Lecture 14:
Hypothesis Testing for Proportions
API-201Z

Maya Sen

Harvard Kennedy School
http://scholar.harvard.edu/msen
Announcements
Announcements

- We will be posting examples of successful final exercise memos from years past
Roadmap

Continuing with different types of hypothesis tests

Last time: Hypothesis tests involving means and difference in means

Today: Hypothesis tests involving proportions

Interpreting hypothesis tests

Practical versus statistical significance

Type I and Type II errors (and power, if time)
Roadmap

- Continuing with different types of hypothesis tests
Roadmap

▶ Continuing with different types of hypothesis tests
▶ Last time: Hypothesis tests involving means and difference in means
Roadmap

▶ Continuing with different types of hypothesis tests
▶ Last time: Hypothesis tests involving means and difference in means
▶ Today: Hypothesis tests involving proportions
Roadmap

- Continuing with different types of hypothesis tests
- Last time: Hypothesis tests involving means and difference in means
- Today: Hypothesis tests involving proportions
- Interpreting hypothesis tests
Roadmap

- Continuing with different types of hypothesis tests
- Last time: Hypothesis tests involving means and difference in means
- Today: Hypothesis tests involving proportions
- Interpreting hypothesis tests
- Practical versus statistical significance
Roadmap

- Continuing with different types of hypothesis tests
- Last time: Hypothesis tests involving means and difference in means
- Today: Hypothesis tests involving proportions
- Interpreting hypothesis tests
- Practical versus statistical significance
- Type I and Type II errors (and power, if time)
Hypothesis Testing for Proportions

- Often not interested in making inferences about a population mean, but a population proportion.
- Proportions commonly used to summarize yes-or-no outcomes:
  - True U.S. presidential approval rating
  - True support in the U.S. population for same-sex marriage
  - Proportion of people in an indigenous population who have a certain genetic marker
  - Proportion of households in a city that are below national poverty levels
- Can extend hypothesis testing to analyze all sorts of proportions
- Tweak existing hypothesis framework only slightly
Hypothesis Testing for Proportions

- Often not interested in making inferences about a population mean, but a population proportion
Hypothesis Testing for Proportions

- Often not interested in making inferences about a population mean, but a *population proportion*
- Proportions commonly used to summarize yes-or-no outcomes:
  - True U.S. presidential approval rating
  - True support in the U.S. population for same-sex marriage
  - Proportion of people in an indigenous population who have a certain genetic marker
  - Proportion of households in a city that are below national poverty levels
Hypothesis Testing for Proportions

- Often not interested in making inferences about a population mean, but a population proportion.
- Proportions commonly used to summarize yes-or-no outcomes:
  - True U.S. presidential approval rating
  - True support in the U.S. population for same-sex marriage
  - Proportion of people in an indigenous population who have a certain genetic marker
  - Proportion of households in a city that are below national poverty levels
- Can extend hypothesis testing to analyze all sorts of proportions.
Hypothesis Testing for Proportions

- Often not interested in making inferences about a population mean, but a *population proportion*

- Proportions commonly used to summarize yes-or-no outcomes:
  - True U.S. presidential approval rating
  - True support in the U.S. population for same-sex marriage
  - Proportion of people in an indigenous population who have a certain genetic marker
  - Proportion of households in a city that are below national poverty levels

- Can extend hypothesis testing to analyze all sorts of proportions
- Tweak existing hypothesis framework only slightly
Hypothesis Testing for Proportions

Can use hypothesis tests to explore these questions

Objective: Use sample of data to make inferences about true population proportion, \( \pi \) (sometimes denoted \( p \))

Same steps as before

Test statistics slightly different
Hypothesis Testing for Proportions

- Can use hypothesis tests to explore these questions
Hypothesis Testing for Proportions

- Can use hypothesis tests to explore these questions
- Objective: Use sample of data to make inferences about true population proportion, $\pi$ (sometimes denoted $p$)
Hypothesis Testing for Proportions

- Can use hypothesis tests to explore these questions
- Objective: Use sample of data to make inferences about true population proportion, \( \pi \) (sometimes denoted \( p \))
- Same steps as before
Hypothesis Testing for Proportions

- Can use hypothesis tests to explore these questions
- Objective: Use sample of data to make inferences about true population proportion, $\pi$ (sometimes denoted $p$)
- Same steps as before
- Test statistics slightly different
Hypothesis Testing for Proportions

Conducted using several steps:
1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a $p$-value
5. Use the $p$-value to decide whether to reject null hypothesis and interpret results
Hypothesis Testing for Proportions

Conducted using several steps:
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your **null** and **alternative hypotheses**
2. Collect sample/s of data
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a \( p \)-value
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a $p$-value
5. Use the $p$-value to decide whether to reject null hypothesis and interpret results
Historically it has been difficult to attract general physicians in the U.S. to rural areas, with 52% leaving rural areas within 1 year of arriving. You work on a new program designed to incentive doctors to stay in rural areas via loan forgiveness, housing allowances. Random sample of physicians in the program shows that 62 out of 130 leave within 1 year of arriving. Based on this sample, what can you conclude about enrollees in the program compared to physicians arriving to rural areas overall?
Historically it has been difficult to attract general physicians in the U.S. to rural areas, with 52% leaving rural areas within 1 year of arriving.
Proportion Example: Physician Training Program

- Historically it has been difficult to attract general physicians in the U.S. to rural areas, with 52% leaving rural areas within 1 year of arriving.
- You work on a new program designed to incentivize doctors to stay in rural areas via loan forgiveness, housing allowances.
Historically it has been difficult to attract general physicians in the U.S. to rural areas, with 52% leaving rural areas within 1 year of arriving.

You work on a new program designed to incentive doctors to stay in rural areas via loan forgiveness, housing allowances.

Random sample of physicians in the program shows that 62 out of 130 leave within 1 year of arriving.
Proportion Example: Physician Training Program

- Historically it has been difficult to attract general physicians in the U.S. to rural areas, with 52% leaving rural areas within 1 year of arriving.
- You work on a new program designed to incentive doctors to stay in rural areas via loan forgiveness, housing allowances.
- Random sample of physicians in the program shows that 62 out of 130 leave within 1 year of arriving.
- Based on this sample, what can you conclude about enrollees in the program compared to physicians arriving to rural areas overall?
Proportion Example: Physician Training Program

Step 1: Null and Alternative Hypotheses

$H_0$: $\pi = 0.52$

$H_a$: $\pi \neq 0.52$ for two-sided test (or $\pi \geq 0.52$, for one-sided test)
Proportion Example: Physician Training Program

- Step 1: Null and Alternative Hypotheses

  \[ H_0: \pi = 0.52 \]
  \[ H_a: \pi \neq 0.52 \text{ for two-sided test} \]
  \[ H_a: \pi \geq 0.52 \text{ for one-sided test} \]
Step 1: Null and Alternative Hypotheses

- $H_0: \pi = 0.52$
Step 1: Null and Alternative Hypotheses

- $H_0$: $\pi = 0.52$
- $H_a$: $\pi \neq 0.52$ for two-sided test (or $\pi \geq 0.52$, for one-sided test)
Proportion Example: Physician Training Program
Proportion Example: Physician Training Program

- Step 2: Collect sample data

Assume this is done for us already! $n = 130$, with 62 doctors leaving within 1 year.
Proportion Example: Physician Training Program

- Step 2: Collect sample data
- Assume this is done for us already!
Step 2: Collect sample data
Assume this is done for us already!
$n = 130$, with 62 doctors leaving within 1 year
Proportion Example: Physician Training Program

Step 3: Conduct statistical test

As before, need to find appropriate test statistic to compare our sample to distribution of data under the null hypothesis.

Another application of CLT.
Proportion Example: Physician Training Program

- Step 3: Conduct statistical test
Step 3: Conduct statistical test

As before, need to find appropriate test statistic to compare our sample to distribution of data under the null hypothesis.
Proportion Example: Physician Training Program

- Step 3: Conduct statistical test
- As before, need to find appropriate test statistic to compare our sample to distribution of data under the null hypothesis
- As before, another application of CLT
Proportion Example: Physician Training Program

We'll work with a sample proportion (denoted $\hat{\pi}$ or $\hat{p}$) → sample mean when all observations are 1s and 0s

$\bar{x} = \hat{\pi} = \frac{\sum x_i}{n}$

As a review, under the CLT:

1. The sums and means of random variables have an approximately normal distribution
2. This distribution becomes more and more normal the more observations are included in the sum or the mean
3. This is the case even though the underlying distribution may not be normal

Thus, under repeated sampling, $\hat{\pi}$’s distribution will follow a normal distribution (see Lecture 11 slides for the simulation)
Proportion Example: Physician Training Program

- We’ll work w/ sample proportion (denoted $\hat{\pi}$ or $\hat{p}$) → sample mean when all observations 1s and 0s
Proportion Example: Physician Training Program

- We’ll work w/ sample proportion (denoted \( \hat{\pi} \) or \( \hat{\rho} \)) → sample mean when all observations 1s and 0s
- \( \bar{x} = \hat{\pi} = \frac{\sum x_i}{n} \)
Proportion Example: Physician Training Program

- We’ll work w/ sample proportion (denoted \( \hat{\pi} \) or \( \hat{p} \)) \( \rightarrow \) sample mean when all observations 1s and 0s
- \( \bar{x} = \hat{\pi} = \frac{\sum x_i}{n} \)
- As a review, under the CLT,
We’ll work with sample proportion (denoted \( \hat{\pi} \) or \( \hat{\rho} \)) → sample mean when all observations 1s and 0s

\[ \bar{x} = \hat{\pi} = \frac{\sum x_i}{n} \]

As a review, under the CLT,

1. The sums and means of random variables have an approximately normal distribution
Proportion Example: Physician Training Program

- We’ll work with sample proportion (denoted $\hat{\pi}$ or $\hat{p}$) → sample mean when all observations 1s and 0s
- $\bar{x} = \hat{\pi} = \frac{\sum x_i}{n}$
- As a review, under the CLT,
  - (1) The sums and means of random variables have an approximately normal distribution
  - (2) This distribution becomes more and more normal the more observations are included in the sum or the mean
We’ll work w/ sample proportion (denoted \( \hat{\pi} \) or \( \hat{\rho} \)) → sample mean when all observations 1s and 0s

\[ \bar{x} = \hat{\pi} = \frac{\sum x_i}{n} \]

As a review, under the CLT,

1. The sums and means of random variables have an approximately normal distribution
2. This distribution becomes more and more normal the more observations are included in the sum or the mean
3. This is the case even though underlying distribution may not be normal
Proportion Example: Physician Training Program

- We’ll work w/ sample proportion (denoted $\hat{\pi}$ or $\hat{p}$) → sample mean when all observations 1s and 0s
- $\bar{x} = \hat{\pi} = \frac{\sum x_i}{n}$
- As a review, under the CLT,
  - (1) The sums and means of random variables have an approximately normal distribution
  - (2) This distribution becomes more and more normal the more observations are included in the sum or the mean
  - (3) This is the case even though underlying distribution may not be normal
- Thus, under repeated sampling, $\hat{\pi}$’s distribution will follow a normal distribution (see Lecture 11 slides for the simulation)
Proportion Example: Physician Training Program

Specifically, with large enough sample, \( \hat{\pi} \) will follow a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by \( n \). Which means we can standardize:

\[
Z = \frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}}
\]

where \( Z \) follows a standard normal.
Specifically, with large enough sample, \( \hat{\pi} \) will

\[
\hat{\pi} \sim N(\pi, \pi(1-\pi)/n)
\]

See proofs for both mean and variance in Appendix.

\[\sum x_i\] follows a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by \( n \).

Which means we can standardize

\[ Z = \frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}} \]

where \( Z \) follows a standard normal.
Proportion Example: Physician Training Program

- Specifically, with large enough sample, \( \hat{\pi} \) will

\[
\hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n})
\]
Proportion Example: Physician Training Program

- Specifically, with large enough sample, \( \hat{\pi} \) will

\[
\hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n})
\]

- See proofs for both mean and variance in Appendix)
Proportion Example: Physician Training Program

- Specifically, with large enough sample, $\hat{\pi}$ will

$$\hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n})$$

- See proofs for both mean and variance in Appendix)

- $(\sum x_i$ follows a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by $n$)
Proportion Example: Physician Training Program

- Specifically, with large enough sample, $\hat{\pi}$ will

$$\hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n})$$

- See proofs for both mean and variance in Appendix

- $\sum x_i$ follows a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by $n$

- Which means we can standardize
Proportion Example: Physician Training Program

- Specifically, with large enough sample, \( \hat{\pi} \) will

\[
\hat{\pi} \sim \mathcal{N}(\pi, \frac{\pi(1-\pi)}{n})
\]

- See proofs for both mean and variance in Appendix)

- \( \sum x_i \) follows a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by \( n \)

- Which means we can standardize

\[
Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}
\]
Proportion Example: Physician Training Program

- Specifically, with large enough sample, $\hat{\pi}$ will

$$\hat{\pi} \sim N(\pi, \frac{\pi(1-\pi)}{n})$$

- See proofs for both mean and variance in Appendix)

- $\left(\sum x_i\right)$ follows a binomial distribution, so parameters for sampling distribution derived from the parameters of a binomial distribution divided by $n$)

- Which means we can standardize

$$Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

- where $Z$ follows a standard normal
Under the assumption that the null is true (here, $\pi = 0.52$), we can calculate the test statistic using our data:

$$Z = \frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}}$$

$$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \frac{62/130 - 0.52}{\sqrt{0.52(1-0.52)/130}} = -0.983$$
Proportion Example: Physician Training Program

- Under assumption that null is true (here, $\pi = 0.52$), can calculate test statistic using our data:

$$Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi}{n}}} = \frac{62}{130} - 0.52 \sqrt{\frac{0.52}{130}} = -0.983$$
Under assumption that null is true (here, \( \pi = 0.52 \)), can calculate test statistic using our data:

\[
Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}
\]

\[
z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}
\]

\[
= \frac{62/130 - 0.52}{\sqrt{0.52(1-0.52)/130}}
\]

\[= -0.983\]
Proportion Example: Physician Training Program

Note: Test statistic does not rely on any sample variance:

\[
    z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0 \left(1 - \pi_0\right)}} \frac{1}{n}
\]

In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null).

Thus, more OK to use standard normal (b/c we are not using sample standard deviation).

But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:

\[
    n \times \pi_0 \geq 10 \quad \text{and} \quad n \times \left(1 - \pi_0\right) \geq 10
\]

If yes, distribution of \( \hat{\pi} \) is approximately normal, ok to use \( z \)-test (Normal).

If not → calculate using binomial distribution directly.
Proportion Example: Physician Training Program

- Note: Test statistic does not rely on any sample variance:

\[
z = \hat{\pi} - \pi_0 \sqrt{\pi_0 \left(1 - \pi_0\right)} \frac{1}{n}
\]

In fact, fully determined by \(n\) and \(\pi_0\) (assuming null)

Thus, more OK to use standard normal (b/c we are not using sample standard deviation)

But still want to check sufficient sample size → above \(n = 30\) and \(\pi\) not close to 0 or 1:

\[
n \times \pi_0 \geq 10 \text{ and } n \times \left(1 - \pi_0\right) \geq 10
\]

If yes, distribution of \(\hat{\pi}\) is approximately normal, ok to use \(z\)-test (Normal)

If not → calculate using binomial distribution directly
Note: Test statistic does not rely on any sample variance:

\[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \]

In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null).

Thus, more OK to use standard normal (b/c we are not using sample standard deviation).

But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:

\[ n \times \pi_0 \geq 10 \text{ and } n \times (1-\pi_0) \geq 10 \]

If yes, distribution of \( \hat{\pi} \) is approximately normal, ok to use \( z \)-test (Normal).

If not → calculate using binomial distribution directly.
Proportion Example: Physician Training Program

- Note: Test statistic does not rely on any sample variance:

\[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

- In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)

Thus, more OK to use standard normal (b/c we are not using sample standard deviation)

But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:

\[ n \times \pi_0 \geq 10 \text{ and } n \times (1-\pi_0) \geq 10 \]

If yes, distribution of \( \hat{\pi} \) is approximately normal, ok to use \( z \)-test (Normal)

If not → calculate using binomial distribution directly
Proportion Example: Physician Training Program

▶ Note: Test statistic does not rely on any sample variance:

\[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

▶ In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)
▶ Thus, more OK to use standard normal (b/c we are not using sample standard deviation)
Proportion Example: Physician Training Program

Note: Test statistic does not rely on any sample variance:

\[
    z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0) / n}}
\]

In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)

Thus, more OK to use standard normal (b/c we are not using sample standard deviation)

But still want to check sufficient sample size \( \rightarrow \) above \( n = 30 \) and \( \pi \) not close to 0 or 1:
Proportion Example: Physician Training Program

- Note: Test statistic does not rely on any sample variance:
  \[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

- In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)
- Thus, more OK to use standard normal (b/c we are not using sample standard deviation)
- But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:
  - \( n \times \pi_0 \geq 10 \) and \( n \times (1 - \pi_0) \geq 10 \)
Proportion Example: Physician Training Program

Note: Test statistic does not rely on any sample variance:

\[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)

Thus, more OK to use standard normal (b/c we are not using sample standard deviation)

But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:

\( n \times \pi_0 \geq 10 \) and \( n \times (1 - \pi_0) \geq 10 \)

If yes, distribution of \( \hat{\pi} \) is approximately normal, ok to use z-test (Normal)
Proportion Example: Physician Training Program

- Note: Test statistic does not rely on any sample variance:
  \[ z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

- In fact, fully determined by \( n \) and \( \pi_0 \) (assuming null)
- Thus, more OK to use standard normal (b/c we are not using sample standard deviation)
- But still want to check sufficient sample size → above \( n = 30 \) and \( \pi \) not close to 0 or 1:
  - \( n \times \pi_0 \geq 10 \) and \( n \times (1 - \pi_0) \geq 10 \)
- If yes, distribution of \( \hat{\pi} \) is approximately normal, ok to use z-test (Normal)
- If not → calculate using binomial distribution directly
Proportion Example: Physician Training Program

Step 4: Once have calculated the test statistic, use that to calculate $p$-value.

Under two-tailed test:

$2 \times P(Z \leq -0.983) = 0.3256$
Proportion Example: Physician Training Program

- Step 4: Once have calculated the test statistic, use that to calculate \( p \)-value

\[
2 	imes P(Z \leq -0.983) = 0.3256
\]
Proportion Example: Physician Training Program

- Step 4: Once have calculated the test statistic, use that to calculate \( p \)-value
- Under two-tailed test:
Proportion Example: Physician Training Program

- Step 4: Once have calculated the test statistic, use that to calculate \( p \)-value
- Under two-tailed test:
  - \( 2 \times P(Z \leq -0.983) = 0.3256 \)
Proportion Example: Physician Training Program
Proportion Example: Physician Training Program

- Step 5: Decide whether or not to reject the null hypothesis and interpret results

Given a two-tailed test with a p-value of 0.3256, what would you do?

Would you reject if a one-tailed test?
Step 5: Decide whether or not to reject the null hypothesis and interpret results

Given two-tailed test with $p$-value of 0.3256, what would you do?
Proportion Example: Physician Training Program

- Step 5: Decide whether or not to reject the null hypothesis and interpret results
- Given two-tailed test with $p$-value of 0.3256, what would you do?
- Would you reject if one-tailed test?
Extending to Comparing Differences in Proportions

This example used a single proportion. Just like in case of a single mean, there is "some value" to compare sample proportion to. Ex) Whether program changes share of doctors leaving rural areas (H₀: π = 0.52). Ex) Whether half of jobs occupied by women (H₀: π = 0.50). However, can extend to explore comparisons of proportions between groups (analogy to difference in means). Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)? Ex) Are Canadians more likely to favor tighter gun control restrictions compared to Americans? Ex) Do police in the U.S. pull over black motorists more frequently than white motorists?
Extending to Comparing Differences in Proportions

- This example used a single proportion

- Just like in case of a single mean, there is "some value" to compare sample proportion to

- Ex) Whether program changes share of doctors leaving rural areas (H\(_0\): \(\pi = 0.52\))

- Ex) Whether half of jobs occupied by women (H\(_0\): \(\pi = 0.50\))

- However, can extend to explore comparisons of proportions between groups (analogy to difference in means)

- Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)?

- Ex) Are Canadians more likely to favor tighter gun control restrictions compared to Americans?

- Ex) Do police in the U.S. pull over black motorists more frequently than white motorists?
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
  - Ex) Whether program changes share of doctors leaving rural areas ($H_0 : \pi = 0.52$)
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
  - Ex) Whether program changes share of doctors leaving rural areas ($H_0 : \pi = 0.52$)
  - Ex) Whether half of jobs occupied by women ($H_0 : \pi = 0.50$)

- However, can extend to explore comparisons of proportions between groups (analogy to difference in means)
  - Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)?
  - Ex) Are Canadians more likely to favor tighter gun control restrictions compared to Americans?
  - Ex) Do police in the U.S. pull over black motorists more frequently than white motorists?
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
  - Ex) Whether program changes share of doctors leaving rural areas ($H_0 : \pi = 0.52$)
  - Ex) Whether half of jobs occupied by women ($H_0 : \pi = 0.50$)
- However, can extend to explore comparisons of proportions between groups (analogy to difference in means)
This example used a single proportion

Just like in case of a single mean, there is “some value” to compare sample proportion to

- Ex) Whether program changes share of doctors leaving rural areas ($H_0 : \pi = 0.52$)
- Ex) Whether half of jobs occupied by women ($H_0 : \pi = 0.50$)

However, can extend to explore comparisons of proportions between groups (analogy to difference in means)

- Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)?
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
  - Ex) Whether program changes share of doctors leaving rural areas ($H_0 : \pi = 0.52$)
  - Ex) Whether half of jobs occupied by women ($H_0 : \pi = 0.50$)
- However, can extend to explore comparisons of proportions between groups (analogy to difference in means)
  - Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)?
  - Ex) Are Canadians more likely to favor tighter gun control restrictions compared to Americans?
Extending to Comparing Differences in Proportions

- This example used a single proportion
- Just like in case of a single mean, there is “some value” to compare sample proportion to
  - Ex) Whether program changes share of doctors leaving rural areas ($H_0: \pi = 0.52$)
  - Ex) Whether half of jobs occupied by women ($H_0: \pi = 0.50$)
- However, can extend to explore comparisons of proportions between groups (analogy to difference in means)
  - Ex) Compare voter turnout on Brexit in Scotland (not competitive) versus England (competitive)?
  - Ex) Are Canadians more likely to favor tighter gun control restrictions compared to Americans?
  - Ex) Do police in the U.S. pull over black motorists more frequently than white motorists?
Randomized Controlled Trial Example

2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes

Looking at six US hospitals, patients randomly assigned into two groups

604 assigned to have volunteers pray for them → 315 developed complications

597 assigned to control (no prayer) → 304 developed complications

Patients in both groups "informed that they may or may not receive prayer"

Based on this data, should we agree with skeptics that say prayer unrelated to surgical outcomes?
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
- Looking at six US hospitals, patients randomly assigned into two groups
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
- Looking at six US hospitals, patients randomly assigned into two groups
  - 604 assigned to have volunteers pray for them → 315 developed complications
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
- Looking at six US hospitals, patients randomly assigned into two groups
  - 604 assigned to have volunteers pray for them → 315 developed complications
  - 597 assigned to control (no prayer) → 304 developed complications

Based on this data, should we agree with skeptics that say prayer unrelated to surgical outcomes?
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
- Looking at six US hospitals, patients randomly assigned into two groups
  - 604 assigned to have volunteers pray for them → 315 developed complications
  - 597 assigned to control (no prayer) → 304 developed complications
- Patients in both groups ”informed that they may or may not receive prayer”
Randomized Controlled Trial Example

- 2006 study to determine healing power of spirituality with coronary surgery patients & their health outcomes
- Looking at six US hospitals, patients randomly assigned into two groups
  - 604 assigned to have volunteers pray for them → 315 developed complications
  - 597 assigned to control (no prayer) → 304 developed complications
  - Patients in both groups ”informed that they may or may not receive prayer”
- Based on this data, should we agree with skeptics that say prayer unrelated to surgical outcomes?
Hypothesis Testing for Proportions

Conducted using several steps:
1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a \( p \)-value
5. Use the \( p \)-value to decide whether to reject null hypothesis and interpret results

Let's review these steps using the prayer and surgery randomized controlled trial (RTC) example
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a $p$-value
5. Use the $p$-value to decide whether to reject null hypothesis and interpret results

Let's review these steps using the prayer and surgery randomized controlled trial (RTC) example
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic

Let's review these steps using the prayer and surgery randomized controlled trial (RTC) example
Hypothesis Testing for Proportions

Conducted using several steps:
1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a $p$-value
Hypothesis Testing for Proportions

Conducted using several steps:

1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a \( p \)-value
5. Use the \( p \)-value to decide whether to reject null hypothesis and interpret results
Hypothesis Testing for Proportions

Conducted using several steps:
1. Generate your null and alternative hypotheses
2. Collect sample/s of data
3. Calculate appropriate test statistic
4. Use that to calculate a probability called a $p$-value
5. Use the $p$-value to decide whether to reject null hypothesis and interpret results

Let’s review these steps using the prayer and surgery randomized controlled trial (RTC) example
Randomized Controlled Trial Example

Step 1: Null and Alternative Hypotheses

Null Hypothesis

\[ H_0: \pi_1 - \pi_2 = 0 \] (or \( \pi_1 = \pi_2 \))

Alternative Hypothesis

\[ H_a: \pi_1 - \pi_2 \neq 0 \] (or \( \pi_1 \neq \pi_2 \))

\( H_a: \pi_1 - \pi_2 \geq 0 \)
Randomized Controlled Trial Example

- Step 1: Null and Alternative Hypotheses

  - Null Hypothesis
    - $H_0$: $\pi_1 - \pi_2 = 0$ (or $\pi_1 = \pi_2$)
  
  - Alternative Hypothesis
    - $H_a$: $\pi_1 - \pi_2 \neq 0$ (or $\pi_1 \neq \pi_2$)
      - ($H_a$: $\pi_1 - \pi_2 \geq 0$)
Randomized Controlled Trial Example

- Step 1: Null and Alternative Hypotheses
  - Null Hypothesis
    - Null Hypothesis: \( H_0: \pi_1 - \pi_2 = 0 \) (or \( \pi_1 = \pi_2 \))
    - Alternative Hypothesis: \( H_a: \pi_1 - \pi_2 \neq 0 \) (or \( \pi_1 \neq \pi_2 \))
Randomized Controlled Trial Example

▶ Step 1: Null and Alternative Hypotheses
▶ Null Hypothesis
  ▶ No difference (prayer not effective)
Randomized Controlled Trial Example

- Step 1: Null and Alternative Hypotheses
- Null Hypothesis
  - No difference (prayer not effective)
  - \( H_0 : \pi_1 - \pi_2 = 0 \) (or \( \pi_1 = \pi_2 \))
Randomized Controlled Trial Example

- Step 1: Null and Alternative Hypotheses
- Null Hypothesis
  - No difference (prayer not effective)
  - $H_0 : \pi_1 - \pi_2 = 0$ (or $\pi_1 = \pi_2$)
- Alternative Hypothesis
Randomized Controlled Trial Example

- Step 1: Null and Alternative Hypotheses
  - Null Hypothesis
    - No difference (prayer not effective)
    - $H_0 : \pi_1 - \pi_2 = 0$ (or $\pi_1 = \pi_2$)
  - Alternative Hypothesis
    - $H_a : \pi_1 - \pi_2 \neq 0$ (or $\pi_1 \neq \pi_2$)
Step 1: Null and Alternative Hypotheses

Null Hypothesis
- No difference (prayer not effective)
- \( H_0 : \pi_1 - \pi_2 = 0 \) (or \( \pi_1 = \pi_2 \))

Alternative Hypothesis
- \( H_a : \pi_1 - \pi_2 \neq 0 \) (or \( \pi_1 \neq \pi_2 \))
- \( (H_a : \pi_1 - \pi_2 \geq 0) \)
Randomized Controlled Trial Example

Step 2: Collect sample data

Done for us:

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prayer</td>
<td>289</td>
<td>293</td>
</tr>
<tr>
<td>Complications</td>
<td>315</td>
<td>304</td>
</tr>
<tr>
<td>Total</td>
<td>604</td>
<td>597</td>
</tr>
</tbody>
</table>
Randomized Controlled Trial Example

- Step 2: Collect sample data
Randomized Controlled Trial Example

- Step 2: Collect sample data
- Done for us:

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prayer</td>
<td>289</td>
<td>293</td>
</tr>
<tr>
<td>Complications</td>
<td>315</td>
<td>304</td>
</tr>
<tr>
<td>Total</td>
<td>604</td>
<td>597</td>
</tr>
</tbody>
</table>
Randomized Controlled Trial Example

- Step 2: Collect sample data
- Done for us:

<table>
<thead>
<tr>
<th>Prayer</th>
<th>No Complications</th>
<th>Complications</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>289</td>
<td>315</td>
<td>604</td>
</tr>
<tr>
<td>No</td>
<td>293</td>
<td>304</td>
<td>597</td>
</tr>
</tbody>
</table>
Randomized Controlled Trial Example

Step 3: Calculated appropriate test statistic

Again, under CLT, if \( n_1 \) and \( n_2 \) are relatively large, then \( \hat{\pi}_1 - \hat{\pi}_2 \) will have an approximately normal distribution.

Also, for independent samples, the standard error will just be the sum of the two standard errors.

So \( \hat{\pi}_1 - \hat{\pi}_2 \) will follow:

\[
\sim N(\pi_1 - \pi_2, \pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2)
\]

and we can standardize to:

\[
Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2}}
\]
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
- Again, under CLT, if $n_1$ and $n_2$ are relatively large, then estimator $\hat{\pi}_1 - \hat{\pi}_2$ will have an approximately normal distribution
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
- Again, under CLT, if $n_1$ and $n_2$ are relatively large, then estimator $\hat{\pi}_1 - \hat{\pi}_2$ will have an approximately normal distribution
- Also, for independent samples, the standard error will just be the sum of the two standard errors
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
- Again, under CLT, if $n_1$ and $n_2$ are relatively large, then estimator $\hat{\pi}_1 - \hat{\pi}_2$ will have an approximately normal distribution
- Also, for independent samples, the standard error will just be the sum of the two standard errors
- So $\hat{\pi}_1 - \hat{\pi}_2$ will follow:
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
  - Again, under CLT, if $n_1$ and $n_2$ are relatively large, then estimator $\hat{\pi}_1 - \hat{\pi}_2$ will have an approximately normal distribution
  - Also, for independent samples, the standard error will just be the sum of the two standard errors
  - So $\hat{\pi}_1 - \hat{\pi}_2$ will follow:

$$\sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})$$
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
- Again, under CLT, if \( n_1 \) and \( n_2 \) are relatively large, then estimator \( \hat{\pi}_1 - \hat{\pi}_2 \) will have an approximately normal distribution
- Also, for independent samples, the standard error will just be the sum of the two standard errors
- So \( \hat{\pi}_1 - \hat{\pi}_2 \) will follow:

  \[
  \sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})
  \]

- and we can standardize to
Randomized Controlled Trial Example

- Step 3: Calculated appropriate test statistic
- Again, under CLT, if \( n_1 \) and \( n_2 \) are relatively large, then estimator \( \hat{\pi}_1 - \hat{\pi}_2 \) will have an approximately normal distribution
- Also, for independent samples, the standard error will just be the sum of the two standard errors
- So \( \hat{\pi}_1 - \hat{\pi}_2 \) will follow:

\[
\sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})
\]

- and we can standardize to

\[
Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}}}
\]
Randomized Controlled Trial Example

Given that the null is true, \( \pi_1 - \pi_2 = 0 \), we can calculate the test statistic

\[
Z = \hat{\pi}_1 - \hat{\pi}_2 - \left( \pi_1 - \pi_2 \right) \sqrt{\hat{\pi}_1 \left( 1 - \hat{\pi}_1 \right) / n_1 + \hat{\pi}_2 \left( 1 - \hat{\pi}_2 \right) / n_2}
\]

Using our example:

\[
Z = \frac{315}{604} - \frac{304}{597} \sqrt{\frac{315}{604} \left( 1 - \frac{315}{604} \right) / 604 + \frac{304}{597} \left( 1 - \frac{304}{597} \right) / 597}
\]

\[
= 0.4637
\]
Randomized Controlled Trial Example

- Given that the null is true, $\pi_1 - \pi_2 = 0$, we can calculate the test statistic

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}_1(1-\hat{\pi}_1)n_1 + \hat{\pi}_2(1-\hat{\pi}_2)n_2}}$$

Using our example:

$$z = \frac{315/604 - 304/597}{\sqrt{315/604(1-315/604)604 + 304/597(1-304/597)597}} = 0.4637$$
Randomized Controlled Trial Example

- Given that the null is true, \( \pi_1 - \pi_2 = 0 \), we can calculate the test statistic

\[
Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1 (1-\pi_1)}{n_1} + \frac{\pi_2 (1-\pi_2)}{n_2}}}
\]

\[
z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\frac{\pi_1 (1-\pi_1)}{n_1} + \frac{\pi_2 (1-\pi_2)}{n_2}}}
\]

Using our example:

\[
z = \frac{315}{604} - \frac{304}{597} \times \sqrt{\frac{315}{604} (1-\frac{315}{604}) + \frac{304}{597} (1-\frac{304}{597})}
\]

\[
z = 0.4637
\]
Randomized Controlled Trial Example

- Given that the null is true, $\pi_1 - \pi_2 = 0$, we can calculate the test statistic

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}}$$

- Using our example:
Randomized Controlled Trial Example

Given that the null is true, $\pi_1 - \pi_2 = 0$, we can calculate the test statistic

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1 (1-\pi_1)}{n_1} + \frac{\pi_2 (1-\pi_2)}{n_2}}}$$

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\frac{\hat{\pi}_1 (1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2 (1-\hat{\pi}_2)}{n_2}}}$$

Using our example:

$$= \frac{315/604 - 304/597}{\sqrt{315/604(1-315/604)/604 + 304/597(1-304/597)/597}}$$

$$= 0.4637$$
Randomized Controlled Trial Example

Step 4: Calculate p-value

Note: Again, with proportions, standard normal works well (no need to use t)

Let's use a two-tailed test

Here, p-value = 0.642813

Step 5: Decide whether or not to reject the null hypothesis and interpret results

What would you think?

Evidence to reject null of no effect of prayer?

Using a one-sided test?
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value

- Note: Again, with proportions, standard normal works well (no need to use $t$).
- Let's use a two-tailed test.
- Here, $p$-value = 0.642813.
- Step 5: Decide whether or not to reject the null hypothesis and interpret results.

What would you think?

Evidence to reject null of no effect of prayer?

Using a one-sided test?
Randomized Controlled Trial Example

- Step 4: Calculate \( p \)-value
- Note: Again, w/ proportions, standard normal works well (no need to use \( t \))
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value
- Note: Again, w/ proportions, standard normal works well (no need to use $t$)
- Let’s use a two-tailed test
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value
- Note: Again, w/ proportions, standard normal works well (no need to use $t$)
- Let’s use a two-tailed test
  - Here, $p$-value = 0.642813
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value
- Note: Again, w/ proportions, standard normal works well (no need to use $t$)
- Let’s use a two-tailed test
  - Here, $p$-value = 0.642813
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value
- Note: Again, w/ proportions, standard normal works well (no need to use $t$)
- Let’s use a two-tailed test
  - Here, $p$-value = 0.642813
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
  - What would you think?
Randomized Controlled Trial Example

- **Step 4:** Calculate *p*-value
- **Note:** Again, w/ proportions, standard normal works well (no need to use *t*)
- **Let’s use a two-tailed test**
  - Here, *p*-value = 0.642813
- **Step 5:** Decide whether or not to reject the null hypothesis and interpret results
  - What would you think?
  - Evidence to reject null of no effect of prayer?
Randomized Controlled Trial Example

- Step 4: Calculate $p$-value
- Note: Again, w/ proportions, standard normal works well (no need to use $t$)
- Let’s use a two-tailed test
  - Here, $p$-value = 0.642813
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
  - What would you think?
  - Evidence to reject null of no effect of prayer?
  - Using a one-sided test?
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Will not guarantee practical significance
3) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
Potential issues with Hypothesis Testing

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Will not guarantee practical significance
3) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
Hypothesis tests are not perfect!
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Will not guarantee practical significance
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

► 1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
► 2) Will not guarantee practical significance
► 3) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
1) Hypotheses Tests test the null hypothesis
1) Hypothesis Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null
- NOT whether you accept the alternative
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null
- NOT whether you accept the alternative
- (A bit like proof by contradiction)
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null
- NOT whether you accept the alternative
- (A bit like proof by contradiction)
- Is $p$-value the probability that the null is true?
2) Practical versus Statistical Significance

A very small $p$-value provides strong evidence to reject the null hypothesis.

But: statistical significance $\neq$ practical significance.

We can often get small $p$-values due to large sample sizes.
2) Practical versus Statistical Significance

- A very small $p$-value provides strong evidence to reject null hypothesis
2) Practical versus Statistical Significance

- A very small $p$-value provides strong evidence to reject null hypothesis
- But: statistical significance $\neq$ practical significance
2) Practical versus Statistical Significance

- A very small $p$-value provides strong evidence to reject null hypothesis
- But: statistical significance $\neq$ practical significance
- We can often get small $p$-values due to large sample sizes
2) Practical versus Statistical Significance

Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

$H_0: \mu_1 - \mu_2 = 0,

H_a: \mu_1 - \mu_2 \neq 0$

$n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value < 0.00001

→ What would you do?

Would it affect your decision if intervention cost $1$ billion? $500,000? Costless?
Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

H_{0}: \mu_{1} - \mu_{2} = 0, \ H_{a}: \mu_{1} - \mu_{2} \neq 0

n = 1 million study yields difference of 0.02 on 100-point test and \ p\ - \ value < 0.00001

→ What would you do?

Would it affect your decision if intervention cost $1 billion? $500,000? Costless?
Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

\[ H_0 : \mu_1 - \mu_2 = 0, \quad H_a : \mu_1 - \mu_2 \neq 0 \]
2) Practical versus Statistical Significance

- Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)
  - $H_0: \mu_1 - \mu_2 = 0, \quad H_a: \mu_1 - \mu_2 \neq 0$
  - $n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value $< 0.00001$

Would it affect your decision if intervention cost $1$ billion? $500,000$? Costless?
Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

- $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$
- $n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value < 0.00001
- $\rightarrow$ What would you do?
Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

- $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$
- $n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value $< 0.00001$
- → What would you do?
- Would it affect your decision if intervention cost $1$ billion?
2) Practical versus Statistical Significance

▶ Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

▶ $H_0 : \mu_1 - \mu_2 = 0, \ H_a : \mu_1 - \mu_2 \neq 0$

▶ $n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value < 0.00001

▶ → What would you do?

▶ Would it affect your decision if intervention cost $1$ billion? $500,000$?
2) Practical versus Statistical Significance

Example: Suppose you have intervention to increase test scores of poor children (Group 1) versus affluent children (Group 2)

- $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$
- $n = 1$ million study yields difference of 0.02 on 100-point test and $p$-value $< 0.00001$
- $\rightarrow$ What would you do?
- Would it affect your decision if intervention cost $1$ billion? $500,000$? Costless?
3) No full certainty about correctly rejecting null hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true. However: Do not allow us to rule out null hypothesis with 100% certainty. Specifically, we're concerned about two scenarios: we reject the null hypothesis, even though the null is true; we fail to reject the null hypothesis, even though the null is false. Sound familiar? (Conditional statements)
3) No full certainty about correctly rejecting null

- Hypothesis tests given us *p*-values, or how extreme our test statistic is given null being true.
3) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
3) No full certainty about correctly rejecting null

- Hypothesis tests given us \( p \)-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios
3) No full certainty about correctly rejecting null

- Hypothesis tests given us \( p \)-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios
- We reject the null hypothesis, even though the null is true
3) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios
  - We reject the null hypothesis, even though the null is true
  - We fail to reject the null hypothesis, even though the null is false
3) No full certainty about correctly rejecting null

- Hypothesis tests given us \( p \)-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios
  - We reject the null hypothesis, even though the null is true
  - We fail to reject the null hypothesis, even though the null is false
- Sound familiar? (Conditional statements)
Type I versus Type II Errors

- $H_0$ is true
- $H_0$ is not true

- Reject $H_0$
- Do not reject $H_0$
Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not reject</td>
<td>$H_0$</td>
<td></td>
</tr>
</tbody>
</table>
# Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td></td>
<td>Good</td>
</tr>
<tr>
<td>Do not reject $H_0$</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>
Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td>Bad!</td>
<td>Good</td>
</tr>
<tr>
<td>Do not reject $H_0$</td>
<td>Good</td>
<td>Bad!</td>
</tr>
</tbody>
</table>
# Type I versus Type II Errors

- **$H_0$ is true**
  - **Reject $H_0$**: Type 1
  - **Do not reject $H_0$**: Good

- **$H_0$ is not true**
  - **Reject $H_0$**: Good
  - **Do not reject $H_0$**: Type 2
Type I versus Type II Errors

Both types of errors can occur with a hypothesis test.

Good practice: Before study carried out, set limits on how much error we would tolerate.
Type I versus Type II Errors

- Both types of errors can occur with a hypothesis test

Type I versus Type II Errors

- Both types of errors can occur with a hypothesis test
- Good practice: Before study carried out, set limits on how much error we would tolerate
Type I Error

▶ Type I: Null hypothesis rejected when in fact it is true
▶ Akin to false negative
▶ Mammogram analogy: Negative test, despite having the disease

▶ $P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha$

▶ $\alpha$: Also referred to as level of significance or the critical value

▶ Rejection region: Values of $\bar{X}$ to left/right of values of $\alpha$ for which the hypothesis test would reject the null

▶ Very common to set $\alpha = 0.05$, $\alpha = 0.10$, or $\alpha = 0.01$

▶ Question: Why would we want to avoid Type I Error?

▶ Question: How do you interpret a level of significance of 0.05?
Type I Error

- Type I: Null hypothesis rejected when in fact it is true

\[ P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \]

\( \alpha \): Also referred to as level of significance or the critical value

Rejection region: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null

Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)

Question: Why would we want to avoid Type I Error?

Question: How do you interpret a level of significance of 0.05?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease

\[
\Pr(\text{Type I error}) = \Pr(\text{Reject } H_0 | H_0 \text{ true}) = \alpha
\]

- \( \alpha \): Also referred to as level of significance or the critical value
- Rejection region: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null
- Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)

**Question**: Why would we want to avoid Type I Error?

**Question**: How do you interpret a level of significance of 0.05?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \)
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- **Akin to false negative**
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \)
- **\( \alpha \)**: Also referred to as level of significance or the critical value
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
- **Rejection region**: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
- **Rejection region**: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null
- Very common to set \( \alpha = 0.05, \alpha = 0.10, \) or \( \alpha = 0.01 \)
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
- **Rejection region**: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null
- Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)
- Question: Why would we want to avoid Type I Error?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \[ P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \]
- \( \alpha \): Also referred to as level of significance or the critical value
- **Rejection region**: Values of \( \bar{X} \) to left/right of values of \( \alpha \) for which the hypothesis test would reject the null
- Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)
- Question: Why would we want to avoid Type I Error?
- Question: How do you interpret a level of significance of 0.05?
Type II Error

- Type II: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
- Mammogram analogy: Positive test, despite not having the disease

\[ P(\text{Type II error}) = P(\text{Not rejecting } H_0 | H_0 \text{ false}) \]

- Often set at \[ P(\text{Type II error}) = 0.20 = \beta \]
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
Type II Error

- Type II: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \[ P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \]
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \( P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \)
- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
Type II Error and Power

$\beta = P(\text{Rejecting } H_0 | H_0 \text{ false})$

Probability of correctly rejecting the null

Known as the power of a test

Sometimes considered less important from substantive perspective

However: consider your specific problem → in some instances, either error may be more costly

Note: Exact power calculation will depend on your alternative $H_a$
Type II Error and Power

$1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false})$
Type II Error and Power

- $1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false})$
- Probability of correctly rejecting the null
Type II Error and Power

- $1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false})$
- Probability of correctly rejecting the null
- Known as the power of a test
Type II Error and Power

1 − β = P(Rejecting $H_0|H_0 \text{ false}$)

- Probability of correctly rejecting the null
- Known as the *power* of a test
- Sometimes considered less important from substantive perspective
Type II Error and Power

- \( 1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false}) \)
- Probability of correctly rejecting the null
- Known as the power of a test
- Sometimes considered less important from substantive perspective
- However: consider your specific problem → in some instances, either error may be more costly
Type II Error and Power

- $1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false})$
- Probability of correctly rejecting the null
- Known as the power of a test
- Sometimes considered less important from substantive perspective
- However: consider your specific problem $\rightarrow$ in some instances, either error may be more costly
- Note: Exact power calculation will depend on your alternative $H_a$
Power Analyses

Power Analysis: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true:

$P(\text{Rejecting } H_0 | H_a \text{ true})$

Usually approached in one of two ways:

1) You are asked to calculate the sample size you will need to detect an effect (difference between null and alternative) of a certain size
   - Need to state a precise alternative null
2) You are asked to calculate what is the smallest effect (difference) you could detect given a sample size
   - Both mean that power analyses usually done before data are collected (and $ spent!$)
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?
Power Analyses

- Power Analysis: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?
- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true: 
  \[ P(\text{Rejecting } H_0 | H_a \text{ true}) \]
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true: $P(\text{Rejecting } H_0 | H_a \text{ true})$

- Usually approached in one of two ways:
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true:
  \[ P(\text{Rejecting } H_0|H_a \text{ true}) \]

- Usually approached in one of two ways:

  1) You are asked to calculate the sample size you will need to detect an effect (difference between null and alternative) of a certain size

  → Both mean that power analyses usually done before data are collected (and $ spent!)
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true: 

  \[ P(\text{Rejecting } H_0 | H_a \text{ true}) \]

- Usually approached in one of two ways:

  1) You are asked to calculate the sample size you will need to detect an effect (difference between null and alternative) of a certain size

     - Need to state a precise alternative null

  2) You are asked to calculate what is the smallest effect (difference) you could detect given a sample size

- Both mean that power analyses usually done before data are collected (and $ spent!)
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true: $P(\text{Rejecting } H_0 | H_a \text{ true})$

- Usually approached in one of two ways:

  1) You are asked to calculate the sample size you will need to detect an effect (difference between null and alternative) of a certain size
     - Need to state a precise alternative null

  2) You are asked to calculate what is the smallest effect (difference) you could detect given a sample size

- Both mean that power analyses usually done before data are collected (and $ spent!)
Power Analyses

- **Power Analysis**: If a given alternative hypothesis were true, how good would our test be at (correctly) rejecting the null?

- Somewhat similar to hypothesis test set up, but for purposes of calculating power, assume alternative true:
  \[ P(\text{Rejecting } H_0|H_a \text{ true}) \]

- Usually approached in one of two ways:
  1) You are asked to calculate the sample size you will need to detect an effect (difference between null and alternative) of a certain size
     - Need to state a precise alternative null
  2) You are asked to calculate what is the smallest effect (difference) you could detect given a sample size

→ Both mean that power analyses usually done before data are collected (and $ spent!)
Power Analysis Example

You are an urban policy expert studying commuting times. Suppose you are considering testing whether average time spent in commuting is 4 hrs/week versus an alternative that it is greater due to recent construction. Suppose further that you know (assume) the true standard deviation to be 2 hrs.

Find power when:

a) sample size equals 16 and
b) alternative hypothesis is that commuting hours equals 6 hrs (so effect of construction is 2 hrs)
Power Analysis Example

- You are an urban policy expert studying commuting times
Power Analysis Example

- You are an urban policy expert studying commuting times
- Suppose you are considering testing whether average time spent in commuting is 4 hrs/week versus alternative that it is greater due to recent construction

- Find power when a) sample size equals 16 and b) alternative hypothesis is that commuting hours equals 6hrs/week so effect of construction is 2hrs
You are an urban policy expert studying commuting times.

Suppose you are considering testing whether average time spent in commuting is 4 hrs/week versus alternative that it is greater due to recent construction.

Suppose further that you know (assume) true standard deviation to be 2hrs.

Find power when:

a) sample size equals 16
b) alternative hypothesis is that commuting hours equals 6hrs (so effect of construction is 2hrs)
Power Analysis Example

▶ You are an urban policy expert studying commuting times
▶ Suppose you are considering testing whether average time spent in commuting is 4 hrs/week versus alternative that it is greater due to recent construction
▶ Suppose further that you know (assume) true standard deviation to be 2hrs
▶ Find power when a) sample size equals 16 and b) alternative hypothesis is that commuting hours equals 6hrs (so effect of construction is 2hrs)
Power Analysis Example

Steps:
▶ Because you'll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard setup (with given $\alpha$ values)
▶ Assume alternative hypothesis is true, $\mu = 6$
▶ Calculate probability of not being in rejection region when this alternative is true (this is $\beta$)
▶ And then finally calculate $1 - \beta$ to get Power
▶ Note: Can repeat for different values of $\mu$ (e.g., $\mu = 6, 7, 8, 9$) to plot a power curve or power function
Power Analysis Example

Steps:

Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values).

Assume alternative hypothesis is true, $\mu = 6$.

Calculate probability of not being in rejection region when this alternative is true (this is $\beta$).

And then finally calculate $1 - \beta$ to get Power.

Note: Can repeat for different values of $\mu$ (e.g., $\mu = 6, 7, 8, 9$) to plot a power curve or power function.
Power Analysis Example

Steps:

▶ Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values)
Power Analysis Example

Steps:

▶ Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values)
▶ Assume alternative hypothesis is true, $\mu = 6$
Power Analysis Example

Steps:

▶ Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values)
▶ Assume alternative hypothesis is true, $\mu = 6$
▶ Calculate probability of not being in rejection region when this alternative is true (this is $\beta$)
Steps:

▶ Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values)
▶ Assume alternative hypothesis is true, $\mu = 6$
▶ Calculate probability of not being in rejection region when this alternative is true (this is $\beta$)
▶ And then finally calculate $1 - \beta$ to get Power
Power Analysis Example

Steps:

- Because you’ll eventually do a hypothesis test assuming the null, first calculate the rejection region under standard set-up (with given $\alpha$ values)
- Assume alternative hypothesis is true, $\mu = 6$
- Calculate probability of not being in rejection region when this alternative is true (this is $\beta$)
- And then finally calculate $1 - \beta$ to get Power
- Note: Can repeat for different values of $\mu$ (e.g., $\mu = 6, 7, 8, 9$) to plot a power curve or power function
Power Analysis Intuition

---

![Power Analysis Intuition](image-url)
Power Analysis Intuition
Power Analysis Intuition

![Diagram showing critical regions for rejecting the null hypothesis. The critical values are at \(-2\) and \(2\) on the standard normal distribution. The shaded areas represent the critical regions for rejecting the null hypothesis (Reject) when the observed value of the test statistic is outside these regions. The middle region (Do not reject) is where the observed value falls, suggesting no rejection of the null hypothesis.]
Power Analysis Intuition

![Graph showing the intuition of power analysis](image)
Power Analysis Intuition

- **Null Hypothesis (H₀)**: The null hypothesis is represented by a normal distribution centered at 0.
- **Alternative Hypothesis (H₁)**: The alternative hypothesis is represented by a normal distribution that is shifted away from the null hypothesis.

- **Type II Error**: This occurs when the null hypothesis is not rejected, even though it is false. It is represented by the area under the alternative hypothesis curve that falls outside the critical region.

- **Power**: This is the probability of correctly rejecting the null hypothesis when it is false. It is represented by the area under the alternative hypothesis curve that falls inside the critical region.

- **Critical Region**: The critical region is the range of values for the test statistic that leads to the rejection of the null hypothesis. It is represented by the shaded area on the right side of the graph.

- **Decision Rule**: The decision rule is to reject the null hypothesis if the test statistic falls within the critical region.

- **Distribution of Test Statistic (z)**: The distribution of the test statistic (z) is used to determine whether the null hypothesis should be rejected. The test statistic is standardized by subtracting the mean and dividing by the standard deviation.
Power Analysis Example

Step 1: Calculate rejection region for eventual (null) hypothesis test.

Remember that the test statistic for a single mean is:

\[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

Assuming a one-tailed test, we reject \( H_0 \) if \( z > 1.645 \) (\( \alpha = 0.05 \)).

Step 2: Calculate rejection regions under null.

\[ 1.645 < \bar{X} - 4 \]
\[ \bar{X} > 4.8825 \]

That is, if the data have a sample mean \( \bar{X} > 4.89 \) you would reject the null.
Power Analysis Example

- Step 1: Calculate rejection region for eventual (null) hypothesis test
Power Analysis Example

- Step 1: Calculate rejection region for eventual (null) hypothesis test
- Remember that test statistic for single mean is

\[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]
Power Analysis Example

- **Step 1**: Calculate rejection region for eventual (null) hypothesis test
- Remember that test statistic for single mean is
  \[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]
- Assuming one-tailed test, we reject \( H_0 \) if \( z > 1.645 \) (\( \alpha = 0.05 \))
Power Analysis Example

- Step 1: Calculate rejection region for eventual (null) hypothesis test
- Remember that test statistic for single mean is
  \[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]
- Assuming one-tailed test, we reject \( H_0 \) if \( z > 1.645 \) (\( \alpha = 0.05 \))
- Step 2: Calculate rejection regions under null
Power Analysis Example

- Step 1: Calculate rejection region for eventual (null) hypothesis test
- Remember that test statistic for single mean is
  \[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

- Assuming one-tailed test, we reject \( H_0 \) if \( z > 1.645 \) (\( \alpha = 0.05 \))
- Step 2: Calculate rejection regions under null
  \[ 1.645 < \frac{\bar{X} - 4}{1/2} \rightarrow \bar{X} > 4.8825 \]
Power Analysis Example

- Step 1: Calculate rejection region for eventual (null) hypothesis test
- Remember that test statistic for single mean is
  \[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]
- Assuming one-tailed test, we reject \( H_0 \) if \( z > 1.645 \) (\( \alpha = 0.05 \))
- Step 2: Calculate rejection regions under null
  \[ 1.645 < \frac{\bar{X} - 4}{1/2} \rightarrow \bar{X} > 4.8825 \]
- That is, if data have a sample mean of \( \bar{X} > 4.89 \) you would reject null
Power Analysis Example

Now, assume alternative true, \( \mu = 6 \).

If \( \mu \) was 6, how often would we get a sample mean in the rejection region of null of \( \mu = 4 \)? That is, what is \( P(\bar{X} > 4.89) \) if \( \mu = 6 \)?

\[
Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{4.89 - 6}{1/2} = -2.22
\]

Finally \( P(Z < -2.22) = 0.013 \). So \( \text{Power} = 1 - P(Z < -2.22) = 0.987 \).
Power Analysis Example

- Now, assume alternative true, \( \mu = 6 \)
Power Analysis Example

- Now, assume alternative true, $\mu = 6$
- If $\mu$ was 6, how often would we get a sample mean in the rejection region of null of $\mu = 4$?

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{4.89 - 6}{1/2} = -2.22$$

$$P(Z < -2.22) = 0.013$$

So, Power $= 1 - P(Z < -2.22) = 0.987$
Now, assume alternative true, $\mu = 6$

If $\mu$ was 6, how often would we get a sample mean in the rejection region of null of $\mu = 4$?

That is, what is $P(\bar{X} > 4.89)$ if $\mu = 6$?
Power Analysis Example

- Now, assume alternative true, $\mu = 6$
- If $\mu$ was 6, how often would we get a sample mean in the rejection region of null of $\mu = 4$?
- That is, what is $P(\bar{X} > 4.89)$ if $\mu = 6$?

\[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]
\[ = \frac{4.89 - 6}{1/2} \]
\[ = -2.22 \]
Power Analysis Example

▶ Now, assume alternative true, $\mu = 6$
▶ If $\mu$ was 6, how often would we get a sample mean in the rejection region of null of $\mu = 4$?
▶ That is, what is $P(\bar{X} > 4.89)$ if $\mu = 6$?

\[
z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}
\]

\[
= \frac{4.89 - 6}{1/2}
\]

\[
= -2.22
\]

▶ Finally $P(Z < -2.22) = 0.013$
Power Analysis Example

- Now, assume alternative true, $\mu = 6$
- If $\mu$ was 6, how often would we get a sample mean in the rejection region of null of $\mu = 4$?
- That is, what is $P(\bar{X} > 4.89)$ if $\mu = 6$?

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{4.89 - 6}{1/2}$$

$$= -2.22$$

- Finally $P(Z < -2.22) = 0.013$
- So $Power = 1 - P(Z < -2.22) = 0.987$
Power

- Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
- Bigger difference between null and alternative you are testing $\rightarrow$ requires less power
- A two-sided hypothesis test has less power than the one-sided hypothesis, since it is more conservative
- Rule of thumb: 80% power
- Higher power may be better, but perhaps not if it comes in changing the type of test (b/c it affects Type 1 error)
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
- Bigger difference between null and alternative you are testing $\rightarrow$ requires less power
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
- Bigger difference between null and alternative you are testing $\rightarrow$ requires less power
- A two-sided hypothesis test has less power than the one-sided hypothesis, since it is more conservative
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
- Bigger difference between null and alternative you are testing $\rightarrow$ requires less power
- A two-sided hypothesis test has less power than the one-sided hypothesis, since it is more conservative
- Rule of thumb: 80% power
Power

- Note: Power dependent on a) sample size, b) your $H_a$ (size of effect), and c) type of test (change $\alpha$ level)
- Larger sample size $\rightarrow$ more power
- Bigger difference between null and alternative you are testing $\rightarrow$ requires less power
- A two-sided hypothesis test has less power than the one-sided hypothesis, since it is more conservative
- Rule of thumb: 80% power
- Higher power may be better, but perhaps not if it comes in changing the type of test (b/c it affects Type 1 error)
Next time

- Interval estimation and confidence intervals
Proof of Expected Value of $\hat{\pi}$

- If $X$ can only take on 0, 1 values with constant probability, $\pi$, then $\sum x_i$ follows a binomial distribution
- Furthermore, the expected value of a binomial is $n\pi$ (or $np$, in our previous notation)
- Therefore:

\[
E[\hat{\pi}] = E\left[\frac{\sum x_i}{n}\right] = \frac{1}{n}E\left[\sum x_i\right] = \frac{np}{n} = p
\]
Proof of Variance of $\hat{\pi}$

- The variance of a binomial is $n\pi(1 - \pi)$ (or $np(1 - p)$, using our previous notation)
- Therefore:

\[
Var[\hat{\pi}] = Var\left[\frac{\sum x_i}{n}\right]
\]

\[
= \frac{1}{n^2} Var\left[\sum x_i\right]
\]

\[
= \frac{1}{n^2} \times n\pi(1 - \pi)
\]

\[
= \frac{\pi(1 - \pi)}{n}
\]