true guess}$
    - $\pi = 0.5$ guessing at random
    - $\pi > 0.5$ psychic
    - $\pi < 0.5$ bad at job

- Null and Alternative hypothesis:
  - $H_0 : \pi \leq 0.5$
- Flip coin and psychic guess
  - If agree then $X_i = 1$
  - If disagree then $X_i = 0$
  - $X = (X_1, X_2, \ldots, X_N)$
  - Parameter of interest $\pi = \text{probability of true guess}$
    - $\pi = 0.5$ guessing at random
    - $\pi > 0.5$ psychic
    - $\pi < 0.5$ bad at job

- Null and Alternative hypothesis:
  - $H_0 : \pi \leq 0.5$
  - $H_1 : \pi > 0.5$
Distribution under Null
We have assumed that $N$ is large.

\[ p(x) \rightarrow \text{Normal}(\pi, \sigma^2 / N) \]

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$
- So this implies that, $X - \pi \hat{\sigma} / \sqrt{N} \sim \text{Normal}(0, 1)$
- Our test statistic $X - \pi \hat{\sigma} / \sqrt{N}$ is used under null hypothesis to fill in value of $\pi$.
- We assume that the null hypothesis is true.

How weird is our test statistic under null?
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.

$$p(\overline{X}) \rightarrow^{d} \text{Normal}(\pi, \sigma^2 / N)$$
Distribution under Null
We have assumed that \( N \) is large. Assume that null is true.

\[
p(\bar{X}) \xrightarrow{d} \text{Normal}(\pi, \sigma^2/N)
\]

- (Either from sampling or maximum likelihood theory)

\[X - \pi \hat{\sigma}/\sqrt{N} \sim \text{Normal}(0, 1)\]

\(X - \pi \hat{\sigma}/\sqrt{N}\) is our test statistic

- We use null hypothesis to fill in value of \( \pi \)

- We assume that the null hypothesis is true

How weird is our test statistic under null?
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.

\[ p(\bar{X}) \xrightarrow{d} \text{Normal}(\pi, \sigma^2 / N) \]

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.

\[ p(\bar{X}) \to^d \text{Normal}(\pi, \sigma^2 / N) \]

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
- So this implies that,

\[ X_i - \pi \hat{\sigma}/\sqrt{N} \sim \text{Normal}(0, 1) \]
Distribution under Null
We have assumed that \( N \) is large. Assume that null is true.

\[
p(\overline{X}) \rightarrow^d \text{Normal}(\pi, \sigma^2/N)
\]

- (Either from sampling or maximum likelihood theory)
- Estimate \( \sigma^2 \) with \( \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})^2 \)
- So this implies that,

\[
\frac{\overline{X} - \pi}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0, 1)
\]
Distribution under Null
We have assumed that $N$ is large. Assume that null is true.

$$p(\bar{X}) \rightarrow d \text{ Normal}(\pi, \sigma^2/N)$$

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
- So this implies that,

$$\frac{\bar{X} - \pi}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0, 1)$$

$\frac{X - \pi}{\hat{\sigma}/\sqrt{N}}$ is our test statistic
Distribution under Null

We have assumed that \( N \) is large. Assume that null is true.

\[
p(\bar{X}) \rightarrow^d \text{Normal}(\pi, \sigma^2/N)
\]

- (Either from sampling or maximum likelihood theory)
- Estimate \( \sigma^2 \) with \( \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \)
- So this implies that,

\[
\frac{\bar{X} - \pi}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0, 1)
\]

\( \frac{\bar{X} - \pi}{\hat{\sigma}/\sqrt{N}} \) is our test statistic

- We use null hypothesis to fill in value of \( \pi \)
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.

$$p(\bar{X}) \rightarrow^d \text{Normal}(\pi, \sigma^2 / N)$$

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
- So this implies that,

$$\frac{\bar{X} - \pi}{\hat{\sigma} / \sqrt{N}} \sim \text{Normal}(0, 1)$$

$\frac{\bar{X} - \pi}{\hat{\sigma} / \sqrt{N}}$ is our test statistic

- We use null hypothesis to fill in value of $\pi$
- We assume that the null hypothesis is true
Distribution under Null

We have assumed that $N$ is large. Assume that null is true.

$$p(\bar{X}) \rightarrow^d \text{Normal}(\pi, \sigma^2/N)$$

- (Either from sampling or maximum likelihood theory)
- Estimate $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N}(X_i - \bar{X})^2$
- So this implies that,

$$\frac{\bar{X} - \pi}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0, 1)$$

$\frac{\bar{X} - \pi}{\hat{\sigma}/\sqrt{N}}$ is our test statistic

- We use null hypothesis to fill in value of $\pi$
- We assume that the null hypothesis is true

How weird is our test statistic under null?
## Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- **Type I Error**: rejecting a null that is true
  - Identify someone as psychic when they are not
- **Type II Error**: failing to reject a null that is false
  - Fail to identify a psychic

**Size of test**: probability of rejecting $H_0$ when $H_0$ is true

**Power of test**: probability of failing to reject $H_0$ when $H_1$ is true

- Set $\alpha$ depending on loss from false discovery
  - Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
  - Conservative values of $\alpha$ guard against paper retractions
  - Nothing magical about 0.05
## Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true

Size of test: probability of rejecting $H_0$ when $H_0$ is true

Power of test: probability of failing to reject $H_0$ when $H_1$ is true

- Set $\alpha$ depending on loss from false discovery
- Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
- Conservative values of $\alpha$ guard against paper retractions
- Nothing magical about 0.05
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true
  - Identify someone as psychic when they are not

Size of test: probability of rejecting $H_0$ when $H_0$ is true

Power of test: probability of failing to reject $H_0$ when $H_1$ is true

- Set $\alpha$ depending on loss from false discovery
  - Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
  - Conservative values of $\alpha$ guard against paper retractions
- Nothing magical about $0.05$. 

Justin Grimmer (Stanford University)
## Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- **Type I Error**: rejecting a null that is true
  - Identify someone as psychic when they are not
- **Type II Error**: failing to reject a null that is false

---

Set $\alpha$ depending on loss from false discovery

- Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
- Conservative values of $\alpha$ guard against paper retractions
- Nothing magical about 0.05.
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true
  - Identify someone as psychic when they are not
- Type II Error: failing to reject a null that is false
  - Fail to identify a psychic
## Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- **Type I Error**: rejecting a null that is true
  - Identify someone as psychic when they are not
- **Type II Error**: failing to reject a null that is false
  - Fail to identify a psychic

**Size of test**: probability of rejecting $H_0$ when $H_0$ is true $\alpha$
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true
  - Identify someone as psychic when they are not
- Type II Error: failing to reject a null that is false
  - Fail to identify a psychic

Size of test: probability of rejecting $H_0$ when $H_0$ is true $\alpha$
Power of test: probability of failing to reject $H_0$ when $H_1$ is true $\beta$
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true
  - Identify someone as psychic when they are not
- Type II Error: failing to reject a null that is false
  - Fail to identify a psychic

Size of test: probability of rejecting $H_0$ when $H_0$ is true $\alpha$

Power of test: probability of failing to reject $H_0$ when $H_1$ is true $\beta$

- Set $\alpha$ depending on loss from false discovery
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- Type I Error: rejecting a null that is true
  - Identify someone as psychic when they are not
- Type II Error: failing to reject a null that is false
  - Fail to identify a psychic

**Size of test**: probability of rejecting $H_0$ when $H_0$ is true $\alpha$

**Power of test**: probability of failing to reject $H_0$ when $H_1$ is true $\beta$

- Set $\alpha$ depending on loss from false discovery
  - Tendency to cheat: if you make $\alpha$ large, you’ll reject often and discover nothing
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- **Type I Error**: rejecting a null that is true
  - Identify someone as psychic when they are not
- **Type II Error**: failing to reject a null that is false
  - Fail to identify a psychic

**Size** of test: probability of rejecting $H_0$ when $H_0$ is true $\alpha$

**Power** of test: probability of failing to reject $H_0$ when $H_1$ is true $\beta$

- Set $\alpha$ depending on loss from false discovery
  - Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
  - Conservative values of $\alpha$ guard against paper retractions
Error Types, Size, and Power

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_1$ True</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- **Type I Error**: rejecting a null that is true
  - Identify someone as psychic when they are not
- **Type II Error**: failing to reject a null that is false
  - Fail to identify a psychic

**Size** of test: probability of rejecting $H_0$ when $H_0$ is true $\alpha$

**Power** of test: probability of failing to reject $H_0$ when $H_1$ is true $\beta$

- Set $\alpha$ depending on loss from false discovery
  - Tendency to cheat: if you make $\alpha$ large, you'll reject often and discover nothing
  - Conservative values of $\alpha$ guard against paper retractions
- **Nothing** magical about 0.05
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:

$$t(X) \equiv \frac{X - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$$

- Critical value ($x_0$): what value is sufficiently weird to reject null?

- $\Pr(t(X) > x_0) = \alpha$

- If $\alpha = 0.1$, critical value is 1.28.
- If $\alpha = 0.05$, critical value is 1.64.
- If $\alpha = 0.01$, critical value is 2.33.
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:

$$-t(X) \equiv \frac{\hat{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$$
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:

- $t(X) \equiv \frac{\overline{X} - 0.5}{\hat{\sigma} / \sqrt{N}} \sim\text{Normal}(0,1)$
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:
- $t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma} / \sqrt{N}} \sim \text{Normal}(0,1)$
- Critical value ($x_0$): what value is sufficiently weird to reject null?
Example of Test Statistic and Critical Region

Assume Null is true, we set \( \pi = 0.5 \).

For our example:

- \( t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1) \)
- Critical value \( (x_0) \): what value is sufficiently weird to reject null?
- \( \Pr(t(X) > x_0) = \alpha \)
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:

- $t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$
- **Critical value** ($x_0$): what value is sufficiently weird to reject null?
- $\Pr(t(X) > x_0) = \alpha$
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:

- $t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$

- Critical value ($x_0$): what value is sufficiently weird to reject null?

- $\Pr(t(X) > x_0) = \alpha$

- If $\alpha = 0.1$ critical value is 1.28
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:
- $t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$
- Critical value ($x_0$): what value is sufficiently weird to reject null?
- $\Pr(t(X) > x_0) = \alpha$

- If $\alpha = 0.1$ critical value is 1.28
- If $\alpha = 0.05$ critical value is 1.64
Example of Test Statistic and Critical Region

Assume Null is true, we set $\pi = 0.5$.

For our example:
- $t(X) \equiv \frac{\bar{X} - 0.5}{\hat{\sigma}/\sqrt{N}} \sim \text{Normal}(0,1)$
- Critical value ($x_0$): what value is sufficiently weird to reject null?
- $\Pr(t(X) > x_0) = \alpha$

\begin{itemize}
  \item If $\alpha = 0.1$ critical value is 1.28
  \item If $\alpha = 0.05$ critical value is 1.64
  \item If $\alpha = 0.01$ critical value is 2.33
\end{itemize}
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard).
What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples
Further suppose that we set $\alpha = 0.05$ (as is standard)
What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value. Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples.
Further suppose that we set $\alpha = 0.05$ (as is standard).
What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples.
Further suppose that we set $\alpha = 0.05$ (as is standard).
What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value. Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true \( \pi = 0.5 \) and we perform repeated samples
Further suppose that we set \( \alpha = 0.05 \) (as is standard)
What happens under repeated sampling?

We reject when \( t(X) > \) critical value
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$.
Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$.
Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$
Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true \( \pi = 0.5 \) and we perform repeated samples. Further suppose that we set \( \alpha = 0.05 \) (as is standard). What happens under repeated sampling?

We reject when \( t(X) > \text{critical value} \). Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value. Fail to reject otherwise.
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) > \text{critical value}$
Fail to reject otherwise
Example of Test Statistic and Critical Region

Suppose null is true $\pi = 0.5$ and we perform repeated samples. Further suppose that we set $\alpha = 0.05$ (as is standard). What happens under repeated sampling?

We reject when $t(X) >$ critical value. Fail to reject otherwise.
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value

Small p-values do not imply:

a) That there is a larger substantive effect (confidence intervals!)
b) That there is a small probability that null hypothesis is false
c) That there is a large probability that alternative hypothesis is true

One sided hypothesis test: (our example)

$\text{area to right of test statistic}$

$p\text{-value} = \int_{\infty}^{t(X)} \exp\left[-\left(Z^2\right)/2\right]/\left(\sqrt{2\pi}\right) dZ$
P-Value

How weird is test statistic?
- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
P-Value

How weird is test statistic?
- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- Small p-values do not imply:
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
  c) That there is a large probability that alternative hypothesis is true

One sided hypothesis test: (our example)
area to right of test statistic
p-value = 
\[ \int_{\infty}^{t(X)} \exp\left[ -\left( Z \right)^2 / 2 \right] / \left( \sqrt{2\pi} \right) \, dZ \]
P-Value

How weird is test statistic?
- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- Small p-values do not imply:
  a) That there is a larger substantive effect (confidence intervals!)
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- Small p-values do not imply:
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- Small p-values do not imply:
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
  c) That there is a large probability that alternative hypothesis is true
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- **Small p-values do not imply:**
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
  c) That there is a large probability that alternative hypothesis is true

One sided hypothesis test: (our example)
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- **Small p-values do not imply:**
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
  c) That there is a large probability that alternative hypothesis is true

One sided hypothesis test: (our example)

area to right of test statistic
P-Value

How weird is test statistic?

- What is the smallest value of $\alpha$ that we would still reject the null?
- Under the null, what is the probability of observing this value of the test statistic, or one more extreme?
- We will call this quantity the p-value
- Small p-values do not imply:
  a) That there is a larger substantive effect (confidence intervals!)
  b) That there is a small probability that null hypothesis is false
  c) That there is a large probability that alternative hypothesis is true

One sided hypothesis test: (our example)
area to right of test statistic
p-value $= \int_{t(x)}^{\infty} \exp\left[-(Z)^2/2\right]/(\sqrt{2\pi}) dZ$
One Sided Hypothesis Tests
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
- Place all rejection region in one tail (like this example)
- Easier to reject (without increasing $\alpha$)

Give me one example where appropriate → Terese (my wife) and I will take you + guest out for very nice dinner
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
- Place all rejection region in one tail (like this example)

Terese (my wife) and I will take you + guest out for very nice dinner

Justin Grimmer (Stanford University)
Methodology I
November 26th, 2012
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
- Place all rejection region in one tail (like this example)
- easier to reject (without increasing $\alpha$)
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
- Place all rejection region in one tail (like this example)
- *easier* to reject (without increasing $\alpha$

Give me one example where appropriate
One Sided Hypothesis Tests

- One sided hypothesis test: no way test statistic can go in other direction.
- Place all rejection region in one tail (like this example)
- easier to reject (without increasing $\alpha$)

Give me one example where appropriate $\rightarrow$ Terese (my wife) and I will take you + guest out for very nice dinner
Example 2: Difference in Normal Means

Observe feeling thermometer scores for member of Congress from two groups (populations)

- Treatment: reminded of pork brought to district
- Control: not reminded of pork

Suppose treated responses, \( n_t \), are iid Normal, with \( \mu_t \) and variance \( \sigma^2_t \)

\[ T_i \sim \text{Normal}(\mu_t, \sigma^2_t) \]

Suppose control responses, \( n_c \), are iid Normal, with \( \mu_c \) and variance \( \sigma^2_c \)

\[ C_i \sim \text{Normal}(\mu_c, \sigma^2_c) \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0: \mu_t = \mu_c \]
\[ H_1: \mu_t \neq \mu_c \]

From previous class we know (if \( n_t \) and \( n_c \) are sufficiently large)

\[ T \sim \text{Normal}(\mu_t, \sigma_t^2 / n_t) \]
\[ \hat{\sigma}_t^2 = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (T_i - T)^2 \]

\[ C \sim \text{Normal}(\mu_c, \sigma_c^2 / n_c) \]
\[ \hat{\sigma}_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (C_i - C)^2 \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0 : \mu_t = \mu_c \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

$H_0 : \mu_t = \mu_c$

$H_1 : \mu_t \neq \mu_c$
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

$H_0 : \mu_t = \mu_c$

$H_1 : \mu_t \neq \mu_c$

From previous class we know (if $n_t$ and $n_c$ are sufficiently large)
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0 : \mu_t = \mu_c \]
\[ H_1 : \mu_t \neq \mu_c \]

From previous class we know (if \( n_t \) and \( n_c \) are sufficiently large)

\[ \bar{T} \sim \text{Normal}(\mu_t, \sigma^2_t/n_t) \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0 : \mu_t = \mu_c \]

\[ H_1 : \mu_t \neq \mu_c \]

From previous class we know (if \( n_t \) and \( n_c \) are sufficiently large)

\[ \bar{T} \sim \text{Normal}(\mu_t, \sigma_t^2/n_t) \]

\[ \hat{\sigma}_t^2 = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (T_i - \bar{T})^2 \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0 : \mu_t = \mu_c \]
\[ H_1 : \mu_t \neq \mu_c \]

From previous class we know (if \( n_t \) and \( n_c \) are sufficiently large)

\[ \bar{T} \sim \text{Normal}(\mu_t, \sigma_t^2/n_t) \]
\[ \hat{\sigma}_t^2 = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (T_i - \bar{T}_i)^2 \]
\[ \bar{C} \sim \text{Normal}(\mu_c, \sigma_c^2/n_c) \]
Example 2: Difference in Normal Means

Does reminding about pork increase support for members of Congress?

\[ H_0 : \mu_t = \mu_c \]
\[ H_1 : \mu_t \neq \mu_c \]

From previous class we know (if \( n_t \) and \( n_c \) are sufficiently large)

\[
\bar{T} \sim \text{Normal}(\mu_t, \sigma_t^2 / n_t)
\]

\[
\hat{\sigma}_t^2 = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (T_i - \bar{T}_i)^2
\]

\[
\bar{C} \sim \text{Normal}(\mu_c, \sigma_c^2 / n_c)
\]

\[
\hat{\sigma}_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (C_i - \bar{C}_i)^2
\]
Example 2: Difference in Normal Means

Assume that Null is true: $\mu_t = \mu_c$
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \text{ why?}
\]
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \text{ why?}
\]
\[
\text{var}(\bar{T} - \bar{C}) = \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C})
\]
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \quad \text{why?}
\]

\[
\text{var}(\bar{T} - \bar{C}) = \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C})
\]

\[
= \frac{\sigma^2_t}{n_t} + \frac{\sigma^2_c}{n_c} \quad \text{why?}
\]
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \text{ why?}
\]

\[
\text{var}(\bar{T} - \bar{C}) = \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C})
\]

\[
= \frac{\sigma^2}{n_t} + \frac{\sigma^2}{n_c} \text{ why?}
\]

\[
= \frac{\hat{\sigma}^2}{n_t} + \frac{\hat{\sigma}^2}{n_c}
\]
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \text{ why?}
\]

\[
\text{var}(\bar{T} - \bar{C}) = \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C})
\]

\[
= \frac{\sigma^2_t}{n_t} + \frac{\sigma^2_c}{n_c} \text{ why?}
\]

\[
= \frac{\hat{\sigma}^2_t}{n_t} + \frac{\hat{\sigma}^2_c}{n_c}
\]

\[
\bar{T} - \bar{C} \sim \text{Normal}(0, \frac{\hat{\sigma}^2_t}{n_t} + \frac{\hat{\sigma}^2_c}{n_c})
\]
Example 2: Difference in Normal Means

Assume that Null is true: $\mu_t = \mu_c$

\[
\begin{align*}
E[\bar{T} - \bar{C}] &= 0 \text{ why?} \\
\text{var}(\bar{T} - \bar{C}) &= \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C}) \\
&= \frac{\sigma^2_t}{n_t} + \frac{\sigma^2_c}{n_c} \text{ why?} \\
&= \frac{\hat{\sigma}^2_t}{n_t} + \frac{\hat{\sigma}^2_c}{n_c} \\
\bar{T} - \bar{C} &\sim \text{Normal}(0, \frac{\hat{\sigma}^2_t}{n_t} + \frac{\hat{\sigma}^2_c}{n_c})
\end{align*}
\]

This implies our test statistic:
Example 2: Difference in Normal Means

Assume that Null is true: \( \mu_t = \mu_c \)

\[
E[\bar{T} - \bar{C}] = 0 \text{ why?}
\]

\[
\text{var}(\bar{T} - \bar{C}) = \text{var}(\bar{T}) + \text{var}(\bar{C}) - 2\text{cov}(\bar{T}, \bar{C})
\]

\[
= \frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c} \quad \text{why?}
\]

\[
= \frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c}
\]

\[
\bar{T} - \bar{C} \sim \text{Normal}(0, \frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c})
\]

This implies our test statistic:

\[
\frac{(\bar{T} - \bar{C}) - 0}{\sqrt{\frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c}}} \sim \text{Normal}(0, 1)
\]
Example 2: Difference in Normal Means

We will set $\alpha = 0.05$.
We need to find critical value, $x_0$ such that, under null,

$$\Pr\left( \left| \frac{T - C}{\sqrt{\frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c}}} \right| > x_0 \right) = 0.05$$

- **Note**: both extreme positive and extreme negative values matter
- **Two sided hypothesis**: allocate rejection area to both tails

$$\Pr\left( \frac{T - C}{\sqrt{\frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c}}} < -x_0 \right) + \Pr\left( x_0 < \frac{T - C}{\sqrt{\frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_c^2}{n_c}}}, \right) = 0.05$$
\[
\text{Pr}\left( \frac{\overline{T} - \overline{C}}{\sqrt{\hat{\sigma}^2_t n_t} + \hat{\sigma}^2_c n_c} < -x_0 \right) + \text{Pr}\left( x_0 < \frac{\overline{T} - \overline{C}}{\sqrt{\hat{\sigma}^2_t n_t} + \hat{\sigma}^2_c n_c} \right) = 0.05
\]

Rejection region:
\[
\frac{\overline{T} - \overline{C}}{\sqrt{\hat{\sigma}^2_t n_t} + \hat{\sigma}^2_c n_c} < -1.96 \quad \text{and} \quad \frac{\overline{T} - \overline{C}}{\sqrt{\hat{\sigma}^2_t n_t} + \hat{\sigma}^2_c n_c} > 1.96
\]
Example 2: Difference in Normal Means

Suppose that we run our experiment:

- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
- Probability of observing value of test statistic or one more extreme, under null
- Calculate area to right of 2.4, to left of -2.4
- p-value = \[ \int_{-\infty}^{-2.4} \exp\left(-\frac{Z^2}{2}\right) \frac{1}{\sqrt{2\pi}\sigma} \, dZ + \int_{2.4}^{\infty} \exp\left(-\frac{Z^2}{2}\right) \frac{1}{\sqrt{2\pi}\sigma} \, dZ \]
- p.value \approx 0.016
- What if we observed test statistic of 1.2?
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?

\[
\text{p-value} \approx 0.016
\]
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?

\[ p\text{-value} \approx 0.016 \]
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
  - Probability of observing value of test statistic or one more extreme, under null

\[
\begin{align*}
\int_{-\infty}^{-2.4} &\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(Z)^2}{2}\right)\,dZ \\
+ \int_{2.4}^{\infty} &\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(Z)^2}{2}\right)\,dZ \\
\end{align*}
\]

\[p\text{-value} \approx 0.016\]
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
  - Probability of observing value of test statistic or one more extreme, under null
  - Calculate area to right of 2.4, to left of -2.4
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
  - Probability of observing value of test statistic or one more extreme, under null
  - Calculate area to right of 2.4, to left of -2.4
  - p-value = 
    \[
    \int_{-\infty}^{-2.4} \exp\left[-\frac{(Z)^2}{2}\right]/\left(\sqrt{2\pi}\sigma\right) dZ + \int_{2.4}^{\infty} \exp\left[-\frac{(Z)^2}{2}\right]/\left(\sqrt{2\pi}\sigma\right) dZ
    \]
  - $p\text{-value} \approx 0.016$
  - What if we observed test statistic of 1.2?
Example 2: Difference in Normal Means

Suppose that we run our experiment:

- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
  - Probability of observing value of test statistic or one more extreme, under null
  - Calculate area to right of 2.4, to left of -2.4
  - p-value =
    \[
    \int_{-\infty}^{-2.4} \exp\left[-\frac{(Z)^2}{2}\right] / \left(\sqrt{2\pi}\sigma\right) dZ + \int_{2.4}^{\infty} \exp\left[-\frac{(Z)^2}{2}\right] / \left(\sqrt{2\pi}\sigma\right) dZ
    \]
  - p.value \approx 0.016
Example 2: Difference in Normal Means

Suppose that we run our experiment:
- Observe test-statistic of 2.4
- What does this imply?
- What is the p-value?
  - Probability of observing value of test statistic or one more extreme, under null
  - Calculate area to right of 2.4, to left of -2.4
  - p-value =
    \[
    \int_{-\infty}^{-2.4} \exp\left[-\frac{(Z)^2}{2}/\sqrt{2\pi\sigma}\right]dZ + \int_{2.4}^{\infty} \exp\left[-\frac{(Z)^2}{2}/\sqrt{2\pi\sigma}\right]dZ
    \]
    - p.value \approx 0.016
- What if we observed test statistic of 1.2?