Hypothesis Testing

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of OpenIntro and John McGreedy of Johns Hopkins University

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

What proportion of Males are promoted? What proportion of Females?

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

What proportion of Males are promoted? What proportion of Females?

$\hat{p}_{\text{males}} = 21 / 24 = 0.88$

$\hat{p}_{\text{females}} = 14 / 24 = 0.58$

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

$$\hat{p}_{males} = 21 / 24 = 0.88$$

$$\hat{p}_{females} = 14 / 24 = 0.58$$

Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.

- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

$$\hat{p}_{males} = 21 / 24 = 0.88$$
$$\hat{p}_{females} = 14 / 24 = 0.58$$

Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.
  → null hypothesis (nothing is going on)
- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.
  → alternative. hypothesis (something is going on)

Gender discrimination experiment:

How do we test which of these explanations is most likely?

Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.
  → null hypothesis (nothing is going on)
- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.
  → alternative. hypothesis (something is going on)

Hypothesis testing

Hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.

Hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.

Hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).

Hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Number of college applications

A survey of Duke students asked how many colleges they applied to. 206 students responded. This sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all Duke students apply to is higher than recommended?

http://www.collegeboard.com/student/apply/the-application/151680.html

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by <u>all</u> Duke students.

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by <u>all</u> Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
    - The true population mean is different.
    - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by <u>all</u> Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
  - The true population mean is different.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability
- We start with the assumption the average number of colleges Duke students apply to is 8 (as recommended)

$$H_0 : \boldsymbol{\mu} = 8$$

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by all Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
  - The true population mean is different.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability
- We start with the assumption the average number of colleges Duke students apply to is 8 (as recommended)

$$H_0 : \boldsymbol{\mu} = 8$$

- We test the claim that the average number of colleges Duke students apply to is greater than 8

$$H_A : \boldsymbol{\mu} > 8$$

Number of college applications - conditions

Which of the following is *not* a condition that needs to be met to proceed with this hypothesis test?

a) Students in the sample should be independent of each other with respect to how many colleges they applied to.
b) Sampling should have been done randomly.
c) The sample size should be less than 10% of the population of all Duke students.
d) There should be at least 10 successes and 10 failures in the sample.
e) The distribution of the number of colleges students apply to should not be extremely skewed.

Number of college applications - conditions

Which of the following is *not* a condition that needs to be met to proceed with this hypothesis test?

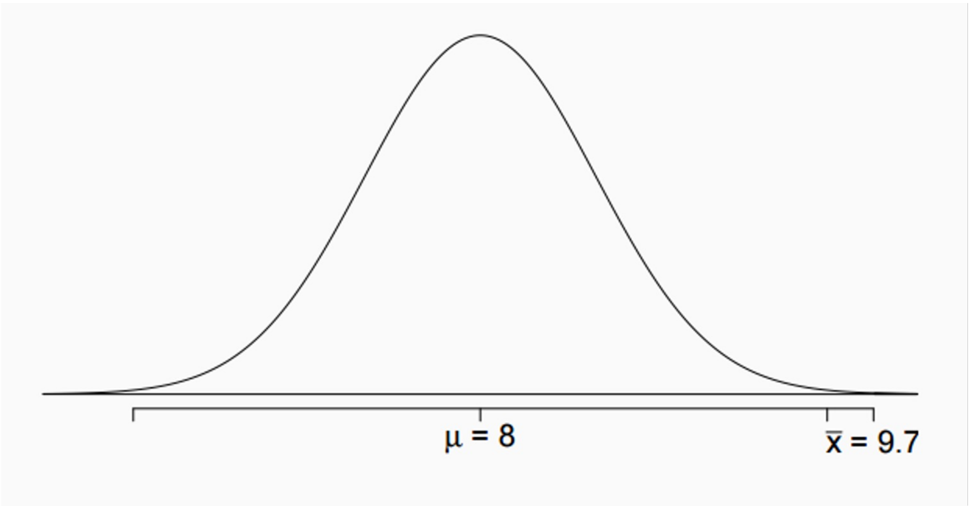
a) Students in the sample should be independent of each other with respect to how many colleges they applied to.
b) Sampling should have been done randomly.
c) The sample size should be less than 10% of the population of all Duke students.
d) *There should be at least 10 successes and 10 failures in the sample.*
e) The distribution of the number of colleges students apply to should not be extremely skewed.

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

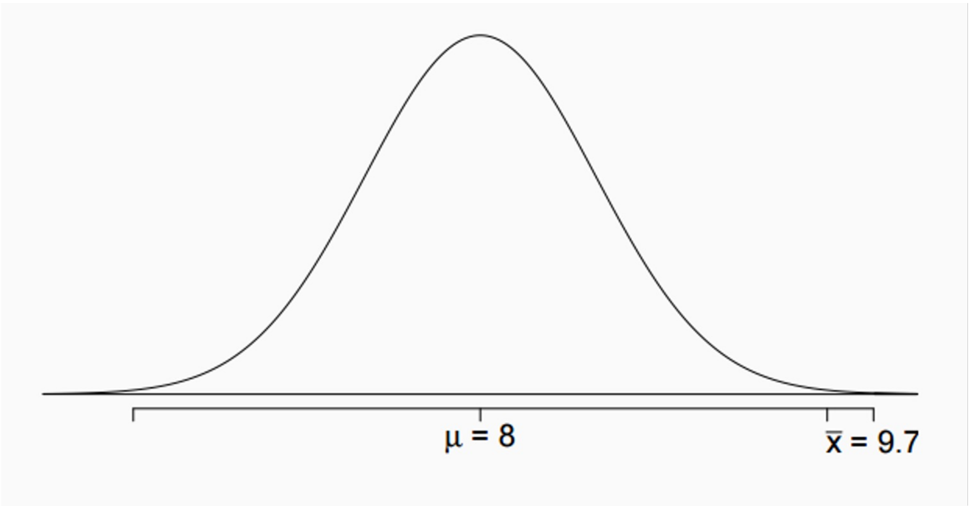
Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} = N(\mu, SE)$$

$$SE = \frac{s}{\sqrt{n}}$$

Number of college applications

A survey of Duke students asked how many colleges they applied to. 206 students responded. This sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all Duke students apply to is higher than recommended?

$$\bar{x} = N(\mu, SE)$$

$$SE = \frac{s}{\sqrt{n}}$$

What is SE?

http://www.collegeboard.com/student/apply/the-application/151680.html

Number of college applications

A survey of Duke students asked how many colleges they applied to. 206 students responded. This sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all Duke students apply to is higher than recommended?

$$\bar{x} = N(\mu, SE)$$

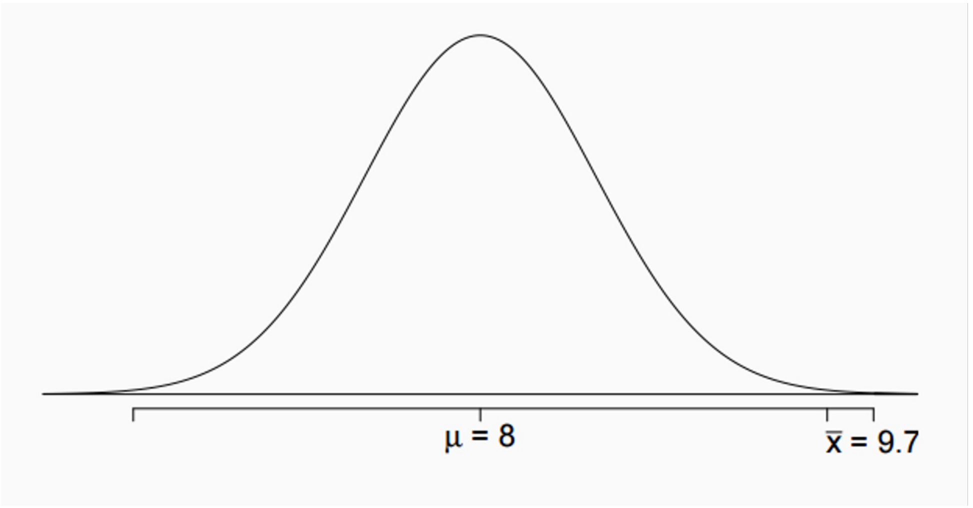
$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$SE = \frac{s}{\sqrt{n}}$$

http://www.collegeboard.com/student/apply/the-application/151680.html

Test Statistic

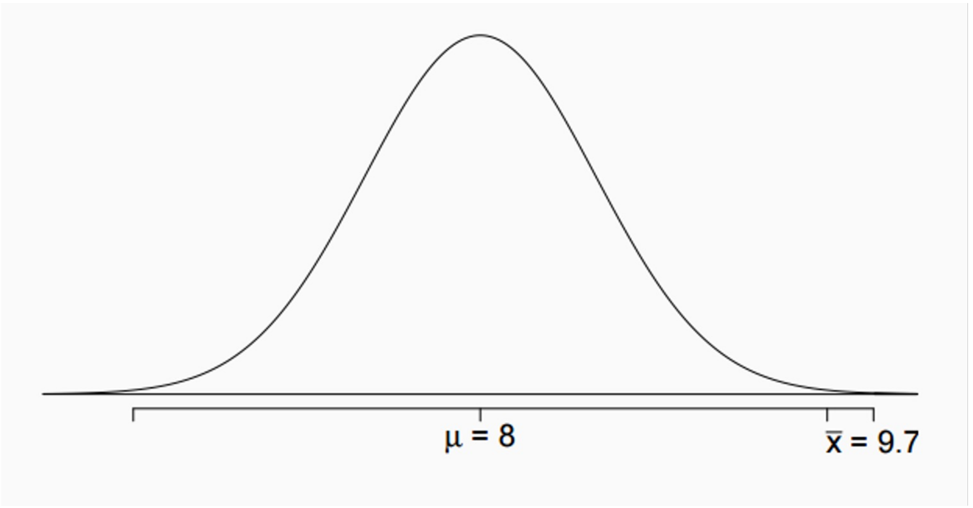
In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

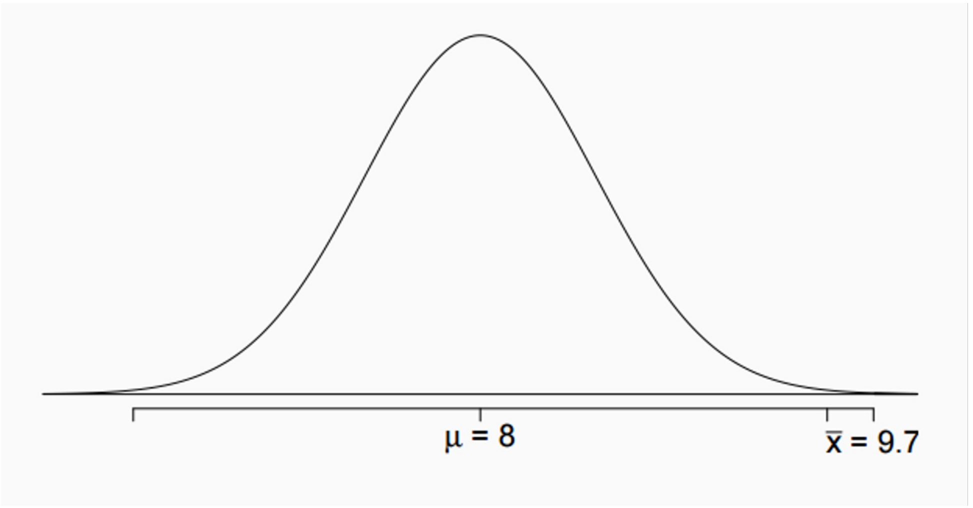


$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

How many SE's from μ is $\bar{x}$?

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



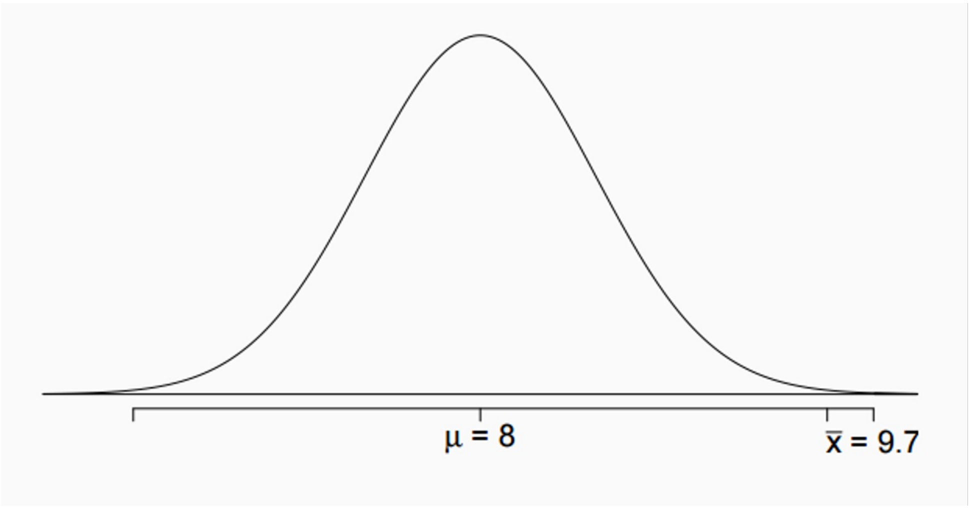
$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

How many SE's from μ is $\bar{x}$?

$$z = \frac{x - \mu}{SE}$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

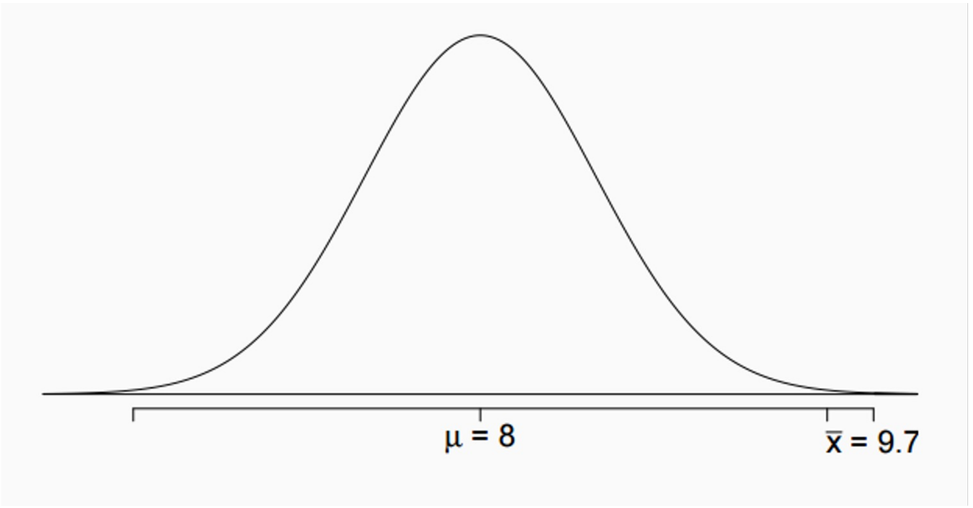
How many SE's from μ is $\bar{x}$?

$$Z = \frac{x - \mu}{SE}$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



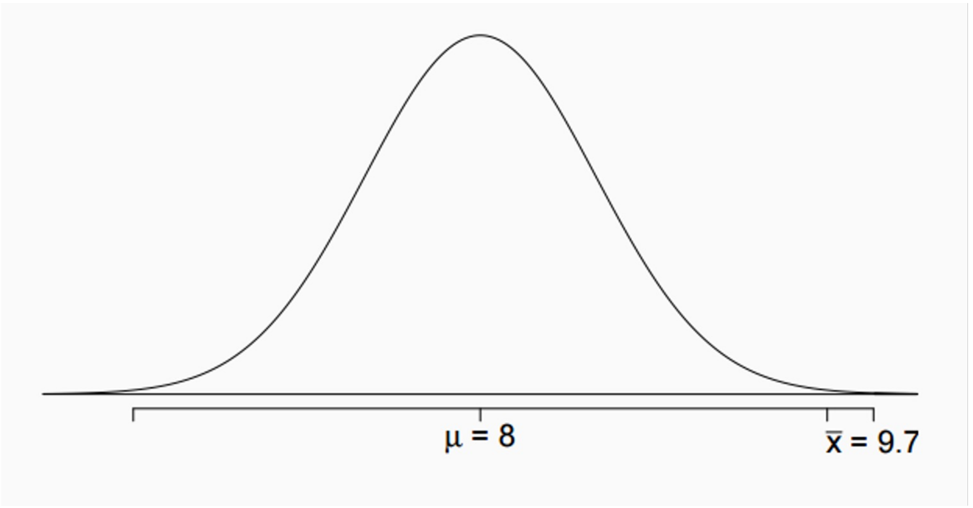
The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually high? That is, is the result ***statistically significant***?

$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually high? That is, is the result ***statistically significant***?

Yes, and we can quantify how unusual it is using a p-value.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

p-values

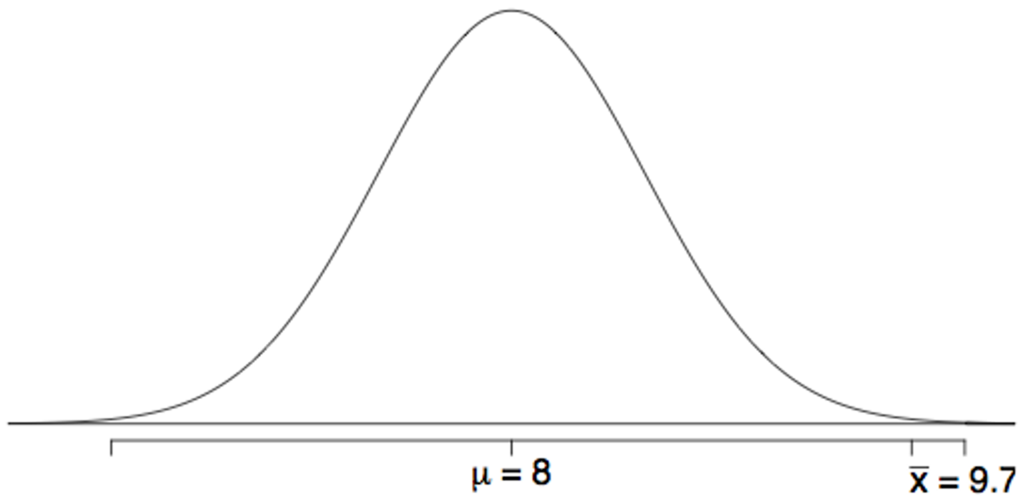
- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject H_0*.
- If the p-value is *high* (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject H_0*.

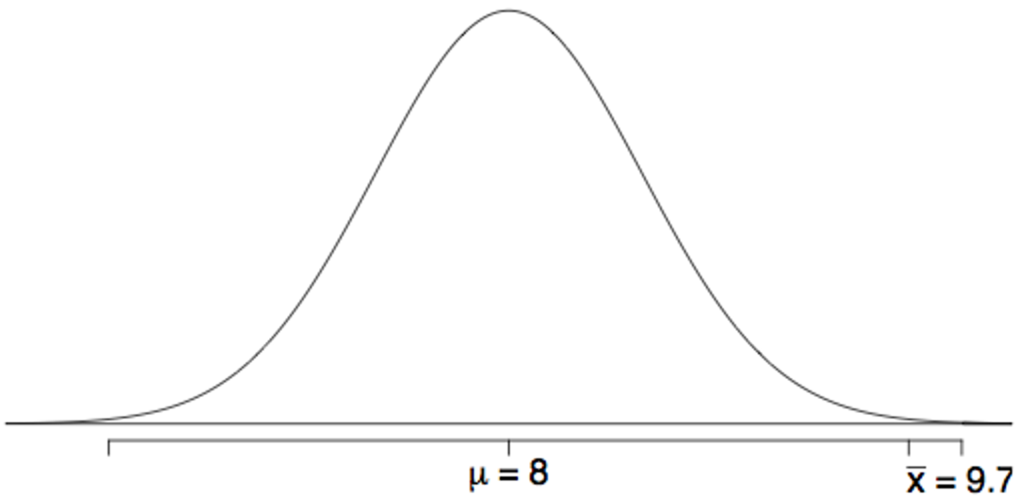
Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



Number of college applications - p-value

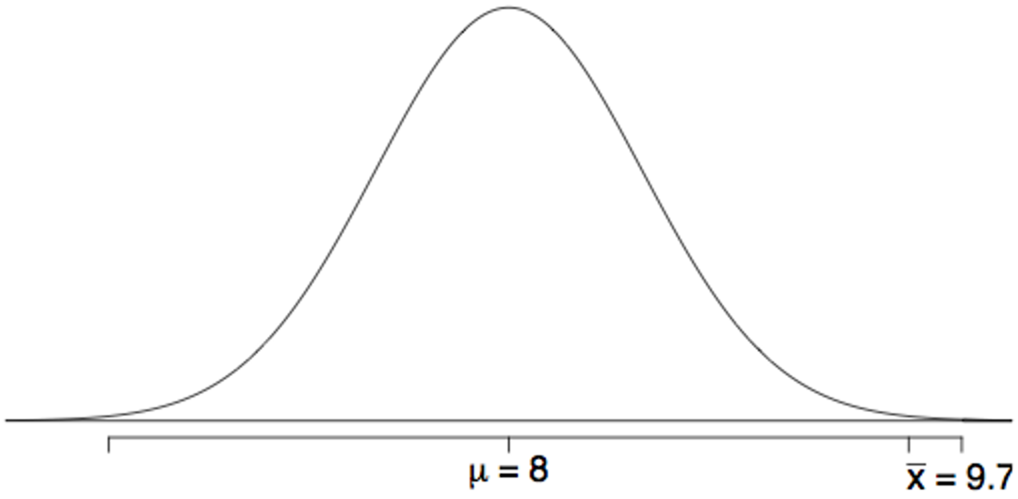
p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4)$$

Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).

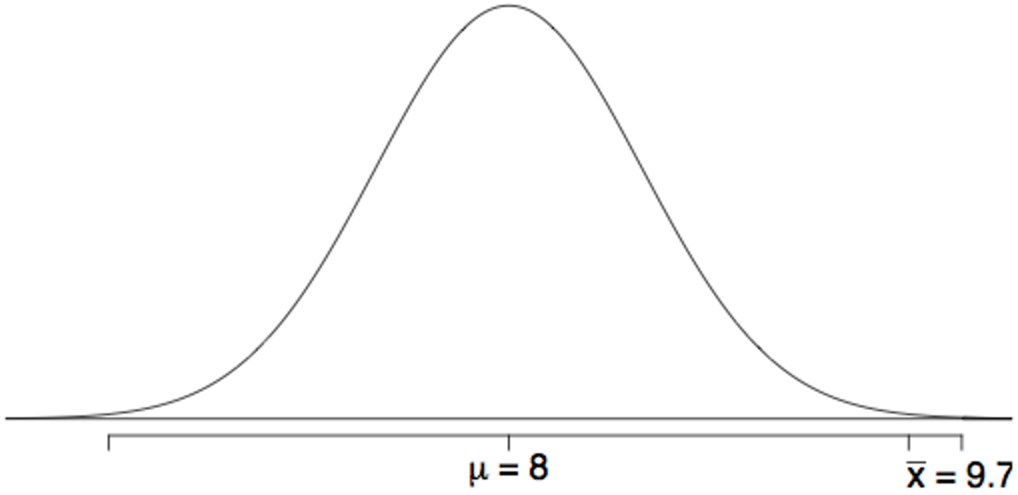


$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4)$$

What is this probability?

Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4) = 0.0003$$

Number of college applications - Making a decision

- p-value = 0.0003

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject H_0*.

Number of college applications - Making a decision

- p-value = 0.0003
    - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
    - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject H_0*.
- The data provide convincing evidence that Duke students apply to more than 8 schools on average.

Number of college applications - Making a decision

- p-value = 0.0003
    - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
    - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject H_0*.
- The data provide convincing evidence that Duke students apply to more than 8 schools on average.
- The difference between the null value of 8 schools and observed sample mean of 9.7 schools is *not due to chance* or sampling variability.

Practice

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 college students taking an introductory statistics class yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all college students *(bit of a leap of faith?)*, a hypothesis test was conducted to evaluate if college students on average sleep less than 7 hours per night. The p-value for this hypothesis test is 0.0485. Which of the following is correct?

a) Fail to reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.
b) Reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.
c) Reject H_0, the data prove that college students sleep more than 7 hours on average.
d) Fail to reject H_0, the data do not provide convincing evidence that college students sleep less than 7 hours on average.
e) Reject H_0, the data provide convincing evidence that college students in this sample sleep less than 7 hours on average.

Practice

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 college students taking an introductory statistics class yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all college students *(bit of a leap of faith?)*, a hypothesis test was conducted to evaluate if college students on average sleep less than 7 hours per night. The p-value for this hypothesis test is 0.0485. Which of the following is correct?

a) Fail to reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.
b) *Reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.*
c) Reject H_0, the data prove that college students sleep more than 7 hours on average.
d) Fail to reject H_0, the data do not provide convincing evidence that college students sleep less than 7 hours on average.
e) Reject H_0, the data provide convincing evidence that college students in this sample sleep less than 7 hours on average.

Two-sided hypothesis testing with p-values

- If the research question was “Do the data provide convincing evidence that the average amount of sleep college students get per night is *different* than the national average?”, the alternative hypothesis would be different

H_0: $\mu = 7$

H_A: $\mu \neq 7$

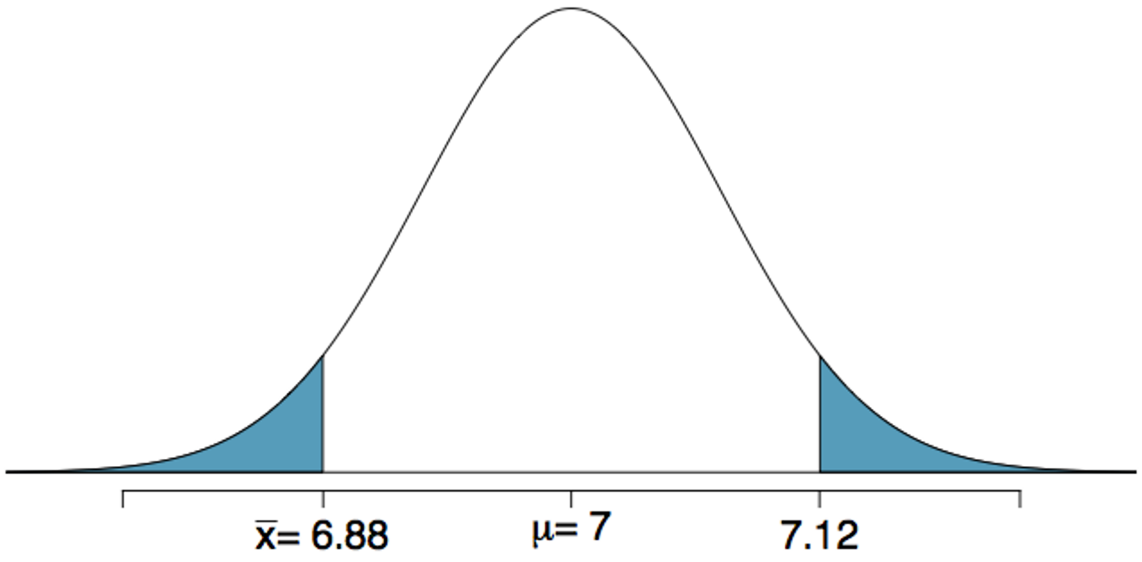
Two-sided hypothesis testing with p-values

- If the research question was “Do the data provide convincing evidence that the average amount of sleep college students get per night is *different* than the national average?”, the alternative hypothesis would be different

$$H_0: \mu = 7$$

$$H_A: \mu \neq 7$$

- Hence the p-value would change as well:



$$\text{p-value} = 0.0485 \times 2 = 0.097$$

Choosing a significance level

- Choosing a significance level for a test is important in many contexts, and the traditional level is 0.05. However, it is often helpful to adjust the significance level based on the application.
- We may select a level that is smaller or larger than 0.05 depending on the consequences of any conclusions reached from the test.

Decision errors

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		✓

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true		✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true	Type 2 Error	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

Hypothesis Test Analogy

If think of a hypothesis test as a medical test for illness X, then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Patient does not have X

H_A: Patient has X

Which type of error is being committed in the following circumstances?

- Declaring the patient healthy when they are actually sick
- Declaring the patient sick when they are actually healthy

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true	Type 2 Error	✓

Hypothesis Test Analogy

If think of a hypothesis test as a medical test for illness X, then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Patient does not have X

H_A: Patient has X

Which type of error is being committed in the following circumstances?

- Declaring the patient healthy when they are actually sick

  Type 2 error

- Declaring the patient sick when they are actually healthy

  Type 1 error

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

Hypothesis Test Analogy

If think of a hypothesis test as a medical test for illness X, then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Patient does not have X

H_A: Patient has X

Which type of error is being committed in the following circumstances?

- Declaring the patient healthy when they are actually sick

  Type 2 error

- Declaring the patient sick when they are actually healthy

  Type 1 error

Which error do you think is the worse error to make?

Hypothesis Test Analogy

If think of a hypothesis test as a medical test for illness X, then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Patient does not have X

H_A: Patient has X

Which type of error is being committed in the following circumstances?

- Declaring the patient healthy when they are actually sick

  Type 2 error

- Declaring the patient sick when they are actually healthy

  Type 1 error

Which error do you think is the worse error to make?
Would a different situation make the worse error change?

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

- This is why we prefer small values of α -- increasing α increases the Type 1 error rate.

Choosing a significance level

- Choosing a significance level for a test is important in many contexts, and the traditional level is 0.05. However, it is often helpful to adjust the significance level based on the application.
- We may select a level that is smaller or larger than 0.05 depending on the consequences of any conclusions reached from the test.
- If making a Type 1 Error is dangerous or especially costly, we should choose a small significance level (e.g. 0.01). Under this scenario we want to be very cautious about rejecting the null hypothesis, so we demand very strong evidence favoring H_A before we would reject H_0.
- If a Type 2 Error is relatively more dangerous or much more costly than a Type 1 Error, then we should choose a higher significance level (e.g. 0.10). Here we want to be cautious about failing to reject H_0 when the null is actually false.

Practice: Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Practice: Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

What are the hypothese?

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Practice: Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

What are the hypothese?

Setting the hypotheses

We start with the assumption that 50% of American Facebook users are comfortable with Facebook creating categories of interests for them

H_0: $p = 0.50$

We test the claim that the proportion of American Facebook users who are comfortable with Facebook creating categories of interests for them is different than 50%.

H_A: $p \neq 0.50$

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook interest categories - conditions

Which of the following is not a condition that needs to be met to proceed with this hypothesis test?

(a) Respondents in the sample should be independent of each other with respect to whether or not they feel comfortable with their interests being categorized by Facebook.

(b) Sampling should have been done randomly.

(c) The sample size should be less than 10% of the population of all American Facebook users.

(d) There should be at least 30 respondents in the sample.

(e) There should be at least 10 expected successes and 10 expected failure.

Facebook interest categories - conditions

Which of the following is not a condition that needs to be met to proceed with this hypothesis test?

(a) Respondents in the sample should be independent of each other with respect to whether or not they feel comfortable with their interests being categorized by Facebook.

(b) Sampling should have been done randomly.

(c) The sample size should be less than 10% of the population of all American Facebook users.

(d) There should be at least 30 respondents in the sample.

(e) There should be at least 10 expected successes and 10 expected failure.

Practice: Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

$$p = Binomial(\hat{p}, s)$$

What is the test statistic?

H_0: $p = 0.50$

H_A: $p \neq 0.50$

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

$$z = \frac{x - \hat{p}}{SE}$$

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Test statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.50, SE = \sqrt{\frac{0.50 \times 0.50}{850}}\right)$$

$$Z = \frac{0.41 - 0.50}{0.0171} = -5.26$$

The sample proportion is 5.26 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result *statistically significant*?

Test statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.50, SE = \sqrt{\frac{0.50 \times 0.50}{850}}\right)$$

$$Z = \frac{0.41 - 0.50}{0.0171} = -5.26$$

The sample proportion is 5.26 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result *statistically significant*?

Yes, and we can quantify how unusual it is using a p-value: p-value < 0.0001

Facebook interest categories
- Making a decision

p-value < 0.0001

- If 50% of all American Facebook users are comfortable with Facebook creating these interest categories, there is less than a 0.01% chance of observing a random sample of 850 American Facebook users where 41% or fewer or 59% of higher feel comfortable with it.
- This is a pretty low probability for us to think that the observed sample proportion, or something more extreme, is likely to happen simply by chance.

Since p-value is *low* (lower than 5%) we *reject* H_0.

The data provide convincing evidence that the proportion of American Facebook users who are comfortable with Facebook creating a list of interest categories for them is different than 50%.

The difference between the null value of 0.50 and observed sample proportion of 0.41 is *not due to chance* or sampling variability.

Recap: Hypothesis testing framework

1. Set the hypotheses.
2. Check assumptions and conditions.
3. Calculate a *test statistic* and a p-value.
4. Make a decision, and interpret it in context of the research question.

Recap: Hypothesis testing for a population mean

1. Set the hypotheses

- H_0: μ = null value
- H_A: μ < or > or ≠ null value

2. Calculate the point estimate

3. Check assumptions and conditions

- Independence: random sample/assignment, 10% condition when sampling without replacement
- Normality: nearly normal population or $n \geq 30$, no extreme skew -- or use the *t* distribution (Ch 5)

4. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{x} - \mu}{SE}, \text{ where } SE = \frac{s}{\sqrt{n}}$$

5. Make a decision, and interpret it in context

- If p-value < α, reject H_0, data provide evidence for H_A
- If p-value > α, do not reject H_0, data do not provide evidence for H_A