Stat 201: Statistics I
Week 8

Metropolitan State University

Week 8
Hypothesis Testing

Section 8.1
More on Confidence Intervals

Sample size

A confidence interval defines a region of probable values for a population parameter. Often, it is desirable to design an experiment which will estimate a population parameter within a specified accuracy. The sample size needed to achieve a desired margin of error can be calculated.

Calculating sample size

Recall, a confidence interval is defined as,

$$CI(1-\alpha)\% = x \pm ME$$

where the margin of error is calculated by,

$$ME = z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Solving for n (sample size) results in, after some algebra,

$$n = \left(\frac{s \times z_{\alpha/2}}{ME}\right)^2$$

Calculating sample size, cont.

$$n = \left(\frac{s \times z_{\alpha/2}}{ME} \right)^2$$

A sample size calculation requires the following values:

Calculating sample size, cont.

$$n = \left(\frac{s \times z_{\alpha/2}}{ME}\right)^2$$

A sample size calculation requires the following values:

- The critical value $z_{\alpha/2}$ specified by the desired confidence level $(1-\alpha)\%$

Calculating sample size, cont.

$$n = \left(\frac{s \times z_{\alpha/2}}{ME}\right)^2$$

A sample size calculation requires the following values:

- The critical value $z_{\alpha/2}$ specified by the desired confidence level $(1-\alpha)\%$
- The desired accuracy of the estimate specified by an acceptable margin of error

Calculating sample size, cont.

$$n = \left(\frac{s \times z_{\alpha/2}}{ME} \right)^2$$

A sample size calculation requires the following values:

- The critical value $z_{\alpha/2}$ specified by the desired confidence level $(1 - \alpha)\%$
- The desired accuracy of the estimate specified by an acceptable margin of error
- The sample standard deviation s. Finding a suitable value for s is often the most difficult part of a sample size calculation.

Standard deviations for means

A reasonable estimate for sample standard deviation must be determined when calculation sample sizes for confidence intervals of means.

Possible sources:

Standard deviations for means

A reasonable estimate for sample standard deviation must be determined when calculation sample sizes for confidence intervals of means.

Possible sources:

- Known population values

Standard deviations for means

A reasonable estimate for sample standard deviation must be determined when calculation sample sizes for confidence intervals of means.

Possible sources:

- Known population values
- Previous published studies

Standard deviations for means

A reasonable estimate for sample standard deviation must be determined when calculation sample sizes for confidence intervals of means.

Possible sources:

- Known population values
- Previous published studies
- Pilot studies

Standard deviations for means

A reasonable estimate for sample standard deviation must be determined when calculation sample sizes for confidence intervals of means.

Possible sources:

- Known population values
- Previous published studies
- Pilot studies
- Sometimes a sample size calculation is performed using a margin of error defined as a proportion of an unknown stand deviation (i.e. a margin of error of half a standard deviation)

Standard deviations for proportions

When working with proportions and binomial distributions, standard deviations are calculated from the sample probability or proportion rather than measured directly from the data. Thus, a reasonable estimate of the sample proportion is required.

Standard deviations for proportions

When working with proportions and binomial distributions, standard deviations are calculated from the sample probability or proportion rather than measured directly from the data. Thus, a reasonable estimate of the sample proportion is required.

- Proportion estimates may be estimated from previous work or known values in a similar manner as standard deviations for means

Standard deviations for proportions

When working with proportions and binomial distributions, standard deviations are calculated from the sample probability or proportion rather than measured directly from the data. Thus, a reasonable estimate of the sample proportion is required.

- Proportion estimates may be estimated from previous work or known values in a similar manner as standard deviations for means
- If no reasonable estimate can be made, a conservative value of $\hat{p} = 0.5$ should be used.

Then, the sample size calculation for a confidence interval of proportions in,

$$n = \left(\frac{\sqrt{\hat{p} \times (1 - \hat{p})} \times z_{\alpha/2}}{ME} \right)^2 = \hat{p} \times (1 - \hat{p}) \times \left(\frac{z_{\alpha/2}}{ME} \right)^2$$

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

- A 95% confidence level means $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

- A 95% confidence level means $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$
- The desired margin of error is 2 inches

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

- A 95% confidence level means $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$
- The desired margin of error is 2 inches
- Previous studies have estimated the mean height of adult males in the U.S. is 69.2 inches with a standard deviation of 5.79.

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

- A 95% confidence level means $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$
- The desired margin of error is 2 inches
- Previous studies have estimated the mean height of adult males in the U.S. is 69.2 inches with a standard deviation of 5.79.

Then,

$$n = \left(\frac{s \times z_{\alpha/2}}{ME}\right)^2 = \left(\frac{5.79 \times 1.96}{2}\right)^2 = 32.197 \Rightarrow 33$$

Calculating sample size, example

Example

Metro State wants to know the mean height of its male students within ± 2 inches with 95% confidence. How large of a sample is required to get the desired results?

- A 95% confidence level means $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$
- The desired margin of error is 2 inches
- Previous studies have estimated the mean height of adult males in the U.S. is 69.2 inches with a standard deviation of 5.79.

Then,

$$n = \left(\frac{s \times z_{\alpha/2}}{ME}\right)^2 = \left(\frac{5.79 \times 1.96}{2}\right)^2 = 32.197 \Rightarrow 33$$

- Calculated sample sizes should always be rounded **up**.

Confidence intervals as inference

A confidence interval from a sample can be used to determine whether population the sample was drawn from has a parameter that matches a value of interest.

Confidence intervals as inference

A confidence interval from a sample can be used to determine whether population the sample was drawn from has a parameter that matches a value of interest.

- If the value of interest **is not** contained within the confidence interval, then there is evidence that the population parameter differs from the value.

Confidence intervals as inference

A confidence interval from a sample can be used to determine whether population the sample was drawn from has a parameter that matches a value of interest.

- If the value of interest **is not** contained within the confidence interval, then there is evidence that the population parameter differs from the value.
- If the value if interest **is** contained within the confidence interval, then there is not evidence the the parameter differs from the value. [Note: this is not the same as saying there is evidence that the parameter is the same as the value.]

Confidence intervals as inference, example

Example

Suppose Metro State conducts a study of height of male students which results in a 95% confidence interval of (63.3, 67.9). What can be said about Metro State students as compared to the general U.S. population?

Confidence intervals as inference, example

Example

Suppose Metro State conducts a study of height of male students which results in a 95% confidence interval of (63.3, 67.9). What can be said about Metro State students as compared to the general U.S. population?

- Since the mean height of U.S. males, 69.2 inches, is not contained in the confidence interval, there is evidence that Metro State students differ from the general U.S. population.

Group work

- For all questions, complete parts (a) and (b).

Section 8.2
Basics of Hypothesis Testing

Statistical inference

Previously, statistics from random samples were used to learn something about populations by estimating population parameters. Knowledge about populations was inferred from the data of the sample.

A similar question can be posed: Is a sample drawn from a population that is the same in an important way to a known population, or is the sample drawn from a population that is significantly different?

Hypothesis tests

An **hypothesis test** is a formal statistical procedure to test claims about population parameters based on samples drawn from populations. Such claims, or **hypotheses**, are often written as simple mathematical statements.

It is important to be clear as to which population the claims or the tests are about,

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students
  - $\mu < 69.2$

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students
  - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
    - Population: Male Metro State students
    - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%
    - Population: Teen drivers

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students
  - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%
  - Population: Teen drivers
  - $p = 0.4$

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students
  - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%
  - Population: Teen drivers
  - $p = 0.4$
- A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A new experimental treatment will extend the survival time.

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
    - Population: Male Metro State students
    - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%
    - Population: Teen drivers
    - $p = 0.4$
- A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A new experimental treatment will extend the survival time.
    - Population: Patients with disease on experimental treatment

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
  - Population: Male Metro State students
  - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%
  - Population: Teen drivers
  - $p = 0.4$
- A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A new experimental treatment will extend the survival time.
  - Population: Patients with disease on experimental treatment
  - $\mu > 36$

Hypotheses

The first step to conduct an hypothesis test is to identify two hypotheses.

Hypotheses

The first step to conduct an hypothesis test is to identify two hypotheses.

The **null hypothesis** is the claim that nothing interesting has occurred, that a sub-population is **not** different than the general population or that population parameters did **not** change after treatment.

Hypotheses

The first step to conduct an hypothesis test is to identify two hypotheses.

The **null hypothesis** is the claim that nothing interesting has occurred, that a sub-population is **not** different than the general population or that population parameters did **not** change after treatment.

Conversely, the **alternative hypothesis** is the claim that something interesting has occurred, that a sub-population is different or that parameters did change after treatment.

Hypotheses

The first step to conduct an hypothesis test is to identify two hypotheses.

The **null hypothesis** is the claim that nothing interesting has occurred, that a sub-population is **not** different than the general population or that population parameters did **not** change after treatment.

Conversely, the **alternative hypothesis** is the claim that something interesting has occurred, that a sub-population is different or that parameters did change after treatment.

Remember, both the null and alternative hypotheses are statements about populations, which will be tested using a sample.

Hypotheses, cont.

Null hypothesis:

- Denoted by $\boldsymbol{H_0}$
- Always a statement that a parameter is **equal to** some value
- That value, denoted p_0 or μ_0, is called the proportion or mean under the null hypothesis

Hypotheses, cont.

Null hypothesis:

- Denoted by $\boldsymbol{H_0}$
- Always a statement that a parameter is **equal to** some value
- That value, denoted p_0 or μ_0, is called the proportion or mean under the null hypothesis

Alternative hypothesis:

- Denoted by $\boldsymbol{H_1}$ or $\boldsymbol{H_a}$
- Can be a statement that a parameter is **less than**, **greater than** or **not equal to** some value
- Is usually a statement representing the research question

One-sided vs. two sided tests

If an alternative hypothesis has the form of a parameter being less than or greater than some value, the hypothesis test is called a **one-sided test**.

If an alternative hypothesis has the form of a parameter being not equal to some value, the hypothesis test is called a **two-sided test**.

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

In the United States, adult men have a mean height of 69.2 inches. The Metro State administration want to do a study to see if male Metro State students are shorter than the general US population.

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

In the United States, adult men have a mean height of 69.2 inches. The Metro State administration want to do a study to see if male Metro State students are shorter than the general US population.

- $H_0 : \mu = 69.2$ $\qquad H_a : \mu < 69.2$ $\qquad$ One-sided

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

In the United States, adult men have a mean height of 69.2 inches. The Metro State administration want to do a study to see if male Metro State students are shorter than the general US population.

- $H_0 : \mu = 69.2$ $\qquad H_a : \mu < 69.2$ $\qquad$ One-sided

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time.

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

In the United States, adult men have a mean height of 69.2 inches. The Metro State administration want to do a study to see if male Metro State students are shorter than the general US population.

- $H_0 : \mu = 69.2 \qquad H_a : \mu < 69.2$ One-sided

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time.

- $H_0 : \mu = 36 \qquad H_a : \mu \neq 36$ Two-sided

Structure of hypothesis tests

Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

Structure of hypothesis tests

Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

- A test statistic is calculated which indicates the location of the sample within the sampling distribution, assuming the null hypothesis is true.

Structure of hypothesis tests

Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

- A test statistic is calculated which indicates the location of the sample within the sampling distribution, assuming the null hypothesis is true.
- The probability of getting a test statistic equal to or more extreme than the statistics belonging to the sample is calculated.

Structure of hypothesis tests

Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

- A test statistic is calculated which indicates the location of the sample within the sampling distribution, assuming the null hypothesis is true.
- The probability of getting a test statistic equal to or more extreme than the statistics belonging to the sample is calculated.
- If the calculated probability is below a pre-specified threshold, the null hypothesis is **rejected** and it is said that there is evidence to support the alternative hypothesis.

Structure of hypothesis tests

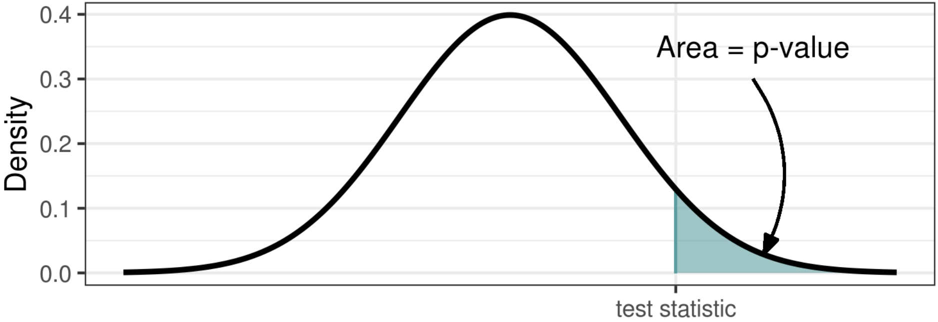
Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

- A test statistic is calculated which indicates the location of the sample within the sampling distribution, assuming the null hypothesis is true.
- The probability of getting a test statistic equal to or more extreme than the statistics belonging to the sample is calculated.
- If the calculated probability is below a pre-specified threshold, the null hypothesis is **rejected** and it is said that there is evidence to support the alternative hypothesis.
- If the calculated probability is not below the pre-specified threshold, the null hypothesis is **not rejected** and it is said that there is not evidence to support the alternative hypothesis.

P-values

In a hypothesis test, the **p-value** is the probability of getting a sample with the test statistic or one more extreme, assuming the null hypothesis is true.

- Not to be confused with the population proportion p or the probability function $P(A)$, though a p-value does represent a probability.



Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

- Identify sampling distribution:
  - z distribution for proportions
  - t distribution for means

Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

- Identify sampling distribution:
  - z distribution for proportions
  - t distribution for means
- Calculate **test statistic**: z-score or t-score

Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

- Identify sampling distribution:
  - z distribution for proportions
  - t distribution for means
- Calculate **test statistic**: z-score or t-score
- Find probability of test statistic or more extreme values in sampling distribution

Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

- Identify sampling distribution:
  - z distribution for proportions
  - t distribution for means
- Calculate **test statistic**: z-score or t-score
- Find probability of test statistic or more extreme values in sampling distribution
- That probability is the p-value

Luckily, all these steps can be accomplished easily with technology.

Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

- Denoted by α

Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

- Denoted by $\boldsymbol{\alpha}$
- This is the same α used for critical values and confidence intervals

Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

- Denoted by $\boldsymbol{\alpha}$
- This is the same α used for critical values and confidence intervals
- Thus, significance level can be thought of as an area in a sampling distribution

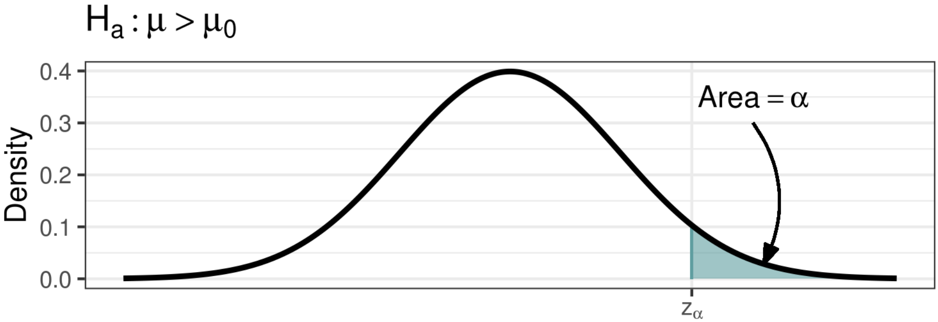
Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

- Denoted by $\boldsymbol{\alpha}$
- This is the same α used for critical values and confidence intervals
- Thus, significance level can be thought of as an area in a sampling distribution
- Sometimes referred to as the rejection region

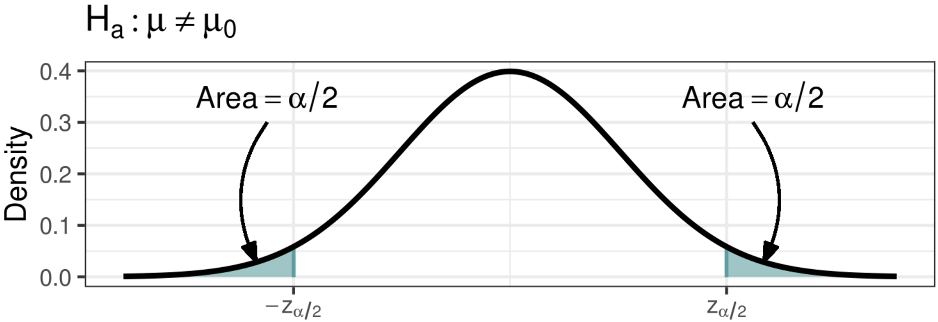
Significance level for one-sided test

In a one-sided test, the entire rejection region is located at one end of the distribution or the other.



Significance level for two-sided test

In a two-sided test, the rejection region is split between the lower and upper extremes of the distribution.



Stating conclusions

When reporting the results of an hypothesis tests, the following elements should be included:

Stating conclusions

When reporting the results of an hypothesis tests, the following elements should be included:

- Report the test statistic and p-value

Stating conclusions

When reporting the results of an hypothesis tests, the following elements should be included:

- Report the test statistic and p-value
- Report a decision on the null hypothesis based on the p-value and significance level (α).
  - State the decision as "Reject H_0" or "Do not reject H_0".
  - The null hypothesis is never "accepted".

Stating conclusions

When reporting the results of an hypothesis tests, the following elements should be included:

- Report the test statistic and p-value
- Report a decision on the null hypothesis based on the p-value and significance level (α).
  - State the decision as "Reject H_0" or "Do not reject H_0".
  - The null hypothesis is never "accepted".
- State the conclusion in terms of the research question
  - "There is evidence for..."
  - "There is not evidence for..."

Stating conclusions, example

Example

A study is conducted to test the claim that male Metro State students are shorter than the general population height of 69.2 inches. The test at a $\alpha = 0.05$ level of significance produces a test statistic $t = -1.859$ and a p-value of 0.0358. State the conclusion of the test.

Stating conclusions, example

Example

A study is conducted to test the claim that male Metro State students are shorter than the general population height of 69.2 inches. The test at a $\alpha = 0.05$ level of significance produces a test statistic $t = -1.859$ and a p-value of 0.0358. State the conclusion of the test.

- $H_0 : \mu = 69.2, \qquad H_a : \mu < 69.2$

Stating conclusions, example

Example

A study is conducted to test the claim that male Metro State students are shorter than the general population height of 69.2 inches. The test at a $\alpha = 0.05$ level of significance produces a test statistic $t = -1.859$ and a p-value of 0.0358. State the conclusion of the test.

- $H_0 : \mu = 69.2, \qquad H_a : \mu < 69.2$
- $p = 0.0358 < \alpha = 0.05$. Reject the null hypothesis. There is evidence to conclude that male Metro students are shorter than the general population.

Stating conclusions, example

Example

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time. The hypothesis test at $\alpha = 0.01$ level of significance produces a p-value of 0.098. State the conclusion of the test.

Stating conclusions, example

Example

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time. The hypothesis test at $\alpha = 0.01$ level of significance produces a p-value of 0.098. State the conclusion of the test.

- $H_0 : \mu = 36 \qquad H_a : \mu \neq 36$

Stating conclusions, example

Example

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time. The hypothesis test at $\alpha = 0.01$ level of significance produces a p-value of 0.098. State the conclusion of the test.

- $H_0 : \mu = 36 \qquad H_a : \mu \neq 36$
- $p = 0.098 > \alpha = 0.01$. Do not reject the null hypothesis. There is not evidence to conclude that the experimental treatment changes survival time.

Steps for hypothesis test

1. Identify null and alternative hypotheses from research question
2. Determine appropriate sampling distribution
3. Calculate test statistic
4. Calculate p-value
5. Compare p-value to significance level α and report decision
6. State conclusion in terms of original research question

Note: Steps 3 and 4 are often accomplished with technology

Making a decision with p-value

A **small** p-value ($p < \alpha$) means a low probability of getting the observed sample if the null hypothesis is true. Thus, the null hypothesis is rejected.

A **large** p-value ($p > \alpha$) then means that the sample is not unlikely under the null hypothesis. Thus, the null hypothesis is not rejected.

Making a decision with p-value

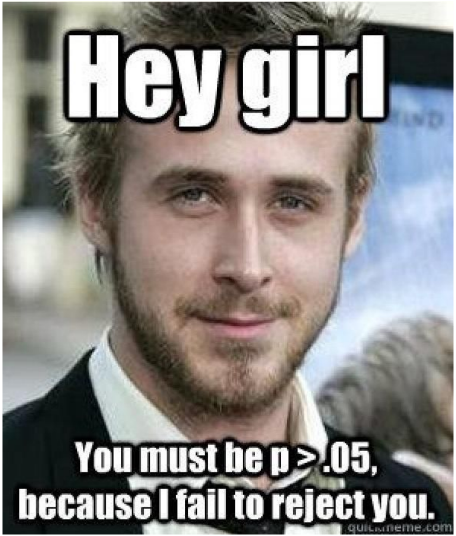
A **small** p-value ($p < \alpha$) means a low probability of getting the observed sample if the null hypothesis is true. Thus, the null hypothesis is rejected.

A **large** p-value ($p > \alpha$) then means that the sample is not unlikely under the null hypothesis. Thus, the null hypothesis is not rejected.

To remember. . .

If p is low, the null must go.

Or if this helps...



Critical value method

The hypothesis test procedure discussed thus far is known as the p-value method. An alternative method, known as the **critical value** method, does not use p-values, but rather compares test statistics directly to appropriate critical values. If the test statistic is more extreme than the critical value, the null hypothesis is rejected.

Critical value method, example

Example

A two-sided test with $\alpha = 0.05$ level of significance is conducted and results in a test statistic $z = -2.8$.

- Recall, critical value $z_{\alpha/2} = \pm 1.96$.
- Since -2.8 is more extreme than -1.96 ($-2.8 < -1.96$), reject the null hypothesis.

Types of errors in hypothesis tests

In an hypothesis test, either the null hypothesis is true or it is not, and either the null hypothesis is rejected or it is not. Thus, there are four possible outcomes.

	Reject H_0	Do not reject H_0
H_0 is true	Type I error (α)	Correct decision
H_0 is not true	Correct decision	Type II error (β)

Types of errors in hypothesis tests

In an hypothesis test, either the null hypothesis is true or it is not, and either the null hypothesis is rejected or it is not. Thus, there are four possible outcomes.

	Reject H_0	Do not reject H_0
H_0 is true	Type I error (α)	Correct decision
H_0 is not true	Correct decision	Type II error (β)

- A type I error occurs when the null hypothesis is rejected when it is in fact true. The probability of making a type I error in a test is designated α.

Types of errors in hypothesis tests

In an hypothesis test, either the null hypothesis is true or it is not, and either the null hypothesis is rejected or it is not. Thus, there are four possible outcomes.

	Reject H_0	Do not reject H_0
H_0 is true	Type I error (α)	Correct decision
H_0 is not true	Correct decision	Type II error (β)

- A type I error occurs when the null hypothesis is rejected when it is in fact true. The probability of making a type I error in a test is designated α.
- A type II error occurs when the null hypothesis is not rejected when it is in fact not true. The probability of making a type I error in a test is designated β.

Types of errors in hypothesis tests, cont.

- The acceptable probability of committing a type I error (α), or level of significance, is chosen prior to conducting a hypothesis test.

Types of errors in hypothesis tests, cont.

- The acceptable probability of committing a type I error (α), or level of significance, is chosen prior to conducting a hypothesis test.
- The probability of committing a type II error (β) is determined by a number of factors, including the chosen α and the sample size.

Types of errors in hypothesis tests, cont.

- The acceptable probability of committing a type I error (α), or level of significance, is chosen prior to conducting a hypothesis test.
- The probability of committing a type II error (β) is determined by a number of factors, including the chosen α and the sample size.
- The probability of **not** committing a type II error ($1 - \beta$) is known as the **power** of a test.

Types of errors in hypothesis tests, cont.

- The acceptable probability of committing a type I error (α), or level of significance, is chosen prior to conducting a hypothesis test.
- The probability of committing a type II error (β) is determined by a number of factors, including the chosen α and the sample size.
- The probability of **not** committing a type II error $(1 - \beta)$ is known as the **power** of a test.
- There is a trade-off between α and β. Smaller α result in larger β (and lower power) and vice versa.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.
- A prosecutor presents **evidence** to show that the defendant should be considered guilty.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.
- A prosecutor presents **evidence** to show that the defendant should be considered guilty.
- The jury considers whether the evidence, when judged against a standard (**beyond a reasonable doubt**), is sufficient to abandon the presumption of innocence.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.
- A prosecutor presents **evidence** to show that the defendant should be considered guilty.
- The jury considers whether the evidence, when judged against a standard (**beyond a reasonable doubt**), is sufficient to abandon the presumption of innocence.
- The jury returns a **verdict**: either guilty or not guilty. A defendant is generally not declared innocent.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work. Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.
- A prosecutor presents **evidence** to show that the defendant should be considered guilty.
- The jury considers whether the evidence, when judged against a standard (**beyond a reasonable doubt**), is sufficient to abandon the presumption of innocence.
- The jury returns a **verdict**: either guilty or not guilty. A defendant is generally not declared innocent.
- A jury can make two kinds of mistakes: they can convict an innocent defendant or they can fail to convict a guilty defendant.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests begins by assuming that there is nothing interesting about the population(s) being studied.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests begins by assuming that there is nothing interesting about the population(s) being studied.
- The evidence is comparable to the **sample** collected to answer the research the question.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests begins by assuming that there is nothing interesting about the population(s) being studied.
- The evidence is comparable to the **sample** collected to answer the research the question.
- The "beyond a reasonable doubt" standard is comparable to the **significance level** or α of the test.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests begins by assuming that there is nothing interesting about the population(s) being studied.
- The evidence is comparable to the **sample** collected to answer the research the question.
- The "beyond a reasonable doubt" standard is comparable to the **significance level** or α of the test.
- The verdict is comparable to the **conclusion** of the test: either to reject the null or to fail to reject the null. The null is never accepted or proven.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests begins by assuming that there is nothing interesting about the population(s) being studied.
- The evidence is comparable to the **sample** collected to answer the research the question.
- The "beyond a reasonable doubt" standard is comparable to the **significance level** or α of the test.
- The verdict is comparable to the **conclusion** of the test: either to reject the null or to fail to reject the null. The null is never accepted or proven.
- A conclusion can be in error one of two ways: a type I error (convict an innocent defendant) or a type II error (fail to convict a guilty defendant).

Group work

- For all questions, complete part (c).